Making Sense of Text: Artificial Intelligence-Enabled Content Analysis

Journal:	*European Journal of Marketing*
Manuscript ID	EJM-02-2019-0219.R1
Manuscript Type:	Original Article
Keywords:	Content analysis, Marketing, Leadership, artificial intelligence, IBM Watson, Qualitative research, Research methods, Topic modeling

Making Sense of Text: Artificial Intelligence-Enabled Content Analysis

Introduction

Content analysis is a widely used research method for systematically and objectively analyzing content, while carefully considering the reliability, validity, and efficiency of this analysis and the development of insights and related theory (Neuendorf, 2017, Krippendorff, 2018). The content data can include the written text, speeches, reviews, images, audio, videos, and hypertext that are found in any form of communication. Marketing researchers use content analysis to examine content central to their field, including: brand-controlled content such as advertising (Gross and Sheth, 1989; Gilly, 1988) and websites (Govers and Go, 2004; Jose and Lee, 2007); marketing content disseminated by media such as press coverage (Harris et al., 2001); and user-generated content such as complaints (Harrison-Walker, 2001), travel blogs (Pan et al., 2007), and employer reviews (Dabirian et al., 2017; Dabirian et al., 2019). Furthermore, marketing researchers use content analysis to analyze interview and survey data (Bitner et al, 1990; Dong et al., 2015) and marketing literature (Helgeson et al., 1984; Leonidou and Leonidou, 2011).

This paper focuses on one type of content, text, which in itself is complex and requires a significant understanding of language and human cognition. Traditionally, researchers used what is now known as the manual approach to carrying out a content analysis of text. This involves humans manually coding and analyzing the text. In the 1980s, the computer-aided approach for content analysis of text was developed for researchers to automate, at least partially, the coding and analysis. Software programs are used to manipulate text and compute word frequency lists, keyword-in-context lists, concordances, classifications of text in terms of content categories, category counts, and so on – results of which human researchers then interpret (Deffner, 1986).

Computer-aided content analysis has been widely adopted as it aids researchers' intuition and reasoning by facilitating the manipulation and presentation of data to help uncover patterns unlikely to be detected by the researcher (Wolfe et al. 1993). And, despite validity concerns about the computer-aided approach's ability to understand the sentiments, opinions, and expressions in content (Morris, 1994; Su et al., 2017), there is a strong view that such downsides are outweighed by the reliability, time, and cost benefits of the approach when dealing with large data sets (Rosenberg et al., 1990, Conway, 2006). Given this trade-off between the validity of manual content analysis versus the reliability and efficiency of computer-aided content analysis, scholars argue that a hybrid of both approaches would capitalize on all benefits (Su et al., 2017; Lewis et al., 2013).

Artificial intelligence (AI) promises to offer the benefits of both manual and computer-aided approaches. In the context of this paper, AI is "the study of knowledge representations (generally) by way of computers and the use of those representations in language performance, reasoning, learning and problem solving" (Sowa, 1984, p. 22). In other words, AI has the potential to improve dramatically our ability to analyze, manipulate, and understand complex content such as natural language. This has led researchers to suggest that AI could deliver the validity of manual content analysis with the reliability and efficiency of computer-aided content analysis by automating aspects of human thinking and actions in a rational way (Hannigan et al., 2019). Assessing the promise of AI-enabled content analysis, specifically the validity, reliability, and efficiency of the approach, is the aim of this paper. This aim is also central to the focus of this Special Issue that deals with the opportunities that computerized content analysis approaches can have for marketing scholars and practitioners.

To achieve this aim, this paper is structured as follows. First, it presents a review of how content analysis is used in marketing research and highlights the types of content studied. It then explains what AI is, why it should be used for content analysis, and offers a roadmap of steps for how to use AI-enabled content analysis. This paper illustrates the application and comparison of AI-enabled content analysis relative to manual and computer-aided (i.e., non-AI) approaches using the text of leadership speeches, which can reflect the brand of a leader. Then, each approach's reliability, validity, and efficiency is compared. Despite the heavy focus on computers and advanced technologies, this paper is largely nontechnical.

Content Analysis in Marketing Research

Content analysis has long been used in the social sciences and is an increasingly important method, rising from 4,765 published studies in 2002 to 33,223 published studies in 2017 (Hannigan et al., 2019). It was introduced as a method for consumer research by Ferber and Wales (1958) and since then, marketing research went on to use content analysis to examine most, if not all forms of marketing communications, including brand-controlled content, content disseminated by media, and user-generated content. Table 1 presents a selection of marketing studies that use content analysis. The studies were chosen to highlight how this method is employed to examine different marketing phenomena from different sources, types of content, and with different content analysis approaches (i.e., manual or computer-aided approaches, with no marketing studies as of yet using AI).

An early example of manual content analysis for brand-controlled content compares the roles of men and women in magazine ads, concluding that they rarely show women in working roles and confirming existing clichés about the roles of women in American society in the 1970s (Courtney and Lockertz, 1971). This type of research was then extended with computer-aided content analysis in the 1990s, for example with Kolbe and Albanese's (1996) study of the physical characteristics of men appearing alone in magazine ads, finding the portrayal of inappropriate male stereotypes.

Insert Table 1 about here.

As websites were developed to communicate marketing messages, scholars began studying them, often using content analysis. Travel websites (such as travel agencies, travel magazines, and travel guides) were analysed using CATPAC II, a tool for computer-aided analysis to identify word frequency to examine destination image (Choi et al., 2007). In terms manual content analysis, the websites of green electricity providers in Germany were analyzed, concluding that utilitarian benefits for customers were conveyed well but psychological and self-expressive benefits were not (Herbes and Ramine, 2014).

In the area of brand packaging, scholars have used manual content analysis to examine anthropomorphism (i.e., the attribution of human characteristics or behavior to a god, animal, or object) in product packaging of the top 100 grocery brands in the U.K. (Triantos et al., 2016). Another study used manual content analysis to examine how misleading environmental information on packaging can be, resulting in seven different informational categories and four accuracy categories for packaging content (Polonsky et al., 1998).

As brand slogans are an important marketing communication tool, content analysis has been used to evaluate their meaning. The General Inquirer text analysis program was used to study and categorize brand slogans based on word frequency (Dowling and Kabanoff, 1996); and manual content analysis was undertaken to examine the brand slogans of Fortune 500 companies for the presence of linguistic devices; finding that 92 percent of brand slogans contained at least one linguistic device (Miller and Tomas, 2016).

Studies of marketing content disseminated by media include a manual content analysis of articles from six Greek newspapers, concluding that the political parties and the press did not have a common view of the relative importance of political issues (Harris et al., 2001). Schultz et al. (2012) used Amsterdam Content Analysis Toolkit, a computer-aided content analysis program to evaluate 1,376 newspaper articles in the U.S. and U.K. covering the BP crisis, finding that BP successfully dissociated itself from being responsible for the cause and presented itself as providing solutions for the crisis.

The rising volume and influence of user-generated content has led to a proliferation of marketing studies examining this content, often using content analysis. For example, Harrison-Walker (2001) manually content analyzed 551 consumer complaints from the "Untied" website, an independent complaint forum against United Airlines, examining the complaints according to seven categories. Pan et al. (2007) used TextAnalyst, a computer-aided content analysis program to examine visitor opinions posted on leading travel blog sites for the Charleston, South Carolina area to generate a semantic network of their experiences to reveal the strengths, weaknesses, and the competitive environment of Charleston as a tourist destination. Table 1 provides further examples of content analysis studies from marketing research in each category of brand, media, and user-generated content.

Finally, and as per the focus of this paper, content analysis has been used to study speeches of leaders who represent corporate or political brands. For example, Oliveira and Murphy (2009) used resonance analysis, a computer-aided text analysis method that identifies the most central words to reveal three distinct frames: profitable multinational, litigation target, and corporate good citizen in the crisis management speeches of the Philip Morris CEO during the 1990s.

In sum, it is clear that content analysis is a widely-used method for marketing research. While computer-aided content analysis tools are increasingly used, they tend to examine manifest content (i.e., content that can be simply counted, such as specific words in a text). When marketing researchers need to examine latent content (i.e., deducing meaning, symbolism and patterns from the content) that require interpretation (Duriau et al., 2007), they still rely on time-consuming manual content analysis. This literature review does not find any marketing studies using AI for content analysis.

Why Use AI-Enabled Content Analysis?

While AI can process content in the form of text, images, audio. and video, this paper focuses on text for comparison purposes, since traditional computer-aided content analysis is limited to text. Natural Language Processing (NLP) is a method for analyzing naturally-occurring human language and processing phrases and dialogs using computers (Demirkan and Delen 2013). NLP has long been used in linguistics to semantically interpret text (Manning and Schütze, 1999). Some computer-aided content analysis programs can conduct Shallow NLP (High, 2012). This type of NLP provides an understanding of the language and sentence structure, but lacks an understanding of the context and sentiments which is an imperative feature of the human cognitive system. For example, with Shallow NLP, “My nose is running.” can be interpreted as a nose that is going on a run, which is nonsensical. On the other hand, Deep NLP, which comes with AI systems such as IBM Watson, Amazon AWS Lex, Salesforce Einstein, uses both sentence structure and context of the text to provide a deeper understanding of the language. This allows such AI systems to correctly intrepret “My nose is running.” as a symptom of an illness.

Another important feature of AI-enabled content analysis systems that perform Deep NLP is their use of machine learning. The AI system receives its information not only from the sentence structure and context of the text but also from a knowledge base (for IBM Watson, the Corpus). Machine learning means that the AI-enabled content analysis approach involves training the program every time it processes a text. As the system gets smarter, it improves its accuracy and context recognition. For example, the system can be applied to medical documents (Diomaiuta et al. 2017) where its knowledge base uses a medical dictionary to analyze the language in the text similar to a human with medical training. Its relative advantage, of course, is that the system can process large documents in a fraction of the time, constantly learns, and never forgets. Therefore, by processing both context and sentence structure from the text and from its previously gathered knowledge, the system achieves a significantly higher level of accuracy than its alternatives.

In the early 2000s, topic modeling developed as a unique NLP-like approach for text analysis. In a comprehensive review of how topic modeling is used in business research, Hannigan et al. (2019) explained how topic modeling uses statistical associations of words in a text to generate topics (i.e., themes) without researchers having to develop and use dictionaries and interpretive rules. This allows researchers to identify important topics that human researchers are unlikely to distinguish. For example, McCarthy and Ruckman (2017) used topic modeling to analyze and compare the text in technology patents that are similar in technological claims to help assess why some patents were licensed while other technologically-similar patents were not. Hannigan et al. (2019) concluded their review by highlighting AI as a promising technology for enhancing text analysis. This is based on its potential to retain more contextual information, handle content that is continuously changing such as online reviews by consumers, and analyze non-text content such as images, audio, and video.

Insert Figure 1 about here.

Figure 1 illustrates how the IBM Watson AI system processes text. In the first step, it takes documents (text files) and performs Deep NLP to generate keywords, which contain one word or a phrase such as nouns, pronouns, noun sequences, and verb-noun phrases. IBM Watson also calculates the frequency of each keyword in the document. The knowledge base then converts the keywords into the subject matter translations (i.e., categorization). For example, if a marketing knowledge base is chosen, expressions like “consumer protest”, “don’t buy”, or “unethical brand” could all translate to “brand boycott”. It is important to note that, at this stage, manual coding is still required. In other words, although the term AI seems to suggest that such content analysis programs could run autonomously, in a fully-automated fashion, this is not the case. NLP and the knowledge base require training, which is commonly undertaken by people, whose role is to connect each keyword to particular constructs. In 2018, for example, medical experts in Japan trained an AI system on hundreds of datasets to see if it could accurately spot instances of stomach cancer. Once trained to differentiate malignant from benign endoscopic images, AI only took 0.004 second to judge whether a patient showed early stage cancer or normal stomach tissue (The Japan Times, 2018). AI correctly detected cancer in 80 percent of cancer images, while the accuracy rate was 95 percent for normal tissue (The Japan Times, 2018). Without training it first, this would not have been possible. For the analysis of text, this training is also required, at least for now, but like the cancer diagnosis example, the potential of vastly improved results is promising. After this overview of artifical intelligence, and its promise for content analysis, the research methods are presented next, including the study context and performance measures for all three content analysis approaches.

Application and Comparison of AI-Enabled Content Analysis

Context

To apply AI-enabled content analysis and compare it to manual and computer-aided approaches, this paper focuses on the levels of charisma in speeches by well-known business leaders. Charisma in speeches is selected not to develop theory about leadership speeches, but to illustrate the application of AI-enabled content analysis relative to the other approaches. The context of leadership speeches is appropriate for such an aim for two reasons.

First, leadership speeches reflect the brands of leaders and organizations, and thus are suited to the topic and readership of this journal. Top executives as human brands represent the brand of their organization and personify the values of the organization (Bendisch et al., 2013; Zerfass et al., 2016). Weber Shandwick (2015) indicates that they can influence the public's opinion about their organization, as an average of 45 percent of a company's reputation can be attributed to the reputation of the chief executive. This paper focuses on the charisma of leaders, as it is believed to be an important quality that enhances employee performance (Humphreys, 2002) and is linked to organizational effectiveness and perceptions of leader effectiveness (Bryman, 1992; Fiol et al., 1999). These impacts in turn can even influence customer orientation and relationship commitment in the purchasing of products and services (Hult et al., 1998). Further, leadership charisma can be assessed by language in general (Conger, 1991; Shamir et al., 2018), and more specifically around leaders' verbal communication style, rhetorical devices, and the use of figurative language and imagery (Willner, 1985).

Second, the text in speeches is an ideal form of content for applying, illustrating, and comparing a new approach to content analysis. The language in a leader's speech can demonstrate their shared values and collective identity (Hermanowicz and Morgan, 1999), which are components of charismatic leadership (Bligh et al., 2004a; Conger and Kanungo, 1987; 1988). Furthermore, speeches by leaders are often long enough for analysis and publicly available. Consequently, this paper can draw upon prior research on the rhetoric content of leadership speeches (e.g., Bligh et al., 2004a, Shamir et al., 2018) and existing dictionaries for the computer-aided approach are available (e.g., Bligh et al., 2004a). The speeches of leaders selected for this paper are based on an Inc. Magazine list of top commencement speeches by business leaders (Murphy, 2014). One male (Bill Gates) and two female (Sheryl Sandberg and Oprah Winfrey) business leaders are chosen. The speeches are analyzed for eight categories of charismatic speech (Bligh et al., 2004a; Shamir et al., 2018), as listed and described in Table 2.

Insert Table 2 about here.

Measures

To contrast and compare manual, computer-aided, and AI-enabled approaches to content analysis, each approach is assessed using three measures: reliability, validity, and efficiency. *Reliability* is the degree of agreement among coders (or within a single coder over time) when classifying content (Riffe et al., 2005). It is an indicator of a researcher's relative subjectivity in coding the content. An additional important aspect of content analysis reliability is category reliability, which is a coder's ability to formulate categories of content that represent key constructs (Kassarjian, 1977; Kolbe and Burnett, 1991). To avoid having to assess this aspect of reliability, the eight pre-formulated categories for speech charisma are used, as outlined in Table 2. Thus, focusing only on inter-coder reliability, this paper calculates and reports this measure as a ratio: total number of coding agreements between coders relative to the total number of coding decisions, using the eight pre-defined categories.

It is anticipated that reliability, when using manual content analysis, would be lowest due to errors of coder judgement and interpretation. The AI-enabled approach relies on the strength of its NLP and its knowledge base, but still requires manual coding of the keywords that are generated. With computer-aided content analysis, the coding is based on pre-defined and relatively fixed dictionaries, which eradicates the issue of inter-coder reliability. While this paper focuses on reliability as a measure of inter-coder agreement, inter-coder reliability assessments reflect the reliability of the coding protocol and a coder's use of this protocol (Lacy et al., 2015). Furthermore, even though computer-aided analysis is anticipated to offer high reliability, this is expected to be at the expense of validity (Su et al., 2017), which is the second of the three measures.

Validity is the degree to which the coding of content echoes the actual meaning of the concepts being measured (Babbie, 1998; Su et al., 2017). This measure reflects the soundness or quality of the analysis and the inferences made from the content, relative to some standard or reference. Validity is related to reliability in that validity can be seen as encompassing reliability (as a measure cannot be valid if it is not reliable) and accuracy (the extent to which the measuring procedure is free from bias) (Neuendorf, 2017). Validity refers both to the coding scheme that is developed and the coding decisions that are made about which content belongs in which category. How validity is established for speech charisma is discussed later on.

Content analysis validity can be limited to manifest content or extended to latent content. This assessment of manual, computer-aided, and AI-enabled approaches to content analysis evaluates the validity of each approach using both manifest and latent content. In other words, validity in this paper refers to the degree to which the language used in the speeches actually represents speech charisma. This evaluation compares how each approach varies in its ability to facilitate counting of content and understanding of complex meanings of the context and content. It is expected that the computer-aided approach will be superior to the manual approach for manifest content analysis, and that manual content analysis will excel over computer-aided for latent content (Conway, 2006; Matthes and Kohring, 2008). Thus, this paper seeks to determine how the AI-enabled approach performs in making sense of underlying latent content.

To establish validity, this paper follows Potter and Levine-Donnerstein (1999). To begin, using the eight categories of speech charisma, a coding scheme is developed to guide coders in the analysis of the content. The coding scheme comprises the list of categories, their definitions, and rules for identifying which words and phrases fit into each category, based not only on the theoretical categories in Table 2 but also on related content analysis studies of speech charisma (Awamleh and Gardner, 1999; Spangler et al., 2012). The coding scheme is the foundation upon which validity is determined, whether it be face or construct (Poole and Folger, 1981; Potter and Levine-Donnerstein, 1999). Face validity is the extent to which the "coding system is logically consistent and the categories clearly defined" (Folger et al., 1984, p. 137), which requires a theory – in this case, speech charisma and its eight theoretical categories. Construct validity is the extent to which a category is related to other categories in a way that is consistent with theory (Carmines and Zeller, 1979) and can be convergent (in which an expected relationship is found between categories) or discriminant (in which the expectation of no relationship is confirmed between two or more categories) (Neuendorf, 2017). A criticism of content analyses is that many use constructs and/or categories that have not been used before and have not been fully validated

(Janis, 1965; Neuendorf, 2017). This is one reason why this paper uses the previously validated construct of speech charisma that has both face and construct validity (Bligh et al., 2004b).

Next, the coding decisions are compared against a "standard" - the "correct" or "accurate" codes (Carlsmith et al., 1976; Wimmer and Dominick, 1991). If the coding matches the "standard for correct decision-making, then the coding is regarded as producing valid data" (Potter and Levine-Donnerstein, 1999, p. 266). Thus, when the coding is accurate, it is considered valid. For this paper, the standard used for validation (i.e., the baseline) is the result from manual coding after undergoing rigorous pre-testing of the coding scheme and training of the coders and follows Bligh et al. (2004b), in which manual coding is used to validate results from computer-aided content analysis.

The third measure is *efficiency*, which this paper defines as the extent to which an approach allows researchers to carry out content analysis quickly relative to other approaches. While the efficiency of any task can be numerically measured and compared, by determining the cost or energy expended to produce an output, this paper focuses on time. This focus on time is chosen because content analysis is a work-intensive activity that can consume significant researcher time, especially when dealing with the large volumes of digitized content. Also, the time it takes to complete an analysis is a relatively straightforward aspect to measure. Relative to manual coding, it is anticipated that computer-aided and AI-enabled approaches to content analysis will be quicker to carry out. Some researchers suggest that such technologies are capable of analyzing large volumes of content at great speed (Conway, 2006; Krippendorff, 2018), and that computer-aided and AI-enabled approaches not only assist in the efficient processing of large volumes of digital content, but also minimize the propensity for mistakes that human coders make during this very lengthy, tedious, and error-prone work (Nacos et al., 2009; Su et al., 2017).

Overall, it is not only each individual measure, but also the combination of the measures that is considered in this comparison of the three approaches. This paper is interested in how manual, computer-aided, and AI-enabled approaches vary with respect to reliability, validity, and efficiency, but also how their combinations will allow for a comparison of the general efficacy of the three approaches.

Application of the Content Analysis Approaches

This section reports on the three different approaches used to analyze the leadership speeches. Based on the traditional content analysis method, nine steps are followed in each approach (discussed below and summarized in Table 3). For each approach (manual, computer-aided, and AI-enabled), this paper specifically outlines and describes every step to allow for the replicability of this study. The steps for AI-enabled content analysis equate to a "roadmap" for how to use this approach to develop theory. Then, this paper explains how the assessment of reliability, validity, and efficiency are determined for each approach.

Manual content analysis is conducted following the steps recommended by Insch et al. (1997) and Neuendorf (2017). Step 1 identifies the research question and construct. The research question is "how charismatic are the speeches of leaders?" and the construct of speech charisma is selected (Bligh et al., 2004a; Shamir et al., 2018). Step 2 identifies the texts to be examined, which are the complete transcribed texts of commencement speeches for Bill Gates, Sheryl Sandberg, and Oprah Winfrey that are publicly available. Step 3 specifies the unit of analysis. Word sense or phrase is selected (which includes single words and phrases versus other possibilities of single words only, sentence, paragraph, or full speech) because human coders are able to discern the meaning of words or phrases based on the context. For instance, the phrase "Angel Network" is recognized as a collective noun rather than two separate nouns.

Step 4 specifies the categories: single versus multiple classification (whether a word or phrase can be assigned to only one or to more than one category), assumed versus inferred categories (deductive or inductive), and the use of existing content analysis dictionaries. For the manual content analysis, a word or phrase can be assigned to only one category. This paper uses the same categories of speech charisma identified by Bligh et al. (2004a) and Shamir et al. (2018). Because the manual content analysis uses an existing coding scheme (Bligh et al., 2004a) with exemplars of words and phrases for each category, existing content analysis dictionaries are not required.

Step 5 generates a sample coding scheme. The coding scheme by Bligh et al. (2004a) is used as a basis for the coding form that is developed. In Step 6, three researchers pretest the coding scheme by coding the three speeches. In Step 7, based on the pretest, the coding scheme and coding form are revised or "purified", with further exemplars of words and phrases that are specific to the context of the speeches (e.g., faculty, graduates). See Appendix A for the final coding scheme and coding form. Step 8 is the actual content analysis, in other words, the data collection. The principal investigator, after augmenting the coding scheme and leading the pretest, trains two researchers who are not involved in the pretesting or revision of the coding scheme, as recommended by Lacy et al. (2015). Manual coding is then conducted, with each speech being coded by two researchers. This involves reading each speech once, then focusing on coding one category with one pass (or more) through the speech before coding the next category. In total, each speech is read at least ten times by a researcher and the coding of each speech takes about two days to complete. To adjust for different speech lengths, the results are reported as percentages. The total score for the charisma in each speech is determined by calculating the sum of the percentages for each of the eight categories of speech charisma. The manual coding indicates that Gates' speech demonstrates more speech charisma than Sandberg or Winfrey.

Step 9 calculates reliability, validity, and efficiency of each content analysis approach. Reliability is calculated as inter-coder agreement for each speech before the differences between the coders are discussed. Due to the length of each speech, at over 3,000 words, there are inaccuracies and inconsistencies (this issue is discussed further in the Results section). After discussion about how the different words and phrases are being interpreted by each coder, the speeches are coded for a second time. The differences between the coders are again checked, resulting in differences of only 4 percent of the coding. The remaining differences are discussed and agreed to, resulting in a final coding that is then used for validity testing. Validity for each content analysis approach is determined by comparing the results for each category of speech charisma against the final results from the manual coding. The final scores from the manual coding (Appendix B) are then used was the baseline for evaluating the validity of the other approaches, consistent with Bligh et al. (2004b) and Neuendorf (2017). Efficiency is reported, comprising the amount of time to complete Steps 5 to 8. See Table 3 for a summary of the steps.

Insert Table 3 about here.

For the *computer-aided content analysis*, two different software programs are used to examine speech charisma using the same three speeches: DICTION 7 and LIWC 2015. DICTION is used extensively in the social sciences (Hart and Curry, 2016) and has 33 different dictionaries containing over 10,000 search words. The search words are single words only and statistical weighting is applied to partially correct for context (Hart, 2000). Users can also create and share their own dictionaries, a feature that is not used in this instance since well-established dictionaries for speech charisma already exist.

Steps 1 to 4 are largely identical to the steps for the manual content analysis (Table 3). The exception is Step 3 in which the unit of analysis for DICTION is individual words only, as DICTION can search only by individual word. For Step 4, single (versus multiple) classification is used, as DICTION does not duplicate words between its dictionaries. For Step 5, the coding scheme developed by Bligh et al. (2004a) and the pre-installed program dictionaries that relate to the categories of speech charisma are used as the base but are updated due to a new version of DICTION being employed (DICTION 7 for this paper versus DICTION 5 being used by Bligh et al., 2004a). The specific dictionaries that are used for each category are shown in Appendix C. Step 6 pretests DICTION using the standard dictionaries and it is determined that there is no need to adjust the coding scheme, thus Step 7 is not undertaken. In Step 8, each of the speeches are run through DICTION and the results shown in Appendix D.

Linguistic Inquiry and Word Count (LIWC) is used widely in psychology and linguistics (Tausczik and Pennebaker, 2009) and was designed to study the emotional, cognitive, and structural components in verbal and written speech using internal dictionaries comprising empirically-defined psychological and structural categories (Pennebaker et al., 2015; Tumasjian et al., 2010). The coding scheme used for LIWC2015 is based on its pre-installed dictionaries with almost 6,400 words, word stems, and select emoticons (Pennebaker et al., 2015). LIWC processes single words only but words can belong in more than one dictionary, which means that words could be counted multiple times (unlike DICTION, in which a word can occur in only one dictionary). Steps 1 to 4 are identical to the steps undertaken for the manual approach. For Step 5, the specific default dictionaries are selected to best match the categories of speech charisma

(Appendix E). Then, Step 6 pretests LIWC using the selected dictionaries. After reviewing the results (Appendix F), it is determined that there are limitations in LIWC's default dictionaries for speech charisma. For example, some words are counted in an inappropriate dictionary (e.g., "class" is counted in the "work" dictionary when it should be counted in the "collective focus" category since "class" refers to a group of graduates) and proper nouns such as names of people, place names (e.g., Harvard), geographical names (e.g. Nashville), and brand names (e.g., Oreos) are not captured in the default dictionaries. These limitations are addressed through using the feature in LIWC that allows users to create custom dictionaries. Thus, in Step 7, three researchers manually code the speeches and use the coding results to create custom dictionaries (i.e., coding schemes) to augment the default dictionaries in four of the categories. These custom dictionaries are added to: (1) "Collective focus" with place names, geographical names, and single nouns connoting plurality e.g. crowd, family. (2) "Followers' worth" with affirmations or abstract values. (3) "Values and moral justification", as there are no dictionaries for values or moral justifications other than one dictionary for religion. To avoid double counting of words, "religion" is deactivated and a custom dictionary for values and moral justification is added. (4) "Action" with words of inactivity. In Step 8, the speeches are run through LIWC with the additional custom dictionaries. For both LIWC and DICTION, reliability, validity, and efficiency are calculated in Step 9.

The *AI-enabled content analysis* approach uses IBM Watson, specifically, the IBM Watson NLP product inside IBM Bluemix, which is a "platform as a service" product developed to build, run, deploy, and manage applications on the cloud. In other words, one uploads the documents into the web interface, without the need to install, configure, and maintain a program or an app. For this product, there are two options. The first option is a free version of IBM

Watson NLP that allows researchers to load the text into the NLP engine and generate a list of up to 50 keywords and eight JavaScript Object Notation (JSON) files: sentiment, emotion, keywords, entities, categories, concept, syntax, and semantic roles. The second option is IBM Watson Explorer, which identifies keywords, phrase constituents, sentiments, emotions, entities, and more. It generates a comma-separated values (CSV) file that can be used for further manipulation. This paper uses the IBM Watson Explorer option to generate up to 500 keywords.

As with the other two approaches, assigning meaning to written text, i.e., creating a semantic representation of the text, is a challenging task given the ambiguity inherent in natural language resulting from contextual circumstances, linguistic styles, or dialog history. A key task in natural language understanding involves analyzing the syntax (i.e., the structure of sentences), semantics (i.e., the relationship between words, phrases, and symbols), and pragmatics (i.e., the context in which words or phrases are used in natural language) (Gill, n.d.). The difference compared to the other approaches is that AI systems can rely on machine learning and understand the context better to extract meaning from text. It can more easily separate the meaning of homonyms - words with the same spelling and pronunciation, but different meanings (e.g., to *book* a criminal versus to *book* a hotel room), homophones - words that share the same pronunciation, regardless of how they are spelled (e.g., to, too, two) and homographs - words that share the same spelling, regardless of how they are pronounced (e.g., to *tear* up versus to *tear* down). Much like computer-aided content analysis, AI applications use a lexicon (a vocabulary) and a set of grammar rules coded into its procedures. However, AI applications then use statistical models and machine learning to apply these rules and determine the most-likely meaning of what was said.

The process of AI-enabled content analysis involves the same Steps 1-4 as for manual and computer-aided content analysis (Table 3). In Step 5, IBM Watson extracts keywords for all three speeches, then one researcher exports the results as a CSV file. The keywords in the CSV file are then connected to the speech charisma categories, just as the process in the manual and computer-aided approaches. To do this, a coding scheme is produced that include the list of keywords in the first column and their frequency in the second column. In Step 6, two researchers compare the coding scheme against the pre-defined categories of speech charisma and determine that it needs to be purified. In Step 7, the coding scheme is purified, similar to the process for manual coding. A third column is added to the coding scheme in which the respective categories of speech charisma are matched to each relevant keyword. Of course, not all keywords are matched to a category of speech charisma. Step 8 is similar to the process for manual coding, with the exception that up to 500 keywords are coded, compared to over 3,000 words for manual coding. It is important to note that Step 8 requires manual coding to connect each keyword (known as mapping) to one of the eight speech charisma categories. In Step 9, inter-coder reliability, validity, and efficiency are calculated. Appendix G shows the results for speech charisma for all three speeches using IBM Watson.

Researchers have two options for analyses. The first option is to stop at Step 8. AI will have helped with the keyword identification, but the rest of the content analysis remains a relatively manual process. The second option is to feed the coded table that mapped the keywords to the respective speech charisma categories back into IBM Watson to train the machine-learning algorithm. Much like the medical example (The Japan Times, 2018) mentioned in the previous section, where oncologists first needed to teach the AI about healthy versus unhealthy tissue samples, researchers can improve the machine learning and NLP. This is particularly valuable when other researchers might want to conduct similar analyses, when datasets are large, when many different keywords are used, or when follow-up studies are required. This training of the machine-learning algorithm would mean that manual coding would eventually not be necessary, as the IBM Watson output would encompass the speech charisma categories and the analysis will continue to become more accurate.

This training of the machine learning process is needed mainly for keyword analyses. In cases of sentiment (positive/negative), emotion (e.g., anger, joy, fear), entities (IBM Watson identifies entities as people, companies, organizations, cities, etc.), categories (IBM Watson provides a hierarchy, for example "parents/children" or "education/homework"), concepts (e.g., in the Gates speech, it calculates three concepts "Harvard University", "Poverty" and "World"), syntax (e.g., "former" is classified as an adjective, "President" is classified as a pronoun), and semantic roles (e.g., where it parses the sentences into subject, verb, and object), machine learning is already built into the cloud-based IBM Watson. Every time IBM Watson is used for text analysis, by any researcher, the algorithm improves for all future analyses.

Results

This section summarizes the results from comparing manual, computer-aided, and AI-enabled approaches to content analysis. It highlights how the interpretation of the coding scheme requires considerably different work, which ultimately contributes to the efficacy, especially the efficiency, of each approach. Then, the three measures of reliability, validity, and efficiency are compared for each of the three approaches.

Interpretation of Coding Scheme

In *manual content analysis*, human coders make interpretation decisions every time they code a word or phrase. While this is a time-consuming task, humans are able to catch misspelled words, anomalies in spelling, or make sense of abbreviations. In *computer-aided content analysis*, researchers make interpretation decisions when deciding which dictionaries best correspond to each category (e.g., of speech charisma). However, the reliance on dictionaries for coding has some drawbacks, as some categories are not well covered by the default dictionaries. With LIWC, for example, the "temporal orientation" category is well represented in the default dictionaries; however, for the category of "collective focus", there is no dictionary for single nouns connoting plurality (e.g., crowd, family). This requires the researchers to add custom dictionaries, as described earlier, which is a time-consuming task. Similarly, spelling mistakes, abbreviations, or other spelling anomalies need to be corrected or else they are not captured by the dictionaries. With *AI-enabled content analysis*, or rather with using an untrained AI content analysis tool (e.g., IBM Watson that is not trained for speech charisma), there is no interpretation of the coding scheme before running the program. Rather, the interpretation decisions are made when researchers map keywords to categories, much like in the manual case.

Approach Reliability

Reliability is the degree of agreement among coders and varies significantly for the three different approaches. In *manual content analysis*, reliability is low, since instances of mis-categorization of words or phrases (wrong category), miscounting words, and missing words altogether are quite high. Furthermore, inconsistencies in interpreting the coding scheme between the different coders or in interpretation of particular words or phrases lead to relatively low inter-coder reliability (64 percent across the three speeches).

In *computer-aided content analysis*, once the appropriate dictionaries (default and custom) are selected and applied, coding reliability is 100 percent. This means that, regardless of who uses the content analysis software program to analyze a particular speech, the matching of the text to the categories in the selected dictionaries will be exactly the same each time.

Somewhat surprising is the fact that *AI-enabled content analysis without training* of the machine-learning algorithm is less reliable than computer-aided approaches. This is because, while the generation of relevant words and phrases is exactly the same when the same speech is processed through IBM Watson, these words and phrases (keywords) still must be coded manually (since IBM Watson is not pre-trained for speech charisma). This means that the manual coding of the IBM Watson output has the same shortcomings as manual coding, albeit with up to 500 keywords to code versus over 3,000. Inter-coder reliability is 96 percent across the three speeches and this is considered high reliability. *AI-enabled content analysis* with training of the machine-learning algorithm on the other hand, relies on a growing knowledge base, such that each additional document benefits from the improved dictionaries and requires less human intervention (e.g., less manual coding), leading to even higher levels of reliability when classifying content. Table 4 summarizes the results from the three approaches.

Insert Table 4 about here.

Approach Validity

In *manual coding*, the validity, or the degree to which the coding of content echoes the actual meaning of the concepts being measured, is high. Researchers can code both manifest and latent content. Moreover, they can place words and phrases into context based on the meaning of the description of the category, even without a dictionary of words. To achieve high validity, coders discuss the differences and then re-code the speeches. The re-coded speeches are then further discussed with a third researcher to agree to the final coding. As a result, the manual coding results are used as the 'gold-standard': the baseline for evaluating the other approaches.

In comparison, *computer-aided content analysis* shows low to moderate validity, with LIWC varying 40 percent from the baseline and with DICTION varying 8 percent from the baseline. LIWC and DICTION are able to code manifest content only. They code single words only, which for instance means that “Angel Network” is counted twice, with “angel” under “values and moral justification” and “network” under “collective focus”. Additionally, proper nouns are not included in the standard dictionaries (e.g., Harvard, Nashville, Oreos) unless they are added to custom dictionaries. There are differences between LIWC and DICTION. LIWC allows words to be counted in more than one dictionary, so certain words are counted multiple times. DICTION’s dictionaries are mutually exclusive and the program uses statistical weighting to determine in which dictionary to place a word. The difference in validity may be due to the size of the dictionaries (DICTION has 10,000 words while LIWC has 6,400) and the ability to match the dictionaries to the categories of speech charisma.

The validity of *AI-enabled content analysis*, compared to the manual approach, is high, varying only 4 percent from the baseline. IBM Watson can process both manifest and latent content and can process keywords in the context of the sentence. However, when using the untrained version of IBM Watson, the keywords are not specific to the coding scheme. which means that manual coding of the keywords is necessary, with the same benefits as the manual coding approach.

Approach Efficiency

Efficiency refers to the extent to which an approach allows researchers to carry out content analysis quickly. Efficiency is assessed, not as an absolute measure, but relative to other approaches, as with validity. As expected, *manual content analysis* exhibits low efficiency, as analyzing content without any technical support is the most uneconomical use of time of the three approaches. As shown in Table 3, Step 6 pretests and purifies the coding scheme. Step 6 requires 96 hours (3 speeches x 16 hours per speech x 2 researchers) while Step 7 requires 4 hours. Step 8 necessitates 114.75 hours: reading each speech and familiarizing oneself with the coding scheme and coding form for 6 hours, coding for 96 hours (3 speeches x 16 hours per speech x 2 researchers), checking for 12 hours (3 speeches x 4 hours per speech), and summarizing of results for 0.75 hour. The grand total for the manual content analysis of three speeches is therefore 214.75 hours.

The efficiency of *computer-aided content analysis* is moderate to high, depending on the program used. For DICTION, the total is 2 hours, which is considered high efficiency. Step 5: 1 hour for the comparison and selection of the existing dictionaries to the categories of speech charisma. Step 6: 1 hour to run the speeches through the program and examine the results. Step 7: no revisions. Step 8: seconds to process the speeches through the program. LIWC is less efficient, requiring 102 hours, which is considered moderate efficiency. After determining that the dictionaries in LIWC do not match the categories of speech charisma well enough, three researchers jointly create new custom dictionaries to augment the default dictionaries to better match the categories of speech charisma. Steps 5 and 6 are the same as for DICTION: 2 hours. Step 7: 100 hours comprising 3 speeches x 16 hours per speech x 2 researchers plus 4 hours for consolidation. Step 8: seconds to process the speeches.

The efficiency of *AI-enabled content analysi*s is moderate, requiring 107 hours. This is mainly because IBM Watson is not pre-trained for speech charisma. Step 5: seconds to generate a set of keywords. Step 6: 1 hour for the researchers to compare the Excel tables for each speech. Step 7: 100 hours for the researchers to jointly revise the coding scheme to prepare for manual coding, comprising 3 speeches x 16 hours per speech x 2 researchers plus 4 hours for consolidation. Step 8: 6 hours comprising 3 speeches x 1 hour per speech x 2 researchers. If *AI-enabled content analysis with training* for machine learning were undertaken, the manual coding would not be as time consuming, as explained earlier.

In summary, manual content analysis has low reliability, high validity, and low efficiency. Of all the approaches, manual content analysis has the highest validity because humans can best detect context and meaning (at this point in time), but has the lowest accuracy. It is also the most time-consuming of all the approaches. Comparatively, computer-aided content analysis has high reliability, low to moderate validity (depending on the software program used), and high efficiency. This approach is the fastest and most reliable of all the approaches but does not detect context and meaning that is inherent in latent content. AI-enabled content analysis has high reliability, high validity, and moderate efficiency. It has slightly lower reliability than computer-aided and slightly lower validity of manual but is several-fold faster than manual coding.

Discussion

While AI is increasingly being developed and used to mimic the cognitive functions that humans use for learning and problem solving, its employment for content analysis studies has not yet taken off. This paper aims to introduce what AI is and to explain and demonstrate how it can be used to conduct content analysis in marketing and other social science fields. Consequently, the first contribution is a review of what content analysis is and its continued importance to marketing research, particularly with the explosive growth of natural language-based, user-generated content. Table 1 and the related discussion identifies the breadth of communication content that have been analyzed in marketing studies and the continued use of manual content analysis, even with the availability of computer-aided content analysis tools.

While both manual and computer-aided content analyses have been used in marketing research, this review also reveals the absence of AI-enabled content analysis studies.

The second contribution is a roadmap for using AI-enabled content analysis. A brief non-technical introduction to AI is provided and its promise for content analysis in marketing research is outlined. Then, Table 3 and the related discussion explain how AI-enabled content analysis adheres to and differs from the established steps for carrying out content analysis. Specifically, AI enabled-content analysis differs in Steps 5-9. With Step 5, NLP is used to generate a set of keywords. Step 6 for AI-enabled content analysis compares the generated keywords to the coding scheme, similar to computer-aided content analysis. Step 7's revision of the coding scheme is similar to manual content analysis. Step 8 is when researchers manually code the reduced set of keywords due to AI's ability to reduce the content to the most relevant, making coding substantially more efficient (taking 1 hour for each speech compared to over 19 hours for the fully manual approach). Consequently, it is clear that when AI-enabled content analysis is applied, it offers a powerful, versatile, and replicable technique for dealing with content that is challenging and rewarding in terms of its volume, variety, and velocity.

The third contribution is that manual, computer-aided, and AI-enabled content analysis are applied and compared using leadership speeches to juxtapose the application of the different approaches and assess each approach for reliability, validity, and efficiency rather than to develop theory about leadership speeches. As reported in Section 5, it is found that AI-enabled content analysis has high reliability, high validity, and moderate efficiency compared to manual and computer-aided. It is important to note though that if the AI-enabled approach is trained to analyze charisma in leadership speeches by examining hundreds of speeches, one would expect the efficiency to be high when analyzing further individual speeches. Manual content analysis relative to AI-enabled has substantial drawbacks for reliability and efficiency while computer-aided relative to AI-enabled has drawbacks for validity.

Building on these contributions, this paper now highlights some of the advantages and disadvantages of AI-enabled content analysis, relative to manual and computer-aided approaches, for marketing research and other social science fields.

Advantages of AI-Enabled Content Analysis

First, one of the main advantages of AI-enabled content analysis over the manual and computer-aided approaches lies in the nature of the data that can be analyzed. The computational and learning capabilities of AI-enabled content analysis make it suitable for handling "big data" which is data that is extremely high in volume (i.e., the amount of content), variety (i.e., the different types of content), and velocity (i.e., the rate and direction at which content is generated) (Dabirian et al., 2017; Dabirian et al., 2019; Gamdomi and Haider, 2015; Paschen et al., 2019). The science and profession of marketing are increasingly able to capture big data due to the vast amount of marketing-related and digitized content that is produced by consumers and firms, typically via social media platforms (Kietzmann et al., 2010). Such big data is "naturally occurring data" (Muller et al., 2016. p. 292) in that it is generated with no specific research purpose in mind and therefore is suited to inductive examinations and theory building offered by AI-enabled content analysis.

Second, the computational strength of AI allows researchers to move away from traditional deductive investigations that would struggle with the unstructured nature of big data to investigations that are more inductive and abductive in nature (Hannigan et al, 2019; Wagner-Pacifici et al., 2015). It enables such investigations with computational features that provide insight into the sentiment tone of text, image, audio, and video content, which in turn facilitates cognition about the meaning of content, much as a human would. Furthermore, like simulation methods, AI programs can act like a "computational laboratory in which researchers can systematically experiment (e.g., unpack constructs, relax assumptions, vary construct values, add new features) in a controlled setting to produce new theoretical insights" (Davis et al., 2007, p. 495). This is particularly the case when the content changes over time and has significant velocity.

Third, AI-enabled content analysis allows researchers to move beyond dictionary-centric content analysis, to more of a process of "rendering" where researchers make contributions by efficiently switching between contrasting data and theory. Rendering produces knowledge by iterating between selecting and trimming content and applying algorithms (Hannigan et al. 2019). In other words, AI-enabled content analysis can be much more open-ended than manual and computer-aided approaches, which helps to delineate the structure and meaning of the content. AI helps researchers to sample content, identify categories, and determine causal links between the categories. For marketing research, this aspect of AI-enabled content analysis would be beneficial to studies dealing with complex and evolving content such as text analysis of customer reviews (Büschken and Allenby, 2016; Lee, 2014; Lee and Bradlow, 2011) or understanding how consumers view brands through social tags (Nam et al., 2017). Furthermore, in contrast to survey-based studies that by design have pre-determined variables and are a typical source of data for research, AI-enabled content analysis can handle similar and greater volumes of more open-ended data in a jointly inductive and qualitative way (Tonidandel et al., 2018, Hannigan et al., 2019).

Disadvantages of AI-Enabled Content Analysis

Like any research method, AI-enabled content analysis also has limitations. The first is the cost or availability of the AI technology. This paper is fortunate to use the IBM Watson NLP product inside IBM Bluemix. However, this technology and similar ones (e.g., Amazon AWS Lex and Salesforce Einstein) are not currently widely available to university researchers, or only at a significant monetary cost. At this time, not many universities have access to sophisticated tools, such as IBM Watson Explorer. However, this is changing. Free and on-demand, pay-per-use versions of these tools are vastly available and can be used for research. However, as with most computing technologies, it is anticipated that both the cost and availability of the AI technologies will be less prohibitive over time, resulting in an accessible and powerful resource for researchers.

A second limitation of AI-enabled content analysis is that its capabilities can be misunderstood. The appeal and hype for AI are such that it can be perceived as a technology with algorithms that miraculously yield meaningful inferences and theory. This is not the case, at least at this point in time. Like most technologies, it is a tool used by humans to support their activities. As shown in Table 3 and discussed earlier, researchers still need to make judgements and decisions for technical issues including (i) how to select, collect, and submit the content; (ii) which keywords are mapped to each category; and (iii) which AI algorithms to use. Furthermore, the meaning and linkages between observations and inferences are decided by researchers. For without appropriate researcher input and guidance, there is a risk of producing decontextualized results based on overly simple indices and counts (Prein and Kelle, 1995). Thus, AI-enabled content analysis does not replace the established steps of content analysis; it automates and augments them.

Limitations

Of course, this paper has its limitations. The first is that only one type of content is analyzed (leadership speeches) and a limited sample is used (three speeches), restricting generalizability. Leadership speeches, as good examples of long-form text content, are chosen for the reasons stated earlier. Other forms of content such as short-form text (e.g., social media text), audio, images, videos, and hypertext and a larger sample could also be studied. Especially with audio-visual content, the ability of AI-enabled content analysis to interpret facial expressions of speakers, tone of voice, etc. promises previously unthinkable richness of data and research insights.

The second limitation is that this paper uses a pre-defined coding scheme because the aim of this paper is to illustrate AI-enabled content analysis compared to manual and computer-aided approaches. However, AI can also be used to develop theory such as through topic modeling to identify themes and through identifying coding schemes in classic content analysis.

Conclusion

This paper opens by highlighting the importance of content analysis for theory development in marketing research and the promise of AI to help with this task. The paper then sets out to introduce, apply, and compare AI-enabled content analysis so researchers will know when and how to use the approach. This is achieved via three contributions. First is the review of studies in marketing that have used content analysis to highlight the marketing phenomena examined, the type of content analyzed, and the content analysis approach used (Table 1). The second is a roadmap of steps for AI-enabled content analysis using NLP for generating keywords that can be mapped to content categories (Table 3). The third is the application of AI-enabled, manual, and computer-aided approaches, illustrated through the content of leadership speeches, to assess and compare the reliability, validity, and efficiency of each approach. It is hoped that these contributions demonstrate the utility and boundaries of AI-enabled content analysis and motivate fellow researchers to use AI-enabled content analysis for studying marketing and other social science phenomena.

References

Awamleh, R. and Gardner, W.L (1999), "Perceptions of leader charisma and effectiveness: The effects of vision content, delivery, and organizational performance", *The Leadership Quarterly*, Vol. 10 No. 3, pp. 345-373.

Babbie, E.R. (1998), *The Practice of Social Research (Vol. 112)*, Belmont, CA: Wadsworth.

Bendisch, F., Larsen, G. and Trueman, M. (2013), "Fame and fortune: A conceptual model of CEO brands", *European Journal of Marketing*, Vol. 47 No. 3/4, pp. 596-614.

Bitner, M.J., Booms, B.H. and Tetreault, M.S. (1990), "The service encounter: diagnosing favorable and unfavorable incidents", *Journal of Marketing*, Vol. 54 No. 1, pp. 71-84.

Bligh, M.C., Kohles, J.C. and Meindl, J.R. (2004a), "Charisma under crisis: Presidential leadership, rhetoric, and media responses before and after the September 11th terrorist attacks", *The Leadership Quarterly*, Vol. 15 No. 2, pp. 211-39.

Bligh, M.C., Kohles, J.C. and Meindl, J.R. (2004b), "Charting the language of leadership: A methodological investigation of President Bush and the crisis of 9/11", *Journal of Applied Psychology*, Vol. 89 No. 3, pp. 562-74.

Bryman, A. (1992), *Charisma and Leadership in Organizations*. Sage Publications.

Büschken, J. and Allenby, G.M. (2016), "Sentence-based text analysis for customer reviews", *Marketing Science*, Vol. 35 No. 6, pp. 953-75.

Carlsmith, J.M., Ellsworth, P.C., and Aronson, E. (1976), *Methods of Research in Social Psychology*, Reading, MA: Addison-Wesley.

Carmines, E.G. and Zeller, R.A. (1979), *Reliability and Validity Assessment* (Vol. 17). Sage Publications.

Choi, S., Lehto, X.Y. and Morrison, A.M. (2007), "Destination image representation on the web: Content analysis of Macau travel related websites", *Tourism Management*, Vol. 28 No. 1, pp. 118-29.

Conger, J.A. (1991), "Inspiring others: The language of leadership", *Academy of Management Perspectives*, Vol. 5 No. 1, pp. 31-45.

Conger, J.A. and Kanungo, R.N. (1987), "Toward a behavioral theory of charismatic leadership in organizational settings", *Academy of Management Review*, Vol. 12 No. 4, pp. 637-47.

Conger, J.A. and Kanungo, R.N. (1988), *Charismatic Leadership: The Elusive Factor in Organizational Effectiveness*. Jossey-Bass.

Conway, M. (2006), "The subjective precision of computers: A methodological comparison with human coding in content analysis", *Journalism and Mass Communication Quarterly*, Vol. 83 No. 1, pp. 186–200.

Courtney, A.E. and Lockeretz, S.W. (1971), "A woman's place: An analysis of the roles portrayed by women in magazine advertisements", *Journal of Marketing Research*, Vol. 8 No. 1, pp. 92-95.

Dabirian, A., Kietzmann, J.H., and Diba, H. (2017), "A great place to work!? Understanding crowdsourced employer branding." *Business Horizons,* Vol. 60 No. 2, pp. 197-205.

Dabirian, A., Paschen, J., and Kietzmann, J. (2019), "Employer Branding: understanding employer attractiveness of IT companies." *IT Professional,* Vol. 21 No. 1, pp. 82-89.

Davis, J.P., Eisenhardt, K.M. and Bingham, C.B. (2007), "Developing theory through simulation methods", *Academy of Management Review*, Vol. 32 No. 2, pp. 480-99.

Deffner, G. (1986), "Microcomputers as aids in Gottschalk-Gleser rating", *Psychiatry Research*, Vol. 18 No. 2, pp. 151-59.

Demirkan, H. and Delen, D. (2013), "Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud", *Decision Support Systems*, Vol. 55 No. 1, pp. 412-21.

Diomaiuta, C., Mercorella, M., Ciampi, M. and De Pietro, G. (2017), "A novel system for the automatic extraction of a patient problem summary", In *2017 IEEE Symposium on Computers and Communications (ISCC)*, pp. 182-186. IEEE.

Dong, B., Sivakumar, K., Evans, K.R. and Zou, S. (2015), "Effect of customer participation on service outcomes: the moderating role of participation readiness", *Journal of Service Research*, Vol. 18 No. 2, pp. 160-76.

Dowling, G. R. and Kabanoff, B. (1996), "Computer-aided content analysis: what do 240 advertising slogans have in common?" *Marketing Letters*, Vol. 7 No. 1, pp. 63-75.

Duriau, V.J., Reger, R.K. and Pfarrer, M.D. (2007), "A content analysis of the content analysis literature in organization studies: Research themes, data sources, and methodological refinements", *Organizational Research Methods*, Vol. 10 No. 1, pp. 5-34.

Ferber, R. and Wales, H.G. (1958), "The effectiveness of pharmaceutical advertising: A case study", *Journal of Marketing*, Vol. 22 No. 4, pp. 398-407.

Fiol, C.M., Harris, D. and House, R. (1999), "Charismatic leadership: Strategies for effecting social change", *The Leadership Quarterly*, Vol. *10 No.* 3, pp. 449-82.

Folger, J. P., Hewes, D. E., and Poole, M. S. (1984), "Coding social interaction", in Dervin, B. and Voigt, M.J. (Eds.), *Progress in Communication Sciences, Volume IV,* Norwood, NJ: Ablex, pp. 115-61.

Gandomi, A. and Haider, M. (2015), "Beyond the hype: Big data concepts, methods, and analytics", *International Journal of Information Management*, Vol *35 No.* 2, pp. 137-44.

Gill, N. S. (n.d.), "Overview of Artificial Intelligence and Natural Language Processing", available at: https://www.upwork.com/hiring/for-clients/artificial-intelligence-and-natural-language-processing-in-big-data/ (retrieved 11 December 2017).

Gilly, M.C. (1988), "Sex roles in advertising: A comparison of television advertisements in Australia, Mexico, and the United States", *Journal of Marketing*, Vol. 52 No. 2, pp. 75-85.

Govers, R. and Go, F.M. (2004), "Projected destination image online: Website content analysis of pictures and text", *Information Technology & Tourism*, Vol. 7 No. 2, pp.73-89.

Gross, B.L. and Sheth, J.N. (1989), "Time-oriented advertising: A content analysis of United States magazine advertising, 1890–1988", *Journal of Marketing*, Vol. 53 No. 4, pp. 76-83.

Hannigan, T., Haans, R.F.J., Vakili, K., Tchalian, H., Glaser, V., Wang, M., Kaplan, S. and Jennings, P.D. (2019), "Topic modeling in management research: Rendering new theory from textual data", *Academy of Management Annals*, May 2019.

Harris, P., Kolovos, I. and Lock, A. (2001), "Who sets the agenda?-An analysis of agenda setting and press coverage in the 1999 Greek European elections", *European Journal of Marketing*, Vol. 35 No. 9/10, pp. 1117-35.

Harrison-Walker, L.J. (2001), "E-complaining: A content analysis of an Internet complaint forum", *Journal of Services Marketing*, Vol. 15 No. 5, pp. 397-412.

Hart, R.P. (2009), *Campaign Talk: Why Elections Are Good for Us*. Princeton University Press.

Hart, R.P. and Curry, A.L. (2016), "The third voice of American politics", *Presidential Studies Quarterly*, Vol. 46 No. 1, pp. 73-97.

Helgeson, J.G., Kluge, E.A., Mager, J. and Taylor, C. (1984), "Trends in consumer behavior literature: a content analysis", *Journal of Consumer Research*, Vol. 10 No. 4, pp. 449-54.

Herbes, C. and Ramme, I. (2014), "Online marketing of green electricity in Germany—A content analysis of providers' websites", *Energy Policy*, Vol. *66* March, pp. 257-66.

Hermanowicz, J.C. and Morgan, H.P. (1999), "Ritualizing the routine: Collective identity affirmation", In *Sociological Forum*, Vol. 14 No. 2, pp. 197-214, Kluwer Academic Publishers-Plenum Publishers.

High, R. (2012), *The Era of Cognitive Systems: An Inside Look at IBM Watson and How It Works.* IBM Corporation. Available at: https://www.ibm.com/developerworks/community/blogs/efc1d8f5-72e5-4c4f-99df-e74fccea10ca/resource/White%20Papers/TheEraofCongnitiveSystemsAnInsideLookatIBMWatsonandHowItWorkswhitepaper.pdf?lang=en (accessed 22 February 2019).

Hult, T.M., Ferrell, O.C., and Schul, P.L. (1998), "The effect of global leadership on purchasing process outcomes", *European Journal of Marketing*, Vol. 32 No. 11/12, pp. 1029-50.

Humphreys, J.H. (2002), "Transformational leader behavior, proximity and successful services marketing", *Journal of Services Marketing*, Vol. 16 No. 6, pp. 487-502.

Insch, G.S., Moore, J.E. and Murphy, L.D. (1997), "Content analysis in leadership research: Examples, procedures, and suggestions for future use", *The Leadership Quarterly*, Vol. 8 No. 1, pp. 1-25.

Janis, I. (1965), "The problem of validating content analysis" *The Content Analysis Reader*, pp. 358-66.

Jose, A. and Lee, S.M. (2007), "Environmental reporting of global corporations: A content analysis based on website disclosures", *Journal of Business Ethics*, Vol. 72 No. 4, pp. 307-21.

Kassarjian, H.H. (1977), "Content analysis in consumer research", *Journal of Consumer Research*, Vol. 4 No. 1, pp. 8-18.

Kietzmann, J.H., Hermkens, K., McCarthy, I.P. and Silvestre, B.S. (2011), "Social media? Get serious! Understanding the functional building blocks of social media", *Business Horizons*, Vol. 54 No. 3, pp. 241-51.

Kolbe, R.H. and Albanese, P.J. (1996), "Man to man: A content analysis of sole-male images in male-audience magazines", *Journal of Advertising*, Vol. 25 No. 4, pp. 1-20.

Kolbe, R.H. and Burnett, M.S. (1991), "Content-analysis research: An examination of applications with directives for improving research reliability and objectivity", *Journal of Consumer Research*, Vol. 18 No. 2, pp. 243-50.

Krippendorff, K. (2018), *Content Analysis: An Introduction to its Methodology*. Sage Publications.

Lacy, S., Watson, B.R., Riffe, D. and Lovejoy, J. (2015), "Issues and best practices in content analysis", *Journalism and Mass Communication Quarterly*, Vol. 92 No. 4, pp. 791-811.

Lee, L.W. (2014), "Advice from creative consumers: a study of online hotel reviews", *International Journal of Technology Marketing*, Vol 9 No. 1, pp. 53-71

Lee, T.Y. and Bradlow, E.T. (2011), "Automated marketing research using online customer reviews", *Journal of Marketing Research*, Vol. 48 No. 5, pp. 881-94.

Leonidou, C.N. and Leonidou, L.C. (2011), "Research into environmental marketing/ management: a bibliographic analysis", *European Journal of Marketing*, Vol. 45 No. 1/2, pp. 68-103.

Lewis, S. C., Zamith, R. and Hermida, A. (2013), “Content analysis in an era of big data: A hybrid approach to computational and manual methods”, *Journal of Broadcasting and Electronic Media*, Vol. 57 No. 1, pp. 34–52.

Manning, C.D. and Schütze, H. (1999), *Foundations of Statistical Natural Language Processing*. MIT Press.

Matthes, J. and Kohring, M. (2008), “The content analysis of media frames: Toward improving reliability and validity”, *Journal of Communication*, Vol. 58 No. 2, pp. 258–79. doi:10.1111/j.1460-2466. 2008.00384.x

McCarthy, I.P. and Ruckman, K. (2017), “Licensing speed: Its determinants and payoffs”, *Journal of Engineering and Technology Management*, Vol. 46 (October), pp. 52-66.

Miller, D.W. and Toman, M. (2016), “An analysis of rhetorical figures and other linguistic devices in corporation brand slogans”, *Journal of Marketing Communications*, Vol. 22 No. 5, pp. 474-93.

Morris, R. (1994), “Computerized content analysis in management research: A demonstration of advantages and limitations”, *Journal of Management*, Vol. *20 No.* 4, pp. 903-31.

Müller, O., Junglas, I., Brocke, J.V. and Debortoli, S. (2016), “Utilizing big data analytics for information systems research: challenges, promises and guidelines”, *European Journal of Information Systems*, Vol. 25 No. 4, pp. 289-302.

Murphy Jr., B. (2014), “14 inspiring quotes about graduations and opportunity”, *Inc. Magazine*, May 5, 2014, available at: https://www.inc.com/bill-murphy-jr/14-inspiring-quotes-about-graduations-and-opportunity.html (accessed 24 January 2019).

Nacos, B.L., Shapiro, R.Y., Young, J.T., Fan, D.P., Kjellstrand, T. and McCaa, C. (2009), “Comparing human coding and a computer-assisted method”, *The Content Analysis Reader*, pp. 243-252.

Nam, H., Joshi, Y.V. and Kannan, P.K. (2017), “Harvesting brand information from social tags”, *Journal of Marketing*, Vol. 81 No. 4, pp. 88-108.

Neuendorf, K. A. (2017). *The Content Analysis Guidebook*. Sage.

Oliveira, M.D.F. and Murphy, P. (2009), “The leader as the face of a crisis: Philip Morris' CEO's speeches during the 1990s”, *Journal of Public Relations Research*, Vol. 21 No. 4, pp. 361-80.

Pan, B., MacLaurin, T. and Crotts, J.C. (2007), “Travel blogs and the implications for destination marketing”, *Journal of Travel Research*, Vol. 46 No. 1, pp. 35-45.

Paschen, J., Kietzmann, J., and Kietzmann, T. C. (2019), “Artificial intelligence (AI) and its implications for market knowledge in B2B marketing”, *Journal of Business & Industrial Marketing*.

Pennebaker, J.W., Boyd, R.L., Jordan, K. and Blackburn, K., (2015), *The Development and Psychometric Properties of LIWC2015*.

Polonsky, M.J., Bailey, J., Baker, H., Basche, C., Jepson, C. and Neath, L. (1998), “Communicating environmental information: are marketing claims on packaging misleading?”, *Journal of Business Ethics*, Vol. 17 No. 3, pp. 281-94.

Poole, M. S. and Folger, J. P. (1981), “Modes of observation and the validation of interaction analysis schemes”, *Small Group Behavior*, Vol. 12 No. 4, pp. 477-93.

Potter, W.J. and Levine-Donnerstein, D. (1999), “Rethinking validity and reliability in content analysis”, *Journal of Applied Communication Research*, Vol. 27 No. 3, pp. 258-84

Prein, G., and Kelle, U. (1995), “Using linkages and networks for theory building”, In U. Kelle (Ed.), *Computer Qualitative Data Analysis: Theory, Methods, and Practice,* pp. 62–68. London, UK: Sage Publications.

Resnik, A. and Stern, B. L. (1977), “An analysis of information content in television advertising”, *Journal of Marketing*, Vol. 41 No. 1, pp. 50-53.

Riffe, D., Lacy, S. and Fico, F.G. (2005), *Analyzing Media Messages: Using Quantitative Content Analysis in Research*. Routledge.

Rokka, J. and Canniford, R. (2016), “Heterotopian selfies: how social media destabilizes brand assemblages”, *European Journal of Marketing*, Vol. 50 No. 9/10, pp. 1789-813.

Rosenberg, S.D., Schnurr, P.P. and Oxman, T.E. (1990), “Content analysis: A comparison of manual and computerized systems”, *Journal of Personality Assessment*, Vol. 54 No. 1-2, pp. 298-310.

Schultz, F., Kleinnijenhuis, J., Oegema, D., Utz, S. and Van Atteveldt, W. (2012), “Strategic framing in the BP crisis: A semantic network analysis of associative frames”, *Public Relations Review*, Vol. 38 No. 1, pp. 97-107.

Shamir, B., Arthur, M.B. and House, R.J. (2018), “The rhetoric of charismatic leadership: A theoretical extension, a case study, and implications for research”, In *Leadership Now: Reflections on the Legacy of Boas Shamir* (pp. 31-49). Emerald Publishing Limited.

Sowa, C.A. (1984), *Traditional Themes and the Homeric Hymns*. Bolchazy-Carducci Publishers.

Spangler, W.D., Gupta, A., Kim, D.H. and Nazarian, S. (2012), “Developing and validating historiometric measures of leader individual differences by computerized content analysis of documents”, *The Leadership Quarterly*, Vol. 23 No. 6, pp. 1152-72.

Su, L.Y.F., Cacciatore, M.A., Liang, X., Brossard, D., Scheufele, D.A. and Xenos, M.A. (2017), “Analyzing public sentiments online: Combining human-and computer-based content analysis”, *Information, Communication & Society*, Vol. 20 No. 3, pp. 406-27.

Tausczik, Y.R. and Pennebaker, J.W. (2010), “The psychological meaning of words: LIWC and computerized text analysis methods”, *Journal of Language and Social Psychology*, Vol. 29 No. 1, pp. 24-54.

The Japan Times (2018), “In breakthrough, Japanese researchers use AI to identify early stage stomach cancer with high accuracy”, 22 July 2018, available at: https://www.japantimes.co.jp/news/2018/07/22/national/science-health/japanese-researchers-use-ai-identify-early-stage-stomach-cancer-high-accuracy/#.XUG8aehKg2w (accessed 31 July 2019).

Thompson, S.A., Loveland, J.M. and Castro, I.A. (2019), “From rumor to release: Does product release influence WOM in brand communities dedicated to technology products?”, *European Journal of Marketing*, Vol. 53 No. 2, pp. 345-65.

Tonidandel, S., King, E.B. and Cortina, J.M. (2018), “Big data methods: Leveraging modern data analytic techniques to build organizational science”, *Organizational Research Methods*, Vol. 21 No. 3, pp. 525-47.

Triantos, A., Plakoyiannaki, E., Outra, E. and Petridis, N. (2016), “Anthropomorphic packaging: is there life on “Mars”?”, *European Journal of Marketing*, Vol. 50 Vol. 1/2, pp. 260-275.

Tumasjan, A., Sprenger, T.O., Sandner, P.G. and Welpe, I.M. (2010), “Predicting elections with twitter: What 140 characters reveal about political sentiment”, In *Fourth international AAAI Conference on Weblogs and Social Media*.

Wagner-Pacifici, R., Mohr, J.W. and Breiger, R.L. (2015), “Ontologies, methodologies, and new uses of Big Data in the social and cultural sciences”, *Big Data & Society*, Vol. 2 No. 2, pp. 1–11

Weber Shandwick (2015). “The CEO reputation premium: A new era of engagement”, available at: https://www.webershandwick.com/news/the-ceo-reputation-premium-a-new-era-of-engagement/ (accessed 24 January 2019).

Willner, A.R. (1985), *The Spellbinders: Charismatic Political Leadership*. Yale University Press.

Wimmer, R.D. and Dominick, J.R. (1991), *Mass Media Research: An Introduction* (ed.). Belmont, CA: Wadsworth.

Wolfe, R.A., Gephart, R.P. and Johnson, T.E. (1993), “Computer-facilitated qualitative data analysis: Potential contributions to management research”, *Journal of Management*, Vol. 19 No. 3, pp. 637-60.

Zerfass, A., Verčič, D. and Wiesenberg, M. (2016), “Managing CEO communication and positioning: A cross-national study among corporate communication leaders”, *Journal of Communication Management*, Vol. 20 No. 1, pp. 37-55.

Table 1: Content analysis in marketing research

Study	Source of Content	Marketing Phenomena	Type of Content	Content Analysis Approach
Courtney and Lockertz, 1971	Brand (advertising)	The portrayed image and role of women in magazine ads	Text and images	Manual
Resnik and Stern, 1977	Brand (advertising)	Informational value of television ads for making buying decisions	Text and images	Manual
Kolbe and Albanese, 1996	Brand (advertising)	Portrayal of men in magazine ads	Images	Computer-aided
Choi et al., 2007	Brand and media (websites)	Destination image as repre-sented on the websites of partner organizations	Text	Computer-aided
Herbes and Ramine, 2014	Brand (websites)	Potential customer benefits of green electricity providers on websites	Images and text	Manual
Polonsky et al., 1998	Brand (packaging)	Environmental claims on dishwashing liquid packaging	Images and text	Manual
Triantos et al., 2016	Brand (packaging)	Anthropomorphic elements in product packaging	Images	Manual
Dowling and Kabanoff, 1996	Brand (slogans)	Commonalties among brand slogans	Text	Computer-aided
Miller and Toman, 2016	Brand (slogans)	Rhetorical figures and linguis-tic devices in brand slogans	Text	Manual
Oliveira and Murphy, 2009	Brand (CEO speeches)	CEO speeches during public relations crisis	Text	Computer-aided
Harris et al., 2001	Media (newspapers)	Press coverage of political brands in newspapers	Text	Manual
Schultz et al., 2012	Media (newspapers)	Differences in framing of news coverage of BP crisis	Text	Computer-aided
Rokka and Canniford, 2016	Customer (Instagram)	Portrayal of champagne brands with consumers in Instagram selfies	Images	Manual
Thompson et al., 2019	Customer (brand forums)	Postings on online brand forums	Text	Computer-aided
Harrison-Walker, 2001	Customer (complaint website)	Consumer complaints about United Airlines on a complaint website	Text	Manual
Pan et al., 2007	Customer (travel blogs)	Visitor opinions about a desti-nation on popular travel blogs	Text	Computer-aided

Figure 1: IBM Watson natural language processing of text

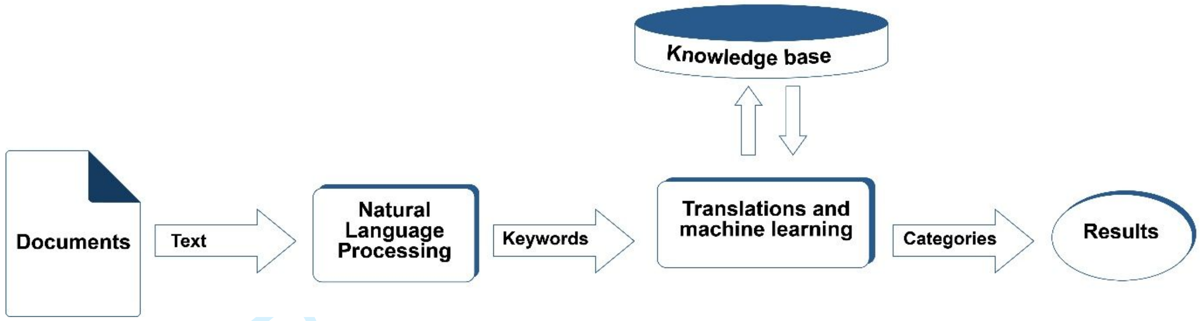

Table 2: Eight categories of speech charisma

Category	*Speeches of charismatic leaders contain...*
1) Collective focus	More references to collectives and fewer references to individual self-interest
2) Temporal orientation	More references to the continuity between past and present
3) Followers' worth	More celebratory terms; desirable moral and personal qualities; positive affirmations of a person, group, or abstract entity; and positive affective states
4) Similarity to followers	More leveling words that ignore individual differences; words that focus on human beings and their activities; and common everyday words
5) Values and moral justifications	More references to values; moral justifications; and patriotism
6) Tangibility	More references to tangibility and materiality and less variety of words (reverse coded)
7) Action	More references to competition; action; and triumph and fewer references to hesitation and uncertainty
8) Adversity	More references to social inappropriateness; evil; unfortunate circumstances; and censurable human behavior

(Adapted from Bligh et al., 2004a; Shamir et al., 2018)

Table 3: Steps and measures for three approaches to content analysis

Step Number	Content Analysis Methods: Manual	Computer-aided (DICTION and LIWC)	Artificial Intelligence enabled (IBM Watson)
Step 1 identifies the research questions and constructs	*Research question:* how charismatic are the speeches of leaders? *Construct:* Speech charisma		
Step 2 identifies the texts to be examined	*Texts:* commencement speeches by Bill Gates, Sheryl Sandberg, and Oprah Winfrey		
Step 3 specifies the unit of analysis	*Unit of analysis:* word or phrase in every speech		
Step 4 specifies the categories	*Categories:* collective focus; temporal orientation; followers' worth; similarity to followers; values and moral justifications; tangibility; action; and diversity		
Step 5 generates a sample coding scheme	The coding scheme from Bligh et al. (2004a) was used. From this, a coding form was developed.	Specific pre-installed program dictionaries were selected to be the coding scheme, based on the categories in step 4.	Natural Language Processing (NLP) generated a set of keywords that were mapped to the categories in order to create a coding scheme.
Step 6 coders pretest the coding scheme by using a sample of text	Three researchers independently pretest the coding scheme by coding each speech.	Two researchers independently ran each program to pretest the coding scheme for each speech.	Two researchers compared the NLP generated coding scheme for each speech.
Step 7 the coding scheme is revised or "purified"	Three researchers jointly revised the coding scheme.	For DICTION, no revisions were made. For LIWC, three researchers jointly created new custom coding scheme.	Three researchers jointly revised the coding scheme.
Step 8 is the actual content analysis, in other words, the data collection	Two researchers use the coding scheme to independently code a speech according to the categories.	Two researchers independently run DICTION and LIWC to code all three speeches according to the categories.	Two researchers run IBM Watson to independently code all three speeches according to the categories.
Step 9 calculate reliability, validity and efficiency for each method.	*Reliability:* inter-coder agreement on how content was coded. *Validity:* A fourth researcher evaluates the quality of results and inferences made from the content analysis method. *Efficiency:* time it takes to complete steps 5 to 8.	*Reliability:* agreement on how content was coded by the program between each run. *Validity:* Comparison of results to the manual approach. *Efficiency:* time it takes to complete steps 5 to 8.	*Reliability:* inter-coder agreement on how content was coded. *Validity:* Comparison of results to the manual approach. *Efficiency:* time it takes to complete steps 5 to 8.

Table 4: Reliability, validity, and efficiency results across content analysis approaches

Approach to Content Analysis	Evaluation Measures		
	Reliability	Validity	Efficiency
Manual	Low	High (baseline)	Low
Computer-aided	High	Low to moderate	Moderate to high
AI-enabled	High	High	Moderate

Appendix A: Codebook and coding form

Category	Description	Sample Words and Phrases
1: Collective focus		
Collectives (+)	Singular nouns connoting plurality that function to decrease specificity, reflecting a dependence on categorical modes of thought. Includes social groupings, task groups, geographical entities, and place names.	Alumni, America, country, Beijing, Charleston, China, City, College, Community, Faculty, First grade class, Group, Hokie, Lean In Circle, LeanIn.org, Mother Emanual church, Paris, Posse, Posse Foundation, school, UVA, Virginia Tech class of 2017, Virginia Tech, university, crowd, choir, team, humanity, army, congress, legislature, staff, county, world, kingdom, republic
People references (+)	Words referring to the citizenry-writ-large, including sociological, political, and generic group designations.	Baristas, colleagues, graduates, leaders, men, neighbors, people, professors, students, women, crowd, residents, constituencies, majority, citizenry, masses, population
Self reference (-)	All first person references that reflect the locus of action residing in the speaker and not in the world at large.	I, I'd, I'll, I'm, I've, me, mine, my, myself
2: Temporal orientation		
Present concern (+)	Present-tense verbs denoting an emphasis on the here and now.	Build, do, go, help, knows, see, talk, take, wait, write, cough, tastes, sing, take, canvass, touch, govern, meet, make, cook, print, paint, today, right now
Past concern (+)	The past-tense forms of the verbs in the present concern dictionary.	Did, done, happened, rejected, thought, coughed, tasted, sang, took, canvassed, touched, governed, met, made, cooked, printed, painted, last year, past year, last month
3: Followers' worth		
Praise (+)	Affirmations of a person, group, or abstract entity.	Amazing, beautiful, brave, esteemed, high potential, outstanding, resilient, thoughtful, dear, delightful, witty, mighty, handsome, beautiful, shred, bright, vigilant, reasonable, successful, conscientious, renowned, faithful, good, noble
Inspiration (+)	Abstract virtues deserving of universal respect and attractive personal qualities.	Decency, excellence, strength, honesty, self-sacrifice, virtue, courage, dedication, wisdom, mercy, patriotism, success, education, justice
Satisfy (+)	Terms associated with positive affective states, moments of undiminished joy, and moments of triumph.	Delighted, grateful, happy, honored, joy, thrilled, triumph, cheerful, passionate, happiness, smile, welcome, excited, fun, lucky, celebrating, pride, proud, secure, relieved
4: Similarity to followers		
Leveling (+)	Words used to ignore individual differences and to build a sense of completeness and assurance.	Absolute, all, each, we, everyone, everybody, anyone, each, fully, always, completely, inevitably, consistently, unconditional, consummate, absolute, open-and-shut
Familiarity (+)	A dictionary of the most common words in the English language. Includes common prepositions, demonstrative pronouns, interrogative pronouns, and particles, conjunctions, and connectives.	A, about, above, across, after, an, and, around, as, at, because, before, but, by, for, from, how, if, in, into, of, off, on, over, so, than, that, the, then, these, this, those, through, throughout, to, what, whether, which, who, with
Human interest (+)	Words that concentrate on people and their activities (but does not include words that belong in "people references" – "human interest" includes humans in relationships with others).	Baby, children, cousin, dad, grandchild, families, family, father, friend, friends, he, him, his, husband, mom, our, ourselves, parents, she, siblings, sister, son, their, them, they, uncle, us, we, wife, you, your, yours, yourselves

Category	Description	Sample Words and Phrases
5: Values and moral justifications		
Spirituality (+)	Broad-based, Judeo-Christian terminology including value-laden terms and theological constructs.	Blessings, hymn, churches, doctrine, sermons, conscience, blessing, god-fearing, spiritual, faith, hope, hopeful, heavenly
Patriotic terms (+)	Standard tokens of Americanism, including constitutional language, celebratory terms, words related to fundamental rights, and historic language.	Inalienable, emancipation, flag waving, homeland, Fourth-of-July, justice, liberty, equality, for-the-people, old-glory
6: Tangibility (reverse)		
Concreteness (+)	A dictionary of words denoting tangibility and materiality, including physical structures, modes of transportation, articles of clothing, household animals, etc.	Books, burger, calling card, campus, caps, computer, corner, elephant, hospital, jail, lobby, meal, muscle, notebook, path, plate, poster, rain, room, stones, stuff, toppings, wings, airplane, ship, bicycle, stomach, eyes, lips, slacks, pants, shirt, cat, insects, horse, wine grain, sugar, oil, silk, sand, courthouse, temple, store
Insistence (+)	A calculated measure reflecting the assumption that repetition of key terms indicates a preference for a limited, ordered world.	A calculation of repetition of key terms
Variety (-)	High scores indicate a speaker's avoidance of overstatement and preference for recise, molecular statements.	Calculated score: divides the number of different words in a passage by the total words
7: Action		
Aggression (+)	Words denoting human competition and forceful action, including physical energy, social domination, and goal directedness.	Beat, challenges, combat, defy, demands, effort, fighting, harassment, killed, pushed, rape, shooting, stole, strike, trauma, violence, blast, crash, explode, collide, conquest, attacking, violation, commanded, challenging, overcome, mastered, pound, shove, dismantle, overturn, prevent, reduce, defend
Accomplishment (+)	Words expressing task completion and organized human behavior.	Breakthrough, founded, establish, finish, influence, proceed, motivated, influence, leader, manage, strengthen, succeed, agenda, enacted, working, leadership
Passivity (-)	Words ranging from neutrality to inactivity, including terms of compliance, docility, and cessation.	Retreat, wait, allow, tame, appeasement, submit, contented, sluggish, arrested, capitulate, refrain, yielding, immobile, unconcerned, nonchalant
Ambivalence (-)	Words expressing hesitation or uncertainty, implying an inability or unwillingness to commit to what is being said.	Maybe, allegedly, perhaps, might, almost, approximate, vague, baffled, puzzling, hesitate, could, would, guess, suppose, seems
8: Adversity		
Blame (+)	Terms designating social inappropriateness and evil, as well as unfortunate circumstances.	Anxious, mean, naive, sloppy, stupid, fascist, repugnant, malicious, bankrupt, rash, morbid, weary, nervous, painful, detrimental, cruel
Hardship (+)	Natural disasters, hostile actions, censurable human behavior,unsavory political outcomes, and human fears.	Attacks, killers, problem, tragedy, earthquake, starvation, killers, bankruptcy, enemies, vices, infidelity, despots, betrayal, injustices, exploitation, grief, death
Denial (+)	Standard negative contractions, negative function words, and null sets.	Can't, didn't, doesn't, none, wasn't, won't, wouldn't, aren't, shouldn't, don't, nor, not, nay, nothing, nobody, none

NOTE: *Each word or phrase to go into only one category*
Adapted from Bligh et al. (2004a)

Speech ID	***2Gates, 3Winfrey, 4Sandberg***		
Speech giver	***Bill Gates or Oprah Winfrey or Sheryl Sandberg***		
Coder	***Name of coder***		
Date & time started	***Feb 2, 2019 08:01am***	***Feb 4, 2019 12:45pm***	***Feb 12, 2019 09:22am***
Duration (hours:minutes)	***3:10***	***3:00***	***3:22***
Total duration (hours:minutes)			***9:32***
Category	**Words or phrases**		**No. Of Instances**
Collective focus	*America* *first grade class*		*xx* *xx*
Temporal orientation	*build* *do*		*xx* *xx*
Followers' worth	*amazing* *beautiful*		*xx* *xx*
Similarity to followers	*all* *everyone*		*xx* *xx*
Values and moral justifications	*blessings* *hymn*		*xx* *xx*
Tangibility	*books* *burger*		*xx* *xx*
Action	*challenges* *combat*		*xx* *xx*
Adversity	*repugnant* *malicious*		*xx* *xx*

Appendix B: Manual coding results for charismatic leadership

	Gates	Sandberg	Winfrey
Words in commencement speech	3,036	3,256	3,808
Charismatic leadership category	Percent of words for each category		
Collective focus	3%	-1%	-2%
Temporal orientation	9%	8%	9%
Followers' worth	2%	2%	1%
Similarity to followers	35%	38%	30%
Values/moral justifications	0.5%	1%	0%
Tangibility (reverse)	8%	7%	7%
Action	0%	-1%	-1%
Adversity	3%	1%	1%
Total score	61%	55%	50%

NOTE: *The percentages are read as the percent of the speech that contained words and phrases of a particular category. For example, 35% of Gates' speech was about "similarity to followers" compared to 38% of Sandberg's speech.*

Appendix C: DICTION dictionaries used for charismatic leadership categories

Category	Dictionaries used
Collective focus	Collectives - self reference
Temporal orientation	Present concern + past concern
Followers' worth	Inspiration + praise + satisfaction
Similarity to followers	Leveling terms + familiarity + human interest
Values/moral justifications	Inspiration
Tangibility (reverse)	Concreteness + insistence + variety
Action	Aggression + accomplishment - ambivalence - passivity
Adversity	Blame + hardship + denial

Appendix D: DICTION coding results for charismatic leadership

	Gates	Sandberg	Winfrey
Charismatic leadership category	Percent of words for each category		
Collective focus	-1%	-1%	-2%
Temporal orientation	7%	4%	5%
Followers' worth	6%	2%	2%
Similarity to followers	28%	26%	23%
Values/moral justifications	0%	0%	0%
Tangibility (reverse)	3%	17%	8%
Action	-4%	1%	-2%
Adversity	1%	3%	2%
Total score	39%	53%	38%

Appendix E: LIWC dictionaries used for charismatic leadership categories

Category	Default dictionaries used	Custom dictionaries added
Collective focus	Affiliation - 1st person singular	+ collective nouns
Temporal orientation	Present focus + past focus	
Followers' worth	Positive emotion	+ affirmations
Similarity to followers	1st person plural + certainty + articles + prepositions + conjunctions + interroga-tives + 3rd person singular + 3rd person plural + social	
Values/moral justifications	Religion	+ morals and value (removed religion)
Tangibility (reverse)	Future focus - body - ingestion - home - common verb	
Action	Power + reward + achievement - tentative	+ inactivity
Adversity	Negative emotion + risk + death + negations	

Appendix F: LIWC coding results for charismatic leadership

	Default Dictionaries Used			Custom Dictionaries Added		
	Gates	Sandberg	Winfrey	Gates	Sandberg	Winfrey
Charismatic leadership category	Percent of words for each category			Percent of words for each category		
Collective focus	0%	0%	0%	1%	1%	0%
Temporal orientation	3%	3%	3%	3%	3%	3%
Followers' worth	1%	1%	1%	1%	1%	1%
Similarity to followers	9%	10%	9%	9%	10%	9%
Values/moral justifications	0%	0%	0%	0%	0%	0%
Tangibility (reverse)	-3%	-3%	-3%	-3%	-3%	-3%
Action	1%	0%	1%	1%	0%	1%
Adversity	1%	1%	1%	1%	1%	1%
Total score	12%	12%	11%	13%	12%	11%

Appendix G: IBM Watson / manual coding results for charismatic leadership

	Gates	Sandberg	Winfrey
Words and phrases in IBM Watson Output	501	461	493
Charismatic leadership category	Percent of words for each category		
Collective focus	16%	19%	13%
Temporal orientation	6%	8%	6%
Followers' worth	5%	11%	5%
Similarity to followers	23%	21%	22%
Values/moral justifications	0%	3%	4%
Tangibility (reverse)	-6%	-5%	-8%
Action	7%	3%	3%
Adversity	9%	2%	2%
Total score	60%	62%	47%