Automated Text Analysis

Ashlee Humphreys

Contents

A. Humphreys (✉)
Integrated Marketing Communications, Medill School of Journalism, Media, and Integrated Marketing Communications, Northwestern University, Evanston, IL, USA
e-mail: a-humphreys@northwestern.edu

C. Homburg et al. (eds), *Handbook of Market Research*,
https://doi.org/10.1007/978-3-319-05542-8_26-1

Abstract

The amount of text available for analysis by marketing researchers has grown exponentially in the last two decades. Consumer reviews, message board forums, and social media feeds are just a few sources of data about consumer thought, interaction, and culture. However, written language is filled with complex meaning, ambiguity, and nuance. How can marketing researchers possibly transform this rich linguistic representation into quantifiable data for statistical analysis and modeling? This chapter provides an introduction to text analysis, covering approaches that range from top-down deductive methods to bottom-up inductive methods for text mining. After covering some foundational aspects of text analysis, applications to marketing research such as sentiment analysis, topic modeling, and studying organizational communication are summarized and explored, including a case study of word-of-mouth response to a product launch.

Keywords

Text analysis · computer-assisted text analysis · automated content analysis · content analysis · topic modeling · sentiment analysis · LDA · word-of-mouth

Introduction

Automated or computer-assisted text analysis describes a family of methods for parsing, classifying, and then quantifying textual data for further statistical analysis. Although automated text analysis using computers dates to the 1960s, the rise of digital technology for communicating has created a deluge of textual data for analysis and increased managerial desire to gain insights from text produced by consumers. Platforms like Twitter and Facebook provide a space for consumer-to-consumer discussion of products, brands, and services. Retail sites like Amazon, Best Buy, and Zappos and review sites like CNET and Yelp! host consumer reviews on a nearly endless array of products and services. Particular brand sites like Sephora, Gap, and Brooks Brothers offer social shopping capabilities such as consumer reviews represented by stars and extensive product reviews that detail fit, material, and quality (Stephen and Toubia 2010). This text from consumers, firms, and the media can provide insight into consumer needs and wants, sentiment, market structure, and transmission of word-of-mouth communication.

This chapter presents a high-level overview of methods for conducting text analysis in market research and provides resources for further investigating the methodological details depending on the approach one takes to text analysis.

Foundations of Text Analysis

History

To understand the implementation of automatic analysis, it will help to first review its relation to and its emergence from traditional content analysis. Content analysis is a method used in the social sciences to systematically assess and analyze the content of a message, usually in the form of text. Although traditions of content analysis go as far back as sixteenth-century monastic life, modern content analysis was first proposed by Max Weber (1924) to study the press. Since then, scholars in sociology and communications have used human-coded content analysis to investigate differences in media content, describe trends in communications over time, reveal patterns of organizational or individual attention, and examine attitudes, interests, intentions, or values of an individual or a group (e.g., Berelson 1971; Gamson and Modigliani 1989).

Traditional content analysis was first introduced to consumer behavior with Kassarjian's (1977) outline of the method and was then updated by Kolbe and Burnett (1991) in an attempt to improve reliability and objectivity, focusing primarily on standards for calculating inter-coder agreement (see also Grayson and Rust 2001). In consumer research and marketing, traditional content analysis has been used to analyze trends in magazine advertisements (Belk and Pollay 1985), direct mail (Stevenson and Swayne 1999), newspaper articles (Garrett 1987), and word-of-mouth communication (Moore 2015; Phelps et al. 2004) to name a few. Although automated text analysis can improve the efficiency and reliability of traditional content analysis, it also has limitations. For instance, computerized text analysis can miss subtleties in the text and cannot code finer shades of meaning. While dealing with negation is possible (Jia et al. 2009; Villarroel Ordenes et al. 2017), it remains somewhat analytically onerous.

Automated text analysis is not radically new, but it has become easier to implement since the widespread of adoption of the personal computer. The General Inquirer (Stone 1966) was one of the first computer content-analytic tools used in consumer research (Kranz 1970). Since then, vast strides have been made in automated text analysis. Kranz's (1970) early three-page treatment of computer-assisted content analysis in marketing deals with dictionary creation, but does not address category creation, validity, or measurement decisions. Since then, a variety of approaches have emerged.

Approaches to Text Analysis

In current practice, there are essentially two orientations toward automated text analysis: top-down vs. bottom-up approaches (Boyd and Pennebaker 2015a; Mehl and Gill 2008). The top-down approach counts concepts of interest, identified either through a list of words or through a set of rules. Top-down, also called dictionary-based, methods are deductively or theoretically driven in the sense that researchers use them to look for hypothesized patterns in text from a known set of concepts. Bottom-up approaches, on the other hand, code all concepts present in the text and then look for patterns (Rayson 2009). These approaches can range considerably from methods of supervised learning, where researchers define some preliminary categories and then train the computer to sort documents based on latent differences, to discovery-oriented approaches such as calculating then flagging statistically significant differences between groups of texts (Rayson 2009), or fully automated processes where a computer identifies topics based on word co-occurrence (Lee and Bradlow 2011). In this way, bottom-up approaches to text analysis become similar to data mining approaches. That is, first the researcher looks at all differences in the data and builds conclusions from those differences.

Top-down, dictionary-based methods have been used extensively in social sciences like consumer research (Humphreys and Wang 2018), psychology (Chung and Pennebaker 2013; Mehl and Gill 2008; Pennebaker and King 1999), sociology (Van de Rijt et al. 2013), and political science (Grimmer and Stewart 2013; Lasswell and Leites 1949) due to their ability to translate theoretical constructs into text and the transparency in reporting results and reliabilities. Bottom-up methods, on the other hand, have been used more extensively in engineering, computer science, and marketing science. Marketing strategy has drawn from both approaches, although dictionary-based approaches appear to be more common (Ertimur and Coskuner-Balli 2015; Humphreys 2010; Ludwig et al. 2013; Packard et al. 2014). This chapter briefly covers the fundamentals of each approach before moving to their application in marketing.

Dictionary-Based Methods

Dictionary-based methods for text analysis are based on a predeveloped word list, or dictionary, for counting the occurrence of words in a text. Standardized dictionaries are available for many constructs such as sentiment (e.g., Hutto and Gilbert 2014), marketing-related constructs like authenticity and brand personality (Kovács et al. 2013; Opoku et al. 2006), as well as many standard concepts in psychology (Pennebaker et al. 2001; Snefjella and Kuperman 2015) and other fields like political science (Dunphy et al. 1974; Stone 1966). In addition to using a standard dictionary, many researchers choose to create their own dictionary to fit the specific context, although this should be done only if a standard dictionary is not available.

There are several methods for dictionary creation ranging from inductive to deductive. The most inductive method of dictionary creation is to work from a concordance, or all words in the document listed in terms of frequency and group words according to relevant categories for the research question and hypothesis (Chung and Pennebaker 2013). If the researcher does not know what categories are relevant a priori, qualitative methods of reading and coding the text prior to dictionary development can be used to create a set of relevant concepts and a list of words for their operationalization in text (Humphreys 2010). For example, to study institutional logics pertaining to the Yoga industry in newspaper articles, Ertimur and

Coskuner-Balli (2015) first open and then axially code a dataset of newspaper articles and other historical texts. Generally, a random sample of 10–20% of the dataset is sufficient for coding (Humphreys and Wang 2018), but researchers should be mindful of unevenness in data quantity according to category or time period and stratify accordingly (Humphreys 2010). The most deductive method for dictionary creation is to create a wordlist from theoretical concepts or categories. However, one should be mindful of the tendency for researchers and writers to pick more abstract words than are generally present in textual data (Palmquist et al. 2009). For this reason, careful postmeasurement validation is necessary to ensure construct validity. After text is cleaned and stored and the dictionary has been created, researchers use a program like Diction, LIWC, WordStat, or R to execute counts. Data can then be saved and analyzed using a traditional statistical package or, for some packages like Wordstat and R, analyzed within the same package.

After calculating word frequencies, postmeasurement validation should be performed, and for this there are a variety of methods ranging from methods that are iterative with dictionary development to stand-alone calculations of inter-rater reliability. Weber (2005) suggests a saturation procedure whereby researchers pull a sample of 10 or 20 instances of a concept and have a research assistant code them as accurately representing the category (or not). If the rate is below 80%, the dictionary category should be revised until the threshold is met. Pennebaker et al. (2001) recommend a method of validating the dictionary, but not the resulting measurements. Here, three research assistants count a word as being representative of the category or not, and words are retained if two of the three coders agree. If they do not, the word should be dropped from the dictionary. Percentage agreements on dictionary categories can then be calculated and reported, and the general threshold is similar to that for Krippendorf's alpha, above 75%. A final option is to compare the computer-coded results with an extensive set of human-coded results from two or more coders. To do this, one selects a random sample from the dataset (the amount may vary depending on the size of the dataset) and human coders code the text according to the category descriptions, calculating reliability as one would in a traditional content analysis. This can then be compared to the additional "coder" of the computer to produce a similarity score. Although this final method has the advantage of comparison with traditional content analysis, it is not always necessary and in some cases can produce misguided results. Human coders pick up on subtle meanings that computers cannot and likewise computers are able to code concepts consistently and evenly over an entire dataset without omission or bias. For this reason, comparing human to computer coding can in some cases be like comparing apples to oranges.

准备一部分数据给编码者编码，通时与词典法自动化标注结果对比，对比结果为准确率。

达到80%以上，该词典构建的较为合理准确。

不过上述的词典验证非必要。

Dictionary-based analyses have studied a wide range of theoretical concepts such as emotion (Berger and Milkman 2012), construal level (Snefjella and Kuperman 2015), institutional logics (Ertimur and Coskuner-Balli 2015), risk (Humphreys and Thompson 2014), 言语行为 speech acts (Ludwig et al. 2016; Villarroel Ordenes et al. 2017), and framing (Fiss and Hirsch 2005; Humphreys and Latour 2013; Jurafsky et al. 2014). A wide variety of contexts can be explored through dictionary-based analysis such as product and restaurant reviews (Barasch and Berger 2014, Jurafsky et al.

2014; Kovács et al. 2013), tweets (Mogilner et al. 2010), customer service calls (Packard et al. 2014), blogs (Arsel and Bean 2013), and news articles (Humphreys 2010; Humphreys and Thompson 2014).

Classification Methods

Bottom up methods include classification and topic modeling. Classification methods of text analysis are based on categorizing documents into different "types" and then further describing what textual elements best predict the likelihood of being a "type." For example, Tirunillai and Tellis (2012) use classification to train a model to recognize positive versus negative reviews based on star rating. Using a training data set, they use both a Naïve Bayes and a support vector machine (SVM) classifier to find which words predict star rating and then use this information to categorize the entire set of reviews, achieving a precision – meaning their algorithm predicts true positives – 68–85% of the time, depending on the product category. Villarroel Ordenes et al. (2017) further refine measures of sentiment by using both explicit and implicit indicators of emotion to measure sentiment and sentiment strength, also testing their framework on a set of starred reviews from Tripadvisor, Amazon, and Barnes and Noble. Classification models vary in sophistication; accuracy of these approaches varies from 55% to 96% for sentiment, for example (Hutto and Gilbert 2014). In general, considerations for model selection are based on the underlying frequency of occurrence of words that one wants to use to make predictions and the clarity of categories one wants to produce. For instance, SVM classification provides clear, mutually-exclusive categories, while LDA produces probabilistic groupings where it is possible for categories to overlap.

Classification models have been used to study reviews (Tirunillai and Tellis 2012; Van Laer et al. 2017), online forums (Homburg et al. 2015), email (Ludwig et al. 2016), and literary texts (Boyd and Pennebaker 2015b; Plaisant et al. 2006). For example, to measure sentiment of message board posts, Homburg et al. (2015) classify a training dataset of unambiguously positive and negative posts. They then use sentiment as a dependent measure to understand how much firm engagement actually increases positive consumer sentiment, finding that there are diminishing returns to engagement.

Topic Modeling

Topic modeling is an approach that begins by parsing text into discrete words, and then finding recurring patterns in co-occurrence that are statistically unlikely if one assumes that word occurrence is independent. In this way, the analysis identifies categories that may be latently represented by the manifest presence of words, and these word groupings are then labeled to represent meaningful concepts or traits in the data as one would in factor analysis. For example, in a study of hotel reviews, Mankad et al. (2016) use latent Dirichlet allocation (LDA) to identify five topics that occur in users' TripAdvisor comments, identifying amenities, location, transactions, value, and experience as key topics mentioned by reviewers. Latent semantic analysis (LSA), k-means clustering (Lee and Bradlow 2011), probabilistic latent semantic analysis (PLSA), and LDA (Blei et al. 2003) are all methods for topic modeling, with LDA being the most recent and common analytical methods for topic modeling.

LSA is based on the relatively straightforward process of generating a matrix that represents word occurrence (0 for nonoccurrence and 1 for occurrence) and then generating a vector of similarity that represents either the similarity between *documents* (the dot product of the rows) or the similarity between two or more *words* (the dot product of the columns). These vectors can then be reduced using singular value decomposition (SVD) to represent the "topics" that tend to occur across documents. PLSA is a similar process; topics are treated as word distributions based on probability.

LDA is a hierarchical Bayesian model for determining the mixture of topics present in a given document. Like PLSA, it assumes topics are probabilistic distributions of words, except it uses a Dirichlet prior for estimation, which reduces overfitting. For LDA, one sets the number of topics prior to running the analysis (other methods such as hierarchical Diriclet Process do not need this assumption). Using assumptions that there is a certain probability distribution for the choice of topic, and a certain distribution within that for choice of words to represent that topic, LDA produces a final list of topics (as represented by a list of words in that topic) and probabilities that a given topic is in the document. Although most approaches are word or phrase based, Büschken and Allenby (2016) conduct an LDA analysis using sentences as the unit of analysis and find that this produces results more predictive of rating than word-based LDA. A sentence-based model assumes that all words in the sentence are part of the same topic, which is reasonable, given Grice's maxims of relation and manner (Grice 1975). Büschken and Allenby (2016) use this model to identify topics for Italian restaurants and hotels from reviews on Expedia and we8there.com.

LDA has been used in a wide range of applications (Büschken and Allenby 2016; Tirunillai and Tellis 2014). As with dictionary approaches, postmeasurement validation, in this case using a hold-out sample or other predictive technique (e.g., external DV) is highly advisable. Machines will only read literal meaning, and therefore homonyms and other colloquialisms including sarcasm can be problematic, as they are overly general and overly specific words. Further, careful cleaning and preparation of the text can reduce errors, as textual markers can sometimes be added during data collection (e.g., headers, footers, etc.).

Market Research Applications of Text Analysis

This section discusses ways that text analysis has been incorporated into marketing research. Although potentially useful for many types of sources and research questions, text analysis has been particularly fruitful for representing consumer sentiment, studying word-of-mouth communication, and creating positioning maps from online text, among other uses.

Sentiment Analysis

Many text analytic programs and practitioners claim to measure sentiment, but it is not always clear what goes into this key metric. Before discussing the text analysis of sentiment, it might first to help to discuss what sentiment is and what it is trying to capture. In most marketing contexts, researchers and practitioners are interested in consumer attitude toward a brand, product, or service. Yet attitudes are complex mental structures composed not only of emotion, but also cognitive beliefs and intentions (Fishbein and Ajzen 1972). Further, the importance an attitude for any given product for ultimate purchase and future behavior like loyalty depends to a large degree on context and involvement (Petty and Cacioppo 1979). Further, people may articulate attitudes online that do not fully reflect their underlying attitude, there may be selection bias in the attitudes they choose to articulate, and they may behave differently than the attitudes they espouse. Nonetheless, discourse online, as expressed in sentiment, can reflect some underlying attitude about a brand, product, or service, and importantly can affect the social consensus shared among other consumers. Sentiment has been shown to predict movie sales (Krauss et al. 2008; Mestyán et al. 2013) and stock market returns (Bollen et al. 2011; De Choudhury et al. 2008; Tirunillai and Tellis 2012), although there may be natural biases in nonreporting of null results. Structurally, most approaches seek to classify or measure text as having positive, negative, or sometimes neutral sentiment, and some approaches transform this into net sentiment, subtracting negative words from positive words (e.g., Ludwig et al. 2013; Homburg et al. 2015). Top-down approaches do this using a dictionary or lexicon of words, while bottom-up approaches use some underlying external classification like human coding of a training set or customer ratings to identify the set of words that indicate sentiment.

对情感的测量算法有：
1. 正、负、中的量
2. 净值，正面减负面

In addition to valence, sentiment can also have strength and certainty. Previous research has used both explicit, semantic indicators of emotion along with implicit, more pragmatic indicators of emotion such as speech acts (commission, assertion, and direction) to successfully measure strength of sentiment (Villarroel Ordenes et al. 2017). Work has further shown that other types of speech such as demonstratives (Potts and Schwarz 2010) and other pragmatic markers 语用标记 can indicate expressive content, commonly expressed in product reviews (Constant et al. 2009).

Using predeveloped, standardized dictionaries is one of the most reliable ways to measure sentiment across contexts, as these wordlists have been developed and tested on a wide range of textual data, and some have themselves been developed through bottom-up approaches.

VADAR, for example, uses a dictionary with a rule-based approach for measuring sentiment. Specifically, Hutto and Gilbert (2014) use a combination of dictionaries based on previous standardized dictionaries like LIWC and General Inquirer but then also develop five rules that take into account syntax and grammar to measure intensity as well. Bottom-up approaches to measure sentiment produce accuracies ranging from 55% to 96%, depending on the context (Hutto and Gilbert 2014). For example, Tirunillai and Tellis (2012) use star rating to create a classification system for sentiment, with an accuracy rate of 68–85%.

Studying Word of Mouth Through Text Analysis

The primary use of text analysis in marketing research to date has been to study online word-of-mouth communication. Consumers have always shared product information through interpersonal communication (Arndt 1967), and this communication has been shown to be more effective than commercial messages (Brown and Reingen 1987; see also Godes and Mayzlin 2004; Money et al. 1998). And yet while word-of-mouth communication was previously communicated face to face or over the telephone, it is now visible and archived on social shopping sites (Stephen and Toubia 2010), social media (Humphreys 2015), and third-party review sites and platforms. Product reviews on Amazon, hotel reviews on TripAdvisor, and restaurant reviews on Yelp! have all provided marketing insights to better understand the relationship of ratings to sales and stock price (Moe and Schweidel 2014; Schweidel and Moe 2014; Moe and Trusov 2011). For example, Moe and Trusov (2011) find that positive reviews have a direct effect on sales, but this effect is somewhat short-lived because of downward convergence as people post more ratings (i.e., the social dynamics of posts result in reviews becoming relatively more negative over time). Further, positivity can vary depending on platform (Schweidel and Moe 2014; Villarroel Ordenes et al. 2017).

评分、评论数、评分的方差

Word of mouth online can be represented by measuring valence, volume, and variance (Godes and Mayzlin 2004). Volume and variance are relatively compatible with existing modeling measures, as volume can be aggregated and variance can be measured through start ratings or other user input. Valence, while partially captured by star measures, is perhaps best measured by sentiment, which requires text analysis as a method for converting the unstructured data of linguistic description into data that can be incorporated into quantitative models. There is also, it should be noted, a wide range of linguistic properties and semantic content beyond valence that usefully informs marketing research (Humphreys and Wang 2018). For instance, Kovács et al. (2013) show that restaurants have higher ratings if reviewers mention 真实性 authenticity in their reviews, even when controlling for restaurant quality.

The role of emotion in the spread of word of mouth is one key topic. In a study of sharing news articles, Berger and Milkman (2012) find that positive emotion increases virality, but so too does the presence of intense negative emotion like anger or anxiety in the article. Effects of the sender and speech context have also been investigated through text analysis using pronouns. Using a standard dictionary for first-person personal pronouns ("I", "me"), Packard and Wooten (2013) find that consumers self-enhance more in word of mouth to signal knowledge about a particular domain. Consumers have also been shown to engage in self-presentation by sharing fewer negative emotions when broadcasting to a large audience versus narrowcasting to a smaller one (Barasch and Berger 2014). When evaluating a product like a movie, consumers are more likely to use pronouns referring to themselves when expressing views about taste vs. their views about quality (Spiller and Belogolova 2016).

Topic Discovery and Creating Positioning Maps from Online Text

Text analysis can be used to create positioning maps for brands, companies, or products and to visualize market structure based on attributes within a particular category. Bottom-up methods such as LDA, LSA, and similar methods like k-means clustering are used to group words in a text (like reviews) into attributes or brands based on common co-occurrence. For example, to create a visualization of market structure for cameras from a set of reviews on Epinions.com, Lee and Bradlow (2011) first extract phrases related to particular attributes (e.g., battery life, photo quality) and then use k-means clustering to group phrases based on their similarity (calculated as cosine similarity between vectors of words). They then go on to show that this kind of analysis reveals attributes mentioned by and important to consumers, but absent from expert reviews such as size, design, and screen brightness. Similarly, using text data from diabetes forums, Netzer et al. (2012) find several side effects commonly mentioned on the forum, but absent from a site like WebMD (e.g., weight gain, kidney problems).

Topic-based models are compatible with psychological theories such as spreading activation in semantic memory (Collins and Loftus 1975). For instance, based on the idea that people talk about brands together that are related in semantic memory, Netzer et al. (2012) produce a perceptual map for car brands using reviews from Edmunds.com and compare that to results from perceptual maps based on more typical survey and brand-switching based on sales approaches. In doing so, they find several notable differences between the results based on text analysis versus those based on sales or survey data. For instance, based on the sales data, Korean brands of cars are not associated with the Japanese brands. However, based on the textual data, these brands are grouped together. This suggests that while text analysis can capture cognitive associations, these may not necessarily translate into behavior such as brand switching (Table 1).

Measurement of the Organization and Firm Environment

Finally, text analysis can be used to measure organizational attention through the analysis of shareholder reports, press releases, and other marketing communication. These studies are primarily based on dictionary-based analysis, and often create dictionaries rather than using standardized dictionaries to fit the industry or original context and research question. For example, scholars have developed dictionaries to study the changes in CSR language over time to reveal differences in developing countries (Gandolfo et al. 2016). In an analysis of annual reports, Lee et al. (2004) find that companies that issued internal reasons for negative events had higher stock

Table 1 Types of text analysis

Type of text analysis	Materials	Theoretical areas	Software/ methods	Relevant examples
Dictionary-based	Reviews, tweets, online forums, news articles, press releases, annual reports	Sentiment/emotion, psychological mindset (e.g., construal level), brand attention and brand value, legitimacy/corporate image, customer service	LIWC, WordStat, Diction	Humphreys (2010), Berger and Milkman (2012), Packard et al. (2018)
Classification	Reviews, online forums, literary texts, tweets, email	Sentiment, deception, product attributes, market structure	SVM, Naïve Bayes, k-nearest neighbor, neural networks, WordStat	Homburg et al. (2015), Van Laer et al. (2018), Tirunillai and Tellis (2012)
Topic modeling	Product or service reviews, online forums	Product attributes, positioning, market structure, customer needs	LDA, LSA, PLSA, K-means clustering, R, WordStat	Netzer et al. (2012), Lee and Bradlow (2006), Buschken and Allenby (2016)

prices a year after the event, suggesting that organizations who attribute blame to firm-controlled factors appear more in control than those who do not and therefore have more favorable impressions from investors. Interactions between firm employees or agents can also be better understood. For example, Ludwig et al. (2016) develop a method for detecting deception in sales emails. They find that deceivers are more likely to use elaborate, superfluous descriptions, and less self-referencing, quickly taking on the linguistic style of their intralocular.

Firm environment can also be captured through measuring media such as newspapers, magazines, and trade publications. For example, Humphreys (2010) shows that changes in the institutional and cultural environment enabled the legitimation of the casino gambling industry in the United States. Humphreys and Thompson (2014) study the environment of risk perceptions following two crises – the Exxon and BP oils spills – and find that the media narratives serve to contain risk perceptions following these disasters. Ertimur and Coskuner-Balli (Ertimur and Coskuner-Balli 2015) trace how the Yoga industry shifted over time, developing distinct institutional logics that impacted branding and positioning within the industry.

Issues in Working with Textual Data

Although language provides a window into many areas of consumer thought and market strategy, there are several issues to consider when analyzing text. Language rarely, if ever, follows patterns of normal distribution (Zipf 1932). For instance, functional words like “a,” “he,” and “there” make up about 40% of all language in normal usage. Common words like nouns and verbs make up another 59%, and only a small fraction of those common words will usually be relevant to the research question. Textual data are often left-skewed (lots of zeros), documents often contain different numbers of words, and the words of interest are often too infrequently or too frequently occurring to make meaningful comparisons. For these reasons, after word frequency has been calculated, researchers will often transform the data prior to statistical analysis. Further, many test such as ANOVA would not be appropriate due to the non-normal distribution of the data.

Text is therefore almost always represented as a percentage of words in the document (e.g., Ludwig et al. 2013), and log transformation to account for skewedness is often commonly employed (Netzer et al. 2012), although there are several possible transformations used (Manning et al. 2008). Tf*idf is a measure often used to account for the term frequency, standardized by the overall frequency of a word in the dataset as a whole (see Salton and McGill 1983 for details in calculating tf*idf, with attendant options for transformation).

Traditional methods for measuring co-occurrence such as Pearson correlation can be problematic due to the large number of zeros in a dataset (Netzer et al. 2012). For this reason, researchers will often use cosine similarity or Jaccard distance to compare words and documents. A series of robustness checks using multiple methods to calculate co-occurrence is often necessary to ensure that results do not occur simply due to infrequently or too-frequently occurring words (Monroe et al. 2009; Netzer et al. 2012). For example, if a word like “him” is very common, it is likely to co-occur with more words than an infrequent word like “airbag.” And yet, the word “airbag” may be more diagnostic of the concept safety than a personal pronoun like “him” even though detecting the co-occurrence will be more likely. Because data are not normally distributed, statistical tests such as the Mann-Whitney test, which tests for significance in rankings rather than absolute number, can serve as a replacement for ANOVA.

Extended Example: Word-Of-Mouth Differences Between Experts and Nonexperts to a Product Launch

Purpose

This section presents a sample text analysis as an illustration of top-down, dictionary-based methods according to the six stages (Table 2) (Reprinted from the Web Appendix to Humphreys and Wang (2018), Automated Text Analysis for Consumer Research, *Journal of Consumer Research, 44*(6), 1 (April), 1274–1306, with permission from Oxford University Press.). Automated text analysis is appropriate for tracking systematic trends in language over time and making comparisons between groups of texts. To illustrate a top-down approach to text analysis, this section presents a short study of consumer response to the product launch of an mp3 player/wireless device, the Apple iTouch. This case has been selected because it

Table 2 Stages of automated content analysis

Stages of automated content analysis (dictionary-based analysis)	
Stage	*Elements of stage*
1. Identify a research question	Select a research topic and a question within that topic
2. Data collection	Identify sources of information Online databases or newspapers Digital converters for printed text Web scraping for internet data Archival materials Field interviews
2a. Data cleaning	Organize the file structure Spell check, if applicable Eliminate problematic characters or words
3. Construct definition	Qualitatively analyze a subsample of the data Create a word list for each concept Have human coders check and refine dictionary Preliminarily implement dictionary to check for false positives and false negatives
4. Operationalization	Conduct computer analysis to compute the raw data Make measurement decisions based on the research question: Percent of all words Percent of words within the time period or category Percent of all coded words Binary ("about" or "not about" a topic)
5. Interpretation and analysis	Make unit of analysis decisions: By article, year, decade Comparison by genre, speaker, etc. Choose the appropriate statistical method for the research question: Analysis of variance (ANOVA) Regression analysis Multidimensional scaling Correlational analysis
6. Validation	Pull a subsample and have coded by a research assistant or researcher
	Calculate Krippendorf's alpha or a hit/miss rate

can be used to illustrate both comparison between groups and change over time and because it is relatively agnostic regarding theoretical framework. One could study word-of-mouth communication from a psychological, sociological, anthropological, or marketing strategy point of view (c.f. Godes and Mayzlin 2004; Kozinets 2010; Phelps et al. 2004; Winer 2009).

Stage 1: Develop a Research Question

This study proposes a specific, strategic research question: After a product launch, do experts respond differently from nonexperts? Further, how does word-of-mouth response change in expert versus nonexpert groups as the product diffuses? Word of mouth from experts can be particularly influential in product adoption, so it is important to know how their views may change over time and in comparison with nonexpert groups. The context chosen for this study, the launch of the Apple iTouch, is a good case to study because both the product category and the criteria for evaluating the product were ambiguous at the time of launch.

Stage 2: Data Collection

Data. Data were collected from two websites, Amazon.com and CNET.com. Consumer comments from Amazon were used to reflect a nonexpert or mixed consumer response, while user comments from CNET were used to measure expert response. Amazon is a website that sells everything from books to toys and has a broad audience. CNET, on the other hand, is a website dedicated exclusively to technology and is likely to have posters with greater expertise. Archival data also suggests that there are differences among visitors to the two sites.

According to Quantcast estimates (Quantcast 2010a, CNET Monthly Traffic (Estimated)) (www.quantcast.com/cnet.com), users to CNET.com are predominantly male and likely to visit websites like majorgeeks.com and read PC World. Amazon users, on the other hand, represent a broader demographic. They are more evenly divided between men and women (48/52), are more likely to have kids, and, visit websites like buy.com (Quantcast 2010b, Amazon monthly traffic (estimated)) (www.quantcast.com/amazon.com). Data were collected on November 2009.

Data were collected with the help of a research assistant from Amazon.com and CNET.com from September 5, 2007 to November 6, 2009. Keyword search for “iPod Touch” was used to gather all customer reviews available for the product at the time of analysis. Reviews for multiple versions of the device (first and second generation) were included and segmented in the analysis according to release date. The first-generation iPod Touch was released on September 5, 2007, and the second-generation was released on September 9, 2008.

Data were scraped from the internet, stored in a spreadsheet, and segmented by post. The comment date, poster name, rating, location of the poster, and the text of the comment itself were all stored as separate variables. Two levels of analysis were chosen. The most basic level of analysis is at the comment level. Each comment was coded for its content so that correlations between the content of that post and the date, poster experience, and location could be assessed. The second level of analysis is the group level, between Amazon and CNET. Comparisons can thus be made between expert and nonexpert groups based on the assumption that Amazon posters are nonexperts or a mix of experts and nonexperts, while dedicated members of the CNET community have more expertise. Lastly, because the time variable exists in the dataset, it will also be possible to periodize the data. This may be relevant in assessing the effects of different product launches (e.g., first- vs. second-generation iPods) on the textual content of posts. About 204 posts were collected from Amazon and 269 posts were collected from CNET, yielding a sample size high enough to make statistical comparisons between groups.

After a file structure was created, data were cleaned by running a spell check on all entries. Slang words (e.g., “kinda”) were replaced with their proper counterparts. Text was scanned for problematic words. For example, “touch” appeared with greater frequency than usual because it was used to refer to the product, not to the sense. For that reason, “touch” was replaced with a noncodable character like “TTT” so that it would not be counted in the haptic category used in the standard dictionary.

Stage 3: Construct Definition

Work in information processing suggests that experts process information differently from novices (Alba and Hutchinson 1987). In general, experts view products more cognitively, evaluating product attributes over benefits or uses (Maheswaran and Sternthal 1990; Maheswaran et al. 1996; Sujan 1985). While novices use only stereotypical information, experts use both attribute information and stereotypical cues (Maheswaran 1994). Experts are able to assimilate categorical ambiguity, which means one would expect for them to adjust to an ambiguous product more quickly than nonexperts (Meyers-Levy and Tybout 1989). They also tend to approach judgment in an abstract, higher level construal than nonexperts (Hong and Sternthal 2010).

From previous research, several working hypotheses can be developed. The strategic comparison we wish to make is about how experts versus nonexperts evaluate the product and whether or not this changes over time. First, one might expect that experts would use more cognitive language and that they would more critically evaluate the device.

H1: Experts will use more cognitive language than novices.

Secondly, one would also expect that experts would attend to features of the device, but nonexperts would attend more to uses of the device (Maheswaran et al. 1996). Note that this is based on the necessary assumption that users discuss or verbally elaborate on what draws their mental attention, which is reasonable according to previous research (Carley 1997).

H2: Experts will discuss features more than nonexperts.

H3: Nonexperts will discuss benefits and uses more than experts.

Thirdly, over time, one might predict that experts would be able to assimilate ambiguous product attributes while nonexperts would not. Because experts can more easily process ambiguous category information and because they have a higher construal level, one would predict that they would like this ambiguous product more than novices and would learn to assimilate the ambiguous information. For example, in this case, the capacity of the device makes it hard to categorize (cell phone vs. mp3 player). One would expect that experts would more quickly understand this ambiguity and that over time their elaboration on this feature would decrease.

H4: Experts will talk about ambiguous attributes (e.g., capacity) less over time, while nonexperts will continue to discuss ambiguous attributes.Lastly, previous research suggests that these differences in focus, experts on features and nonexperts on benefits, would differentially influence product ratings. That is, ratings for non-experts will depend on evaluation of benefits such as entertainment, but expert ratings would be influenced more by features.

H5: Ratings will be driven by benefits for nonexperts.

H6: Ratings will be driven by features by experts.

These are only a few of the many potential hypotheses that could be explored in an analysis of online word-of-mouth communication. One could equally explore the cultural framing of new technologies (Giesler 2008) or the co-production of brand communications by seeding product reviews with bloggers (Kozinets 2010). The question posed here – do experts respond differently to new products than non-experts over time? – is meant to be illustrative of what can be done with automated text analysis rather than a rigorous test of the psychological properties of expertise.

In this illustrative example, the key constructs in examining H1 through H6 are known: expert and nonexperts, cognitive expressions, affect, product features, and benefits. We therefore proceed with a top-down approach. Operationalization for some of the constructs – cognitive and affective language – is available through a standardized measure (LIWC; Pennebaker et al. 2001), and we can therefore use a standardized dictionary for their operationalization. However, some constructs such as features and benefits are context-specific, and a custom dictionary will be necessary for operationalization. In addition, there may be other characteristics that distinguish experts from nonexperts. We will therefore also perform a bottom-up approach of classification.

Stage 4: Operationalization

For this analysis, the standard LIWC dictionary developed by Pennebaker et al. (2001) was used in addition to a custom dictionary. Table 3 presents the categories used from both the standardized and the custom dictionaries. The standard dictionary includes categories for personal pronouns such as “I,” parts of speech such as adjectives, psychometrically pretested categories such as positive and negative emotion, and content-related categories such as leisure, family, and friend-related language.

A custom dictionary was also developed to identify categories specific to the product word-of-mouth data analyzed here. Ten comments from each website were selected and open coded, with the researcher blind to the site from which they came. Then, ten more comments from each website were selected and codes were added until saturation was reached (Weber 2005). In all, the subsample required to develop the custom dictionary was 60 comments, 30 from each website, about 11% of all comments. Fourteen categories were created, each containing six words on average.

The qualitative analysis of comments revealed posters tended to talk about the product in terms of features or aesthetics. Dictionary categories were therefore created for words associated with features (e.g., GPS, camera, hard drive, battery) and for aesthetics (e.g., sharp, clean, sexy, sleek). Posters also had recurring concerns about the capacity of the device, the cost of the product, and reported problems they

Table 3 Standard and custom dictionaries

Category	Abbv	Words	No. of words	Alpha*
Social processes	Social	Mate, talk, they, child	455	97%
Affective processes	Affect	Happy, cried, abandon	915	97%
Positive emotion	Posemo	Love, nice, sweet	406	97%
Negative emotion	Negemo	Hurt, ugly, nasty	499	97%
Cognitive processes	Cogmech	Cause, know, ought	730	97%
Past tense	Past	Went, ran, had	145	94%
Present tense	Present	Is, does, hear	169	91%
Future tense	Future	Will, gonna	48	75%
Discrepancy	Discrep	Should, would, could	76	80%
Exclusive	Excl	But, without, exclude	17	67%
Perceptual processes	Percept	Observing, heard, feeling	273	96%
Relativity	Relativ	Area, bend, exit, stop	638	98%
Space	Space	Down, in, thin	220	96%
Time	Time	End, until, season	239	94%
Work	Work	Job, majors, xerox	327	91%
Aesthetics	Aesth	Sleek, cool, shiny, perfect	9	83%
Capacity	Cap	Capacity, space, storage	7	93%
Cost	Cost	Price, cost, dollars	6	100%
Big	Big	Large, huge, full	5	83%
Problems	Prob	Bugs, crash, freeze	7	100%
Competitors	Comp	Zune, Microsoft, Archos	4	67%
Apple	Apple	Nano, iPod, iPhone	4	100%
Entertainment	Ent	Music, video, fun	9	85%
Job	Job	Work, commute, conference	9	100%
Connectability	Connect	Wifi, internet, web	9	95%
Features	Feat	GPS, camera, battery	5	87%
Love	Love	Amazing, best, love	7	100%
Small	Small	Empty, small, tiny	4	100%
Expertise	Expert	Jailbreak, jailbroke, keynote	4	67%

*Alpha is the percent agreement of three coders on dictionary words in the category

experienced using the product. Categories were created for each of these concerns. Because there might be some researcher-driven interest in product uses and because posters frequently mentioned entertainment and work-related uses, categories were created for each type of use. Categories of "big" versus "small" were included because previous theorization in sociology has suggested that the success of the iPod comes from its offerings of excess – large screen, excess capacity, etc. (Sennett 2006). Two categories were created to count when competitive products were mentioned, either within the Apple brand or outside of it.

The dictionary categories were validated by three coders who suggested words for inclusion and exclusion. Percent agreements between coders on each dictionary category can be found in Table 3. Average agreement was 90%. Text files were run through the LIWC program, first using the standard dictionary, then using the custom dictionary. A spreadsheet was created from three sets of data: (1) the comment data collected directly from the website (e.g., date of post, rating of product), (2) the computer output from the standard dictionary, and (3) the output from the custom dictionary.

Validation. Once rough findings were gleaned, the coding was validated. Twenty instances from each category were pulled from the dataset and categorized. “Hits” and “false hits” were then calculated. This yielded an average hit rate of 85% and a “false hit” rate of 15%. The least accurate category was aesthetics, with a hit rate of 70% and a false hit rate of 30%. The most accurate category was “small,” which had a hit rate of 95% and a false hit rate of 5%.

Stage 5: Interpretation and Analysis

Overall, the findings indicate that there are systematic differences between the way experts and nonexperts interpret the new device. As with most textual data, there are many potential variables and measures of interest. The standard LIWC dictionary contains 61 categories, and in the dataset studied here, 28 of these categories were significantly different among text from the three websites. We will report some of the most notable differences, including those needed to test the hypotheses.

Comparison between groups. First, we assessed differences among the two groups of comments. This was done by comparing differences in the percent of words coded in each category between groups using the Mann-Whitney test due to the skewed distribution of the data. Tables 4 and 5 show the differences by category. With the standard dictionary, several important differences between the word of mouth of nonexperts and experts can be discerned.

First, experts use more cognitive words ($M_{cog|CNET} = 16.57$, $M_{cog|Amazon} = 15.64$, Mann-Whitney $U = 30{,}562$, $z = 2.12$ $p < 0.05$) than nonexperts, but they also use more affective (both positive and negative) language ($M_{affect|CNET} = 7.3$ vs. $M_{affect|Amazon} = 6.53$, $U = 30, 581$, $z = 2.14$, $p < 0.05$) as well. The finding that experts evaluate the product cognitively is congruent with previous research (Maheswaran et al. 1996), and the highly affective tone indicates that they are likely more involved in product evaluation (Kelting and Duhacheck 2009). However, CNET posters use more negation ($M_{neg|CNET} = 2.47$, $M_{neg|Amazon} = 1.74$, $U = 34{,}487$, $z = 4.81$, $p < 0.001$). Together with the presence of cognitive language, this indicates that they may be doing more critical evaluation. The first hypothesis was therefore supported.

Secondly, nonexperts focus on distal rather than proximate uses, while experts focus on device-related issues like features. Nonexperts on Amazon use more distal social, time-, family-related language (e.g., $M_{social|Amazon} = 5.55$ vs. $M_{scoial|NET} = 4.23$, $U = 22{,}259.5$, $z = -3.52$, $p < 0.001$ and $M_{time|Amazon} = 5.65$, $M_{time|CNET} = 3.89$, $U = 18{,}527$ $z = -6.01$, $p < 0.001$). Experts on CNET, on the other hand, focus on features ($M_{features|CNET} = 0.61$ vs. $M_{features|Amazon} = 0.41$, $U = 30{,}012.5$, $z = 2.10$, $p < 0.05$) and capacity ($M_{connect|CNET} = 1.08$ vs. $M_{connect|Amazon} = 0.756$, $U = 35{,}819$, $z = 6.14$, $p < 0.001$), but also on aesthetics

Table 4 Amazon vs. CNET differences in means, standard dictionary

	Amazon	CNET
WC	160.99	149.11
Social***	5.55	4.23
Affect†	6.53	7.20
Posemo	5.50	5.94
Negemo	1.10	1.31
Cogmech*	15.64	16.57
Past***	3.58	2.13
Present	8.91	9.22
Future*	0.76	1.01
Certain	1.66	1.87
Excl**	2.68	3.20
Percept***	3.34	4.86
Relativ***	11.26	9.53
Space*	4.06	4.64
Time***	5.65	3.89
Work	2.08	1.92
Achieve	2.24	2.58
Leisure†	3.28	3.80

†$p < 0.10$
*$p < 0.05$
**$p < 0.01$
***$p < 0.001$

Table 5 Differences in means, custom dictionary

	Amazon	CNET
Aesthetics***	0.168	0.833
Capacity***	0.538	1.408
Cost*	0.384	0.641
Big**	0.070	0.178
Problems†	0.286	0.165
Competitors	0.080	0.104
Apple*	1.461	1.927
Entertainment**	1.377	1.838
Job†	0.164	0.087
Connect*	0.756	1.075
Features†	0.413	0.606
Love***	0.746	1.470
Small*	0.054	0.135
Expert*	0.009	0.028

†$p < 0.10$
*$p < 0.05$
**$p < 0.01$
***$p < 0.001$

($M_{aesth|CNET} = 0.833$ vs. $M_{aesth|Amazon} = 0.168$, $U = 33{,}518$, $z = 5.02$, $p < 0.001$). Experts discussed aesthetics about eight times more than the mixed group on Amazon. These differences indicate that, in general, experts focus on the device itself while nonexperts focus on uses. This lends convergent evidence to support to H2 and H3.

One other finding not specified by the hypotheses is notable. Nonexperts use more past-oriented language ($M_{past|Amazon} = 3.58$ vs. $M_{past|CNET} = 2.13$, $U = 21{,}289$, $z = -4.20$, $p < 0.001$), while expert posters use more future-oriented language ($M_{future|CNET} = 1.01$, $M_{future|Amazon} = 0.76$, $U = 31{,}446$, $z = 2.83$, $p < 0.01$). This suggests that experts might frame the innovation in the future while nonexperts focus on the past. Recent research suggests experts and novices differ in temporal construal (Hong and Sternthal 2010). Experts focus on the far future while novices focus on the near future. The results here provide convergent evidence that supports previous research and suggests a further hypothesis – that novices focus on past-related information – for future experimental research (Table 6).

In an extended analysis, adding a third group could help the researcher draw more rigorous conclusions through techniques of analytic induction (Mahoney 2003; Mill 1843). That is, if an alternative explanation is possible, the researcher could include a comparison set to rule out the alternative explanation. For example, one might propose that the difference in "cost" discourse is because Amazon.com users make less money than CNET users, on average, and are therefore more concerned about price. One could then include an expert website where the users are known to have a lower income than the posters on Amazon to address this explanation. If the same results are found, this would rule out the alternative hypothesis.

Trends over time. Because the product studied here is an innovation, the change of comments over time as the product diffuses is of interest. Time was analyzed first as a continuous variable in a correlation analysis and then as a discrete variable in ordinary least squares regression analyses, where the release of the first and second generation of iTouch marked each period.

A correlation analysis was used to analyze time as a continuous variable (Table 7). We find that affect increases over time in the expert group, which indicates that group becomes more involved ($r_{(affect,\ Date|CNET)} = 0.144$, $p < 0.01$). Experts become less concerned with capacity ($r_{(capacity,\ Date|CNET)} = -0.203$ $p < 0.01$) while Amazon users do not change in their concern for capacity. This indicates that experts learn something about the product category: the limited capacity was initially a shock to reviewers, as it was unorthodox for an mp3 player. But, over time, experts learned that this new category segment – mp3 wireless devices – did not offer as much memory. This supports Hypothesis 4 (Fig. 1).

Besides the correlation analysis, we also did ordinary least square linear regression analyses to analyze whether reviewers' expressions changed over time (Table 8). We created a binary variable, which is set to "1″ if the review is posted after the second generation of iTouch is released, and "0″ if the review is for the first generation of iTouch. To account for asymmetry in their distributions due to non-normality, we log-transformed the term frequency measurements of affect and capacity, our variables of interest. The results from the OLS analyses are congruent

Table 6 Correlation table, Amazon vs. CNET

Correlations														
Statistics = Pearson correlation														
	Site	Rating	Date	Affect	Posemo	Negemo	Aesth	Capacity	Ent	Connect	Feat	Love	Big	Small
Rating	Amazon	1	0.009	0.282[a]	0.387[a]	−0.200[a]	0.061	0.064	0.216[a]	0.002	0.128	0.273[a]	0.015	−0.024
	CNET	1	−0.012	0.095	0.319[a]	−0.433[a]	0.024	−0.058	0.044	0.145[b]	−0.118	0.373[a]	0.091	−0.053
Date	Amazon	0.009	1	−0.087	−0.046	−0.118	−0.082	0.013	0.073	0.008	−0.040	0.022	−0.156[b]	−0.095
	CNET	−0.012	1	0.144[b]	0.145[b]	0.011	−0.009	−0.203[a]	0.114	0.127[b]	−0.102	−0.006	−0.106	−0.001
Affect	Amazon	0.282[a]	−0.087	1	0.910[a]	0.350[a]	−0.049	−0.098	−0.043	−0.187[a]	0.049	0.450[a]	−0.001	−0.036
	CNET	0.095	0.144[b]	1	0.865[a]	0.263[a]	0.367[a]	−0.036	0.111	0.036	0.108	0.411[a]	−0.096	0.034
Posemo	Amazon	0.387[a]	−0.046	0.910[a]	1	−0.056	0.005	−0.052	0.032	−0.164[b]	0.064	0.473[a]	0.006	−0.015
	CNET	0.319[a]	0.145[b]	0.865[a]	1	−0.253[a]	0.409[a]	−0.019	0.156[b]	0.106	0.104	0.514[a]	−0.038	−0.056
Negemo	Amazon	−0.200[a]	−0.118	0.350[a]	−0.056	1	−0.117	−0.140[b]	−0.194[a]	−0.104	−0.030	−0.013	0.026	−0.050
	CNET	−0.433[a]	0.011	0.263[a]	−0.253[a]	1	−0.086	−0.026	−0.087	−0.139[b]	0.000	−0.205[a]	−0.119	0.167[a]
Aesth	Amazon	0.061	−0.082	−0.049	0.005	−0.117	1	0.131	−0.019	0.016	0.005	−0.055	0.126	0.003
	CNET	0.024	−0.009	0.367[a]	0.409[a]	−0.086	1	−0.025	0.040	−0.052	0.291[a]	0.015	−0.072	−0.053
Capacity	Amazon	0.064	0.013	−0.098	−0.052	−0.140[b]	0.131	1	0.055	0.052	−0.044	−0.010	−0.046	0.144[b]
	CNET	−0.058	−0.203[a]	−0.036	−0.019	−0.026	−0.025	1	0.079	−0.177[a]	−0.079	−0.048	−0.025	0.020
Ent	Amazon	0.216[a]	0.073	−0.043	0.032	−0.194[a]	−0.019	0.055	1	0.139[b]	−0.022	−0.061	0.069	0.063
	CNET	0.044	0.114	0.111	0.156[b]	−0.087	0.040	0.079	1	0.023	−0.141[b]	0.072	0.055	−0.012

(*continued*)

Table 6 (continued)

Correlations														
Connect	Amazon	0.002	0.008	−0.187[a]	−0.164[b]	−0.104	0.016	0.052	0.139[b]	1	0.007	−0.055	−0.077	−0.009
	CNET	0.145[b]	0.127[b]	0.036	0.106	−0.139[b]	−0.052	−0.177[a]	0.023	1	0.008	0.139[b]	0.038	−0.056
Feat	Amazon	0.128	−0.040	0.049	0.064	−0.030	0.005	−0.044	−0.022	0.007	1	0.000	−0.019	−0.024
	CNET	−0.118	−0.102	0.108	0.104	0.000	0.291[a]	−0.079	−0.141[b]	0.008	1	−0.086	−0.045	−0.096
Love	Amazon	0.273[a]	0.022	0.450[a]	0.473[a]	−0.013	−0.055	−0.010	−0.061	−0.055	0.000	1	−0.016	−0.048
	CNET	0.373[a]	−0.006	0.411[a]	0.514[a]	−0.205[a]	0.015	−0.048	0.072	0.139[b]	−0.086	1	0.078	0.044
Big	Amazon	0.015	−0.156[b]	−0.001	0.006	0.026	0.126	−0.046	0.069	−0.077	−0.019	−0.016	1	0.055
	CNET	0.091	−0.106	−0.096	−0.038	−0.119	−0.072	−0.025	0.055	0.038	−0.045	0.078	1	0.059
Small	Amazon	−0.024	−0.095	−0.036	−0.015	−0.050	0.003	0.144[b]	0.063	−0.009	−0.024	−0.048	0.055	1
	CNET	−0.053	−0.001	0.034	−0.056	0.167[a]	−0.053	0.020	−0.012	−0.056	−0.096	0.044	0.059	1

[a]Correlation is significant at the 0.01 level (2-tailed)
[b]Correlation is significant at the 0.05 level (2-tailed)

Table 7 OLS regression coefficient estimates. Affect and capacity by time and Amazon vs. CNET

Dependent variable		B	Std. error
ln(capacity)	(Intercept)***	0.275	0.058
	Is 2nd Gen	0.024	0.081
	Is CNET***	0.407	0.069
	Is 2nd Gen × CNET***	−0.546	0.158
ln(affect)	(Intercept)***	1.916	0.048
	Is 2nd Gen	−0.043	0.068
	Is CNET	0.063	0.057
	Is 2nd Gen × CNET*	0.275	0.132

$p < 0.10$
*$p < 0.05$
**$p < 0.01$
***$p < 0.001$

with the correlation analysis. We observe that in general expert reviewers discussed capacity more than nonexperts ($\widehat{\beta} = 0.407$, $p < 0.001$). However, as predicted by Hypothesis 4, such discussions decreased after the release of the second-generation iPod ($\widehat{\beta} = -0.546$, $p < 0.001$).

Affect also changes differentially in each group (Fig. 2). The OLS analysis (Table 7) shows that in the first time-period, affective language is roughly equivalent, but experts on CNET use more affective language in the second time-period than they do in the first time-period ($\widehat{\beta} = 0.275$, $p < 0.05$). In short, site and period have a positive interactive effect on affective expressions. These are just two examples of how automated content analysis can be used to assess changes in word-of-mouth communication.

Regression with ratings. Now that relationships between semantic elements in the text have been discerned, their relationship to other, nonsemantic variables is of interest. For example, what factors impact ratings for experts vs. nonexperts? To test the impact of discourse on rating, an OLS regression was run with rating as the dependent variable and the discursive categories as the independent variables. Several discursive variables were significant predictors of ratings overall ($F_{Amazon} = 2.55$, $p < 0.05$; $F_{CNET} = 2.30$, $p < 0.05$). Results are shown in Table 8. These reveal that the ratings of nonexperts were influenced by entertainment and features, while the ratings of experts were affected by connectability and by the (negative) evaluation of the features. This provides support for H5 and H6. However, they also indicate a more complicated relationship. Features are correlated with both expert and nonexpert ratings. However, for nonexperts, features are positively correlated with ratings while for experts, they are negatively correlated. Problems and cost, although much discussed in the posts, appeared to have little effect on ratings. The unimportance of cost may be explained by the fact that the ratings data are nonbehavioral, that is, most posters had already purchased the device.

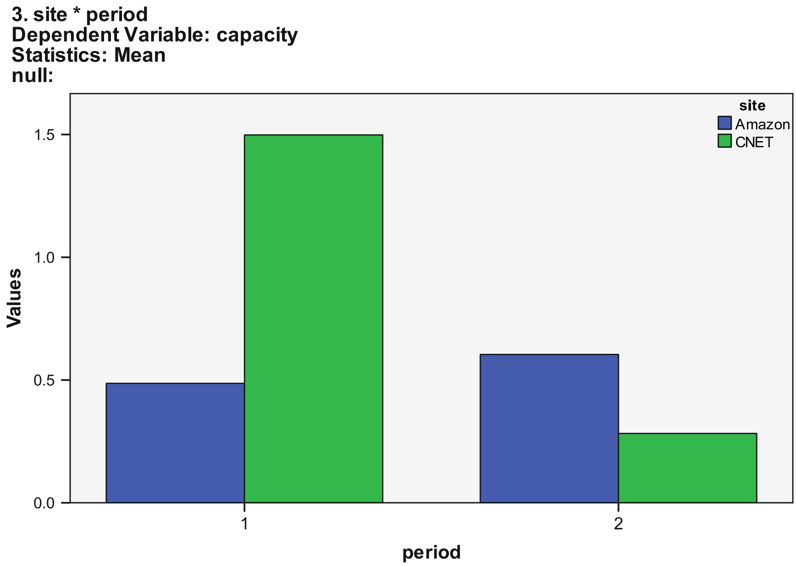

Fig. 1 Mean number of capacity words by site and time period

Table 8 Regression coefficients: predictors of product rating for experts vs. nonexperts

Coefficients						
Site	Category	Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
Amazon	(constant)	3.839	0.137		27.932	0.000
	Aesthetics	0.145	0.175	0.058	0.833	0.406
	Capacity	0.064	0.087	0.051	0.732	0.465
	Problems	−0.015	0.086	−0.012	−0.174	0.862
	Entertainment	**0.150**	**0.047**	**0.221**	**3.178**	**0.002**
	Connect	−0.035	0.073	−0.033	−0.476	0.635
	Features	**0.174**	**0.088**	**0.136**	**1.972**	**0.050**
CNET	(constant)	3.799	0.144		26.373	0.000
	Aesthetics	0.031	0.031	0.062	0.978	0.329
	Capacity	−0.029	0.042	−0.043	−0.697	0.486
	Problems	−0.290	0.195	−0.091	−1.484	0.139
	Entertainment	0.011	0.040	0.017	0.277	0.782
	Connect	**0.100**	**0.049**	**0.128**	**2.062**	**0.040**
	Features	**−0.126**	**0.059**	**−0.137**	**−2.138**	**0.033**

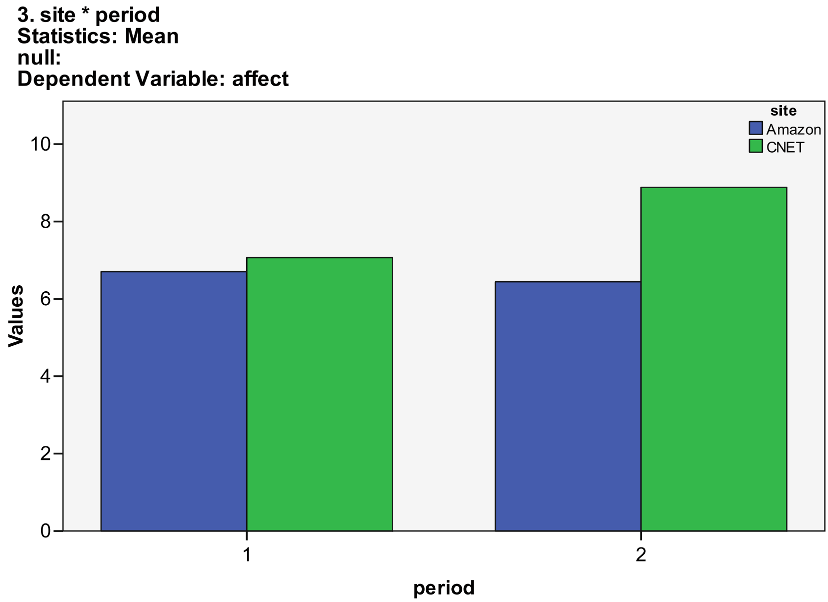

Fig. 2 Mean number of affect words by site and time period

Stage 6: Validation

The previous analyses revealed there were systematic differences in the number of words used between experts and nonexperts. To assess construct validity, we used a triangulation approach to explore the relationships between the concepts through a correlation analysis of word association within comment (Table 7). This means that we are looking for how the dictionary categories occur together within one post. To assess construct validity of affect, we included another operationalization of affect, star rating, in the correlational analysis. We calculated Pearson correlations for all categories in the set and compared them with cosine similarities. Both tables produced directionally similar results, and here we report Pearson correlations, as it accounts for both presence and absence of collocation. First, a few expected correlations between categories were checked. For both sites, positive emotion is correlated with rating ($r_{(posem,\ rating)} = 0.335$, $p < 0.01$), as one would expect. Negative emotion is negatively correlated with positive emotion ($r_{(negemo,\ posemo)} = -0.348$, $p < 0.01$). More can be learned, however, by comparing word association in expert versus nonexpert groups.

In general, nonexperts use positive language alongside distal uses for the iPod such as work and family ($r_{(work,posem|Amazon)} = 0.243$, $p < 0.01$ and $r_{(family,\ posemo|Amazon)} = 0.190$, $p < 0.01$). For the non-experts, negative emotion is correlated with problems, as one would expect ($r_{(problems,negem|Amazon)} = .470$). For experts, positive emotion occurs alongside aesthetics ($r_{(aesth,posem|CNET)} = 0.409$, $p < 0.01$). For experts, there is also a positive correlation between Apple and love ($r_{(Apple,\ love|}$

Table 9 Confusion matrix from tenfold cross-validation. Accuracy = 0.8013. *p*-Value [accuracy > no information rate] = < 2e-16

Prediction		Expert	Not expert
	Expert	237	62
	Not expert	32	142

$_{\text{CNET})}$ = 0.203, $p < 0.01$) that does not exist for nonexperts. These correlations indicate that aesthetics are viewed positively by experts and that they are involved with not only the device but the brand as well. Cosine similarities produce directionally similar results.

Secondly, features are interpreted differentially between the two groups. Novices interpret some features using standards of other categories (like an mp3 player), while experts are more willing to judge them relative to the standards for a new category. For example, from the correlation between small and capacity among the nonexpert group ($r_{(\text{capacity,small}|\text{Amazon})} = 0.144$, $p < 0.01$), one can conclude that posters feel the capacity is too small. No such correlation exists for experts. This could be because the iTouch is a product without a known category. Experts can interpret size for this ambiguous product, but novices are uncertain about what capacity is appropriate for the device. These are just a few of the findings that can be gleaned using a correlation Table. A full spatial analysis might compare the network of meanings in the Amazon group to the network of meanings in the CNET group.

For the binary logistic classification, k-fold cross-validation was performed, and per convention, we set $k = 10$. The resulting comparisons between predicted values based on our model and the real values show that overall the model is 80.13% accurate (95% accuracy confidence interval = [0.7624, 0.8363]). Table 9 shows the confusion matrix.

In sum, the automated text analysis presented here shows that that experts evaluate new products in a systematically different way from nonexperts. Using comparison between groups, we show that experts evaluate products by focusing on features while nonexperts focus on the uses and benefits of the devices. Using correlation analysis, we find that experts associate aesthetics with positive emotion while nonexperts associate positive emotion with uses of the device and negative emotion with problems. Further, the correlation analysis provides some validation for the method of automated content analysis by demonstrating the correlation between positive emotion and ratings, a variable used in previous studies of online word-of-mouth communication (Godes and Mayzlin 2004, 2009). We find that, over time, experts focus less on problematic features like capacity and speak more affectively about the product. A regression analysis of the elements of discourse on ratings demonstrates that ratings for experts are driven by features, while ratings by nonexperts are better predicted by both features and the amount of talk about entertainment, a benefit. Note that, like field research, these findings make sense in convergence with previous findings from experimental data and provide ecological validity to previous findings obtained in laboratory settings. These are not meant to be a rigorous test of expertise, but rather an illustration of the way in which text analysis can provide convergent evidence that is meaningful to consumer researchers.

Conclusion and Future Directions

Developments in text analysis have opened a large and fascinating arena for marketing research. Theoretically, marketing research can now incorporate linguistic theory to understand consumer attitudes, interaction, and culture (Humphreys and Wang 2018). While most approaches have focused on analyzing word frequencies, a vast world of looking at text structure at higher, conversational levels remain open. For example, understanding where a word like "great" falls within the text itself (early, middle, or late in a sentence or paragraph) may shed light on the importance of the word in predicting, for example, consumer sentiment. Drawing inferences on the sentence or paragraph level may yield more meaningful results in some contexts (Büschken and Allenby 2016). Lastly, pragmatics, the area of linguistic research aimed at understanding the effect of context on word meaning may help marketing researchers capture more about the nature of consumer communication online.

Practically, incorporating this kind of data allows researchers and managers to integrate the abundance of textual data with existing and growing datasets of behavioral data collected online or through devices. And yet one must be aware of the many limitations of using machines to interpret a human language that has developed socially in face-to-face contexts over 100,000 years. Text analysis can often be used to gather information about top-line patterns of attention or relatively wrote patterns of interaction, but capturing the subtly of human communication remains allusive to machines. Further, due to the ambiguity of language, careful and transparent analysis and interpretation are required at each step of text analysis, from cleaning textual markers that may be misleading to correctly interpreting correlations and differences. Despite these challenges, marketing researchers have clearly shown the theoretical, practical, and managerial insight that can be distilled through the seemingly simple process of counting words.

Cross-References

► Network Analysis
► Return on Media Models
► Social Media Tracking

References

Alba, J. W., & Hutchinson, J. W. (1987). Dimensions of consumer expertise. *Journal of Consumer Research, 13*(4), 411–454.

Arndt, J. (1967). Role of product-related conversations in the diffusion of a new product. *Journal of Marketing Research, 4*, 291–295.

Arsel, Z., & Bean, J. (2013). Taste regimes and market-mediated practice. *Journal of Consumer Research, 39*(5), 899–917.

Arvidsson, A., & Caliandro, A. (2016). Brand public. *Journal of Consumer Research, 42*(5), 727–748.

Barasch, A., & Berger, J. (2014). Broadcasting and narrowcasting: How audience size affects what people share. *Journal of Marketing Research, 51*(3), 286–299.

Belk, R. W., & Pollay, R. W. (1985). Images of ourselves: The good life in twentieth century advertising. *Journal of Consumer Research, 11*(4), 887.

Berelson, B. (1971). *Content analysis in communication research*. New York: Hafner.

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research, 49*(2), 192–205.

Blei, David M., Andrew Y. Ng, & Michael I. Jordan. (2003). Latent dirichlet allocation. *Journal of machine Learning research 3*, 993–1022.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computer Science, 2*(1), 1–8.

Boyd, R. L., & Pennebaker, J. W. (2015a). Away with words. In *Consumer psychology in a social media world* (p. 222). Abingdon: Routledge.

Boyd, R. L., & Pennebaker, J. W. (2015b). Did Shakespeare write double falsehood? Identifying individuals by creating psychological signatures with text analysis. *Psychological Science, 26* (5), 570–582.

Brown, J. J., & Reingen, P. H. (1987). Social ties and word-of-mouth referral behavior. *Journal of Consumer Research, 14*(3), 350–362.

Büschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Marketing Science, 35*(6), 953–975.

Carley, K. (1997). Network text analysis: The network position of concepts. In C. W. Roberts (Ed.), *Text analysis for the social sciences: Methods for drawing statistical inferences from texts and transcripts*. Mahwah: Lawrence Erlbaum.

Chung, C. K., & Pennebaker, J. W. (2013). Counting little words in Big Data. In *Social cognition and communication* (p. 25). New York: Psychology Press.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407.

Constant, N., Davis, C., Potts, C., & Schwarz, F. (2009). The pragmatics of expressive content: Evidence from large corpora. *Sprache und Datenverarbeitung, 33*(1–2), 5–21.

De Choudhury M., Sundaram H., John A., & Seligmann D. D. (2008). Can blog communication dynamics be correlated with stock market activity? In *Proceedings of the nineteenth ACM conference on hypertext and hypermedia*, ACM, pp. 55–60

Duhachek, Adam, and Katie Kelting. (2009). Coping repertoire: Integrating a new conceptualization of coping with transactional theory. *Journal of Consumer Psychology 19*(3), 473–485.

Dunphy, D. M., Bullard, C.G., & Crossing, E.E.M. (1974). Validation of the general inquirer Harvard Iv Dictionary. Paper presented at the 1974 Pisa conference on content analysis, Pisa, Italy.

Ertimur, B., & Coskuner-Balli, G. (2015). Navigating the institutional logics of markets: Implications for strategic brand management. *Journal of Marketing, 79*(2), 40–61.

Fishbein, M., & Ajzen, I. (1972). Attitudes and opinions. *Annual Review of Psychology, 23*(1), 487–544.

Fiss, P. C., & Hirsch, P. M. (2005). The discourse of globalization: Framing and sensemaking of an emerging concept. *American Sociological Review, 70*(1)., 24p.

Gamson, W. A., & Modigliani, A. (1989). Media discourse and public opinion on nuclear power: A constructionist approach. *The American Journal of Sociology, 95*(1), 1–37.
Gandolfo, A., Tuan, A., Corciolani, M., & Dalli, D. (2016). What do emerging economy firms actually disclose in their CSR reports? A longitudinal analysis. In *CSR-HR Project (Corporate Social Responsability and Human Rights Project). Research Grant of University of Pisa (PRA_2015_0082).*
Garrett, D. E. (1987). The effectiveness of marketing policy boycotts: Environmental opposition to marketing. *Journal of Marketing, 51*(2), 46–57.
Giesler, M. (2008). Conflict and compromise: drama in marketplace evolution. *Journal of Consumer Research, 34*(6), 739–753.
Godes, D., & Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing Science, 23*(4), 545–560.
Godes, D., & Mayzlin, D. (2009). Firm-created word-of-mouth communication: Evidence from a field test. *Marketing Science, 28*(4), 721–739.
Grayson, K., & Rust, R. (2001). Interrater reliability assessment in content analysis. *Journal of Consumer Psychology, 10*(1/2), 71–73.
Grice, H. P. (1975). *Logic and Conversation*. Syntax and Semantics, vol.3 edited by P. Cole and J. Morgan, Academic Press. Reprinted as ch.2 of Grice 1989, 22–40.
Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis, 21*(3), 267–297.
Homburg, C., Ehm, L., & Artz, M. (2015). Measuring and managing consumer sentiment in an online community environment. *Journal of Marketing Research, 52*(5), 629–641.
Hong, J., & Sternthal, B. (2010). The effects of consumer prior knowledge and processing strategies on judgments. *Journal of Marketing Research, 47*(2), 301–311.
Humphreys, A. (2010). Megamarketing: The creation of markets as a social process. *Journal of Marketing, 74*(2), 1–19.
Humphreys, A., & Latour, K. A. (2013). Framing the game: Assessing the impact of cultural representations on consumer perceptions of legitimacy. *Journal of Consumer Research, 40*(4), 773–795.
Humphreys, A., & Thompson, C. J. (2014). Branding disaster: Reestablishing trust through the ideological containment of systemic risk anxieties. *Journal of Consumer Research, 41*(4), 877–910.
Humphreys, A. (2015). *Social media: Enduring principles*. New York/Oxford: Oxford University Press.
Humphreys, A., & Wang, R. J.-H. (2018). Automated text analysis for consumer research. *Journal of Consumer Research, 44*(6), 1274–1306. https://doi.org/10.1093/jcr/ucx104
Hutto, C. J., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.
Jia, L., Clement, Y., & Meng, W. (2009). The effect of negation on sentiment analysis and retrieval effectiveness. In *Proceedings of the 18th ACM conference on information and knowledge management*: ACM, pp. 1827–1830.
Jurafsky, D., Chahuneau, V., Routledge, B. R., & Smith, N. A. (2014). Narrative framing of consumer sentiment in online restaurant reviews. *First Monday, 19*(4). https://doi.org/10.5210/fm.v19i4.4944.
Kassarjian, H. H. (1977). Content analysis in consumer research. *Journal of Consumer Research, 4*(1), 8–19.
Kolbe, R. H., & Burnett, M. S. (1991). Content-analysis research: An examination of applications with directives for improving research reliability and objectivity. *Journal of Consumer Research, 18*(2), 243–250.
Kovács, B., Carroll, G. R., & Lehman, D. W. (2013). Authenticity and consumer value ratings: Empirical tests from the restaurant domain. *Organization Science, 25*(2), 458–478.
Kozinets, R. V. (2010). Networked narratives: Understanding word-of-mouth marketing in online communities. *Journal of Marketing, 74*(2), 71–89.

Kranz, P. (1970). Content analysis by word group. *Journal of Marketing Research, 7*(3), 377–380.
Krauss, J., Nann, S., Simon, D., Gloor, P. A., & Fischbach, K. (2008). Predicting movie success and academy awards through sentiment and social network analysis. In *ECIS*, pp. 2026–2037.
Lasswell, H. D., & Leites, N. (1949). *Language of politics; studies in quantitative semantics*. New York: G. W. Stewart.
Lee, F., Peterson, C., & Tiedens, L. Z. (2004). Mea culpa: Predicting stock prices from organizational attributions. *Personality and Social Psychology Bulletin, 30*(12), 1636–1649.
Lee, T. Y., & Bradlow, E. T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research, 48*(5), 881–894.
Ludwig, S., Ko, d. R., Friedman, M., Brüggen, E. C., Wetzels, M., & Pfann, G. (2013). More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. *Journal of Marketing, 77*(1), 87–103.
Ludwig, S., Van Laer, T., De Ruyter, K., & Friedman, M. (2016). Untangling a web of lies: Exploring automated detection of deception in computer-mediated communication. *Journal of Management Information Systems, 33*(2), 511–541.
Maheswaran, D., & Sternthal, B. (1990). The effects of knowledge, motivation, and type of message on ad processing and product judgments. *Journal of Consumer Research, 17*(1), 66–73.
Maheswaran, D. (1994). Country of origin as a stereotype: Effects of consumer expertise and attribute strength on product evaluations. *Journal of Consumer Research, 21*(2), 354–365.
Maheswaran, D., Sternthal, B., & Gurhan, Z. (1996). Acquisition and impact of consumer expertise. *Journal of Consumer Psychology, 5*(2), 115.
Mahoney, J. (2003). Strategies of causal assessment in comparative historical analysis. In J. Mahoney & D. Rueschemeyer (Eds.), *Comparative historical analysis in the social sciences*. Cambridge, UK/New York: Cambridge University Press. pp. xix, 444.
Mankad, S., Han, H. S., Goh, J., & Gavirneni, S. (2016). Understanding online hotel reviews through automated text analysis. *Service Science, 8*(2), 124–138.
Mehl, M. R., & Gill, A. J. (2008). Automatic text analysis. In S. D. G. J. A. Johnson (Ed.), *Advanced methods for behavioral research on the internet*. Washington, DC: American Psychological Association.
Mestyán, M., Yasseri, T., & Kertész, J. (2013). Early prediction of movie box office success based on Wikipedia activity big data. *PLoS One, 8*(8), e71226.
Meyers-Levy, J., & Tybout, A. M. (1989). Schema congruity as a basis for product evaluation. *Journal of Consumer Research, 16*(1), 39–54.
Mill, J. S. (1843). *A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence, and methods of scientific investigation*. London: J.W. Parker.
Moe, Wendy W., and Michael Trusov. (2011). The value of social dynamics in online product ratings forums. *Journal of Marketing Research 48*(3), 444–456.
Moe, W. W., & Schweidel, D. A. (2014). *Social media intelligence*. Cambridge, UK/New York: Cambridge University Press.
Mogilner, C., Kamvar, S. D., & Aaker, J. (2010). The shifting meaning of happiness. *Social Psychological and Personality Science, 2*(4), 395–402.
Money, R. B., Gilly, M. C., & Graham, J. L. (1998). Explorations of national culture and word-of-mouth referral behavior in the purchase of industrial services in the United States and Japan. *Journal of Marketing, 62*, 76–87.
Monroe, B. L., Colaresi, M. P., & Quinn, K. M. (2009). Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis, 16*(4), 372–403.
Moore, S. G. (2015). Attitude predictability and helpfulness in online reviews: The role of explained actions and reactions. *Journal of Consumer Research, 42*(1), 30–44.
Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science, 31*(3), 521–543.

Opoku, R., Abratt, R., & Pitt, L. (2006). Communicating brand personality: Are the websites doing the talking for the top South African business schools? *Journal of Brand Management, 14*(1–2), 20–39.
Packard, G., Moore, S. G., & McFerran, B. (2014). How can "I" help "you"? The impact of personal pronoun use in customer-firm agent interactions. MSI report, pp. 14–110.
Packard, G. M., & Wooten, D. B. (2013). Compensatory knowledge signaling in consumer word-of-mouth. *Journal of Consumer Psychology 23*(4), 434–450.
Palmquist, M. E., Carley, K., & Dale, T. (2009). Analyzing maps of literary and non-literary texts. In K. Krippendorff & M. A. Bock (Eds.), *The content analysis reader* (pp. 4120–4415). Thousand Oaks: Sage.
Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: Liwc 2001* (Vol. 71). Mahway: Lawrence Erlbaum Associates.
Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77*(6)., 17p.
Petty, R. E., & Cacioppo, J. T. (1979). Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses. *Journal of Personality and Social Psychology, 37*(10), 1915.
Phelps, J. E., Lewis, R., Mobilio, L., Perry, D., & Raman, N. (2004). Viral marketing or electronic word-of-mouth advertising: Examining consumer responses and motivations to pass along email. *Journal of Advertising Research, 44*(4), 333–348.
Plaisant, C., Rose, J., Bei, Y., Auvil, L., Kirschenbaum, M. G., Smith, M. N., Clement T., & Lord G. (2006). Exploring erotics in emily Dickinson's correspondence with text mining and visual interfaces. In *Proceedings of the 6th ACM/IEEE-CS joint conference on digital libraries*, ACM, pp. 141–150.
Potts, C., & Schwarz, F. (2010). Affective 'this'. *Linguistic Issues in Language Technology, 3*(5), 1–30.
Quantcast. (2010a) Cnet monthly traffic (estimated). (www.quantcast.com/cnet.com).
Quantcast. (2010b). Amazon monthly traffic (estimated). (www.quantcast.com/amazon.com).
Rayson, P. (2009). Wmatrix: A web-based corpus processing environment. Edited by C. Department, Lancaster University, UK.
Salton, Gerard, and Michael J. McGill. (1983). Introduction to modern information retrieval McGraw-Hill. New York.
Schweidel, D. A., & Moe, W. W. (2014). Listening in on social media: A joint model of sentiment and venue format choice. *Journal of Marketing Research, 51*(4), 387–402.
Sennett, R. (2006). *The culture of the new capitalism*. New Haven: Yale University Press.
Snefjella, B., & Kuperman, V. (2015). Concreteness and psychological distance in natural language use. *Psychological Science, 26*(9), 1449–1460.
Spiller, S. A., & Belogolova, L. (2016). On consumer beliefs about quality and taste. *Journal of Consumer Research, 43*(6), 970–991.
Stephen, A. T., & Toubia, O. (2010). Deriving value from social commerce networks. *Journal of Marketing Research, 47*(2), 215–228.
Stevenson, T. H., & Swayne, L. E. (1999). The portrayal of African-Americans in business-to-business direct mail: A benchmark study. *Journal of Advertising, 28*(3), 25–35.
Stone, P. J. (1966). *The general inquirer; a computer approach to content analysis*. Cambridge: MIT Press.
Sujan, M. (1985). Consumer knowledge: Effects on evaluation strategies mediating consumer judgments. *Journal of Consumer Research, 12*(1), 31–46.
Tirunillai, S., & Tellis, G. J. (2012). Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science, 31*(2), 198–215.
Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *Journal of Marketing Research, 51*(4), 463–479.

Van de Rijt, A., Shor, E., Ward, C., & Skiena, S. (2013). Only 15 minutes? The social stratification of fame in printed media. *American Sociological Review, 78*(2), 266–289.

Van Laer, T., Escalas J. E., Ludwig S., & Van den Hende E. A. (2017). What happens in Vegas stays on TripAdvisor? Computerized text analysis of narrativity in online consumer reviews.

Ordenes, V., Francisco, S. L., Ko, D. R., Grewal, D., & Wetzels, M. (2017). Unveiling what is written in the stars: Analyzing explicit, implicit, and discourse patterns of sentiment in social media. *Journal of Consumer Research, 43*(6), 875–894.

Weber, K. (2005). A toolkit for analyzing corporate cultural toolkits. *Poetics, 33*(3/4), 26p.

Weber, M. (1924). Towards a sociology of the press. Paper presented at the first congress of sociologists, Frankfurt.

Winer, R. S. (2009). New communications approaches in marketing: Issues and research directions. *Journal of Interactive Marketing, 23*(2), 108–117. https://doi.org/10.1016/j.intmar.2009.02.004.

Zipf, G. K. (1932). *Selected studies of the principle of relative frequency in language*. Cambridge, MA: Harvard University Press.