Automated Text Analysis for Consumer Research

Automated Text Analysis for Consumer Research

ASHLEE HUMPHREYS
REBECCA JEN-HUI WANG

The amount of digital text available for analysis by consumer researchers has risen dramatically. Consumer discussions on the internet, product reviews, and digital archives of news articles and press releases are just a few potential sources for insights about consumer attitudes, interaction, and culture. Drawing from linguistic theory and methods, this article presents an overview of automated text analysis, providing integration of linguistic theory with constructs commonly used in consumer research, guidance for choosing amongst methods, and advice for resolving sampling and statistical issues unique to text analysis. We argue that although automated text analysis cannot be used to study all phenomena, it is a useful tool for examining patterns in text that neither researchers nor consumers can detect unaided. Text analysis can be used to examine psychological and sociological constructs in consumer-produced digital text by enabling discovery or by providing ecological validity.

Keywords: automated text analysis, computer-assisted text analysis, automated content analysis, computational linguistics

O ver the last two decades, researchers have seen an explosion of text data generated by consumers in the form of text messages, reviews, tweets, emails, posts, and blogs. Some part of this rise is attributed to an increase in

sites like Amazon.com, CNET.com, and thousands of other product websites that offer forums for consumer comment. Another part of this growth comes from consumer-generated content, including discussions of products, hobbies, or brands on feeds, message boards, and social networking sites. Researchers, consumers, and marketers swim in a sea of language, and more and more of that language is recorded in the form of text. Yet within all of this information lies knowledge about consumer decision making, psychology, and culture that may be useful to scholars in consumer research. Blogs can be used to study opinion leadership; message boards can tell us about the development of consumer communities; feeds like Twitter can help us unpack social media firestorms; and social commerce sites like Amazon can be mined for details about word-of-mouth communication.

Correspondingly, ways of doing social science are also changing. Because data has become more readily available and the tools and resources for analysis are cheaper and more accessible, researchers in the material sciences, humanities, and social sciences are developing new methods of data-driven discovery to deal with what some call the "data deluge" or "big data" (Bell, Hey, and Szalay 2009; Borgman 2015). Just as methods for creating, circulating, and storing online discussion have grown more sophisticated, so too

have tools for analyzing language, aggregating insight, and distilling knowledge from this overwhelming amount of data. Yet despite the potential importance of this shift, consumer research is only beginning to incorporate methods for collecting and systematically measuring textual data to support theoretical propositions and make discoveries.

In light of the recent influx of available data and the lack of an overarching framework for doing consumer research using text, the goal of this article is to provide a guide for research designs that incorporate text and to help researchers assess when and why text analysis is useful for answering consumer research questions. We provide an overview of both deductive top-down, dictionary-based approaches as well as inductive and abductive bottom-up approaches such as supervised and unsupervised learning to incorporate discovery-oriented as well as theoretically guided methods. These designs help make discoveries and expand theory by allowing computers to detect and display patterns that humans cannot and by providing new ways of "seeing" data through aggregation, comparison, and correlation. We further offer guidance for choosing amongst different methods and address common issues unique to text analysis, such as sampling internet data, developing word lists to represent a construct, and analyzing sparse, non-normally distributed data. We also address validity, reliability, generalizability, and ethical issues for research using textual data.

Although there are many ways to incorporate automated text analysis into consumer research, there is not much agreement on the standard set of methods, reporting procedures, steps of data inclusion, exclusion, and sampling, and, where applicable, dictionary development and validation. Nor has there been an integration of the linguistic theory on which these methods are based into consumer research, which can enlighten us to the multiple dimensions of language that can be used to measure consumer thought, interaction, and culture. While fields like psychology provide some guidance for dictionary-based methods (Tausczik and Pennebaker 2010) and for analysis of certain types of social media data (Kern et al. 2016), they don't provide grounding in linguistics, cover the breadth of methods available for studying text, or provide criteria for deciding amongst approaches. In short, most of the existing literature examines only a handful of aspects of discourse that pertain to the research questions of interest, does not address why one method is chosen over others, and does not discuss the unique methodological issues consumer researchers face when dealing with text.

This article therefore offers three contributions to consumer research. First, we detail how linguistic theory can inform theoretical areas common in consumer research, such as attention, processing, interpersonal interaction, group dynamics, and cultural characteristics. Second, we outline a practical roadmap for researchers who want to use textual data, particularly unstructured text obtained from real-world settings, such as tweets, newspaper articles, or online reviews. Lastly, we examine what can and cannot be done with text analysis and provide guidance for validating results and interpreting findings in non-experimental contexts.

The rest of the article is organized around the roadmap in figure 1. This chart presents a series of decisions a researcher faces when analyzing text. We outline six stages: (1) developing a research question, (2) identifying the constructs, (3) collecting data, (4) operationalizing the constructs, (5) interpreting the results, and (6) validating the results. Although text analysis need not necessarily unfold in this order (for instance, construct definition will sometimes occur after data collection), researchers have generally followed this progression (Lee and Bradlow 2011).

ROADMAP FOR AUTOMATED TEXT ANALYSIS

Methods of automated text analysis come from the field of computational linguistics (Kranz 1970; Stone 1966). The relationship between computational linguistics and text analysis is analogous to that of biology to medicine or of physics to engineering. That is, computational linguistics, at its core, emphasizes advancing linguistics theory and often focuses on the accuracy of prediction as an end in itself (Hausser 1999, 8). Computer-assisted or automated text analysis, on the other hand, refers to a set of techniques that use computing power to answer questions related to psychology (Chung and Pennebaker 2013; Tausczik and Pennebaker 2010), political science (Grimmer and Stewart 2013), sociology (Mohr 1998; Shor et al. 2015), and other social sciences (Carley 1997; Weber 2005). In these fields, language represents some focal construct of interest, and computers are used to measure those constructs, provide systematic comparisons, and sometimes find patterns that neither human researchers nor subjects of the research can detect. In other words, while computational linguistics is a field that is primarily concerned with language in the text, for consumer researchers, text analysis is merely a lens through which to view consumer thought, behavior, and culture. Analyzing texts, in many contexts, is not the ultimate goal of consumer researchers, but is instead a precursor for testing the relationship between or amongst the constructs or variables of interest.

Therefore, we use the term "automated text analysis" or "computer-assisted text analysis" over "computational linguistics" (Brier and Hopp 2011). Although we follow convention by using the term "automated," this should not imply that human intervention is absent. In fact, many of the tasks—such as dictionary construction, validation, and cluster labeling—are iterative processes that require human design, modification, and interpretation. Some prefer the term "computer-assisted text analysis" (Alexa 1997) to explicitly encompass a broad set of methods that take advantage of computation in

FIGURE 1

STAGES OF AUTOMATED TEXT ANALYSIS
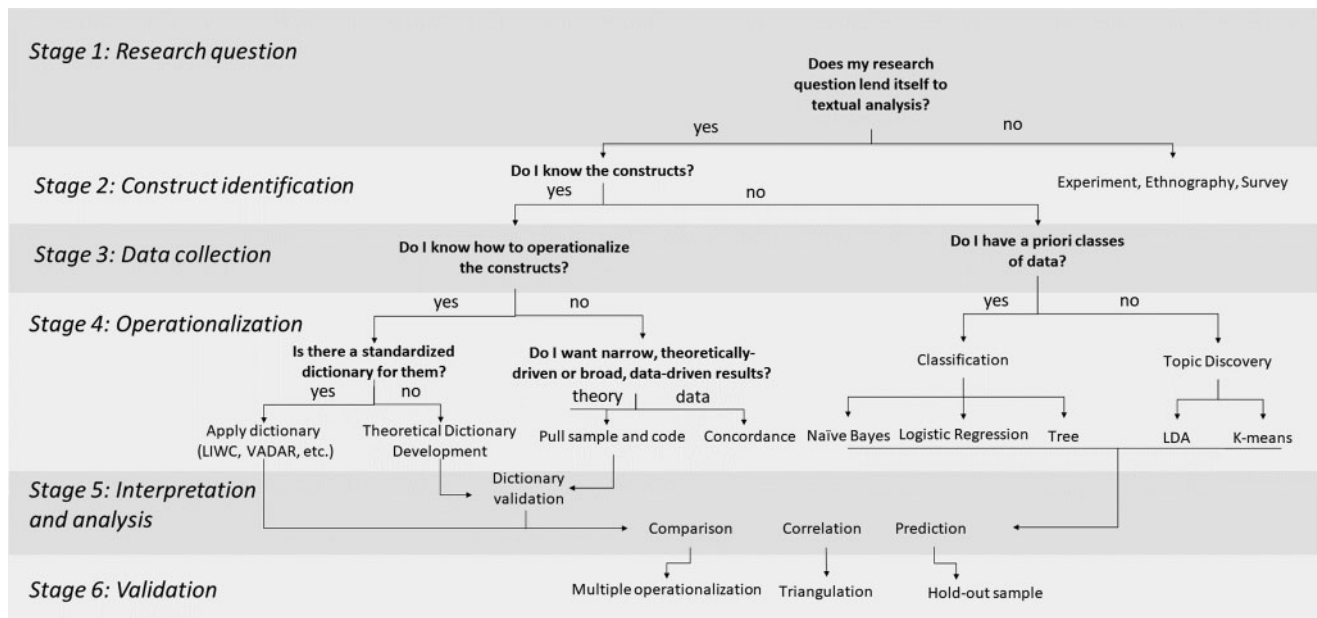
varying amounts ranging from a completely automated process using machine learning to researcher-guided approaches that include manual coding and word list development. In the following sections, we discuss the design and execution of automated text analysis in detail, beginning with selection of a research question and connecting linguistic aspects to important constructs in consumer research.

STAGE 1: DEVELOP A RESEARCH QUESTION

As with any research, the first step is developing a research question. To understand the implementation of automated text analysis, one should start by first considering if the research question lends itself to text analysis. Contemplating whether text analysis is suitable for the research context is perhaps the most important decision to consider, and there are at least three purposes for which text analysis would be inappropriate.

First, much real-world textual content is observational data that occurs without the controlled conditions of an experiment or even a field test. Depending on the context and research question, automated text analysis alone would not be the best method for inferring causation when studying a psychological mechanism.[1] If the researcher needs precise

control to compare groups, introduce manipulations, or rule out alternative hypotheses through random assignment protocols (Cook, Campbell, and Day 1979), textual analysis would be of limited use.[2]

Secondly, if the research question concerns data at the behavioral or unarticulated level (e.g., response time, skin conductance, consumer practices), text analysis would not be appropriate. Neural mechanisms that govern perception or attention, for example, would be ill suited for the method. Equally, if one needs a behavioral dependent variable, text analysis would not be appropriate to measure it. For example, when one is studying self-regulation, it is clearly important to include behavioral measures to examine not just behavioral intention—what people say they will do—but action itself. This restriction applies to sociologically oriented research as well. For example, with practice theory (Allen 2002; Schatzki 1996) or ethnography (Belk, Sherry, and Wallendorf 1988; Schouten and McAlexander 1995), observation of consumer practices is vital because consumer behavior may diverge markedly from discourse (Wallendorf and Arnould 1988). Studying text is simply no substitute for studying behavior. Not all constructs lend themselves to examination through text, and these constructs tend to be behaviorally oriented.

---

1   For alternative perspectives of studying causation with historical case data and macro-level data, see Mahoney and Rueschemeyer (2003) and Jepperson and Meyer (2011).

2   However, text analysis has been used to code thought protocols in experimental settings (Hsu et al. 2014).

Lastly, there are many contexts in which some form of text analysis would be valuable, but automated text analysis would be insufficient. Identifying finer shades of meaning, such as sarcasm, and differentiating amongst complex concepts, rhetorical strategies, or complex arguments are often not possible via automated processes. Additionally, studies that employ text analysis often sample data from public discourse in the form of tweets, message boards, or posts, and there is a wide range of expression that consumers may not pursue in these media because of stigma or social desirability. There is a rich tradition of text analysis in consumer research, such as discourse analysis (Holt and Thompson 2004; Thompson and Hirschman 1995), hermeneutic analysis (Arnold and Fischer 1994; Thompson, Locander, and Pollio 1989), and human content analysis (Kassarjian 1977) for uncovering rich, deep, and sometimes personal meaning of consumer life in the context in which it is lived. Although automated text analysis could be a companion to these methods, it cannot be a standalone approach for understanding this kind of richer, deeper, and culturally laden meaning.

So, when is automated text analysis appropriate? In general, it is good for analyzing data in a context where humans may be limited or partial. Computers can sometimes see patterns in language that humans cannot detect, and they are impartial in the sense that they measure textual data evenly and precisely over time or in comparisons between groups without preconception. Further, by quantifying constructs in text, computers provide new ways of aggregating and displaying information to uncover patterns that may not be obvious at the granular level. There are at least four types of problems where these advantages can be leveraged.

First, automated text analysis can lead to discoveries of systematic relationships in text and hence amongst constructs that may be overlooked by researchers or consumers themselves. Patterns in correlation, notable absences, and relationships amongst three or more textual elements are all things that are simply hard for a human reader to see. For example, in medical research, Swanson (1988) finds a previously unrecognized relationship between migraine headaches and magnesium levels through the text analysis of other, seemingly unrelated research. Automated text analysis may also provide alternative ways of "reading" the text to make new discoveries (Kirschenbaum 2007). For instance, Jurafsky et al. (2014) find expected patterns in negative restaurant reviews, such as negative emotion words, but they also discover words like "after," "would," and "should" in these reviews, which are used to construct narratives of interpersonal trauma primarily based on norm violations. Positive restaurant reviews, on the other hand, contain stories of addiction rather than simple positive descriptions of food or service. These discoveries, then, theoretically inform researchers' understanding of negative and positive sentiment, particularly in the context of consumer experiences.

Using text analysis, researchers have also discovered important differences between expert and consumer discourse when evaluating products (Lee and Bradlow 2011; Netzer et al. 2012). In the case of cameras, for example, systematic linguistic comparison of expert reviews to consumer reviews reveals that there is a significant disconnect between what each of these groups considers important. For example, in their reviews, consumers value observable attributes like camera size and design, while experts stress less visible issues like flash range and image compression (Lee and Bradlow 2011). In the case of prescription drugs, the differences between consumers and experts take on heightened meaning, as textual comparison of patient feedback of drugs on WebMD shows that consumers report side effects missing from the official medical literature (Netzer et al. 2012). In this way, text analysis can reveal discoveries that would be hard to detect on a more granular level. Further, The scope and systematicity of the analysis can grant more validity and perhaps power to consumers' point of view.

Second, researchers can use computers to execute rules impartially in order to measure changes in language over time, compare between groups, or aggregate large amounts of text. These tasks are more than mere improvements in efficiency in that they present an alternative way of "seeing" the text through conceptual maps (Martin, Pfeffer, and Carley 2013), timelines (Humphreys and Latour 2013), or networks (Arvidsson and Caliandro 2016), and provide information about rate and decay. For example, using features like geolocation and timestamps along with textual data from Twitter, Snefjella and Kuperman (2015) develop new knowledge about construal level such as its rate of change given a speaker's physical, temporal, social, or topical proximity.

Providing an explicit rule set and having a computer execute the rules over the entire dataset, reduces the possibility that texts will be analyzed unevenly or incompletely. This is especially important when researchers are making statistical inferences about changes in concepts over time, because it ensures that measurement is consistent throughout the dataset. For example, by aggregating and placing counts of hashtags on a timeline, Arvidsson and Caliandro (2016) demonstrate how networks of concepts used to discuss Louis Vuitton handbags peak at particular times and in accordance with external events, highlighting attention for a particular public. If a researcher wants to study a concept like brand meaning, text analysis can help to create conceptual or positioning maps that represent an aggregated picture of consumer perceptions that can then be used to highlight potential gaps or tensions in meaning amongst different constituencies or even for one individual (Lee and Bradlow 2011; Netzer et al. 2012).

Third, text analysis can be a valuable companion to experimental research designs by adding ecological validity to lab results. For example, Mogilner et al. (2011) find robust support for changes in the frame of happiness that

correspond with age by looking at a large dataset of personal blogs, patterns they also find in a survey and laboratory experiment. In a study of when and why consumers explain choices as a matter of taste versus quality, Spiller and Belogolova (2016) use text analysis first to code a dependent variable in experimental analysis, but then add robustness to their results by demonstrating the effect in the context of online movie reviews. In this way, text analysis is valuable beyond its more traditional uses for coding thought protocols, but also useful for finding and measuring psychological and sociological constructs in naturally occurring consumer discourse.

Lastly, there are some relationships for which observational data is the most natural way to study the phenomenon. Interpersonal relationships and group interaction can be hard to study in the lab, but they can be examined through text analysis of online interaction or transcripts of recorded conversation (Jurafsky, Ranganath, and McFarland 2009). For example, Barasch and Berger (2014) combine laboratory studies with dictionary-based text analysis of consumer discussions to show that consumers share different information depending on the size of their audience.

Given these considerations, once it is established that text analysis is appropriate for some part of the research design, the next question is what role it will play. Text could be used to represent the independent variable (IV), dependent variable (DV), or both. For example, Tirunillai and Tellis (2012) operationalize "chatter" using quantity and valence of product reviews to represent the IV, which predicts firm performance in the financial stock market. Conversely, Hsu et al. (2014) experimentally manipulate distraction, the IV, and measure thoughts as the DV using text analysis. Other studies use text as both the IV and the DV. For example, Humphreys (2010) examines how terms related to casino gambling and entertainment in newspaper articles converged over time along with a network of other concepts such as luxury and money, while references to illegitimate frames like crime fell. As these cases illustrate, text analysis is a distinct component of the research design, to be executed and then incorporated into the overall design.

More generally, text analysis can occupy different places in the scientific process, depending on the interests and orientation of the researchers. It is compatible with both theory testing and discovery-oriented designs. For some, text analysis is a way of first discovering patterns that are later verified through laboratory experiments (Barasch and Berger 2014; Berger and Milkman 2012; Packard and Berger 2016). Others use text analysis to enrich findings after investigating a psychological or social mechanism (Mogilner et al. 2011; Spiller and Belogolova 2016). In the same way, sociological work has used text analysis to illustrate findings after an initial discovery phase through qualitative analysis (Arsel and Bean 2013; Humphreys 2010), or to set the stage by presenting sociocultural discourses prior to individual or group-level analysis (Arsel and Thompson 2011).

STAGE 2: IDENTIFY THE CONSTRUCT

After one has decided that text analysis might be appropriate for the research question, the next step is to identify the construct. Doing so, however, entails recognizing that text is ultimately based on language. To build sound hypotheses and make valid conclusions from text, one must first understand the underpinnings of language.

Language indelibly shapes how humans view the world (Piaget 1959; Quine 1970; Vico 1725/1984; Whorf 1944). It can be both representative of thought and instrumental in shaping thought (Kay and Kempton 1984; Lucy and Shweder 1979; Sapir 1929; Schmitt and Zhang 1998; see also Graham 1981; Whorf 1944). For example, studies have shown that languages with gendered nouns, like Spanish and French, are more likely to make speakers think of physical objects as having a gender (Boroditsky, Schmidt, and Phillips 2003; Sera, Berge, and del Castillo Pintado 1994). Languages like Mandarin that speak of time vertically rather than horizontally shape native speakers' perceptions of time (Boroditsky 2001), and languages like Korean that emphasize social hierarchy reflect this value in the culture (McBrian 1978). These effects underscore the fact that by studying language, consumer researchers are studying thought and that language is conversely important because it shapes thought.

As a sign system, language has three aspects—semantic, pragmatic, and syntactic (Mick 1986; Morris 1994)—each of which provides a unique window into a slightly different part of consumer thought, interaction, or culture. Semantics concerns word meaning that is explicit in linguistic content (Frege 1892/1948), while pragmatics addresses the interaction between linguistic content and extra-linguistic factors like context or the relationship between speaker and hearer (Grice 1975). Syntax focuses on grammar, the order in which linguistic elements are presented (Chomsky 1957/2002). By understanding and appreciating these linguistic underpinnings, researchers can develop sounder operationalizations of the constructs and more insightful, novel hypotheses. We will discuss semantics, pragmatics, and syntax in turn as they are relevant to constructs in consumer research. Extensive treatment of these properties can be found in Mick (1986), although previous use of semiotics in consumer research has been focused primarily on objects and images as signs (Grayson and Shulman 2000; McQuarrie and Mick 1996; Mick 1986; Sherry, McGrath, and Levy 1993) rather than on language itself. In discussing construct identification, we link linguistic theory to topics of interest in consumer research—attention, processing, social influence, and group properties.

To more fully understand what kinds of problems might be fruitfully studied through text analysis, we detail four theoretical areas of consumer research that link with linguistic dimensions of semantics, pragmatics, and syntax.

Specifically, attention can be examined through semantics, processing through syntax, interpersonal dynamics through pragmatics, and group-level characteristics through semantics and higher-order combinations of these dimensions.

Attention

The first area where text analysis is potentially valuable to consumer research is in the study of attention. Consumer attention is important in the evaluation of products and experiences, self-awareness, attitude formation, and attribution, to name only a few domains. Language represents attention in two ways. When consumers are thinking of or attending to an issue, they tend to express it in words. Conversely, when consumers are exposed to a word, they are more likely to attend to it. In this way, researchers can measure what concepts constitute attention in a given context, study how attention changes over time, and evaluate how concepts are related to others in a semantic network. Through semantics, researchers can measure temporal, spatial, and self-focus, and in contrast to self-reports, text analysis can reveal patterns of attention or focus of which the speaker may not be conscious (Mehl 2006).

Semantics, the study of word meaning, links language with attention. From the perspective of semantics, a word carries meaning over multiple, different contexts, and humans store that information in memory. Word frequency, measuring how frequently a word occurs in text, is one way of measuring attention and then further mapping a semantic network. For example, based on the idea that people discuss the attributes that are top of mind when thinking of a particular car, Netzer et al. (2012) produce a positioning map of car attributes from internet message board data using supervised learning.

Researchers can infer the meaning of the word, what linguists and philosophers call the sense (Frege 1892/1948), through its repeated and systematic co-occurrence with a system of other words based on the linguistic principle of holism (Quine 1970). For example, if the word "Honda" is continually and repeatedly associated with "safety," one can infer that these concepts are related in consumers' minds such that Honda *means* safety to a significant number of consumers. In this way, one can determine the sense through the context of words around it (Frege 1892/1948; Quine 1970), and this holism is a critical property from a methodological perspective because it implies that one can derive the meaning of a word by studying its collocation with surrounding words (Neuman, Turney, and Cohen 2012; Pollach 2012). Due to the inherent holism of language, semantic analysis is a natural fit with spreading activation models of memory and association (Collins and Loftus 1975).

Text analysis can also measure implicit rather than explicit attention through semantics. The focus of consumer attention on the self as opposed to others (Spiller and Belogolova 2016) and temporal focus, such as psychological distance and construal (Snefjella and Kuperman 2015), are patterns that may not be recognized by consumers themselves, but can be made manifest through text analysis (Mehl 2006). For example, a well-known manipulation of self-construal is the "I" versus "we" sentence completion task (Gardner, Gabriel, and Lee 1999). Conversely, text analysis can help detect differences in self-construal using measures for these words.

Language represents the focus of consumer attention, but it can also direct consumer attention through semantic framing (Lakoff 2014; Lakoff and Ferguson 2015). For example, when Oil of Olay claims to "reverse the signs of aging" in the United States, but the same product claims to "reduce the signs of aging" in France, the frame activates different meaning systems—"reversing" being more associated with agency, and "reduction" being a more passive framing. As ample research in framing and memory has shown, consumers' associative networks can be activated when they see a particular word, which in turn may affect attitudes (Humphreys and Latour 2013; Lee and Labroo 2004; Valentino 1999), goal pursuit (Chartrand and Bargh 1996; Chartrand et al. 2008), and regulatory focus (Labroo and Lee 2006; Lee and Aaker 2004).

Language not only represents the cognitive components of attention, but also reflects the emotion consumers may feel in a particular context. Researchers have used automated text analysis to study the role of emotional language in the spread of viral content (Berger and Milkman 2012), response to national tragedies (Doré et al. 2015), and well-being (Settanni and Marengo 2015). As we will later discuss, researchers use a broad range of sentiment dictionaries to measure emotion and evaluate how consumer attitudes may change over time (Hopkins and King 2010), in certain contexts (Doré et al. 2015), or due to certain interpersonal groupings. Building on these approaches, researchers studying narrative have used the flow of emotional language (e.g., from more to less emotional words) to code different story arcs, such as comedy (positive to negative to positive) versus tragedy (negative to positive to negative) (Van Laer et al. 2017).

Processing

The structure of language, or syntax, can provide evidence of different kinds of processing for senders and can prompt different kinds of responses from readers. Syntax refers to the structure of phrases and sentences in text (Morris 1938). In any language, there are many ways to say something without loss of meaning, and these differences in grammatical construction can indicate differences in footing (Goffman 1979), complexity (Gibson 1998), or assertiveness (Kronrod, Grinstein, and Wathieu 2012). For example, saying "I bought the soap" rather than "The soap was bought" has different implications for attribution of agency, which could have consequences for satisfaction and attribution in the case of product success or failure.

Passive versus active voice is one key difference that can be measured through syntax, indicated by word order or by use of certain phrases or verbs. Active versus passive voice, for instance, affects persuasiveness of the message. Specifically, consumer-to-consumer word of mouth that is expressed in a passive voice may be more persuasive than active voice, particularly when language contains negative sentiment or requires extensive cognitive processing (Bradley and Meeds 2002; Carpenter and Henningsen 2011; see also Kronrod et al. 2012). When speakers use passive sentences, they shift the attention from the self to the task or event at hand (Senay, Usak, and Prokop 2015), and passive voice may therefore further signify lower power or a desire to elude responsibility. For instance, literature in accounting suggests that companies tend to report poor financial performance with passive voice (Clatworthy and Jones 2006).

Syntactic complexity (Gibson 1998; Wong, Ormiston, and Haselhuhn 2011) can influence the ease of processing. Exclusion words like "but" and "without" and conjunctions like "and" and "with" are used in more complex reasoning processes, and the frequency of these words can therefore be used to represent the depth of processing in consumer explanations, reviews, or thought listings. Sentence structures also lead to differences in recall and memorability. For example, Danescu-Niculescu-Mizil et al. (2012a) examine the syntactic characteristics using quotations from the IMDB (Internet Movie Database) website, finding that memorable quotations tend to have less common word sequence but common syntax.

Similarly, exclusions (without, but, or) are used to make distinctions (Tausczik and Pennebaker 2010), while conjunctions (and, also) are often used to tell a cohesive story (Graesser et al. 2004). Syntax can be further used to identify narrative versus non-narrative language (Jurafsky et al. 2009; Van Laer et al. 2017), which could be used to study transportation, a factor that has been shown to affect consumers' processing of advertising and media (Green and Brock 2002; Wang and Calder 2006). Categories like exclusion and conjunctive words also potentially provide clues as to decision strategy—those using exclusion (e.g., "or") might be using a disjunctive strategy, while those using words such as "and" may be using a conjunctive strategy. Certainty can be measured by occurrences of tentative language, passive voice, and hedging phrases such as "I think" or "perhaps."

In sum, theories of semantics help consumer researchers link language with thought such that they can use language to study different aspects of attention and emotion. Using theories of syntax, on the other hand, sheds light on the complexity of thought, as it looks for markers of structure that indicate the nature—complexity, order, or extent—of thinking, which has implications for processing and persuasion (Petty, Cacioppo, and Schumann 1983). Here, text analysis can be used to test predictions or hypotheses about attention and processing in real-world data, even if it cannot necessarily determine the cognitive mechanism underlying the process.

Interpersonal Dynamics

The study of interpersonal dynamics—including the role of status, power, and social influence in consumer life—can be meaningfully informed by linguistic theory and text analysis. Social interaction and influence are key parts of consumer life, but can be difficult to study in the lab. Consumers represent a lot about their relationships through language they use, and we can use this knowledge to understand more about consumer relationships on both the dyadic and group level.

The theory for linking language with social relationships comes from the field of pragmatics, which studies the interactions between extra-linguistic factors and language. Goffman (1959) and linguists following in the field of pragmatics such as Grice (1975) argue that people use linguistic and nonlinguistic signs to both signal and govern social relationships indicating status, formality, and agreement. By understanding when, how, and why people tend to use these markers, we can understand social distance (McTavish et al. 1995), power (Danescu-Niculescu-Mizil et al. 2012b), and influence (Gruenfeld and Wyer 1992). Pragmatics is used to study how these subtle yet pervasive cues structure human relationships and represent the dynamics of social interaction in turn. In fact, about 40% of language is composed of these functional markers (Zipf 1932).

One way to capture pragmatic elements is through the analyses of pronouns (e.g., "I," "me," "they") and demonstratives (i.e., "this," "these," "that," and "those"), words that are the same over multiple contexts but whose meaning is indexical or context dependent (Nunberg 1993). Pronoun use can be guided by different sets of contextual factors—such as intimacy, authority, or self-consciousness—and pragmatic analyses can be usefully applied to research that pertains to theories of self and interpersonal interaction, particularly through the measurement of pronouns (Packard, Moore, and McFerran 2018; Pennebaker 2011). Pronouns can detect the degree to which a speaker is lying (Newman et al. 2003), feeling negative emotion (Rude, Gortner, and Pennebaker 2004), and collaborating in a social group (Gonzales, Hancock, and Pennebaker 2010). Similarly, linguistic theories suggest that demonstratives ("this" or "that") mark solidarity and affective functions and can therefore be effective in "achieving camaraderie" and "establishing emotional closeness between speaker and addressee" (Lakoff 1973, 351; Potts and Schwarz 2010). Demonstratives have social effects, as shown in both qualitative and quantitative analyses of politicians' speeches (Acton and Potts 2014), and can be used for emphasis. For example, product and hotel reviews with demonstratives (e.g., "that" or "this" hotel) have more polarized ratings (Potts and Schwarz 2010).

Through pragmatics, speakers also signify differences in status and power. For example, people with high status use more first-person plural ("we") and ask fewer questions (Sexton and Helmreich 2000), while those with low status use more first-person singular like "I" (Hancock et al. 2010; Kacewicz et al. 2014). Language also varies systematically according to gender, and many argue this is due to socialization into differences in power reflected in tentative language, self-referencing, and the use of adverbs and other qualifiers (Herring 2000, 2003; Lakoff 1973).

Because norms are important in language, dyadic components prove to be an important part of the analysis when one is studying interpersonal interaction. For example, by incorporating dyadic interaction versus analyzing senders and receivers in isolation, Jurafsky et al. (2009) improve accuracy in their identification of flirtation, awkwardness, and friendliness from a range of 51–72% to 60–75%, with the prediction for women being the most improved.

Language is social, and pragmatics illustrate that not all words are meant to carry meaning. Phatic expressions, for example, are phrases in which the speaker's intention (i.e., what is meant) is not informative, but rather social or representational (Jakobson 1960; Malinowski 1972). For example, an expression like "How about those Cubs?" is an invitation to talk about the baseball team, not a sincere question. A tweet like "I can't believe Mariah Carey's album comes out on Monday!" is not intended to communicate information about a personal belief or even the release date, but is an exclamation of excitement (Marwick and boyd 2011). The phatic function can be informative in text analysis when one is interested simply in a word's ability to represent a category or concept to make it accessible or to form a bond with others, irrespective of semantic content. Here, the mention of the name is not used as a measure of semantics or meaning but rather of presence versus absence, and hence mere accessibility and, more broadly, cultural awareness (Arvidsson and Caliandro 2016).

Group- and Cultural-Level Characteristics

Lastly, and of particular interest to scholars of sociology and culture, language can be used to represent constructs at the group, cultural, and corpus level. At this level, group attention, differences amongst groups, the collective structure of meaning or agreement shared by groups, and changes in cultural products over time can be measured. Further, the ability of text analysis to span levels of analysis from individuals to dyadic, small group, and subcultural interaction is particularly apt for a multidisciplinary field like consumer research.

In sociocultural research, semantics is again key because words can represent patterns of cultural or group attention (Gamson and Modigliani 1989; McCombs and Shaw 1972; Schudson 1989). For example, Shor et al.'s (2015) study of gender representation in the news measures the frequency of female names in national newspapers to understand changes in the prominence of women in public discourse over time, and van de Rijt et al. (2013) similarly use name mentions to measure the length of fame. These are matters of attention, but this time public, collective attention rather than individual attention. Historical trends in books (Twenge, Campbell, and Gentile 2012) and song lyrics have also been discovered through text analysis. For example, in a study of all text uploaded to Google Books (4% of what has been published), Michel et al. (2011) find a shift from first-person plural pronouns (we) to first-person singular (I, me), and interpret this as reflecting a shift from collectivism to individualism (see also DeWall et al. 2011). Merging these approaches with an extra-linguistic DV, researchers can sometimes predict book sales, movie success (Mestyán, Yasseri, and Kertész 2013), and even stock market price using textual data from social media (Bollen, Mao, and Zeng 2011; De Choudhury et al. 2008; Gruhl et al. 2005).

Studies of framing and agenda setting naturally use semantic properties to study the social shaping of public opinion (Benford and Snow 2000; Gamson and Modigliani 1989; McCombs and Shaw 1972). For example, measuring the diffusion of terms such as "illegal immigrants" versus "undocumented workers" helps sociologists and sociolinguists understand the role of social movements in setting the agenda and shaping public discourse (Lakoff and Ferguson 2015). Humphreys and Thompson (2014), for example, use text analysis to understand how news narratives culturally resolve anxiety felt by consumers in the wake of a crisis such as an oil spill.

However, some caution is warranted when cultural products are used to represent the attitudes and emotions of a social group. Sociologists and critical theorists acknowledge a gap between cultural representation and social reality (Holt 2004; Jameson 1981). That is, the presence of a concept in public discourse does not mean that it directly reflects attitudes of all individuals in the group. In fact, many cultural products often depict fantasy or idealized representations that are necessarily far from reality (Jameson 1981). For this reason, as we will discuss, sampling and an awareness of the source's place in the larger media system are particularly important when one is conducting this kind of sociocultural analysis.

In addition to using words to measure patterns of individual and collective attention, researchers can put together textual elements to code patterns such as narrative (Van Laer et al. 2017), style matching (Ludwig et al. 2013; Ludwig et al. 2016), and linguistic cohesiveness (Chung and Pennebaker 2013). For example, people tend to use the same proportion of function words in cities where there is more even income distribution (Chung and Pennebaker 2013). In studies of consumption this could be used to study agreement in co-creation (Schau, Muniz, and Arnould 2009) and subcultures of consumption (Schouten and McAlexander 1995), and perhaps even to

predict fissioning of a group (Parmentier and Fischer 2015). One might speculate that homophilous groups will display more linguistic cohesiveness, and this may even affect other factors like strength of group identity, participation in the group, and satisfaction with group outcomes. Words associated with assent like "yes" and "I agree" can be used to measure group agreements (Tausczik and Pennebaker 2010). In this way, text analysis can be used for studying group interactional processes to predict quality of products and satisfaction with participation in peer-to-peer production (Mathwick, Wiertz, and De Ruyter 2008).

These four domains—attention, processing, interpersonal interaction, and group- and cultural-level properties—provide rich fodder for posing research questions that can be answered by studying language and by defining constructs through language. By linking linguistic theory pertaining to semantics, pragmatics, and syntax, researchers can formulate more novel, interesting, and theoretically rich research questions, and they can develop new angles on constructs key to understanding consumer thought, behavior, and culture. Per our roadmap in figure 1, we now proceed with data collection.

STAGE 3: COLLECT THE DATA

Once a researcher identifies the research question and related constructs, the next step is to collect the data. There are four steps to data collection: identifying, unitizing, preparing, and storing the data.

Identifying Data Sources

One virtue—and perhaps also a curse—of text analysis is that many data sources are available. Query through experiment, survey or interview, web scraping of internet content (Newsprosoft 2012; Pagescrape 2006; Velocityscape 2006; Wilson 2009), archival databases (ProQuest; Factiva), digital conversion of printed or spoken text, product websites like Amazon.com, expert user groups like Usenet, and internet subcultures or brand communities are all potential sources of consumer data. In addition to data collection through scraping, some platforms like Twitter offer access to a 10% random sample of the full "firehose" or APIs for collecting content according to keyword, hashtag, or user types.

Sampling is likely the most important consideration in the data collection stage. In the abstract, any dataset will consist of some sample from the population, and the sample can be biased in various but important ways. For example, Twitter users are younger and more urban than a representative sample of the US population (Duggan and Brenner 2013; Mislove et al. 2011). Generally, only public Facebook posts are available to study, and these users may have different characteristics than those who restrict posts to private. In principle, these concerns are no different from those present in traditional content analysis (see Krippendorff 2004 for a discussion of sampling procedures). However, sampling from internet data presents two unique issues.

First, filtering and promotion on websites can push some posts to the top of a list. On the one hand, this makes them more visible—and perhaps more influential—but on the other hand, they may be systematically different from typical posts. These selection biases are known to media scholars and political scientists, and researchers have a variety of ways for dealing with these kinds of enduring but inevitable biases. For example, Earl et al. (2004) draw from methods previously used in survey research to measure nonresponse bias through imputation (Rubin 1987). Researchers can sample evenly or randomly from categories on the site to obviate the problem of filtering on the website itself (Moore 2015).

The second issue is that keyword search can also introduce systematic bias. Reflecting on the issues introduced by semantic framing, researchers may miss important data because they have the wrong phrasing or keyword. For instance, Martin, Pfeffer, and Carley (2013) find that although the conceptual map for interviews and the text of newspaper articles about a given topic largely overlaps, article keywords—which are selected by the authors and tend to be more abstract—do not. There are at least two ways to correct for this. First, researchers can skip using keyword search altogether and sample using site architecture or time (Krippendorff 2004). Second, if keywords are necessary, researchers can first search multiple keywords and provide information about search numbers and criteria for selection, providing, if required, analyses of alternative keyword samples.

In addition to considering these two unique issues, researchers should employ a careful and defensible sampling strategy that is congruent with the research question. If, for example, categories such as different groups or time periods are under study, a random stratified sampling procedure should be considered. If the website offers categories, the researcher may want to stratify by these existing categories. A related issue is the importance of sound sampling when one is using cultural products or discourse to represent a group or culture of interest. For instance, DeWall et al. (2011) use top 10 songs and Humphreys (2010) uses newspapers with the largest circulation based on the inference that they will be the most widely shared cultural artifacts. As a general guide, being aware of the place and context of the discourse—including its senders, medium, and receivers (Hall 1980)—can resolve many issues (Golder 2000).

Additionally, controls may be available. For example, although conducting traditional content analysis, Moore (2015) compares nonfiction to fiction books when evaluating differences between utilitarian versus hedonic products rather than sampling from two different product categories. At the data collection stage, metadata can also be collected and later used to test alternative hypotheses introduced by selection bias (e.g., age).

Researchers need to account for sample size, and in the case of text analysis, there are two size issues to consider—the number of documents available, and the amount of content (e.g., number of words and sentences) in each document. Depending on the context as well as the desired statistical power, the requirements will differ. One method to avoid overfitting or biases due to small samples is the Laplace correction, which starts to stabilize a binary categorical probabilistic estimate when a sample size reaches 30 (Provost and Fawcett 2013, 74). As a starting rule of thumb, having at least 30 units is usually needed to make statistical inferences, especially since text data is non-normally distributed. However, this is an informal guideline, and depending on the detectable effect size, a power analysis would be required to determine the appropriate sample size (Corder and Foreman 2014). One should be mindful of the virtues of having a tightly controlled sample for making better inferences (Pauwels 2014). Big data is not always better data (Borgman 2015).

Regarding the number of words per unit, if one is using a measure that accounts for length of the unit (e.g., words as a percent of the total words), data can be noisy if units are short, as in tweets. Tirunillai and Tellis (2012), for example, discard online reviews that have fewer than 10 words. However, the number of words required per unit is largely dependent on the base frequency of dictionary or concept-related keywords, as we will later discuss. For psychometric properties like personality, Kern et al. (2016) suggest having at least 1,000 words per person and a sufficiently large dataset of users to cover variation in the construct. In their case, extraversion could be reliably predicted using a set of 4,000 Facebook users with only 500 words per person.

Preparing Data

After the data is identified and stored as a basic text document, it needs to be cleaned and segmented into units that will be analyzed. Spell-checking is often a necessary step because text analysis assumes correct, or at least consistent, spelling (Mehl and Gill 2008). Problematic characters such as wingdings, emoticons, and asterisks should be eliminated or replaced with characters that can be counted by the program (e.g., "smile"). On the other hand, if the research question pertains to fluency (Jurafsky et al. 2009) or users' linguistic habits, spelling mistakes and special characters should be kept and analyzed by custom programming. Data cleaning—which includes looking through the text for irrelevant text or markers—is important, as false inferences can be made if there is extraneous text in the document. For example, in a study of post-9/11 text, Back, Küfner, and Egloff (2010) initially reported a falsely inflated measurement of anger because they did not clean automatically generated messages ("critical" error) in their data (Back, Küfner, and Egloff 2011; Pury 2011).

Languages other than English can also pose unique challenges. Most natural-language processing tools and methodologies that exist today focus on English, a language from a low-context culture composed of individual words that, for the most part, have distinct semantics, specific grammatical functions, and clear markers for discrete units of meaning based on punctuation. However, grammar and other linguistic aspects can meaningfully affect unitization decisions. For example, analysis of character-based languages like Chinese requires first segmenting characters into word units and then dividing sentences into meaningful sequences before a researcher can perform part-of-speech tagging or further analyses (Fullwood 2015; Sun et al. 2015).

Unitizing and Storing the Data

After cleaning is complete, a clearly organized file structure should be created. One straightforward way to achieve this organization is to use one text file for each unit of analysis or "document." If, for example, the unit is one message board post, a text file can be created for each post. Data should be segregated into the smallest units of comparison because the output can always be aggregated upward. If, for example, the researcher is conducting semantic analysis of book reviews, a text file can be created for each review, and then aggregated to months or years to assess historical trends or star groupings to compare valence.

Two technological solutions are available to automate the process of unitizing. First, file separation can be automated via a custom program such as a Word macro, to cut text between identifiable character strings, paste it to a new document, and then save that document as a new file. Secondly, many text analysis programs are able to segregate data within the file by sentence, paragraph, or a common, unique string of text, or this can be done through code. If units are separated by one of these markers, separate text files will not be required. If, for example, each message board post is uniquely separated by a hard return, the researcher can use one file containing all of the text and then instruct the program to segment the data by paragraph. If the research is comparing by groups (say, by website), separate files should be maintained for each site and then segregated within the file by a hard return between each post.

Researchers might also want to consider using a database management system (DBMS) in addition to a file-based structure in order to support reproducible or future research, especially for big datasets and for information other than discourse, such as speakers' attributes (e.g., age, gender, and location). If the text document that stores all the raw data starts to exceed the processing machine's random access memory (RAM) available (some 32-bit text software caps the file size at 2GB), it may be challenging to work with directly, and depending on the way the content is structured in the text file, writing into a database may be necessary. For those experienced with coding, a variety of tools exist for extracting data from text files into a

database (e.g., the base, quanteda, and tm packages in R, or Python's natural-language processing toolkit, NLTK).

STAGE 4A: CHOOSE AN OPERATIONALIZATION APPROACH

Once data has been collected, prepared, and stored, the next decision is choosing the appropriate research approach for operationalizing the constructs. Next to determining if text analysis is appropriate, this is the most important impasse in the research. We discuss the pros and cons of different approaches and provide guidance as to when and how to choose amongst methods (figure 1). The web appendix presents a demonstration of dictionary-based and classification approaches, applied to understand word of mouth following a product launch.

If the construct is relatively clear (e.g., positive affect), one can use a dictionary or rule set to measure the construct. Standard, psychometrically tested dictionaries are available for measuring a variety of constructs (see table 1). Researchers may want to consult this list or one like it first to determine if the construct can be measured through an existing word list.

If the operationalization of the construct in words is not yet clear or the researcher wants to make a posteriori discoveries about operationalization, one should use a classification approach in which the researcher first identifies two or more categories of text and then analyzes recurring patterns of language within these sets. For example, if the researcher wants to study brand attachment by examining the texts produced by brand loyalists versus nonloyalists, but does not know exactly how they differ or wants to be open to discovery, classification would be appropriate. Here, the researcher preprocesses the documents and uses the resulting word frequency matrix as the IV, with loyalty as the existing DV. This leaves one open to surprises about which words may reflect loyalty, for example.

At the extreme, if the researcher does not know the categories at play but has some interesting text, one could use unsupervised learning to have the computer first detect groups within the text and then further characterize the differences in those groups through language patterns, somewhat like multidimensional scaling or factor analysis (Lee and Bradlow 2011). As shown in figure 1, selecting an approach depends on whether the constructs can be clearly defined a priori, and we discuss these decisions in detail.

Top-Down Approaches

Top-down approaches involve analyzing occurrences of words based on a dictionary or a set of rules. If the construct is relatively clear—or can be made clear through human analysis of the text (Corbin and Strauss 2008)—it makes sense to use a top-down approach. We discuss two types of top-down approaches: dictionary-based and rule-based approaches. In principle, a dictionary-based approach can be considered a type of rule-based approach; it is a set of rules for counting concepts based on the presence or absence of a particular word. We will treat the two approaches separately here, but thereafter will focus on dictionary-based methods, as they are most common. Methodologically, after operationalization, the results can be analyzed in the same way, although interpretation of the results may differ.

Dictionary-based Approaches. Although certainly some of the most basic methods available, dictionary-based approaches have remained one of the most enduring methods of text analysis and are still used as a common tool in the text analysis toolkit to produce new knowledge (Boyd and Pennebaker 2015; Eichstaedt et al. 2015; Shor et al. 2015; Snefjella and Kuperman 2015; see appendix).

Dictionary-based approaches have three advantages for research in consumer behavior that draws from psychological or sociological theories. First, they are easy to implement and comprehend, especially for researchers that have limited programming or coding experience. Second, combined with the fundamentals of linguistics, they allow intuitive operationalization of constructs and theories directly from sociology or psychology. Finally, the validation process of dictionary-based approaches is relatively straightforward for nonspecialists, and findings are relatively transparent to reviewers and readers.

For a dictionary-based analysis, researchers define and then calculate measurements that summarize the textual characteristics that represent the construct. For example, positive emotion can be captured by the frequency of words such as "happy," "excited," and "thrilled." The approach is best suited for semantic and pragmatic markers, and attention, interaction, and group properties have all been studied using this approach (see appendix).

One of the simplest measurements used in dictionary-based analysis is frequency, which is calculated based on the assumption that word order does not matter. These methods, also called the "bag of words" approach, assume that the meaning of a text depends only on word occurrence, as if the words are drawn randomly from a bag. While these methods are based on the strong assumption that word order is irrelevant, they can be powerful in many circumstances for marking patterns of attentional focus and mapping semantic networks. For pragmatics and syntax, counting frequency of markers in text can produce measurement of linguistic style or complexity in the document overall. Note that when a dictionary-based approach is used, tests will be conservative. That is, by predetermining a word list, one may not pick up all instances of what one wants to measure, but if meaningful patterns emerge, one can argue that there is an effect, despite the omissions.

A variety of computer programs can be used to conduct top-down automated text analysis and as auxiliaries for

TABLE 1

STANDARDIZED DICTIONARIES

Dictionary	Description	Example categories	Year	No. of words	No. of categories	Citation
Consumer						
Authenticity	A weighted word list used to measure authenticity developed from 35 participants	NA	2013	90	1	Kovács et al. 2013
Brand Personality	Word list to measure Aaker's (1997) five traits of brand personality	Sincerity, excitement, competence, sophistication, ruggedness	2006	833	5	Opoku et al. 2006
Environmental	Measures value systems related to discussion of the natural environment	Aesthetic, utilitarian, life support, moral	1995	612	4	Xu and Bengston 1997
General						
LIWC	A broad dictionary to measure parts of speech, but also psychological and social categories and processes	Sadness, negative emotion, overall affect, verb, and past focus	2015	6,400	13 categories, 68 subcategories	Pennebaker et al. 2015
Diction	Measures verbal tone, developed to study public communication such as political speeches	Certainty, activity, optimism, realism, commonality	2015	10,000	5 categories, 36 subcategories	Yadav et al. 2007; Zachary et al. 2011
Roget	Word list based on Roget's (1911) thesaurus, which includes six broad categories such as words related to space or matter	Space matter, intellectual facilities, voluntary powers, spiritual or moral powers	1999	100,685	6 categories, 1,042 subcategories	Roget 1911; Yarowsky 1992
WordNet	A large lexical dictionary that categories a variety of objects, feelings, and processes	Feeling, adverbs, possession, animal	2005	117,659 synsets; 147,278 unique words	44	Fellbaum 2005; Miller 1995
Harvard IV Psychological Dictionary	A large dictionary that captures a wide variety of semantic and pragmatic aspects of language as well as individually relevant and institutionally relevant words	Positive, negative, active, passive, pleasure, pain, virtue, vice, economy, legal, academic	1977	11,789	184	Stone 1966
Values	Often referred to as "Lasswell's dictionary," this word list contains words associated with deference and welfare; it is often combined with the Harvard IV	Power, respect, affiliation, wealth, well-being	1969/1998	10,628	8 categories, 68 subcategories	Lasswell and Namenwirth 1969
Psychological						
Concreteness	A weighted word list to measure concreteness based on 4,000 participants' ratings of the concreteness of many common words	NA	2013	37,058 words and 2,896 two-word expressions	8	Brysbaert et al. 2014

STANDARDIZED DICTIONARIES (*CONTINUED*)

Dictionary	Description	Example categories	Year	No. of words	No. of categories	Citation
Regressive Imagery Dictionary	Measures words associated with Martindale's (1975) primary or secondary processes	Need, sensation, abstract thought	1975, 1990	3,200	43	Martindale 1975
Communication Vagueness	Measures vagueness in verbal and written communication	Anaphore, probability and possibility, admission of error, ambiguous designation	1971	196	10	Hiller 1971
Body Type	Operationalizes Fisher and Cleveland's (1958) theory of self-boundary strength	Transparency, container, protective surface	2006	515	12	Wilson 2005
Sentiment						
Loughran and McDonald Sentiment	Word lists for textual analysis in financial analyses. Each word is labeled with sentiment categories, whether the word is an irregular verb, and number of syllables. Dictionary was constructed from all 10k documents from 1994 to 2014.	Negative, positive, uncertainty, litigious, modality	2014	85,132	9	Loughran and McDonald 2014
VADER	A dictionary- and rule-based tool for measuring sentiment	Negative, neutral, positive, compound	2015	9,000	4	Hutto and Gilbert 2014
ANEW	Affective ratings of everyday words based on surveys	Pleasure, pain, arousal, dominance	1999	1,033	4	Bradley and Lang 1999
Senti WordNet	A large sentiment dictionary based on the WordNet dictionary. Each mark is marked by a positive sentiment and negative sentiment score	Positive sentiment, negative sentiment, objective language	2010	117,659 synsets	3	Baccianella et al. 2010
Sociological						
Orders of Worth	Operationalizes Boltanski and Thevenot's (2006) orders of worth	Civic, market, industrial, spiritual, and domestic orders of worth	2012	344	5	van Bommel 2014; Weber 2005
Policy Position	Measures political position in public policy or other politically oriented texts	Liberal, conservative, pro-environment, con-environment	2000	415	19	Laver and Garry 2000

cleaning and analyzing the data. A word processing program is used to prepare the text files, an analysis program is needed to count the words, and a statistical package is often necessary to analyze the output. WordStat (Péladeau 2016), Linguistic Inquiry and Word Count (LIWC; Pennebaker, Francis, and Booth 2007), Diction (North, Iagerstrom, and Mitchell 1999), Yoshikoder (Lowe 2006), and Lexicoder (Daku, Young, and Soroka 2011) are all commonly used programs for dictionary-based analysis, although it is also possible with more advanced packages such as R and Python.

Rule-Based Approaches. Rule-based approaches are based on a set of criteria that indicate a particular operationalization. By defining and coding a priori rules according to keywords, sentence structures, punctuations, styles, readability, and other predetermined linguistic elements, a researcher can quantify unstructured texts. For example, a researcher interested in examining passive voice can write a program that, after tagging the part of speech (POS) of the text, counts the number of instances of a subject followed by an auxiliary and a past participle (e.g., "are used"). Van Laer et al. (2017) use a rule-based approach to classify sentences in terms of genre, using patterns in emotion words to assign a categorical variable that classifies a sentence as having rising, declining, comedic,

or tragic action. Rule-based approaches are also often used in readability measures (Bailey and Hahn 2001; Li 2008; Ghose, Ipeirotis, and Li 2012; Tan, Gabrilovich, and Pang 2012) to operationalize fluency of a message.

Bottom-Up Approaches

In contrast to top-down approaches, bottom-up approaches involve examining patterns in text first, and then proposing or interpreting more complex theoretical explanations and patterns. Bottom-up approaches are used in contexts where the explanatory construct or the operationalization of constructs is unclear. In some cases where word order is important (for example, in syntactic analyses), bottom-up approaches via unsupervised learning may also be helpful (Chambers and Jurafsky 2009). We discuss two common approaches used in text analysis: classification and topic discovery.

Classification.   In contrast to dictionary-based or rule-based approaches, where the researcher explicitly identifies the words or characteristics that represent the construct, classification approaches are used to deal with constructs that may be more latent in the text, meaning that the operationalization of a construct in text cannot be hypothesized a priori. Instead of manually classifying every document of interest, supervised classification allows researchers to group texts into predefined categories based on a subset or "training" set of the data. For example, Eliashberg, Hui, and Zhang (2007) classify movies based on their return on investment (ROI) and then, using the movie script, determine the most important factors in predicting a film's ROI, such as action genre, clear and early statement of the setting, and clear premise. After discovering these patterns, the researchers theorize as to why they occur.

There are two advantages to using classification. First, it reduces the amount of human coding required yet produces clear distinctions between texts. While dictionary-based approaches provide information related to magnitude, classification approaches provide information about type and likelihood of being of a type, and researchers can go a step further by understanding what words or patterns lead to being classified as a type. Second, the classification model itself can reveal insights or test hypotheses that may be otherwise buried in a large amount of data. Because classification methods do not define a word list a priori, latent elements, such as surprising combinations of words or patterns that may have been excluded in a top-down analysis, may be revealed.

Researchers use classification when they want to know where one text stands in respect to an existing set or when they want to uncover meaningful yet previously unknown patterns in the texts. In digital humanities research, for example, Plaisant et al. (2006) use a multinomial Naïve Bayes classifier to study word associations commonly associated with spirituality in the letters of Emily Dickenson.

They find that, not surprisingly, words such as "Father and Son" are correlated with religious metaphors, but they also uncover the word "little" as a predictor, a pattern previously unrecognized by experienced Dickenson scholars. This discovery then leads to further hypothesizing about the meaning of "little" and its relationship to spirituality in Dickenson's poems. In consumer research, in studying loyalists versus nonloyalists, researchers might find similarly surprising words such as "hope," "future," and "improvement," and these insights might provoke further investigation into self-brand attachment and goal orientation.

Topic Discovery.   If a researcher wants to examine the text data without a priori restrictions on words, rules, or categories, a topic discovery model is appropriate. Discovery models such as Latent Dirichlet Allocation (LDA) are analyses that recognize patterns within the data without predefined categories. In the context of text analysis, discovery models are used to identify whether certain words tend to occur together within a document, and such patterns or groupings are referred to as "topics." Given its original purpose, topic discovery is used primarily to examine semantics. Topic discovery models typically take a word frequency matrix and output groupings that identify co-occurrences of words, which can then predict the topic of a given text. They can be helpful when researchers want to have an overview of the text beyond simple categorization or to identify patterns.

Topic discovery models are especially useful in situations where annotating even a subset of the texts has a high cost due to complexity, time or resource constraints, or a lack of distinct, a priori groupings. In these cases, a researcher might want a systematic, computational approach that can automatically discover groups of words that tend to occur together. For example, Mankad et al. (2016) use unsupervised learning and find that hotel reviews mainly consist of five topics, which, according to the groups of words for each topic, they label as "amenities," "location," "transactions," "value," and "experience." Once topics have been identified, one can go on to study their relationship with each other and with other variables such as rating.

STAGE 4B: EXECUTE OPERATIONALIZATION

After an approach is chosen, the next step is to make some analytical choices within the approach pertaining to either dictionary or algorithm type. These again depend on the clarity of the construct, the existing methods for measuring it, and the researcher's propensity for theoretically driven versus data-driven results. Within top-down approaches, decisions entail choosing one or more standardized dictionaries versus creating a custom dictionary or rule set. Within bottom-up methods of classification and topic modeling, analytic decisions entail choosing a

technique that fits suitable assumptions and the clarity of output one seeks (e.g., mutually exclusive vs. fuzzy or overlapping categories).

Dictionary- and Rule-based Approaches

Standardized Dictionary. If one chooses a dictionary-based approach, the next question is whether to use a standardized dictionary or to create one. Dictionaries exist for a wide range of constructs in psychology, and less so, sociology (table 1). Sentiment, for example, has been measured using many dictionaries: Linguistic Inquiry Word Count (LIWC), ANEW (Affective Norms for English Words), the General Inquirer (GI), SentiWordNet, WordNet-Affect, and VADER (Valence Aware Dictionary for Sentiment Reasoning). While some dictionaries, like LIWC, are based on existing psychometrically tested scales such as PANAS (Positive and Negative Affect Schedule), others, such as ANEW (Bradley and Lang 1999), have been created based on previous classification applied to offline and/or online texts and human scoring of sentences (Nielsen 2011). VADER (Hutto and Gilbert 2014) includes the word banks of established tools like LIWC, ANEW, and GI, as well as special characters such as emoticons and cultural acronyms (e.g., LOL), which makes it advantageous for social media jargon. Additionally, VADER's model incorporates syntax and punctuation rules, and is validated with human coding, making its sentence prediction 55–96% accurate, which is on par with Stanford Sentiment Treebank, a method that incorporates a more complex computational algorithm (Hutto and Gilbert 2014). However, a dictionary like LIWC bases affect measurement on underlying psychological scales, which may provide tighter construct validity. If one is measuring a construct such as sentiment that has multiple standard dictionaries, it is advisable to test the results using two or more measures, as one might employ multiple operationalizations.

In addition to standardized dictionaries for measuring sentiment, there are a range of psychometrically tested dictionaries for concepts like construal level (Snefjella and Kuperman 2015), cognitive processes, tense, and social processes (LIWC; Pennebaker, Francis, and Booth 2001), pleasure, pain, arousal, motivation (Harvard IV Psychological Dictionary; Dunphy, Bullard, and Crossing 1974) primary versus secondary cognitive processes (Regressive Imagery Dictionary; Martindale 1975), and power (Lasswell's Value Dictionary; Lasswell and Leites 1949; Namenwirth and Weber 1987; table 1). These dictionaries have been validated with a large and varied number of text corpora, and because operationalization does not change, standard dictionaries enable comparison across research, enhancing concurrent validity amongst studies.

For this reason, if a standard dictionary exists, researchers should use it if at all possible to enhance the replicability of their study. If they wish to create a new dictionary for an existing construct, researchers should run and compare the new dictionary to any existing dictionary for the construct, just as one would with a newly developed scale (Churchill 1979).

Dictionary Creation. In some cases, a standard dictionary may not be available to measure the construct, or semantic analyses may require greater precision to measure culturally or socially specific categories. For example, Eritimur and Coskuner-Balli (2015) use a custom dictionary to measure the presence of different institutional logics in the market emergence of yoga in the United States.

To create a dictionary, researchers first develop a word list, but here there are several potential approaches (figure 1). For theoretical dictionary development, one can develop the word list from previous operationalization of the construct, scales, and by querying experts. For example, Pennebaker et al. (2007) use the Positive and Negative Affect Schedule, or PANAS (Watson, Clark, and Tellegen 1988), to develop dictionaries for anger, anxiety, and sadness. To ensure construct validity, however, it is crucial to examine how these constructs are expressed in the text during post-measurement validation.

If empirically guided, a dictionary is created from reading and coding the text. The researcher selects a random subsample from the corpus in order to create categories using the inductive method (Katz 2001). If the data is skewed (i.e., if there are naturally more entries from one category than others), a stratified random sampling should be used to ensure that categories will evenly apply to the corpus. Generally sampling 10–20% of the entire corpus for qualitative dictionary development is sufficient (Humphreys 2010). Alternatively, one can determine the size of the subsample as the dictionary is developed using a saturation procedure (Weber 2005). To do this, one codes 10 entries at a time until a new set of 10 entries yields no new information. Corbin and Strauss (1990) discuss methods of grounded theory development that can be applied here for dictionary creation.

If the approach to dictionary development is purely inductive, researchers can build the word list from a concordance of all words in the text, listed according to frequency (Chung and Pennebaker 2013). In this way, the researcher acts as a sorter, grouping words into common categories, a task that would be performed by the computer in bottom-up analysis. One advantage of this approach is that it ensures that researchers do not miss words that occur in the text that might be associated with the construct.

After dictionary categories are developed, the researcher should expand the category lists to include relevant synonyms, word stems, and tenses. The dictionary should avoid homonyms (e.g., river "bank" vs. money "bank") and other words where the reference is unclear (see Rothwell 2007 for a guide). Weber (2005) suggests using the semiotic square to check for completeness of concepts included. For example, if "wealth" is included in the dictionary,

perhaps "poverty" should also be included. Because measurement is taken from words, one must attend to and remove words that are too general and thus produce false positives. For example, "pretty" can be used as a positive adjective (e.g., "pretty shirt") or for emphasis (e.g., "that was pretty awful"). Alternatively, a rule-based approach can be used to work around critical words that cause false positives in a dictionary. It is then important that rules for inclusion and exclusion be reported in the final analysis.

Languages other than English can also produce challenges in dictionary creation. If researchers are developing a dictionary in a language or vernacular where there are several spellings or terms for one concept, for example, they should include those in the dictionary. Arabic, for example, is characterized by three different vernaculars within the same language—Classical Arabic in religious texts, Modern Standard Arabic, and a regional dialect (Farghaly and Shaalan 2009). Accordingly, researchers should be mindful of these multivalences, particularly when developing dictionaries in other languages or even other vernaculars within English (e.g., internet discourse).

Dictionary Validation. After one has developed a preliminary dictionary, its construct validity should be assessed. Does each word accurately represent the construct? Researchers have used a variety of validation techniques. One method of dictionary validation is to use human coders to check and refine the dictionary (Pennebaker et al. 2007). To do this, the dictionary is circulated to three research assistants who vote to either include or exclude a word from the category and note words they believe should be included in the category. Words are included or excluded based on the following criteria: (1) if two of the three coders vote to include it, the word is included; (2) if two of the three coders vote to exclude it, the word is excluded; (3) if two of the three coders offer a word that should be included, it is added to the dictionary.

A second option for dictionary validation is to have participants play a more involved role in validating the dictionary through survey-based instruments. Kovács et al. (2013), for example, develop a dictionary by first generating a list of potential synonyms and antonyms to their focal construct, authenticity, and then conducting a survey in which they have participants choose the word closest to authenticity. They then use this data to rank words from most to least synonymous, assigning each a score from 0 to 1. This allows dictionary words to be weighted as more or less part of the construct rather than either-or indicators. Another option for creating and validating a weighted dictionary is to regress textual elements on a dependent variable like star rating to get predictors of, say, sentiment. This approach would be similar to the bottom-up approach of classification (Tirunillai and Tellis 2012).

Post-Measurement Validation. After finalizing the dictionary and conducting a preliminary analysis, the researcher should examine the results to ensure that operationalization of the construct in words occurred as expected, and this can be an iterative process with dictionary creation. The first method of post-measurement validation uses comparison with a human coder. To do this, select a subsample of the data, usually about 20 entries per concept, and compare the computer coding with ratings by a human coder. Calculate Krippendorff's alpha to assess agreement between the human coder and the computer coder (Krippendorff 2007, 2010). Traditional criteria for reliability apply; Krippendorff's alpha for each category should be no lower than 70%, and the researcher should calculate Krippendorff's alpha for each category and as an average for all categories (Weber 2005). Packard and Berger (2016) conduct this type of validation, finding 94% agreement between computer- and human-coded reviews.

The advantages of using a human coder for post-measurement validation are that results can be compared to other traditional content analyses and that this method separates validation from the researcher. However, there are several disadvantages. First, it is highly variable because it depends on the expertise and attentiveness of one or more human coders. Secondly, traditional measures of intercoder reliability, such as Krippendorf's alpha, were intended to address the criteria of replicability (Hughes and Garrett 1990; Krippendorff 2004), the chance of getting the same results if the analysis were to be repeated. Because replicability is not an issue with automated text analysis—the use of a specific word list ensures that repeated analyses will have exactly the same results—measures of intercoder agreement are largely irrelevant. While it is important to check the output for construct validity, the transparency of the analysis means that traditional measures of agreement are not always required or helpful. Lastly, and perhaps most importantly, the human coder will likely be more sensitive to subtleties in the text, and may therefore overcode categories or may miscode due to unintentional mistakes or biases. After all, one reason the researcher selects automated text analysis is to capture aspects humans cannot detect.

The second alternative for validation is to perform a check oneself or to have an expert perform a check on categories using a saturation procedure. Preliminarily run the dictionary on the text and examine 10 instances at a time, checking for agreement with the construct or theme of interest and noting omissions and false positives (Weber 2005). The dictionary can then be iteratively revised to reduce false positives and include observed omissions. A hit rate, the percent of accurately coded categories, and a false hit rate, the percent of inaccurately coded categories, can be calculated and reported. Thresholds for acceptability using this method of validation are a hit rate of at least 80% and a false hit rate of less than 10% (Wade, Porac, and Pollock 1997; Weber 2005).

As with any quantitative research technique, there will always be some level of measurement error. Undoubtedly, words will be occasionally miscategorized; such is the

nature of language. The goal of validation is to ensure that measurement error is low enough relative to the systematic variation so that the researcher can make reliable conclusions from the data.

Classification

After choosing a bottom-up approach, the researcher must next determine whether a priori classifications are available. If the answer is yes, the researcher can use classification or supervised-learning methods. Here we discuss Naïve Bayes classification, logistic regressions, and classification trees because of the ease of their implementation and interpretability. We will also discuss neural networks and k-nearest neighbor classifications, which, as we will describe below, are more suited for predicting categories of new texts than for deriving theories or revealing insights.

Naïve Bayes (NB) predicts the probability of a text belonging to a category given its attributes with Bayes rules and the "naïve" assumption that each attribute in the word frequency matrix is independent from the others. NB has been applied in various fields, such as marketing, information, and computer science. Examining whether online chatter affects a firm's stock market performance, Tirunilai and Tellis (2012) use NB to classify a user-generated review as positive or negative. Using star rating to classify reviews as positive or negative a priori, they investigate language that is associated with these positive or negative reviews. Since there is no complex algorithm involved, NB is very efficient with respect to computational costs. However, in situations where words are highly correlated with each other, NB might not be suitable.

Logistic regression is another classification method, and similar to NB, it takes a word frequency or characteristic matrix as input. It is especially useful when the dataset is large and when the assumption of conditional independence of word occurrences cannot be taken for granted. For example, Thelwall et al. (2010) use it to predict positive and negative sentiment strength for short online comments from MySpace.

In contrast, a classification tree is based on the concept of examining word combinations in a piecewise fashion. Namely, it first splits the text with the word or category that can distinguish the most variation, and then within each resulting "leaf," it splits the subsets of the data again with another parameter. This inductive process iterates until the model achieves the acceptable error rate that is set by the researcher beforehand (see later sections for guidelines on model validation). Because of its conceptual simplicity, the classification tree is also a "white box" that allows for easy interpretation.

There are other classification methods, such as neural networks (NN) or k-nearest neighbor (k-NN), that are more suitable for prediction purposes, but less so for interpreting insights. However, these types of "black box" methods can

be considered if the researcher requires prediction (e.g., positive or negative sentiments), but not enumeration of patterns underlying the prediction.

In classifying a training set, researchers apply some explicit meaning based on the words contained within the unit. Classification is therefore used primarily to study semantics, while applications of classificatory, bottom-up techniques for analyzing pragmatics and syntax remain a nascent area (Kuncoro et al. 2016). However, more recent research has demonstrated the utility of these approaches to study social factors, such as detecting politeness (Danescu-Niculescu-Mizil et al. 2013), predicting lying or deceit (Markowitz and Hancock 2015; Newman et al. 2003), and analyzing sentiment in a way that accounts for sentence structures (Socher et al. 2013).

Topic Discovery

If there is no a priori classification available, topic models, implemented via unsupervised learning, are more suitable. Predefined dictionaries are not necessary since unsupervised methods inherently calculate the probabilities of a text being similar to another text and group them into topics. Some methods, such as Latent Dirichlet Allocation (LDA), assume that a document can present multiple topics and estimate the conditional probabilities of topics that are unobserved (i.e., latent), given the observed words in documents. This can be useful if the researcher prefers "fuzzy" categories to the strict classification of the supervised learning approach. Other methods, such as k-means clustering, use the concept of distance to group documents that are most similar to each other based on co-occurrence of words or other types of linguistic characteristics. We will discuss the two methods, LDA and k-means, in more detail.

LDA is one of the most common topic discovery models (Blei 2012), and it can be implemented in software packages or libraries such as R and Python. LDA (Blei, Ng, and Jordan 2003) is a modeling technique that identifies whether and why a document is similar to another document and specifies the words underlying the unobserved groupings (e.g., topics). Its algorithm is based on the assumptions that (1) there is a mixture of topics in a document, and this mixture follows a Dirichlet distribution; (2) words in the document follow a multinomial distribution; and (3) the total of N words in a given document follows a Poisson distribution. Based on these assumptions, the LDA algorithm estimates the most likely underlying topic structure by comparing observed word groupings with these probabilistic distributions and then outputs K groupings of words that are related to each other. Since a document can belong to multiple topics, and a word can be used to express multiple topics, the resulting groupings may have overlapping words. LDA reveals the underlying topics of a given set of documents, and the meanings are interpreted by the researcher.

Yet sometimes this approach will produce groupings that don't semantically hang together or groupings that are too obviously repetitive. To resolve this issue, researchers will sometimes use word embedding, a technique for reducing and organizing a word matrix based on similarities and dissimilarities in semantics, syntax, and part of speech that are taken from previously observed data. The categories taken from large amounts of previously observed data can be more comprehensive as well as more granular than the categories specified by existing dictionaries such as LIWC's sentiments. Further, in addition to training embeddings from the existing dataset, a researcher can download pretrained layers such as word2vec by Google (Mikolov et al. 2013) or GloVe by Stanford University (Pennington, Socher, and Manning 2014). In these cases, the researcher skips the training stage and jumps directly to text analysis. These packages provide a pretrained embedding structure as well as functions for a researcher to customize the categories depending on the research context in question. As a supplement to LDA, once word embeddings have been learned, they can potentially be reused.

In consumer research, LDA is useful in determining ambiguous constructs such as consumer perceptions, particularly if the corpus is large. For example, Tirunillai and Tellis (2014) analyze 350,000 consumer reviews with LDA to group the contents into product dimensions that reviewers care about. In the context of mobile phones, for example, they find that the dimensions are "portability," "signal receptivity," "instability," "exhaustible," "discomfort," and "secondary features." LDA allows Tirunillai and Tellis (2014) to simultaneously derive product dimensions and review valence by labeling the grouped words as "positive" or "negative" topics. The canonical LDA algorithm is a bag-of-word model, and one potential area of future research is to relax the LDA assumptions. For instance, Büschken and Allenby (2016) extend the canonical LDA algorithm by identifying not just words, but whole sentences, that belong to the same topic. If applying this method to study consumer behavior, one could use topic discovery to identify tensions in a brand community or social network, which could lead to further theorizations about underlying discourse or logic present in a debate.

In cases where a researcher wants to consider linguistic elements beyond word occurrences, conceptually simpler approaches, such as clustering, may be more appropriate (Lee and Bradlow 2011). In addition to word occurrences, a researcher can first code the presence of syntax or pragmatic characteristics of interest, and then perform analyses such as k-means clustering, which is a method that identifies "clusters" of documents by minimizing the distance between a document and its neighbors in the same cluster. After obtaining the clustering results, the researcher can then profile each cluster, examine its most distinctive characteristics, and further apply theory to explain topic groupings and look for further patterns through abduction (Peirce 1957).

Labeling the topics is the last, and perhaps the most critical, step in topic discovery. It is important to note that, despite the increasing availability of big data and machine learning algorithms and tools, the results obtained from these types of discovery models are simply sets of words or documents grouped together to indicate that they constitute a topic. However, the researcher can determine what that topic is or represents only by applying theory and context-specific knowledge or expertise when interpreting the results.

STAGE 5: INTERPRET AND ANALYZE THE RESULTS

After operationalizing the constructs through text analysis, the researcher next must analyze and interpret the results. There are two distinct phases of analysis: the text analysis itself and the statistical analysis, already familiar to many researchers. In this section we discuss three common ways of incorporating the results of text analysis into research design: (1) comparison between groups, (2) correlation between textual elements, and (3) prediction of variables outside the text.

Comparison

Comparison is the most common research design amongst articles that use text analysis in the social sciences, and is particularly compatible with top-down, dictionary-based techniques (see appendix). Comparing between groups or over time is useful for answering research questions that relate directly to the theoretical construct of interest. That is, some set of text is used to represent the construct and then comparisons are made to assess statistically meaningful differences between texts. For example, Kacewicz et al. 2014 compare the speech of high-power versus low-power individuals (manipulated rather than measured), finding that high-power people use fewer personal pronouns (e.g., "I"). Investigating the impact of religiosity, Ritter et al. (2013) compare Christians to atheists, finding that Christians express more positive emotion words than atheists, which the authors attribute to a different thinking style. Holoein and Fiske (2014) compare the word use of people who were told to be warm versus people who were told to appear competent, finding a compensatory relationship whereby people wanting to appear warm also select words that reflect low competence.

Other studies use message type rather than source to represent the construct and thus as the unit of comparison. Bazarova (2012), for example, compare public to private Facebook messages to understand differences in the style of public and private communication. One can also compare observed frequency in a dataset to a large corpus such as the standard Corpus of American English or the Brown Corpus (Conrad 2002; Neuman et al. 2012; Pollach 2012; Wood and Kroger 2000). In this way, researchers can

assess if frequencies are higher than "typical" usage in English, not just relative to other conditions in their own text.

Comparisons over space and time are also common and valuable for assessing how a construct can change in magnitude based on some external variable. In contrast to group comparisons, these studies tend to focus on semantic aspects over pragmatic or syntactic ones. For example, Doré et al. (2015) trace changes in emotional language following the Sandy Hook Elementary School shooting, finding that sadness words decreased with spatial and physical distance while anxiety words increased according to distance. Comparing different periods of regulation, Humphreys (2010) shows how discourse changes over time as the consumer practice of casino gambling becomes legitimate.

Issues with Comparison. Because word frequency matrices can contain a lot of zeros (i.e., each document may contain only a few instances of a keyword), researchers should use caution when making comparisons between word frequencies of different groups. In particular, the lack of normally distributed data violates the assumptions for tests like analysis of variance (ANOVA), and simple comparative methods like Pearson's Chi-squared tests and z-score tests might yield biased results. Alternative comparative measures, such as likelihood methods or linear regression, may be more appropriate (Dunning 1993). Another alternative is using nonparametric tests that do not rely on the normality assumption. For instance, the nonparametric equivalent of a one-way ANOVA is the Kruskal-Wallis test, whose test statistic is based on ordered rankings rather than means.

Many text analysis algorithms take word counts or the term-frequency (*tf*) matrix as an input, but because word frequencies do not follow a normal distribution (Zipf 1932), many researchers transform the data prior to statistical analysis. Transformation is especially helpful in comparison because often the goal is to compare ordinally, as opposed to numerically (e.g., document A contains more pronouns than document B). Typically, a Box-Cox transformation, which is a general class of power-based transformation function ($x' = \frac{x^\lambda - 1}{\lambda}$), can reduce a variable's distribution skewness. One easy transformation is arbitrarily setting $\lambda = 0$, which is equivalent to taking the logarithmic of the variable for any x greater than 0 (Box and Cox 1964; Osborne 2010).

To further account for the overall frequency of words in the text, researchers will also often transform the word- or term-frequency matrix into a normalized measure such as the percent of words in the unit (Kern et al. 2016; Pennebaker and King 1999) or a Term-Frequency Inverse Document Frequency (tf-idf; Spärck Jones 1972).

Common words may not be very diagnostic, and so researchers will often want to weight rare words more heavily because they are more predictive (Netzer et al 2012).

To address this, tf-idf accounts for the total frequency of a word in the dataset. Specifically, one definition of tf-idf is:

$$
\begin{aligned}
\textit{tf-idf} \; &(\text{word } w, \text{ document } d, \text{ corpus } D) \\
&= \; [1 + \log(\text{number of occurrences of } w \text{ in } d)] \\
&\times \; \log\left(\frac{\text{total number of documents in } D}{\text{number of documents containing } w}\right),
\end{aligned}
$$

If the number of occurrences of a word is 0, then it is set to 0 (Manning and Hinrich 1999). After the tf-idf is calculated for all keywords in every document, the resulting matrix is used as measures of (weighted) frequency for statistical comparison. This method gives an extra boost to rare word occurrences in an otherwise sparse matrix, and thus statistical comparisons can leverage the additional variability for testing the hypotheses.

Tf-idf is useful for correcting for infrequently occurring words, but there are other methods one may want to use to compare differences in frequently occurring words like function words. For example, Monroe et al. (2009) compare references to gender in speeches from Republican and Democratic candidates. In this context, eliminating all function words may lead to misleading results because a function word like "she" or "her" can be indicative of the Democratic Party's policies on women's rights. Specifically, Monroe et al. (2009) first observe distribution of word occurrences in their entire dataset of Senate speeches that address a wide range of topics to form prior benchmarks for how often a word should occur. Then they combine the log-odds-ratio method with that prior belief to examine the differences between Republican and Democrat speeches on the topic of abortion. Such methods that incorporate priors account for frequently occurring words and thus complement tf-idf.

Correlation

Co-occurrence helps scholars see patterns of association that may not be otherwise observed, either between textual elements or between textual elements and nontextual elements such as survey responses or ratings. Researchers often report correlations between textual elements as a preliminary analysis before further comparison either between groups or over time, in order to gain a sense of discriminant and convergent validity (Humphreys 2010; Markowitz and Hancock 2015). For example, to study lying Markowitz and Hancock (2015) create an "obfuscation index" composed of multiple measures with notable correlations, including jargon, abstraction (positively indexed), positive emotion, and readability (negatively indexed), and find that these combinations of linguistic markers are indicators of deception. In this way, correlations are used to build higher-order measures or factors such as linguistic style (Ludwig et al. 2013; Pennebaker and King 1999).

When considered on two or more dimensions, co-occurrence between words takes on new meaning as relationships between textual elements can be mapped. These kinds of spatial approaches can include network analysis, where researchers use measures like centrality to understand the importance of some concepts in linking a conceptual network (Carley 1997) or to spot structural holes where concepts may be needed to link concepts. For example, Netzer et al. (2012) study associative networks for different brands using message board discussion of cars, based on co-occurrence of car brands within a particular post. Studying correlation between textual elements gives researchers insights about semantic relationships that may co-occur and thus be linked in personal or cultural associations. For example, Neuman et al. (2012) use similarity scores to understand metaphorical associations for the words "sweet" and "dark," as they're related to other, more abstract words and concepts (e.g., sweetness and darkness).

In addition to using correlations between textual elements in research design, researchers will often look at correlations between linguistic and nonlinguistic elements on the way to forming predictions. For example, Brockmeyer et al. (2015) study correlations between pronoun use and patient-reported depression and anxiety, finding that depressed patients use more self-focused language when recalling a negative memory. Ireland et al. (2011) observe correlations between linguistic style and romantic attachment and use this as support for the hypothesis of linguistic style matching.

Issues with Correlation. Well-designed correlation analysis requires a series of robustness checks—that is, performing similar or related analyses using alternative methodologies to ensure results from these latter analyses are congruent with the initial findings. Some of the robustness checks include: (1) using a random subset of the data and repeating the analyses; (2) examining or checking for any possible effects due to heterogeneity; and (3) running additional correlation analyses using various types of similarity measures, such as lift, Jaccard distance, cosine distance, tf-idf co-occurrence, Pearson correlation (Netzer et al. 2012), Euclidean distance, Manhattan distance, and edit distance.

Generally speaking, results should be congruent regardless of which subset of data or distance measure is used. However, some distance measures may inherently be more appropriate than others, depending on the underlying assumption the distance represents. Netzer et al. (2012) provide an instructive example of robustness check within the context of mapping automobile brands to product attributes. Using and comparing multiple methods of similarity, they find that Jaccard, cosine, and tf-idf co-occurrence distance measures yield similar results as their original findings. Pearson correlation, on the other hand, yields less meaningful results due to sparseness of the data.

It is also important to note that the interpretation of co-occurrence as a measure of correlation can be biased toward frequent words—words that occur in more documents may inherently co-occur more frequently than other words. Therefore, methods such as *z*-scores or simple co-occurrence counts may be inappropriate, and extant literature suggests normalizing the occurrence counts by calculating lift or point-wise mutual information (PMI) using *relative* frequencies of occurrences, for example (Netzer et al. 2012). However, one criticism against mutual information type measurements is that, particularly in smaller datasets, they may overcorrect for word frequency and thus bias the analysis toward rare words. In these cases, a log-likelihood test provides a balance "between saliency and frequency" (Pollach 2012, 8).

Another issue that arises in correlation, particularly with a large number of categories, is that there may be many correlations, but not all of them theoretically meaningful. To account for the presence of multiple significant correlations, some of which may be spurious or due to chance, Kern et al. (2016) suggest calculating Bonferroni-corrected *p*-values and including only correlations with small *p*-values (e.g., $p < .001$).

Prediction

Prediction using text analysis usually goes beyond correlational analysis in that it takes other nontextual variables into account. For example, Ludwig et al. (2016) use elements of email text (e.g., flattery and linguistic style matching) to predict deception, where they have a group of known deceptions. In examining Kiva loan proposals, Genevsky and Knutson (2015) operationalize affect with percentages of positive and negative words, and they then incorporate these two variables as independent variables in a linear regression to predict lending rates.

In other contexts, researchers may have access to readily available data—such as ratings, likes, or some other variable—to corroborate their prediction and incorporate this information into the model. Textual characteristics can also be used as predictors of other content elements, particularly in answering empirical questions. Using a dataset from a clothing store, Anderson and Simester (2014) identify a set of product reviews that are written by 12,000 "users" who did not seem to have purchased the products. Using logistic and ordinary least square (OLS) models, they then find that textual characteristics—such as word count, average word length, occurrences of exclamation marks, and customer ratings—predict whether a review is "fake," controlling for other factors.

Issues with Prediction. When using textual variables for prediction, researchers should recognize endogeneity due to selection bias, omitted variable bias, and heterogeneity issues. As previously discussed, samples of text can be biased in various ways and therefore may not generalize if the sample differs markedly from the population.

When analyzing observational data such as tweets or review posts, a researcher almost certainly encounters

selection bias because the text is not generated by a random sample of the population nor is a random set of utterances. For instance, reviewers may decide to post their negative opinions online when they see positive reviews that go against their perspective (Sun 2012). A researcher who wants to discover consumer sentiment toward a smartphone from CNET, for example, may need to consider when and how the reviews are generated in the first place. Are they posted right after a product has been launched or months afterward? Are they written when the brand is undergoing a scandal? By identifying possible external shocks that may cause a consumer to act in a certain way, a researcher can compare the behaviors before and after the shock to examine the effects. Combining these contexts with methodological frameworks such as regression discontinuity (i.e., comparing responses right before and after the treatment) or matching (i.e., a method that creates a pseudo-control group using observational data) may reduce some of the biases. Future research using controlled lab experiments or field studies to predict hypothesized changes in written text can further bolster confidence in using text to measure certain constructs.

Overfitting is another common problem with prediction in text analysis. Because there are often many independent variables (i.e., words or categories) relative to the number of observations, results can be overly specific to the data or training set. Kern et al. (2016) have suggestions for addressing the issue by reducing the number of predictors, such as applying principle component analysis (PCA) to the predictors and k-fold cross-validation on hold-out sample(s). In general, developing and reducing a model on a training set and then testing on a sufficient hold-out sample can increase generalizability and reduce problems with overfitting.

STAGE 6: VALIDATE THE RESULTS

Automated text analysis, like any method, has strengths and weaknesses. While lab studies may be able to achieve internal validity in that they can control for a host of alternative factors in a lab setting, they are, of course, somewhat weaker on external validity (Cook et al. 1979). Automated text analysis, on the other hand, lends researchers claim to external validity, and particularly ecological validity, as the data is observed in organically produced consumer texts (Mogilner et al. 2011). Beyond this, other types of validity, such as construct, convergent, concurrent, discriminant, and predictive validity, are addressable through a variety of techniques (McKenny, Short, and Payne 2013).

Types of Validity

Construct validity can be addressed in a number of ways. Because text analysis is relatively new for measuring social and psychological constructs, it is important to be sure that constructs are operationalized in ways consistent with their conceptual meaning and previous theorization. Through dictionary development, one can have experts or human coders evaluate word lists for their construct validity in pretests. More elaborately, pretests of the dictionary using a larger sample or survey could also help ensure construct validity (Kovács et al. 2013). Using an iterative approach, one can equally pull coded instances from the data to ensure that operationalization through the dictionary words makes sense (Weber 2005 suggests using a saturation procedure to reach 80% accuracy in a training set). In classification, the selection or coding of training data is another place to address construct validity. For example, does the text pulled and attributed to brand loyalists *actually* represent loyalty? One can use external validation or human ratings for calibration. For example, Jurafsky et al. (2009) use human ratings of awkwardness, flirtation, and assertiveness to classify the training data.

One can assess convergent validity, the degree to which measures of the construct correlate to each other, by measuring the construct using different linguistic aspects, and by comparing linguistic analysis with measurements external to text. For example, one could measure construal level using a semantics-based dictionary (Snefjella and Kuperman 2015) or through pragmatic markers available through LIWC.

Beyond convergent validity in any particular study, concurrent validity, the ability to draw inferences over many studies, is improved when researchers use standard, previously used, and thoroughly tested dictionaries. This allows researchers to draw conclusions across studies, knowing that constructs have been measured with the same list of words. Bottom-up, classificatory analysis does not afford researchers the same assurance.

Discriminant and convergent validity are relatively easy for researchers to assess after conducting the text analysis through factor analysis. Here, bottom-up methods of classification and similarity are invaluable for measuring the likeness of groups of text and placing this likeness on more than one dimension. Researchers can then observe consistent patterns of difference to ascertain discriminant validity.

Predictive validity, the ability of the constructs measured via text to predict other constructs in the nomological net, is perhaps one of the most important types of validity to establish the usefulness of text analysis in social science. Studies have found relationships between language and stock price (Tirunillai and Tellis 2012), personality type (Pennebaker and King 1999), and box office success (Eliashberg, Hui, and Zhang 2007). A hold-out sample can be helpful in investigating whether the hypothesized model is generalizable to new data. There are a variety of ways to do hold-out sampling and validation, such as k-fold cross-validation, which splits the dataset into k parts and, for each iteration, uses $k - 1$ subset for training and one subset for testing. The process is iterated until each part has been used as a test subset. For instance, Jurafsky et al. (2014) hold out 20% of the sample for testing; van Laer et al.

(2017) save about 10% for testing. The accuracy rate should be greater than the no-information rate, and it should also be relatively consistent across all iterations.

Further validation depends on the particular method of statistical analysis used. For comparison, multiple operationalizations using different measures (linguistic and nonlinguistic) can help support the results. If using correlation, a researcher can accomplish triangulation by looking at correlations in words or categories that one would expect (Humphreys 2010; Pennebaker and King 1999).

Lastly, text analysis conducted with many categories on a large dataset can potentially yield many possible correlations and many statistically significant comparisons, some of which may not be actionable, and some of which may even be spurious. For research designs that use hypothesis testing, Bonferroni-corrected p-values can be used where there is the possibility of spurious correlation from testing multiple hypotheses (Kern et al 2016). However, some argue that the test is too stringent and offer other alternatives (Benjamini and Hochberg 1995). While text analysis provides ample information, giving meaning to it requires theory. Without theory, findings can be too broad and relatively unexplained, and sheer computing power is never a replacement for asking the right type of research questions, framing the right type of constructs, collecting and merging the right types of dataset(s), or choosing the right operationalization approach (Crawford, Miltner, and Gray 2014).

As we demonstrate, choosing the most appropriate method depends on those high-level thought processes that cannot be performed by computers or artificial intelligence alone. Designing the right type of top-down research requires carefully operationalized constructs and implementation, and analyzing results and predictions from bottom-up learning requires interpretation, both of which rely on theory and expertise in a knowledge domain. In short, while datasets and computing power are as abundant as ever, one still does not gain insight without a clear theoretical framework. Only through repeated testing through multiple operationalizations can we separate the spurious from systematic findings.

Ethics

Because the data for text analysis often comes from the internet rather than traditional print sources, the ethics of collecting, storing, analyzing, and presenting findings from such data are critical issues to consider and yet still somewhat in flux. Although it can seem depersonalized, text data comes from humans, and per the Belmont Report (1978), researchers should minimize harm to these individuals, being mindful of the respect, beneficence, and justice to those who provide underlying data for research. The Association of Internet Researchers provides an overview of ethical considerations for conducting internet research that usefully apply to most text analyses (Markham and Buchanan 2012), and Townsend and Wallace (2016, 8)

provide succinct guidelines for the ethical decisions one faces in conducting social media research. In general, these guidelines advocate for a case-based approach informed by the context in which the data exists and is collected, analyzed, and circulated. While few organizations offer straightforward rules, three issues deserve particular consideration when one is conducting textual analysis.

The first ethical question is one of access and jurisdiction—do researchers have legitimate jurisdiction to collect the textual data of interest? Here, the primary concern is the boundary between public and private information. Given the criteria laid out by the Common Rule, internet discourse is usually deemed to be public because individuals cannot "reasonabl[y] expect that no observation or recording is taking place" (HHS 2013, 5). Summarizing a report from the US Department of Health and Human Services (1979), Hayden (2013) says, "The guidelines also suggest that, in general, information on the Internet should be considered public, and thus not subject to IRB review— even if people falsely assume that the information is anonymous" (411). However, technical and socially constructed barriers, such as password-protected groups, gatekeepers, and interpersonal understandings of trust, also define participants' expectations of privacy (Nissenbaum 2009; Townsend and Wallace 2016; Whiteman 2012). Guidelines therefore also suggest that "investigators should note expressed norms or requests in a virtual space, which—although not technically binding—still ought to be taken into consideration" (HHS 2013, 5). The boundary between public and private information is not always clear, and researcher judgments should be made based on what participants themselves would expect based on a principle of "contextual integrity" (Marx 2001; Nissenbaum 2009).

A second question concerns the control, storage, and presentation of textual data. There's an important distinction between the ethics of the treatment of aggregated versus individualized data, with aggregated data being less sensitive than individualized data and individualized data of vulnerable populations being the most sensitive. Identification such as screen name should be deleted for those who are not public figures (Townsend and Wallace 2016). Even if anonymized, individualized textual data is often searchable, and therefore attributable to the original source, even if the name or other identifying information is removed. For this reason, when presenting individual excerpts, some researchers will choose to remove words or paraphrase in order to reduce searchability (Townsend and Wallace 2016).

The durability of textual data also means that comments made by consumers long ago may be damaging if discovered and circulated. With large enough sample sizes, aggregated data is less vulnerable to identifiability, although in cases where there are extreme outliers or sparse data, individuals may be identifiable, and researchers should take care when presenting these cases. Further, even in large datasets, researchers have demonstrated that anonymized data can be

made identifiable through metadata such as location, time, or other variables (Narayanan and Shmatikov 2008, 2010), so data security is of primary importance. Current guidelines suggest that text data be treated as any human subject data, under password protection, and, when possible, de-identified (Markham and Buchanan 2012).

A final matter is legitimate ownership or control of the data. Many terms of service (ToS) agreements prohibit scraping of data, and while some platforms offer APIs to facilitate paid or metered access, others do not. While the legalities of ownership are relatively clear, albeit legally untested, some communications researchers have argued that control over this data constitutes an unreasonable obstacle to research that is in the public interest (Waters 2011). Contemporary legal configurations also mean that consumers who produce the data may not themselves have access to or control of it. For this reason, it's important that researchers make efforts to share research with the population from which the data was used. For many researchers, this also means clearing permission with the service provider to use data, although requirements for this vary depending on field and journal.

CONCLUSION

Summary

In this article, we contribute to the growing need for methodologies that incorporate analysis of textual data in consumer research. Although there is a considerable set of phenomena that cannot be measured by text, computers can be used to discover, measure, and represent patterns of language that elude both researchers and consumers. Our goal has been to provide insight into the relationship between language and consumer thought, interaction, and culture to inspire research that uses language; and then to further provide a roadmap for executing text-based research to help researchers make decisions inherent to any text analysis.

We propose that automated text analysis can be a valuable methodology for filling the gap between the data available and theories commonly used in consumer research. And while previous work in psychology has developed a variety of standalone approaches advocated by particular researchers, it has not incorporated linguistic theory into constructs that are meaningful to consumer researchers, provided guidance about what method to use, discussed when a particular method might be appropriate, or tailored approaches to the unique perspective of consumer researchers.

We therefore make three contributions to consumer research in light of previous treatments of text analysis in the social sciences. First, we provide an overview of methods available, guidance for choosing an approach, and a roadmap for making analytical decisions. Second, we address in depth the methodological issues unique to studies of text, such as sampling considerations when one is using

internet data, approaches to dictionary development and validation, and statistical issues such as dealing with sparse data and normalizing textual data. Lastly, and more broadly, we hope to contribute to growing incorporation of communications and linguistic theory into consumer research and provide a tool for linking the multiple levels of analysis (individual, group, and cultural dimensions) so crucial to a multidisciplinary field such as consumer research.

The focus of this article has been written text. However, visual data (such as video) contains tonality and emotion of words as well as visual information such as facial expression and gesture in addition to textual content. Video data therefore have the potential to provide additional measurement of constructs and may require more sophisticated techniques and analyses (Cristani et al. 2013; Jain and Li 2011; Lewinski 2015). Although this topic is beyond the scope of our methodology, some procedures we discuss here may still apply. For example, if studying gesture, one will need to define a "dictionary" of gestures that represent a construct or to code all gestures and then group them into meaningful categories.

Future Directions

Overall, where will text analysis lead us? Consumers are more surrounded by and produce more textual communication than ever before. Developing methodological approaches that can incorporate textual data into traditional analyses can help consumer researchers understand the influence of language and to use language as a measure of consumer thought, interaction, and culture. Equally, methods of social science are changing as capacities for collecting, storing, and analyzing both textual and nontextual (e.g., behavioral) data expand. In a competitive landscape of academic disciplines, it makes sense that consumer research should incorporate some of this data, as it is useful to exploring the theories and questions driving the field. As we have argued, questions of attention, processing, interaction, groups, and culture can all be informed by text analysis.

Text analysis also supports the interdisciplinary nature of consumer research in two ways. First, it further points to theories in linguistics and communications that inform questions common to marketplace interaction. There is a growing acknowledgment of the role of language and communication theory in social influence and consumer life (Barasch and Berger 2014; Moore 2015; Packard and Berger 2016, van Laer et al. 2017), and text analysis methodologically supports the inclusion of these types of theories into consumer research.

Second, text analysis can link different levels of analysis, which is particularly important in a field that incorporates different theoretical traditions. Studying the self, for example, requires understanding not only individual psychological processes like cognitive dissonance (Festinger 1962), but also an awareness of how the self interacts with objects (Belk 1988; Kleine, Kleine, and Allen 1995) and

cultural meanings surrounding these objects, meanings that change due to shifts in culture (Luedicke, Thompson, and Giesler 2010). Similarly, studying social influence can be informed by psychological theories of power (Ng and Bradac 1993) and processing (Petty et al. 1983), but also normative communication and influence (McCracken 1986; McQuarrie, Miller, and Phillips 2013). As Deighton (2007) argues, consumer research is distinct from core disciplines in that, although it is theoretically based, it is also problem-oriented in the sense that researchers are interested in the middle ground between abstract theory and instrumental implication. Text analysis can lend ecological validity to rigorously conducted lab studies that illustrate a causal effect or can point toward new discoveries that can be further investigated.

Our intention is not to suggest that text analysis can stand alone, but rather that it is a valuable companion for producing insight in an increasingly textual world. Social science research methods are changing, and while labora-tory experiments are considered the gold standard for making causal inference, other forms of data can be used to make discoveries and show the import of theoretical relationships in consumer life. If we limit ourselves to data that can be gathered and analyzed in a lab, we are discarding zettabytes of data in today's digital age, including millions of messages transmitted per minute online (Marr 2015). And yet automated text analysis is a relatively new method in the social sciences. It will likely change over the next decade due to the advances in computational linguistics, the increasing availability of digital text, and interest amongst marketing professionals. Our aim has been to provide a link between linguistic elements in text and constructs in consumer behavior and guidance for executing research using textual data. In a world where consumer texts grow more numerous each day, automated text analysis, if done correctly, can yield valuable insights about consumer life.

APPENDIX

Research question	Text	Linguistic aspect	Source
Dictionary-based—Comparison			
How does temporal and spatial distance affect emotions after a tragic event?	Twitter	Semantic	Doré et al. 2015
How do power and affiliation vary by political ideology?	Transcripts (chatrooms, State of the Union), news websites	Semantic	Fetterman et al. 2015
What explains representational gender bias in the media?	Newspapers	Phatic	Shor et al. 2015
How does personal pronoun use in firm-customer interactions impact customer attitude?	Transcripts	Pragmatic	Packard, Moore, and McFerran 2016
Why don't major crises like oil spills provoke broad changes in public discourse concerning the systemic risks inherent to a carbon-dependent economy?	Newspaper articles	Semantic	Humphreys and Thompson 2014
Do people modify warmth to appear competent (and vice versa) when doing impression management?	Emails	Semantic	Holoien and Fiske 2013
Does social hierarchy affect language use? In what ways?	Emails	Pragmatic	Kacewicz et al. 2014
Do Christians and atheists vary in their language use?	Twitter	Semantic	Ritter et al. 2013
How does someone's communication style change based on private versus public communication?	Facebook wall posts and private messages	Semantic, pragmatic	Bazarova 2012
How do letters to shareholders differ in a period of economic growth versus recession?	Letters to shareholders	Semantic	Pollach 2012
Are people with the same linguistic style more likely to form a romantic relationship?	Transcripts, instant messages	Stylistic, pragmatic	Ireland et al. 2011
How does happiness change throughout the lifecycle?	Personal blogs	Semantic	Mogilner et al. 2011
Dictionary-based—Correlation			
Do depressed patients use more self-focused language?	Written essays	Semantic	Brockmeyer et al. 2015

PRIOR TEXTUAL ANALYSES (*CONTINUED*)

Research question	Text	Linguistic aspect	Source
Is construal level (physical, temporal, and social) represented in language and what are the mathematical properties of concrete versus abstract?	Twitter	Semantic, pragmatic	Snefjella and Kuperman 2015
What neural affective mechanisms prompt altruism?	Loan requests (Kiva)	Semantic	Genevsky and Knutson 2015
Does audience size affect the way people share?	Lab experiment (fake emails; conversations; elaborations)	Semantic, pragmatic	Barasch and Berger 2014
Is fame stable or fleeting?	Newspapers, blogs, television	Semantic, phatic	Rijt et al. 2013
How does disease advocacy shape medical policy?	Transcripts	Semantic	Best 2012
How did Alan Greenspan's language change during periods of economic expansion versus decline?	Transcripts	Semantic	Abe 2011
How are markets created?	Newspaper articles	Semantic	Humphreys 2010
Does language use reflect personality?	Written essays, journal articles	Stylistic, pragmatic	Pennebaker and King 1999
Dictionary-based—Prediction			
How do cultural artifacts affect social and political change?	Internet searches, newspapers, Twitter	Semantic	Vasi et al. 2015
Does affect and linguistic style matching in online reviews lead to purchase?	Book reviews	Semantic; stylistic	Ludwig et al. 2013
How does cognitive complexity relate to the relationship between CEO facial structure and performance? (Secondary RQ)	Letters to shareholders	Semantic, syntax	Wong et al. 2011
What makes an article go viral?	NYT articles and top emailed lists	Semantic	Berger and Milkman 2012
Does syntax influence ad recognition or comprehension?	Lab experiment	Syntax	Bradley and Meeds 2002
Classification—Comparison			
Did Shakespeare write this rediscovered play?	Plays	Stylistic, pragmatic	Boyd and Pennebaker 2015
How do consumers frame positive and negative reviews online?	Yelp restaurant reviews	Semantic	Jurafsky et al. 2014
What's the process through which a concrete word becomes an abstract concept ("hypostatic abstraction")?	Text from website	Semantic, pragmatic	Neuman et al. 2012
Classification—Correlation			
How to design a ranking system by mining UGC?	Online reviews	Semantic, stylistic	Ghose, Ipeirotis, and Li 2012
How does sentiment change over time during an election?	Blogs	Semantic	Hopkins and King 2010
What word associations exist in the poems of Emily Dickenson that have gone unnoticed by experts?	Poems	Semantic	Plaisant et al. 2006
Classification—Prediction			
How do customers react to actively managed online communities?	Field experiments as well as observational online forum data	Semantic	Homburg, Ehm, and Artz 2015
Which movies generate the most ROI?	IMDB movie summaries	Semantic	Eliashberg, Hui, and Zhang (2007)
Can a firm predict a customer's churn using their complaints?	Call center emails	Semantic	K. Coussement and D. Van den Poel 2008
Can UGC predict stock performance?	Reviews from Amazon.com, Epinions.com, and Yahoo! Shopping; six markets; and 15 firms	Semantic	Tirunillai and Tellis 2012
Topic Discovery - All			
Can a brand track its perceived quality using online reviews?	Same as Tirunillai and Tellis 2012	Semantic	Tirunillai and Tellis 2014
What do consumers talk about in hotel reviews?	Hotel review data	Semantic	Mankad et al. 2016
What associative networks do consumers have for different brands? / How to construct a perceptual map using web-available text data?	Forums	Semantic	Netzer et al. 2012
How to analyze market structure using text mining of UGC?	Online reviews; pros and cons lists (explicitly stated by the reviewers)	Semantic	Lee and Bradlow 2011
Does analyzing text data by sentence rather than word improve prediction of sentiment?	Online reviews from Expedia and we8there.com	Semantic/syntax	Büschken and Allenby 2016

REFERENCES

Abe, Jo Ann A. (2011), "Changes in Alan Greenspan's Language Use Across the Economic Cycle: A Text Analysis of His Testimonies and Speeches," *Journal of Language and Social Psychology*, 30 (2), 212–23.

Acton, Eric K. and Christopher Potts (2014), "That Straight Talk: Sarah Palin and the Sociolinguistics of Demonstratives," *Journal of Sociolinguistics*, 18 (1), 3–31.

Alexa, Melina (1997), *Computer Assisted Text Analysis Methodology in the Social Sciences*, Mannheim, Germany: ZUMA.

Allen, Douglas E (2002), "Toward a Theory of Consumer Choice as Sociohistorically Shaped Practical Experience: The Fits-Like-a-Glove (Flag) Framework," *Journal of Consumer Research*, 28 (4), 515–32.

Anderson, Eric T. and Duncan I. Simester (2014), "Reviews without a Purchase: Low Ratings, Loyal Customers, and Deception," *Journal of Marketing Research*, 51 (3), 249–69.

Arnold, Stephen J. and Eileen Fischer (1994), "Hermeneutics and Consumer Research," *Journal of Consumer Research*, 21 (1), 55–70.

Arsel, Zeynep and Jonathan Bean (2013), "Taste Regimes and Market-Mediated Practice," *Journal of Consumer Research*, 39 (5), 899–917.

Arsel, Zeynep and Craig J. Thompson (2011), "Demythologizing Consumption Practices: How Consumers Protect Their Field-Dependent Identity Investments from Devaluing Marketplace Myths," *Journal of Consumer Research*, 37 (5), 791–806.

Arvidsson, Adam and Alessandro Caliandro (2016), "Brand Public," *Journal of Consumer Research*, 42 (5), 727–48.

Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani (2010), "Sentiwordnet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," in *LREC*, Vol. 10, 2200–04.

Back, Mitja D., Albrecht C. P. Küfner, and Boris Egloff (2010), "The Emotional Timeline of September 11, 2001," *Psychological Science*, 21 (10), 1417–19.

—— (2011), "'Automatic or the People?' Anger on September 11, 2001, and Lessons Learned for the Analysis of Large Digital Data Sets," *Psychological Science*, 22 (6), 837–38.

Barasch, Alixandra and Jonah Berger (2014), "Broadcasting and Narrowcasting: How Audience Size Affects What People Share," *Journal of Marketing Research*, 51 (3), 286–99.

Bazarova, Natalya N. (2012), "Public Intimacy: Disclosure Interpretation and Social Judgments on Facebook," *Journal of Communication*, 62 (5), 815–32.

Bailey, Todd M. and Ulrike Hahn (2001), "Determinants of Wordlikeness: Phonotactics or Lexical Neighborhoods?" *Journal of Memory and Language*, 44 (4), 568–91.

Belk, Russell W. (1988), "Property, Persons, and Extended Sense of Self," in *Proceedings of the Division of Consumer Psychology, American Psychological Association 1987 Annual Convention*, ed. L. F. Alwitt, Washington, DC: American Psychological Association, 28–33.

Belk, Russell W., John F. Sherry, and Melanie Wallendorf (1988), "A Naturalistic Inquiry into Buyer and Seller Behavior at a Swap Meet," *Journal of Consumer Research*, 14 (4), 449–70.

Bell, Gordon, Tony Hey, and Alex Szalay (2009), "Beyond the Data Deluge," *Science*, 323 (5919), 1297–98.

Benford, Robert D. and David A. Snow (2000), "Framing Processes and Social Movements: An Overview and Assessment," *Annual Review of Sociology*, 26 (1), 611–40.

Benjamini, Yoav and Yosef Hochberg (1995), "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, 57 (1), 289–300.

Berger, Jonah and Katherine L Milkman (2012), "What Makes Online Content Viral?" *Journal of Marketing Research*, 49 (2), 192–205.

Best, Rachel K. (2012), "Disease Politics and Medical Research Funding: Three Ways Advocacy Shapes Policy," *American Sociological Review*, 77 (5), 780–803.

Blei, David M. (2012), "Probabilistic Topic Models," *Communications of the ACM*, 55 (4), 77–84.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003), "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3 (January), 993–1022.

Bollen, Johan, Huina Mao, and Xiaojun Zeng (2011), "Twitter Mood Predicts the Stock Market," *Journal of Computational Science*, 2 (1), 1–8.

Borgman, Christine L. (2015), *Big Data, Little Data, No Data: Scholarship in the Networked World*, Cambridge, MA: MIT Press.

Boroditsky, Lera (2001), "Does Language Shape Thought?: Mandarin and English Speakers' Conceptions of Time," *Cognitive Psychology*, 43 (1), 1–22.

Boroditsky, Lera, Lauren A. Schmidt, and Webb Phillips (2003), "Sex, Syntax, and Semantics," in *Language in Mind: Advances in the Study of Language and Thought*, ed. Susan Goldin-Meadow and Dedre Gentner, Cambridge, MA: MIT Press, 61–79.

Box, G. E. P. and D. R. Cox (1964), "An Analysis of Transformations," *Journal of the Royal Statistical Society. Series B (Methodological)*, 26 (2), 211–52.

Boyd, Ryan L. and James W. Pennebaker (2015), "Did Shakespeare Write *Double Falsehood*? Identifying Individuals by Creating Psychological Signatures with Text Analysis," *Psychological Science*, 26 (5), 570–82.

Bradley, Margaret M. and Peter J. Lang (1999), "Affective Norms for English Words (Anew): Instruction Manual and Affective Ratings," Technical report C-1, the center for research in psychophysiology, University of Florida.

Bradley, Samuel D. and Robert Meeds (2002), "Surface–Structure Transformations and Advertising Slogans: The Case for Moderate Syntactic Complexity," *Psychology & Marketing*, 19 (7–8), 595–619.

Brier, Alan and Bruno Hopp (2011), "Computer Assisted Text Analysis in the Social Sciences," *Quality & Quantity*, 45 (1), 103–28.

Brockmeyer, Timo, Johannes Zimmermann, Dominika Kulessa, Martin Hautzinger, Hinrich Bents, Hans-Christoph Friederich, Wolfgang Herzog, and Matthias Backenstrass (2015), "Me, Myself, and I: Self-Referent Word Use as an Indicator of Self-Focused Attention in Relation to Depression and Anxiety," *Frontiers in Psychology*, 6, 1564.

Brysbaert, Marc, Amy Beth Warriner, and Victor Kuperman (2014), "Concreteness Ratings for 40 Thousand Generally Known English Word Lemmas," *Behavior research methods*, 46 (3), 904–11.

Büschken, Joachim and Greg M. Allenby (2016), "Sentence-based Text Analysis for Customer Reviews," *Marketing Science*, 35 (6), 953–75.

Carley, Kathleen (1997), "Network Text Analysis: The Network Position of Concepts," in *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts*, ed. Carl W. Roberts, Mahwah, NJ: Lawrence Erlbaum.

Carpenter, Christopher J. and David Dryden Henningsen (2011), "The Effects of Passive Verb-Constructed Arguments on Persuasion," *Communication Research Reports*, 28 (1), 52–61.

Chambers, Nathanael and Dan Jurafsky (2009), "Unsupervised Learning of Narrative Schemas and Their Participants," in *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, ed. Keh-Yih Su, Stroudsburg, PA: Association for Computational Linguistics, 602–10.

Chartrand, Tanya L. and John A. Bargh (1996), "Automatic Activation of Impression Formation and Memorization Goals: Nonconscious Goal Priming Reproduces Effects of Explicit Task Instructions," *Journal of Personality and Social Psychology*, 71 (3), 464.

Chartrand, Tanya L., Joel Huber, Baba Shiv, and Robin J. Tanner (2008), "Nonconscious Goals and Consumer Choice," *Journal of Consumer Research*, 35 (2), 189–201.

Chomsky, Noam (1957/2002), *Syntactic Structures*, Berlin: Walter de Gruyter.

Chung, Cindy K. and J. W. Pennebaker (2013), "Counting Little Words in Big Data," in *Social Cognition and Communication*, ed Joseph P. Forgas, Orsolya Vincze, and János László, New York: Psychology Press, 25.

Churchill, Gilbert A. (1979), "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research*, 16 (1), 64–73.

Clatworthy, Mark A. and Michael J. Jones (2006), "Differential Patterns of Textual Characteristics and Company Performance in the Chairman's Statement," *Accounting, Auditing & Accountability Journal*, 19 (4), 493–511.

Collins, Allan M. and Elizabeth F. Loftus (1975), "A Spreading-Activation Theory of Semantic Processing," *Psychological Review*, 82 (6), 407.

Conrad, Susan (2002), "4. Corpus Linguistic Approaches for Discourse Analysis," *Annual Review of Applied Linguistics*, 22 (March), 75–95.

Cook, Thomas D., Donald Thomas Campbell, and Arles Day (1979), *Quasi-Experimentation: Design & Analysis Issues for Field Settings*, Vol. 351, Boston: Houghton Mifflin.

Corbin, Juliet M. and Anselm L. Strauss (2008), *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, Los Angeles: Sage.

Corder, Gregory W. and Dale I. Foreman (2014), *Nonparametric Statistics: A Step-by-Step Approach, Hoboken*, NJ: John Wiley & Sons.

Coussement, Kristof and Dirk Van den Poel (2008), "Improving Customer Complaint Management by Automatic Email Classification Using Linguistic Style Features as Predictors," *Decision Support Systems*, 44 (4), 870–82.

Crawford, Kate, Miltner Kate, and Mary L. Gray (2014), "Critiquing Big Data: Politics, Ethics, Epistemology—Special Section Introduction," *International Journal of Communication*, 8, 1663–72.

Cristani, Marco, R. Raghavendra, Alessio Del Bue, and Vittorio Murino (2013), "Human Behavior Analysis in Video Surveillance: A Social Signal Processing Perspective," *Neurocomputing*, 100, 86–97.

Daku, Mark, Lori Young, and Stuart Soroka (2011), "Lexicoder, Version 3.0" (software), www.lexicoder.com.

Danescu-Niculescu-Mizil, Cristian, Justin Cheng, Jon Kleinberg, and Lillian Lee (2012a), "You Had Me at Hello: How Phrasing Affects Memorability," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, ed. Haizhou Li, Stroudsburg, PA: Association for Computational Linguistics, 892–901.

Danescu-Niculescu-Mizil, Cristian, Lillian Lee, Bo Pang, and Jon Kleinberg (2012b), "Echoes of Power: Language Effects and Power Differences in Social Interaction," in *Proceedings of the 21st International Conference on World Wide Web*, ed. Michael Rabinovich and Steffan Staab, New York: ACM, 699–708.

Danescu-Niculescu-Mizil, Cristian, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts (2013), "A Computational Approach to Politeness with Application to Social Factors," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, ed. Hinrich Schuetze, Pascale Fung, and Massimo Poesio, Stroudsburg, PA: Association for Computational Linguistics, 250–59.

De Choudhury, Munmun, Hari Sundaram, Ajita John, and Dorée Duncan Seligmann (2008), "Can Blog Communication Dynamics Be Correlated with Stock Market Activity?" in *Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia*, ed. Peter Brusilovsky, New York: ACM, 55–60.

Deighton, John (2007), "From the Editor: the Territory of Consumer Research: Walking the Fences," *Journal of Consumer Research*, 34 (3), 279–82.

DeWall, C. Nathan, Richard S. Pond Jr., W. Keith Campbell, and Jean M. Twenge (2011), "Tuning in to Psychological Change: Linguistic Markers of Psychological Traits and Emotions over Time in Popular US Song Lyrics," *Psychology of Aesthetics, Creativity, and the Arts*, 5 (3), 200.

Doré, Bruce, Leonard Ort, Ofir Braverman, and Kevin N. Ochsner (2015), "Sadness Shifts to Anxiety over Time and Distance from the National Tragedy in Newtown, Connecticut," *Psychological Science*, 26 (4), 363–73.

Duggan, Maeve and Joanna Brenner (2013), "The Demographics of Social Media Users—2012," http://www.pewinternet.org/2013/02/14/the-demographics-of-social-media-users-2012/.

Dunning, Ted (1993), "Accurate Methods for the Statistics of Surprise and Coincidence," *Computational Linguistics*, 19 (1), 61–74.

Dunphy, D. M., C. G. Bullard, and E. E. M. Crossing (1974), "Validation of the General Inquirer Harvard IV Dictionary," paper presented at the 1974 Pisa Conference on Content Analysis, Pisa, Italy.

Earl, Jennifer, Andrew Martin, John D. McCarthy, and Sarah A. Soule (2004), "The Use of Newspaper Data in the Study of Collective Action," *Annual Review of Sociology*, 30, 65–80.

Eichstaedt, Johannes C., Hansen Andrew Schwartz, Margaret L. Kern, Gregory Park, Darwin R. Labarthe, Raina M. Merchant, Sneha Jha, Megha Agrawal, Lukasz A. Dziurzynski, and Maarten Sap (2015), "Psychological

Language on Twitter Predicts County-Level Heart Disease Mortality," *Psychological Science*, 26 (2), 159–69.

Eliashberg, Jehoshua, Sam K. Hui, and Z. John Zhang (2007), "From Story Line to Box Office: A New Approach for Green-Lighting Movie Scripts," *Management Science*, 53 (6), 881–93.

Ertimur, Burçak and Gokcen Coskuner-Balli (2015), "Navigating the Institutional Logics of Markets: Implications for Strategic Brand Management," *Journal of Marketing*, 79 (2), 40–61.

Farghaly, Ali and Khaled Shaalan (2009), "Arabic Natural Language Processing: Challenges and Solutions," *ACM Transactions on Asian Language Information Processing (TALIP)*, 8 (4), 14.

Fellbaum, Christiane (2005), "Wordnet and Wordnets."

Festinger, Leon (1962), *A Theory of Cognitive Dissonance*, Vol. 2, Palo Alto, CA: Stanford University Press.

Fetterman, Adam K., Ryan L. Boyd, and Michael D. Robinson (2015), "Power Versus Affiliation in Political Ideology: Robust Linguistic Evidence for Distinct Motivation-Related Signatures," *Personality and Social Psychology Bulletin*, 41 (9), 1195–1206.

Frege, Gottlob (1892/1948), "Sense and Reference," *Philosophical Review*, 57 (3), 209–30.

Fullwood, Michelle (2015), "Parsing Chinese Text with Stanford NLP," http://michelleful.github.io/code-blog/2015/09/10/parsing-chinese-with-stanford/.

Gamson, William A. and Andre Modigliani (1989), "Media Discourse and Public Opinion on Nuclear Power: A Constructionist Approach," *American Journal of Sociology*, 95 (1), 1–37.

Gardner, Wendi L., Shira Gabriel, and Angela Y. Lee (1999), "'I' Value Freedom, but 'We' Value Relationships: Self-Construal Priming Mirrors Cultural Differences in Judgment," *Psychological Science*, 10 (4), 321–26.

Genevsky, Alexander and Brian Knutson (2015), "Neural Affective Mechanisms Predict Market-Level Microlending," *Psychological Science*, 26 (9), 1411–22.

Ghose, Anindya, Panagiotis Ipeirotis, and Beibei Li (2012), "Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content," *Marketing Science*, 31 (3), 493–520.

Gibson, Edward (1998), "Linguistic Complexity: Locality of Syntactic Dependencies," *Cognition*, 68 (1), 1–76.

Golder, Peter N. (2000), "Historical Method in Marketing Research with New Evidence on Long-Term Market Share Stability," *Journal of Marketing Research*, 37 (2), 156–72.

Goffman, Erving (1959), *The Presentation of Self in Everyday Life*, Garden City, NY: Doubleday.

—— (1979), "Footing," *Semiotica*, 25 (1–2), 1–30.

Gonzales, Amy L., Jeffrey T. Hancock, and James W. Pennebaker, (2010), "Language style matching as a predictor of social dynamics in small groups." *Communication Research* 37 (1), 3–19.

Graesser, Arthur C., Danielle S. McNamara, Max M. Louwerse, and Zhiqiang Cai (2004), "Coh-Metrix: Analysis of Text on Cohesion and Language," *Behavior Research Methods, Instruments, & Computers*, 36 (2), 193–202.

Graham, Robert J. (1981), "The Role of Perception of Time in Consumer Research," *Journal of Consumer Research*, 7 (4), 335–42.

Grayson, Kent and David Shulman (2000), "Indexicality and the Verification Function of Irreplaceable Possessions: A Semiotic Analysis," *Journal of Consumer Research*, 27 (1), 17–30.

Green, Melanie C. and Timothy C. Brock (2002), "In the Mind's Eye: Transportation-Imagery Model of Narrative Persuasion," in *Narrative Impact: Social and Cognitive Foundations*, ed. Melanie C. Green, Jeffrey J. Strange, and Timothy C. Brock, Mahwah, NJ: Lawrence Erlbaum Associates, 315–41.

Grice, Herbert P. (1975), "Logic and Conversation," in *Syntax and Semantics, Vol. 3, Speech Acts*, ed. Peter Cole and Jerry L. Morgan, New York: Academic Press, 41–58.

Grimmer, Justin and Brandon M. Stewart (2013), "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts," *Political Analysis*, 21 (3), 267–97.

Gruenfeld, Deborah H. and Robert S. Wyer (1992), "Semantics and Pragmatics of Social Influence: How Affirmations and Denials Affect Beliefs in Referent Propositions," *Journal of Personality and Social Psychology*, 62 (1), 38.

Gruhl, Daniel, Ramanathan Guha, Ravi Kumar, Jasmine Novak, and Andrew Tomkins (2005), "The Predictive Power of Online Chatter," in *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, ed. Robert Grossman, Roberto Bayardo, Kristin Bennett, and Jaideep Vaidya, New York: ACM, 78–87.

Hall, Stuart (1980), "Encoding/Decoding," in *Culture, Media, Language: Working Papers in Cultural Studies 1972–1979*, ed. Centre for Contemporary Cultural Studies, London: Hutchinson, 128–38.

Hancock, Jeffrey T., David I. Beaver, Cindy K. Chung, Joey Frazee, James W. Pennebaker, Art Graesser, and Zhiqiang Cai (2010), "Social Language Processing: A Framework for Analyzing the Communication of Terrorists and Authoritarian Regimes," *Behavioral Sciences of Terrorism and Political Aggression*, 2 (2), 108–32.

Hausser, Roland R. (1999), *Foundations of Computational Linguistics*, Berlin: Springer.

Hayden, Erika Check (2013), "Guidance Issued for US Internet Research: Institutional Review Boards May Need to Take a Closer Look at Some Types of Online Research," *Nature*, 496 (7446), 411–12.

Herring, Susan C. (2000), "Gender Differences in CMC: Findings and Implications," *Computer Professionals for Social Responsibility Journal*, 18 (1).

—— (2003), "Gender and Power in Online Communication," in *Handbook of Language and Gender*, ed. Janet Holmes and Miriam Meyerhoff, Malden, MA: Blackwell, 202–28.

HHS (2013), "Considerations and Recommendations Concerning Internet Research and Human Subjects Research Regulations, with Revisions," https://www.hhs.gov/ohrp/sites/default/files/ohrp/sachrp/mtgings/2013%20March%20Mtg/internet_research.pdf.

Hiller, Jack H (1971), "Verbal Response Indicators of Conceptual Vagueness," *American Educational Research Journal*, 8 (1), 151–61.

Homburg, Christian, Laura Ehm, and Martin Artz (2015), "Measuring and Managing Consumer Sentiment in an Online Community Environment," *Journal of Marketing Research*, 52 (5), 629–41.

Holoien, Deborah S. and Susan T. Fiske (2013), "Downplaying Positive Impressions: Compensation between Warmth and

Competence in Impression Management," *Journal of Experimental Social Psychology*, 49 (1), 33–41.

Holt, Douglas B. (2004), *How Brands Become Icons: The Principles of Cultural Branding*, Boston, MA: Harvard Business Press.

Holt, Douglas B. and Craig J. Thompson (2004), "Man-of-Action Heroes: The Pursuit of Heroic Masculinity in Everyday Consumption," *Journal of Consumer Research*, 31 (2), 425–40.

Hopkins, Daniel J. and Gary King (2010), "A Method of Automated Nonparametric Content Analysis for Social Science," *American Journal of Political Science*, 54 (1), 229–47.

Hsu, Kean J., Kalina N. Babeva, Michelle C. Feng, Justin F. Hummer, and Gerald C. Davison (2014), "Experimentally Induced Distraction Impacts Cognitive but Not Emotional Processes in Think-Aloud Cognitive Assessment," *Frontiers in Psychology*, 5, 474.

Hughes, Marie Adele and Dennis E. Garrett (1990), "Intercoder Reliability Estimation Approaches in Marketing: A Generalizability Theory Framework for Quantitative Data," *Journal of Marketing Research*, 27 (2), 185–95.

Humphreys, Ashlee (2010), "Semiotic Structure and the Legitimation of Consumption Practices: The Case of Casino Gambling," *Journal of Consumer Research*, 37 (3), 490–510.

Humphreys, Ashlee and Kathryn A. Latour (2013), "Framing the Game: Assessing the Impact of Cultural Representations on Consumer Perceptions of Legitimacy," *Journal of Consumer Research*, 40 (4), 773–95.

Humphreys, Ashlee and Craig J. Thompson (2014), "Branding Disaster: Reestablishing Trust through the Ideological Containment of Systemic Risk Anxieties," *Journal of Consumer Research*, 41 (4), 877–910.

Hutto, C. J. and Eric Gilbert (2014), "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," paper presented at the Eighth International AAAI Conference on Weblogs and Social Media, Ann Arbor, MI.

Ireland, Molly E., Richard B. Slatcher, Paul W. Eastwick, Lauren E. Scissors, Eli J. Finkel, and James W. Pennebaker (2011), "Language Style Matching Predicts Relationship Initiation and Stability," *Psychological Science*, 22 (1), 39–44.

Jain, Anil K. and Stan Z. Li (2011), *Handbook of Face Recognition*, London: Springer Verlag London.

Jakobson, Roman (1960), "Closing Statement: Linguistics and Poetics," in *Style in Language*, ed. Thomas Albert Sebeok and John W. Ashton, Cambridge, MA: MIT, 350, 377.

Jameson, Fredric (1981), *The Political Unconscious: Narrative as a Socially Symbolic Act*, Ithacam NY: Cornell University Press.

Jepperson, Ronald and John W. Meyer (2011), "Multiple Levels of Analysis and the Limitations of Methodological Individualisms," *Sociological Theory*, 29 (1), 54–73.

Jurafsky, Dan, Victor Chahuneau, Bryan R. Routledge, and Noah A. Smith (2014), "Narrative Framing of Consumer Sentiment in Online Restaurant Reviews," *First Monday*, 19 (4).

Jurafsky, Dan, Rajesh Ranganath, and Dan McFarland (2009), "Extracting Social Meaning: Identifying Interactional Style in Spoken Conversation," in *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, ed. Fred Popowich and Michael Johnston, Stroudsburg, PA: Association for Computational Linguistics, 638–46.

Kacewicz, Ewa, James W. Pennebaker, Matthew Davis, Moongee Jeon, and Arthur C. Graesser (2014), "Pronoun Use Reflects Standings in Social Hierarchies," *Journal of Language and Social Psychology*, 33 (2), 125–43.

Kassarjian, Harold H. (1977), "Content Analysis in Consumer Research," *Journal of Consumer Research*, 4 (1), 8–19.

Katz, Jack (2001), "Analytic Induction," in *International Encyclopedia of the Social & Behavioral Sciences*, ed. Neil J. Smelser and Paul B. Baltes, Amsterdam and Oxford, UK: Elsevier, xii, 13937–4605.

Kay, Paul and Willett Kempton (1984), "What Is the Sapir-Whorf Hypothesis?" *American Anthropologist*, 86 (1), 65–79.

Kern, Margaret L., Gregory Park, Johannes C. Eichstaedt, H. Andrew Schwartz, Maarten Sap, Laura K. Smith, and Lyle H. Ungar (2016), "Gaining Insights from Social Media Language: Methodologies and Challenges," *Psychological Methods*, 21 (4), 507–25.

Kirschenbaum, Matthew G. (2007), "The Remaking of Reading: Data Mining and the Digital Humanities," paper presented at the National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation, Baltimore, MD.

Kleine, Susan Schultz, Robert E. Kleine, and Chris T. Allen (1995), "How Is a Possession 'Me' or 'Not Me'? Characterizing Types and an Antecedent of Material Possession Attachment," *Journal of Consumer Research*, 22 (3), 327–43.

Kovács, Balázs, Glenn R. Carroll, and David W. Lehman (2013), "Authenticity and Consumer Value Ratings: Empirical Tests from the Restaurant Domain," *Organization Science*, 25 (2), 458–78.

Kranz, Peter (1970), "Content Analysis by Word Group," *Journal of Marketing Research*, 7 (3), 377–80.

Krippendorff, Klaus (2004), *Content Analysis: An Introduction to Its Methodology*, Thousand Oaks, CA: Sage.

—— (2007), "Computing Krippendorff's Alpha Reliability," Departmental Papers, Annenburg School for Communication, University of Pennsylvania, Philadelphia, PA, 43.

—— (2010), *On Communicating: Otherness, Meaning, and Information*, ed. Fernando Bermejo, New York: Routledge.

Kronrod, Ann, Amir Grinstein, and Luc Wathieu (2012), "Go Green! Should Environmental Messages Be So Assertive?" *Journal of Marketing*, 76 (1), 95–102.

Kuncoro, Adhiguna, Miguel Ballesteros, Lingpeng Kong, Chris Dyer, Graham Neubig, and Noah A. Smith (2016), "What Do Recurrent Neural Network Grammars Learn About Syntax?" https://arxiv.org/abs/1611.05774.

Labroo, Aparna A. and Angela Y. Lee (2006), "Between Two Brands: A Goal Fluency Account of Brand Evaluation," *Journal of Marketing Research*, 43 (3), 374–85.

Lakoff, George (2014), *The All New Don't Think of an Elephant!: Know Your Values and Frame the Debate*, White River Junction, VT: Chelsea Green Publishing.

Lakoff, George and Sam Ferguson (2015), "The Framing of Immigration," http://www.huffingtonpost.com/george-lakoff-and-sam-ferguson/the-framing-of-immigratio_b_21320.html

Lakoff, Robin (1973), "Language and Woman's Place," *Language in Society*, 2 (1), 45–79.

Lasswell, Harold D. and Nathan Leites (1949), *Language of Politics; Studies in Quantitative Semantics*, New York: G. W. Stewart.

Lasswell, Harold D and J Zvi Namenwirth (1969), "The Lasswell Value Dictionary," *New Haven*.

Laver, Michael and John Garry (2000), "Estimating Policy Positions from Political Texts," *American Journal of Political Science*, 619–34.

Lee, Angela .Y and Jennifer L. Aaker (2004), "Bringing the Frame into Focus: The Influence of Regulatory Fit on Processing Fluency and Persuasion," *Journal of Personality and Social Psychology*, 86 (2), 205.

Lee, Angela Y. and Aparna A. Labroo (2004), "The Effect of Conceptual and Perceptual Fluency on Brand Evaluation," *Journal of Marketing Research*, 41 (2), 151–65.

Lee, Thomas Y. and Eric T. Bradlow (2011), "Automated Marketing Research Using Online Customer Reviews," *Journal of Marketing Research*, 48 (5), 881–94.

Lewinski, Peter (2015), "Automated Facial Coding Software Outperforms People in Recognizing Neutral Faces as Neutral from Standardized Datasets," *Frontiers in Psychology*, 6 (1386).

Li, Feng (2008), "Annual Report Readability, Current Earnings, and Earnings Persistence," *Journal of Accounting and Economics*, 45 (2), 221–47.

Lowe, Will (2006), "Yoshikoder: An Open Source Multilingual Content Analysis Tool for Social Scientists," paper presented at the annual meeting of the American Political Science Association, Philadelphia, PA.

Loughran, Tim and Bill McDonald (2014), "Measuring Readability in Financial Disclosures," *The Journal of Finance*, 69 (4), 1643–71.

Lucy, John A. and Richard A. Shweder (1979), "Whorf and His Critics: Linguistic and Nonlinguistic Influences on Color Memory," *American Anthropologist*, 81 (3), 581–615.

Ludwig, Stephan, Ko de Ruyter, Mike Friedman, Elisabeth C. Brüggen, Martin Wetzels, and Gerard Pfann (2013), "More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates," *Journal of Marketing*, 77 (1), 87–103.

Ludwig, Stephan, Tom Van Laer, Ko de Ruyter, and Mike Friedman (2016), "Untangling a Web of Lies: Exploring Automated Detection of Deception in Computer-Mediated Communication," *Journal of Management Information Systems*, 33 (2), 511–41.

Luedicke, Marius K., Craig J. Thompson, and Markus Giesler (2010), "Consumer Identity Work as Moral Protagonism: How Myth and Ideology Animate a Brand-Mediated Moral Conflict," *Journal of Consumer Research*, 36 (6), 1016–32.

Mahoney, James and Dietrich Rueschemeyer, eds. (2003), *Comparative Historical Analysis in the Social Sciences*, Cambridge, UK: Cambridge University Press.

Malinowski, Bronislaw (1972), "Phatic Communion," in *Communication in Face-to-Face Interaction*, ed. John Laver and Sandy Hutcheson, New York: Penguin, 146–52.

Manning, Christopher D. and Hinrich Schuetze, *Foundations of Statistical Natural Language Processing*, Cambridge, MA: MIT Press, 1999.

Mankad, Shawn, Hyunjeong "Spring" Han, Joel Goh, and Srinagesh Gavirneni (2016), "Understanding Online Hotel Reviews through Automated Text Analysis," *Service Science*, 8 (2), 124–38.

Markham, Annette and Elizabeth Buchanan (2012), "Ethical Decision-Making and Internet Research: Recommendations from the Aoir Ethics Working Committee (Version 2.0)," https://aoir.org/reports/ethics2.pdf.

Markowitz, David M. and Jeffrey T. Hancock (2015), "Linguistic Obfuscation in Fraudulent Science," *Journal of Language and Social Psychology*, 35 (4), 435–45.

Martin, Michael K., Juergen Pfeffer, and Kathleen M. Carley (2013), "Network Text Analysis of Conceptual Overlap in Interviews, Newspaper Articles, and Keywords," *Social Network Analysis and Mining*, 3 (4), 1165–77.

Martindale, Colin (1975), *The Romantic Progression: The Psychology of Literary History*, New York: Halsted Press.

Marwick, Alice E. and danah boyd (2011), "I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience," *New Media & Society*, 13 (1), 114–33.

Marx, Gary T. (2001), "Murky Conceptual Waters: The Public and the Private," *Ethics and Information Technology*, 3 (3), 157–69.

Mathwick, Charla, Caroline Wiertz, and Ko De Ruyter (2008), "Social Capital Production in a Virtual P3 Community," *Journal of Consumer Research*, 34 (6), 832–49.

McBrian, Charles D. (1978), "Language and Social Stratification: The Case of a Confucian Society," *Anthropological Linguistics*, 320–26.

McCombs, Maxwell E. and Donald L. Shaw (1972), "The Agenda-Setting Function of Mass Media," *Public Opinion Quarterly*, 36 (2), 176–87.

McCracken, Grant (1986), "Culture and Consumption: A Theoretical Account of the Structure and Movement of the Cultural Meaning of Consumer Goods," *Journal of Consumer Research*, 13 (1), 71–84.

McKenny, Aaron F., Jeremy C. Short, and G. Tyge Payne (2013), "Using Computer-Aided Text Analysis to Elevate Constructs an Illustration Using Psychological Capital," *Organizational Research Methods*, 16 (1), 152–84.

McQuarrie, Edward F. and David Glen Mick (1996), "Figures of Rhetoric in Advertising Language," *Journal of Consumer Research*, 22 (4), 424–38.

McQuarrie, Edward F., Jessica Miller, and Barbara J. Phillips (2013), "The Megaphone Effect: Taste and Audience in Fashion Blogging," *Journal of Consumer Research*, 40 (1), 136–58.

McTavish, Donald G., Kenneth C. Litkowski, and Susan Schrader (1995), *A Computer Content Analysis Approach to Measuring Social Distance in Residential Organizations for Older People*, Mannheim, Germany: Society for Content Analysis by Computer.

Mehl, Matthias R. (2006), "Quantitative Text Analysis," in *Handbook of Multimethod Measurement in Psychology*, ed. Ed Diener and Michael Eid, Washington, DC: American Psychological Association, 141–56.

Mehl, Matthias R. and Alastair J. Gill (2008), "Automatic Text Analysis," in *Advanced Methods for Behavioral Research on the Internet*, ed. Samuel D. Gosling and John A. Johnson, Washington, DC: American Psychological Association.

Mestyán, Márton, Taha Yasseri, and János Kertész (2013), "Early Prediction of Movie Box Office Success Based on Wikipedia Activity Big Data," *PloS One*, 8 (8), e71226.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books TeamPickett Joseph P., Dale Hoiberg et al., (2011), "Quantitative Analysis of Culture Using Millions of Digitized Books," *Science*, 331 (6014), 176–82.

Mick, David Glen (1986), "Consumer Research and Semiotics: Exploring the Morphology of Signs, Symbols, and

Significance," *Journal of Consumer Research*, 13 (2), 196–213.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013), "Efficient Estimation of Word Representations in Vector Space," https://arxiv.org/abs/1301.3781.

Miller, George A (1995), "Wordnet: A Lexical Database for English," *Communications of the ACM*, 38 (11), 39–41.

Mislove, Alan, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and J. Niels Rosenquist (2011), "Understanding the Demographics of Twitter Users," in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, ed. Nicolas Nicolov and James G. Shanahan, Menlo Park, CA: AAAI Press.

Mogilner, Cassie, Sepandar D. Kamvar, and Jennifer Aaker (2011), "The Shifting Meaning of Happiness," *Social Psychological and Personality Science*, 2 (4), 395–402.

Mohr, John W. (1998), "Measuring Meaning Structures," *Annual Review of Sociology*, 24, 345–70.

Monroe, Burt L., Michael P. Colaresi, and Kevin M. Quinn (2009), "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict," *Political Analysis*, 16 (4), 372–403.

Moore, Sarah G. (2015), "Attitude Predictability and Helpfulness in Online Reviews: The Role of Explained Actions and Reactions," *Journal of Consumer Research*, 42 (1), 30–44.

Morris, Charles W. (1938), *Foundations of the Theory of Signs*, Chicago: University of Chicago Press.

Morris, Rebecca (1994), "Computerized Content Analysis in Management Research: A Demonstration of Advantages & Limitations," *Journal of Management*, 20 (4), 903–31.

Namenwirth, J. Zvi and Robert Philip Weber (1987), *Dynamics of Culture*, Boston: Allen & Unwin.

Narayanan, Arvind and Vitaly Shmatikov (2008), "Robust De-Anonymization of Large Sparse Datasets," in *SP '08 IEEE Symposium on Security and Privacy*, Washington, DC: IEEE Computer Society, 111–25.

—— (2010), "Myths and Fallacies of Personally Identifiable Information," *Communications of the ACM*, 53 (6), 24–26.

Netzer, Oded, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko (2012), "Mine Your Own Business: Market-Structure Surveillance through Text Mining," *Marketing Science*, 31 (3), 521–43.

Neuman, Yair, Peter Turney, and Yohai Cohen (2012), "How Language Enables Abstraction: A Study in Computational Cultural Psychology," *Integrative Psychological and Behavioral Science*, 46 (2), 129–45.

Newman, Matthew L., James W. Pennebaker, Diane S. Berry, and Jane M. Richards (2003), "Lying Words: Predicting Deception from Linguistic Styles," *Personality & Social Psychology Bulletin*, 29 (5), 665–75.

Newsprosoft (2012), "Web Content Extractor," http://www.newprosoft.com/web-content-extractor.htm.

Ng, Sik Hung and James J. Bradac (1993), *Power in Language: Verbal Communication and Social Influence*, Newbury Park, CA: Sage.

Nielsen, Finn Årup (2011), "A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs," in *Proceedings of the ESWC2011 Workshop on "Making Sense of Microposts": Big Things Come in Small Packages*, ed. Matthew Rowe, Milan Stankovic, Aba-Sah Dadzie, and Mariann Hardie, http://ceur-ws.org/Vol-718/msm2011_proceedings.pdf, 93–8.

Nissenbaum, Helen (2009), *Privacy in Context: Technology, Policy, and the Integrity of Social Life*, Palo Alto, CA: Stanford University Press.

North, Robert, Richard Iagerstrom, and William Mitchell (1999), *Diction Computer Program, Ann Arbor*, MI: Inter-university Consortium for and Political and Social Research.

Nunberg, Geoffrey (1993), "Indexicality and Deixis," *Linguistics and Philosophy*, 16 (1), 1–43.

Opoku, Robert, Russell Abratt, and Leyland Pitt (2006), "Communicating Brand Personality: Are the Websites Doing the Talking for the Top South African Business Schools?" *Journal of Brand Management*, 14 (1–2), 20–39.

Osborne, Jason W. (2010), "Improving Your Data Transformations: Applying the Box-Cox Transformation," *Practical Assessment, Research & Evaluation*, 15 (12), 1–9.

Packard, Grant and Jonah Berger (2016), "How Language Shapes Word of Mouth's Impact," *Journal of Marketing Research*, 54 (4), 572–88.

Packard, Grant, Sarah G. Moore, and Brent McFerran (2018), "How Can 'I' Help "You"? The Impact of Personal Pronoun Use in Customer-Firm Agent Interactions," *Journal of Marketing Research*, forthcoming.

—— (2016), "(I'm) Happy to Help (You): The Impact of Personal Pronoun Use in Customer-Firm Interactions," working paper.

Pagescrape (2006), "Pagescrape," https://www.npmjs.com/package/pagescrape.

Parmentier, Marie-Agnès and Eileen Fischer (2015), "Things Fall Apart: The Dynamics of Brand Audience Dissipation," *Journal of Consumer Research*, 41 (5), 1228–51.

Pauwels, Koen (2014), *It's Not the Size of the Data—It's How You Use It: Smarter Marketing with Analytics and Dashboards*, New York: AMACOM.

Peirce, Charles Sanders (1957), "The Logic of Abduction," in *Charles S. Peirce: Essays in the Philosophy of Science*, ed. Vincent Tomas, New York: Liberal Arts Press.

Péladeau, Normand (2016), *WordStat: Content Analysis Module for Simstat*, Montreal: Provalis Research.

Pennebaker, James W. (2011), "The Secret Life of Pronouns," *New Scientist*, 211 (2828), 42–45.

Pennebaker, James W., Martha E. Francis, and Roger J. Booth (2001), *Linguistic Inquiry and Word Count: LIWC 2001, Mahway*, NJ: Lawrence Erlbaum Associates, 71, 2001.

—— (2007), *Linguistic Inquiry and Word Count (LIWC): Liwc2007, Mahwah*, NJ: Lawrence Erlbaum Associates.

Pennebaker, James W., Ryan L. Boyd, Kayla Jordan, and Kate Blackburn (2015), *The Development and Psychometric Properties of LIWC2015*, Austin, TX: University of Texas at Austin.

Pennebaker, James W. and Laura A. King (1999), "Linguistic Styles: Language Use as an Individual Difference," *Journal of Personality & Social Psychology*, 77 (6), 1296–312.

Pennington, Jeffrey, Richard Socher, and Christopher D. Manning (2014), "GloVe: Global Vectors for Word Representation," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, ed. Alessandro Moschitti, Stroudsburg, PA: Association for Computational Linguistics, 1532–43.

Petty, Richard E., John T. Cacioppo, and David Schumann (1983), "Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement," *Journal of Consumer Research*, 10 (2), 135–46.

Piaget, Jean (1959), *The Language and Thought of the Child*, Vol. 5, Abingdon, UK: Psychology Press.

Plaisant, Catherine, James Rose, Bei Yu, Loretta Auvil, Matthew G. Kirschenbaum, Martha Nell Smith, Tanya Clement, and Greg Lord (2006), "Exploring Erotics in Emily Dickinson's Correspondence with Text Mining and Visual Interfaces," in *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries*, ed. Michael L. Nelson, Cathy Marshall, and Gary Marchionini, New York: ACM, 141–50.

Pollach, Irene (2012), "Taming Textual Data: The Contribution of Corpus Linguistics to Computer-Aided Text Analysis," *Organizational Research Methods*, 15 (2), 263–87.

Potts, Christopher and Florian Schwarz (2010), "Affective 'This,'" *Linguistic Issues in Language Technology*, 3 (5), 1–30.

Provost, Foster and Tom Fawcett (2013), *Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking*, Sebastopol, CA: O'Reilly Media.

Pury, Cynthia L. S. (2011), "Automation Can Lead to Confounds in Text Analysis Back, Küfner, and Egloff (2010) and the Not-So-Angry Americans," *Psychological Science*, 22 (6), 835–36.

Quine, Willard Van Orman and J. S. Ullian (1970), *The Web of Belief*, New York: Random House.

Ritter, Ryan S., Jesse L. Preston, and Ivan Hernandez (2013), "Happy Tweets: Christians Are Happier, More Socially Connected, and Less Analytical than Atheists on Twitter," *Social Psychological and Personality Science*, 5 (2), 243–49.

Roget, Peter Mark (1911), *Roget's Thesaurus of English Words and Phrases*: TY Crowell Company.

Rothwell, David (2007), *Wordsworth Dictionary of Homonyms, Hertfordshire*, UK: Wordsworth Editions.

Rubin, Donald B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley.

Rude, Stephanie, Eva-Maria Gortner, and James W. Pennebaker (2004), "Language Use of Depressed and Depression-Vulnerable College Students," *Cognition & Emotion*, 18 (8), 1121–33.

Sapir, Edward (1929), "The Status of Linguistics as a Science," *Language*, 5 (4), 207–14.

Schatzki, Theodore R. (1996), *Social Practices: A Wittgensteinian Approach to Human Activity and the Social*, Cambridge, UK: Cambridge University Press.

Schau, Hope Jensen, Albert M. Muniz, and Eric J. Arnould (2009), "How Brand Community Practices Create Value," *Journal of Marketing*, 73 (5), 30–51.

Schmitt, Bernd H. and Shi Zhang (1998), "Language Structure and Categorization: A Study of Classifiers in Consumer Cognition, Judgment, and Choice," *Journal of Consumer Research*, 25 (2), 108–22.

Schouten, John W. and James H. McAlexander (1995), "Subcultures of Consumption: An Ethnography of the New Bikers," *Journal of Consumer Research*, 22 (1), 43–61.

Schudson, Michael (1989), "How Culture Works," *Theory and Society*, 18 (2), 153–80.

Senay, Ibrihim, Muhammet Usak, and Pavol Prokop (2015), "Talking about Behaviors in the Passive Voice Increases Task Performance," *Applied Cognitive Psychology*, 29 (2), 262–70.

Sera, Maria D., Christian A. H. Berge, and Javier del Castillo Pintado (1994), "Grammatical and Conceptual Forces in the Attribution of Gender by English and Spanish Speakers," *Cognitive Development*, 9 (3), 261–92.

Settanni, Michele and Davide Marengo (2015), "Sharing Feelings Online: Studying Emotional Well-Being Via Automated Text Analysis of Facebook Posts," *Frontiers in Psychology*, 6, 1045.

Sexton, J. Bryan and Robert L. Helmreich (2000), "Analyzing Cockpit Communications: The Links between Language, Performance, Error, and Workload," *Journal of Human Performance in Extreme Environments*, 5 (1), 6.

Sherry, John F., Mary Ann McGrath, and Sidney J. Levy (1993), "The Dark Side of the Gift," *Journal of Business Research*, 28 (3), 225–44.

Shor, Eran, Arnout van de Rijt, Alex Miltsov, Vivek Kulkarni, and Steven Skiena (2015), "A Paper Ceiling Explaining the Persistent Underrepresentation of Women in Printed News," *American Sociological Review*, 80 (5), 960–84.

Snefjella, Bryor and Victor Kuperman (2015), "Concreteness and Psychological Distance in Natural Language Use," *Psychological Science*, 26 (9), 1449–60.

Socher, Richard, Perelygin Alex, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts (2013), "Recursive Deep Models for Semantic Compositionality over a Sentiment Treebank," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (EMNLP), Vol. 1631, Stroudsburg, PA: ACL, 1642.

Spärck Jones, Karen (1972), "A Statistical Interpretation of Term Specificity and Its Application in Retrieval," *Journal of Documentation*, 28, 11–21.

Spiller, Stephen A. and Lena Belogolova (2016), "On Consumer Beliefs About Quality and Taste," *Journal of Consumer Research*, 43 (6), 970–91.

Stone, Philip J. (1966), *The General Inquirer: A Computer Approach to Content Analysis*, Cambridge, MA: MIT Press.

Sun, Maosong, Yang Liu, Zhiyuan Liu, and Min Zhang, eds. (2015), *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*, Cham, Switzerland: Springer International.

Sun, Monic (2012), "How Does the Variance of Product Ratings Matter?" *Management Science*, 58 (4), 696–707.

Swanson, Don R. (1988), "Migraine and Magnesium: Eleven Neglected Connections," *Perspectives in Biology and Medicine*, 31 (4), 526–57.

Tan, Chenhao, Evgeniy Gabrilovich, and Bo Pang (2012), "To Each His Own: Personalized Content Selection Based on Text Comprehensibility," in *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, New York: ACM, 233–42.

Tausczik, Yla R. and James W. Pennebaker (2010), "The Psychological Meaning of Words: Liwc and Computerized Text Analysis Methods," *Journal of Language and Social Psychology*, 29 (1), 24–54.

Thelwall, Mike, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas (2010), "Sentiment Strength Detection in Short Informal Text," *Journal of the American Society for Information Science and Technology*, 61 (12), 2544–58.

Thompson, Craig J. and Elizabeth C. Hirschman (1995), "Understanding the Socialized Body: A Poststructuralist Analysis of Consumers' Self-Conceptions, Body Images, and Self-Care Practices," *Journal of Consumer Research*, 22 (2), 139–53.

Thompson, Craig J., William B. Locander, and Howard R. Pollio (1989), "Putting Consumer Experience Back into Consumer Research: The Philosophy and Method of Existential-Phenomenology," *Journal of Consumer Research*, 16 (2), 133–46.

Tirunillai, Seshadri and Gerard J. Tellis (2012), "Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance," *Marketing Science*, 31 (2), 198–215.

—— (2014), "Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation," *Journal of Marketing Research*, 51 (4), 463–79.

Townsend, Leanne and Claire Wallace (2016), "Social Media Research: A Guide to Ethics," http://www.dotrural.ac.uk/socialmediaresearchethics.pdf.

Twenge, Jean M., W. Keith Campbell, and Brittany Gentile (2012), "Changes in Pronoun Use in American Books and the Rise of Individualism, 1960–2008," *Journal of Cross-Cultural Psychology*, 44 (3), 406–15.

US Department of Health and Human Services (1979), *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*, National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 45. Washington, DC: US Government Printing Office.

Valentino, Nicholas A. (1999), "Crime News and the Priming of Racial Attitudes During Evaluations of the President," *Public Opinion Quarterly*, 63 (3), 293–320.

van Bommel, Koen (2014), "Towards a Legitimate Compromise? An Exploration of Integrated Reporting in the Netherlands," *Accounting, Auditing & Accountability Journal*, 27 (7), 1157–89.

Van de Rijt, Arnout, Eran Shor, Charles Ward, and Steven Skiena (2013), "Only 15 Minutes? The Social Stratification of Fame in Printed Media," *American Sociological Review*, 78 (2), 266–89.

Van de Rijt, Arnout, Eran Shor, Charles Ward, and Steven Skiena (2013), "Only 15 Minutes? The Social Stratification of Fame in Printed Media," *American Sociological Review*, 78 (2), 266–89.

Van Laer, Tom, Jennifer Edson Escalas, Stephan Ludwig, and Ellis A. Van den Hende (2017), "What Happens in Vegas Stays on TripAdvisor? Computerized Analysis of Narrativity in Online Consumer Reviews," Vanderbilt Owen Graduate School of Management Research Paper No. 2702484, Nashville, TN, https://ssrn.com/abstract=2702484 or http://dx.doi.org/10.2139/ssrn.84.

Vasi, Ion Bogdan, Edward T. Walker, John S. Johnson, and Hui Fen Tan (2015), "'No Fracking Way!' Documentary Film, Discursive Opportunity, and Local Opposition against Hydraulic Fracturing in the United States, 2010 to 2013," *American Sociological Review*, 80 (5), 934–59.

Velocityscape (2006), "Webscraper," http://www.velocityscape.com/.

Vico, Giambattista (1725/1984), *The New Science*, Ithaca, NY: Cornell University Press.

Wade, James B., Joseph F. Porac, and Timothy G. Pollock (1997), "Worth, Words, and the Justification of Executive Pay," *Journal of Organizational Behavior*, 18 (S1), 641–64.

Wallendorf, Melanie and Eric J. Arnould (1988), "'My Favorite Things': A Cross-Cultural Inquiry into Object Attachment, Possessiveness, and Social Linkage," *Journal of Consumer Research*, 14 (4), 531–47.

Wang, Jing and Bobby J. Calder (2006), "Media Transportation and Advertising," *Journal of Consumer Research*, 33 (2), 151–62.

Waters, Audrey (2011), "How Recent Changes to Twitter's Terms of Service Might Hurt Academic Research," http://readwrite.com/2011/03/03/how_recent_changes_to_twitters_terms_of_service_mi/.

Watson, David, Lee Anna Clark, and Auke Tellegen (1988), "Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales," *Journal of Personality and Social Psychology*, 54 (6), 1063–70.

Weber, Klaus (2005), "A Toolkit for Analyzing Corporate Cultural Toolkits," *Poetics*, 33 (3–4), 227–52.

Whiteman, Natasha (2012), *Undoing Ethics: Rethinking Practice in Online Research*, New York: Springer, 1–23.

Whorf, Benjamin Lee (1944), "The Relation of Habitual Thought and Behavior to Language," *ETC: A Review of General Semantics*, 1 (4), 197–215.

Wilson, Andrew (2005), "Development and Application of a Content Analysis Dictionary for Body Boundary Research," *Literary and Linguistic Computing*, 21 (1), 105–10.

Wilson, Todd (2009), "Screen-Scraper," https://www.screen-scraper.com/.

Wong, Elaine M., Margaret E. Ormiston, and Michael P. Haselhuhn (2011), "A Face Only an Investor Could Love: CEOs' Facial Structure Predicts Their Firms' Financial Performance," *Psychological Science*, 22 (12), 1478–83.

Wood, Linda A. and Rolf O. Kroger (2000), *Doing Discourse Analysis: Methods for Studying Action in Talk and Text*, Thousand Oaks, CA: Sage.

Xu, Zhi and David N. Bengston (1997), "Trends in National Forest Values among Forestry Professionals, Environmentalists, and the News Media, 1982–1993," *Society & Natural Resources*, 10 (1), 43–59.

Yadav, Manjit S, Jaideep C Prabhu, and Rajesh K Chandy (2007), "Managing the Future: Ceo Attention and Innovation Outcomes," *Journal of Marketing*, 71 (4), 84–101.

Yarowsky, David (1992), "Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora," in *Proceedings of the 14th conference on Computational linguistics-Volume 2*: Association for Computational Linguistics, 454–60.

Zachary, Miles A, Aaron McKenny, Jeremy Collin Short, and G Tyge Payne (2011), "Family Business and Market Orientation: Construct Validation and Comparative Analysis," *Family Business Review*, 24 (3), 233–51.

Zipf, George Kingsley (1932), *Selected Studies of the Principle of Relative Frequency in Language*, Cambridge, MA: Harvard University Press.