

The writing is on the wall: predicting customers' evaluation of customer-firm interactions using computerized text analysis

The writing is
on the wall

309

Caitlin Ferreira

*Graduate School of Business, University of Cape Town, Cape Town, South Africa and
Luleå University of Technology, Luleå, Sweden*

Jeandri Robertson

*Luleå University of Technology, Luleå, Sweden and
University of Cape Town, Cape Town, South Africa*

Raesah Chohan

University of Cape Town, Cape Town, South Africa

Leyland Pitt

Beedie School of Business, Simon Fraser University, Burnaby, Canada, and

Tim Foster

Luleå University of Technology, Luleå, Sweden

Received 29 April 2022
Revised 10 October 2022
28 January 2023
Accepted 28 January 2023

Abstract

Purpose – This methodological paper demonstrates how service firms can use digital technologies to quantify and predict customer evaluations of their interactions with the firm using unstructured, qualitative data. To harness the power of unstructured data and enhance the customer-firm relationship, the use of computerized text analysis is proposed.

Design/methodology/approach – Three empirical studies were conducted to exemplify the use of the computerized text analysis tool. A secondary data analysis of online customer reviews ($n = 2,878$) in a service industry was used. LIWC was used to conduct the text analysis, and thereafter SPSS was used to examine the predictive capability of the model for the evaluation of customer-firm interactions.

Findings – A lexical analysis of online customer reviews was able to predict evaluations of customer-firm interactions across the three empirical studies. The authenticity and emotional tone present in the reviews served as the best predictors of customer evaluations of their service interactions with the firm.

Practical implications – Computerized text analysis is an inexpensive digital tool which, to date, has been sparsely used to analyze customer-firm interactions based on customers' online reviews. From a methodological perspective, the use of this tool to gain insights from unstructured data provides the ability to gain an understanding of customers' real-time evaluations of their service interactions with a firm without collecting primary data.

Originality/value – This research contributes to the growing body of knowledge regarding the use of computerized lexical analysis to assess unstructured, online customer reviews to predict customers' evaluations of a service interaction. The results offer service firms an inexpensive and user-friendly methodology to assess real-time, readily available reviews, complementing traditional customer research.



© Caitlin Ferreira, Jeandri Robertson, Raesah Chohan, Leyland Pitt and Tim Foster. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

Journal of Service Theory and
Practice
Vol. 33 No. 2, 2023
pp. 309-327
Emerald Publishing Limited
2055-6225
DOI 10.1108/JSTP-04-2022-0100

A tool has been used to transform unstructured data into a numerical format, quantifying customer evaluations of service interactions.

Keywords Digital technologies, Computerized text analysis, Online customer reviews, Customer experience, LIWC

Paper type Research paper

Don't waste customers' time asking them questions unless you are prepared to act on what they say.
- Bruce Temkin, Head of Qualtrics XM Institute (Charan, 2019)

Introduction

New technologies have revolutionized nearly every aspect of 21st century human existence, including how firms engage and build relationships with customers (Grewal *et al.*, 2020; Jain *et al.*, 2017). This technological revolution has resulted in fundamental changes in many service industries (Pemer, 2021), providing scholars the opportunity to look at the service industry at “an inflection point” (Wirtz *et al.*, 2018, p. 908). As technology expands in many service sectors (Al Awadhi *et al.*, 2021), the nature of customer-firm interactions are also transformed (Zouari and Abdelhedi, 2021), compelling marketers to get ahead of the curve (Grewal *et al.*, 2020). Staying ahead of the curve typically means finding new and innovative ways of interacting and learning from your customers (Li *et al.*, 2021). As customer-firm relationships are increasingly characterized by the combination of human interaction with technology (Singh *et al.*, 2022), firms have had to rethink how they can gain insights from those interactions. Among the most important areas for firms to consider is how they communicate with and listen to their customers (Al Awadhi *et al.*, 2021).

To effectively use technology in customer interactions, firms must not only become more innovative in how they engage with and connect to their customers, but also have a firm grasp on how to best make use of the data generated from technology-mediated interactions (Niu and Fan, 2018). As technology continues to become better, smarter and cheaper (Wirtz *et al.*, 2018), new ways of investigating and interpreting customer-firm interactions emerge (Nguyen *et al.*, 2020). One such opportunity is the use of computerized text analysis to analyze unstructured data in the form of verbatim, online customer reviews.

The words customers use when constructing online reviews reflect who they are and the social relationships that they are in, as “language is the most common and reliable way for people to translate their internal thoughts and emotions into a form that others can understand” (Tausczik and Pennebaker, 2010, p. 25). These online reviews thus provide a valuable source of information for firms (Balducci and Marinova, 2018; Yang and Fang, 2004), and their “unstructured, verbatim nature renders traditional market research methods (e.g. surveys, experiments, interviews and focus groups) ineffective” (Ludwig and De Ruyter, 2016, p. 124). Furthermore, with the increased challenge of survey respondent fatigue (O'Reilly-Shah, 2017; Viswanathan and Kayande, 2012) and low survey response rates (Luo, 2009), analyzing verbatim online customer reviews as opposed to self-reported survey data, provides an opportunity to better understand customers authentic sentiment towards a firm (Felbermayr and Nanopoulos, 2016).

From a methodological perspective, text mining is fundamentally based on the supposition that the frequency of certain words and ideas in a text is indicative of their relative significance, focus or emphasis (Krippendorff, 2018). In short, this means that psychologically relevant information is conveyed, beyond the words' literal meaning (Pennebaker *et al.*, 2003). The analysis of online customer reviews using computerized text analysis thus provides marketers with several opportunities. First, as a form of unstructured data, online reviews provide a flexible extension to numerical scores with which to

understand and evaluate customer-firm interactions (Korfiatis *et al.*, 2019). Second, in the absence of existing numerical scores, computerized text analysis enables the qualitative prose of an online customer review to be transformed into quantitative form for subsequent analysis (Pitt *et al.*, 2018; Tausczik and Pennebaker, 2010). Third, online customer reviews provide customers' natural language use. Certain computerized text analysis tools primarily rely on algorithms powered by artificial intelligence (AI) which makes use of natural language processing to understand *what* a customer has said in a particular online review or written text and *how* – providing psychological meaning and understanding to the text (Paschen *et al.*, 2019).

In service settings, the amount of unstructured customer data is estimated to be five times greater than the amount of structured customer data available (Balducci and Marinova, 2018). While service marketers acknowledge the importance of the correct analysis and interpretation of this data as a source of customer intelligence (Berger *et al.*, 2020; Boon *et al.*, 2013; Wang *et al.*, 2023), the question is no longer *whether* to use computer-aided text analysis tools, but rather *how* to approach a given dataset (Pollach, 2012). To obtain comprehensive and robust marketing insights from unstructured data in the form of customer-generated online text, recent calls have encouraged marketers to diversify their methodological approaches when analyzing technology-mediated customer interactions (Marketing Science Institute, 2022). A particular identified gap is our understanding of the conceptual linguistic foundation of the various performative functions of language and communication present in these online reviews (Ludwig and De Ruyter, 2016). In addition, limited research has had a methodological, empirical focus in showing *how* we can use computerized tools to generate evaluative, quantitative measures from unstructured, qualitative customer-firm interaction data (Rooney *et al.*, 2021).

This paper intends to begin to fill these gaps. To this end, the aim of this methodological study is to examine how service firms can use the computerized text analysis tool, Linguistic Inquiry and Word Count (LIWC), to analyze unstructured online data. More specifically, through the lens of Speech Act Theory (SAT) (Austin, 1975), we study key linguistic components of the language used in online customer reviews and demonstrate how a computerized text analysis of these qualitative online customer reviews can be used to quantitatively measure and predict the evaluations of customer-firm interactions. Two research objectives are developed. The first objective seeks *to explore how a computerized text analysis of online customer reviews can be used to evaluate customer-firm service interactions*. To contextualize the use of linguistic analysis as a tool for the measurement and prediction of online customer-firm service interactions, we present a second objective that practically applies the tool based on three cases. As such, the second objective seeks *to explore which linguistic components of a computerized text analysis are the best predictors of a customer's evaluation of their customer-firm interaction*. The section to follow provides a review of pertinent literature and the methodology used in the research. Thereafter the results are presented and discussed, followed by the implications and limitations of the research. Finally, recommendations for future research are suggested.

Literature review

Speech act theory and the language customers use in online reviews

People use language and words to communicate information about who they are, their relationship with their audience and their intentions (Pennebaker *et al.*, 2003). As one of the most influential linguistic theories to study language-in-use, SAT conceptualizes “all forms of speech as acts and suggests interpretations of communicated words require recognition of a higher-order linguistic context” (Ludwig and De Ruyter, 2016, p. 125). At its core, SAT proposes that the language we use represents performative acts on the side of the user

(i.e. speech act), which conveys intentions, thought and feelings (Austin, 1975; Searle, 1979). These speech acts may then elicit responses or trigger behavioral changes from recipients (Ludwig and De Ruyter, 2016). The word categories and sentence constructions, as evidenced in people's everyday language use, can provide insights into their intentions, perceptions and identities (Bagozzi *et al.*, 2007).

In the context of online generated content, content creators strategically use language, i.e. perform an act (Argyris *et al.*, 2021), to present themselves and communicate with others (Carr *et al.*, 2012). Thus, online content creators use language to encode their intentions, thoughts and feelings on online platforms for the purpose of communicating with those who read or consume the content. To respond to what they read online, content readers, in turn, also rely on language to decode the meanings associated with the created content (Labrecque *et al.*, 2020). Online customer reviews then, are a digital representation of traditional word-of-mouth (Goncalves *et al.*, 2018; Tsao, 2019), where customers use language to express (and share) their authentic thoughts, feelings and interactions with a firm with a wider audience of consumers (Wang *et al.*, 2023).

As an example of unstructured textual data, research has suggested that online customer reviews provide a fitting linguistic context within which SAT can be used to help structure and decode intended meaning (Grewal *et al.*, 2022). Extant research suggests that speech acts derive from multilevel language construction, which includes micro-level speech acts (word usage), macro-level speech acts (sentence developments) and meta-level speech acts (between-content exchanges) (Grewal *et al.*, 2022; Ludwig and De Ruyter, 2016; Wang *et al.*, 2023). In the context of online customer reviews, the micro-level acts would encompass how the customer combines words in their written review, the macro-level would reflect the intention and rationale of what the customer writes in the review, and the meta-level relates to the interaction between the customer (as the sender) and the firm (as the receiver), e.g. an online exchange of communication between the two parties. Every level of language construction relies on multilevel language usage to convey social and psychological meaning in the context of customer-firm interactions.

Online customer reviews and customer-firm interaction

Developments in digital, mobile and social technology have become increasingly vital in capturing and monitoring customer-firm interactions (Rooney *et al.*, 2021). Customer-firm interactions take place over time and include "direct contacts with a firm, its representatives or its offering, and indirect interactions outside the control of the firm" (Rooney *et al.*, 2021, p. 45). Not only do positive customer-firm interactions lead to higher customer retention (Anderson and Mittal, 2000), they also drive customers' long-term commitment to a firm (Keiningham *et al.*, 2017). Customer-firm interactions provide opportunities to foster engagement and facilitate the promotion of favorable customer-firm relationships (de Haan and Menichelli, 2020). One such opportunity for engagement is the increased availability of voluntarily provided online customer reviews.

Online customer reviews offer marketers a flourishing new stream of unstructured data (Key and Keel, 2020), that, with the correct analysis, can be transformed into structured, quantitative insights about customers' authentic evaluations of a firm (Pitt *et al.*, 2018; Robertson *et al.*, 2021). Researchers, however, face two challenges. The first challenge is the analysis of the data as the majority of the data are in the form of "unstructured qualitative prose" (Lord Ferguson *et al.*, 2021, p. 2426), such as comments on social media platforms, reviews on websites, or online posts on community forums. The second challenge is the sheer volume of data (Pitt *et al.*, 2018), both in the number of reviews as well as the large amount of text used in these reviews, which could be overwhelming to any firm wanting to glean relational insights into customers' experiences.

However, the sheer volume of qualitative online customer reviews posted daily provides marketers with the opportunity to accurately gauge the experiences, opinions, feelings and concerns of customers (Ludwig *et al.*, 2013; He *et al.*, 2017). The collection and subsequent analysis of these online reviews serves as a valuable tool through which firms can gain a more nuanced understanding of their customers' expectations of the firm (Yang and Fang, 2004). Online reviews also represent the voice of the customer, as customers primarily use online platforms to share their experiences (Duncan *et al.*, 2019), including thoughts and perceptions about services (Song *et al.*, 2016). Leveraging the content of online reviews, as assessed through the language used, not only serves to improve marketers' knowledge and understanding of customer preferences, but also the competitiveness of their service offering (Rossetti *et al.*, 2016). As such, these online customer reviews enable a firm to obtain insights about how to improve current service systems to deliver a better customer experience and improve customer-firm interactions for positive benefits (Song *et al.*, 2016). However, these benefits are reserved for firms that are able to effectively analyze and extract insights from online customer reviews.

Online customer reviews and computerized text analysis

Online customer reviews act as a source of information for both current and potential customers, with extant research indicating that a growing number of customers rely on online customer reviews to guide their own purchase behavior in global service industries such as tourism, hospitality (Wu *et al.*, 2016) and healthcare (Lord Ferguson *et al.*, 2021). Drawing on customers' online reviews, coupled with computerized text analysis tools, massive amounts of text can be analyzed to link customers' everyday language use with self-reported measures of behavior (Niu and Fan, 2018).

Computerized text analysis tools such as LIWC, IBM Watson, DICTION and T-lab, have gained popularity as powerful research instruments with which to conduct customer research (Humphreys and Wang, 2018). Some of these tools are AI-based, which means it requires a large volume of text in order to complete an effective analysis (Balducci and Marinova, 2018). A further type of computerized text analysis relies on the use of built-in dictionaries, where a piece of text is analyzed by comparing words contained in the text with those in a particular predefined dictionary (e.g. Pitt *et al.*, 2018). Dictionary-based text analysis tools are effective with smaller datasets, such as online customer reviews, where each review typically comprises only a small number of words (Robertson *et al.*, 2021). Making use of a computerized text analysis tool with predefined dictionaries, allows for the creation of specific summary variables designed to identify psychological states, thinking styles, emotional tone and social concerns (Pennebaker *et al.*, 2015). Additionally, it is suitable for the analysis of multilevel language use, as per SAT, offering service firms an alternative tool with which to measure and predict customers' evaluation of their interactions with a firm, using verbatim online customer reviews.

Using computerized text analyses to predict customers' evaluation of customer-firm interactions

Extant literature offers a multitude of ways in which firms can gain a better understanding of their customers' evaluations of their service experiences with a firm through online reviews – this phenomenon in itself is not novel. Computerized text analysis has been used to predict customer behavior and facilitate more effective interaction with customer target groups (Ludwig *et al.*, 2013; Palese and Piccoli, 2016), as it enables marketers to systematically analyze specific aspects of customer-firm interactions in real time (Robertson *et al.*, 2021). However, research using computerized text analysis tools to measure and predict customers' evaluation of their customer-firm service interactions is limited. Table 1 provides an overview of key articles, including the respective computerized text analysis tools used.

Table 1.
An overview of articles that have used online customer reviews for computerized text analysis

Author/s	Data type used in analysis	Computerized text analysis tool used	Research purpose
Boon <i>et al.</i> (2013)	Online customer reviews (<i>unstructured data</i>)	RapidMiner	The purpose of this research was to perform word frequency analysis on qualitative online comments and customer reviews to measure and quantify hotel service quality
Büschken and Allenby (2016)	Online customer reviews (<i>unstructured data</i>)	Sentence-Constrained Latent Dirichlet allocation (SC-LDA) model	The research developed a model for text analysis, making use of the sentence structure contained in customer reviews, showing that it leads to improved inference and prediction of consumer ratings of hotels and restaurants
Chatterjee <i>et al.</i> (2021)	Online customer reviews (<i>unstructured data</i>)	NRC Word-Emotion Association Lexicon (EmoLex)	Using text-mining, machine-learning and econometric techniques, the purpose of this research was to find which core and augmented service aspects and which emotions are more important in which service contexts in terms of reflecting and predicting customer satisfaction
de Haan and Michelli (2020)	Three different data sources: customer database data (<i>structured data</i>), survey data (<i>structured data</i>), and textual customer data, i.e. reviews, comments on social media, verbal or written customer-firm interactions (<i>unstructured data</i>)	Latent Dirichlet allocation (LDA)	The research investigated the extent to which three structured and unstructured data, can predict customer churn, showing that the inclusion of unstructured data significantly improves the estimation of customer retention and churn
Oh <i>et al.</i> (2022)	Written customer reviews, hotel information and images (<i>unstructured data</i>)	Valence Aware Dictionary and Sentiment Reasoner (VADER)	Using deep learning techniques and approaches, coupled with the theoretical principles of expectation-confirmation theory, the research used unstructured data to predict customer satisfaction in hospitality services
Song <i>et al.</i> (2016)	Online customer reviews (<i>unstructured data</i>)	Conducted part-of-speech tagging on customer reviews, whereafter a service-feature word dictionary and sentiment word dictionary was developed with which to measure all the constructs and predict service quality	This study developed an analytic framework and procedures ("customer review-based gap analysis") to diagnose service quality from online customer reviews
Zhu <i>et al.</i> (2021)	Online customer reviews (<i>unstructured data</i>)	Heuristic Processing, Linguistic Feature Analysis, and Deep Learning-based Natural Language Processing (NLP)	This research proposed and tested mechanisms and AI-based technology for defining and identifying the critical online consumer reviews that firms could prioritize to optimize their online response strategies

Based on the studies represented (Table 1), extant research suggests that analyzing the lexical and semantic content and language style properties of online customer reviews, can assist firms to quantifiably measure service quality (Boon *et al.*, 2013; Song *et al.*, 2016), predict customer satisfaction and ratings (Büschken and Allenby, 2016; Chatterjee *et al.*, 2021; Oh *et al.*, 2022), assist in prioritizing customer response strategies (Zhu *et al.*, 2021) and improve the predictive accuracy of customer retention and churn (de Haan and Menichelli, 2020). With the exception of three studies (de Haan and Menichelli, 2020; Chatterjee *et al.*, 2021; Song *et al.*, 2016), all studies have focused on tourism and hospitality as the research context (Boon *et al.*, 2013; Büschken and Allenby, 2016; Oh *et al.*, 2022; Zhu *et al.*, 2021). From a customer-firm interaction perspective, with the exception of Zhu *et al.* (2021), none of the studies approached their analysis from a multilevel language use perspective. As such, this research sought to further enhance this body of knowledge by harnessing the power of unstructured data using computerized text analysis to enhance the customer-firm relationship.

Methodology

The purpose of this study is to examine how service firms can use computerized text analysis tools to analyze unstructured online data. To do this, we analyze the linguistic components of qualitative online customer reviews using LIWC, to quantify these consumers' reviews and predict their evaluations of their customer-firm interactions. LIWC was chosen on the basis that it offers its users a user-friendly and inexpensive means through which to analyze unstructured data, thereby aligning with the research purpose. In light of the fact that this methodological paper seeks to provide a mechanism through which service firms can analyze unstructured data by using available digital technology, it was imperative that the tool had broad market appeal. In addition, there is strong empirical evidence to support the use of LIWC to discover meaningful linguistic insights relating to social relationships, emotionality, thinking styles and individual-level differences, based on natural language use (Pennebaker *et al.*, 2015; Manchaiah *et al.*, 2021).

Computerized text analysis tool – LIWC

LIWC was developed on the premise that the words that individuals use are able to offer insight into their motives, social status and demographics beyond the underlying emotion present in the text (Pennebaker *et al.*, 2003). LIWC is a text analysis program with two central components used as part of the text analysis, namely a language processing component and built-in dictionaries (Tausczik and Pennebaker, 2010). These two components work together to open each individual text file, in this case each individual review, to compare each word in the file to a predetermined list of words in the built-in dictionary (Tausczik and Pennebaker, 2010). The dictionaries comprise a collection of words that define a particular category. In doing this, LIWC is able to categorize and count different words used in the text, such as the number of articles, adjectives and personal pronouns used. This information is then presented as the percentage of words within a given piece of text that can clearly be linked to each category (i.e. 17% of words in a particular text may be identified as personal pronouns). The software counts the number of words when it reads a passage of text, which then reflects different psychological states, including thinking styles, emotional tone and social concerns (Pennebaker *et al.*, 2015). The software then uses this categorization to create four summary language variables namely, *analytical thinking* (Pennebaker *et al.*, 2014), *clout* (Kacewicz *et al.*, 2013), *authenticity* (Newman *et al.*, 2003) and *emotional tone* (Cohn *et al.*, 2004). Proprietary algorithms built into the LIWC software, developed on the basis of empirical research, are able to compute the summary language variables. While the exact computation behind the

algorithm is proprietary information, users are able to compute the summary language variables using the research articles upon which they are based (LIWC, 2023). The specific research articles, as they pertain to the development of each summary language variable, are identified in the sections below:

- (1) *Analytical thinking*: Analytical thinking is representative of the extent to which individuals use words that suggest the use of formal and hierarchical thinking patterns (Pennebaker *et al.*, 2014). According to Jordan *et al.* (2010) analytical thinking is most apparent in an individual's use of articles which typically signals concepts and prepositions which conveys the relationships between the concepts. Individuals naturally differ in the extent to which they engage in analytical thinking, which is typically formalized and associated with more formal settings, as opposed to intuitive thinking. Interestingly, researchers have examined the presence of the four summary language variables amongst both high-rated (4- or 5-star ratings out of a possible 5 stars) and low-rated reviews (1- or 2-star ratings out of a possible 5 stars) (Robertson *et al.*, 2021). When evaluating the analytical thinking presented in online reviews, they identified that high-rated reviews indicated greater levels of analytical thinking – suggesting that high-rated reviews were more formal, logical and present hierarchical thinking. Customer reviews that exhibit a higher level of analytical thinking are assumed to be more helpful (Park, 2018). Examples of the critical empirical research that underpins the algorithmic computation of the analytical thinking summary language variable is Pennebaker *et al.* (2014) and Jordan *et al.* (2010).
- (2) *Clout*: The value assigned to clout indicates the social status or confidence that is apparent in an individual's written text (Pennebaker *et al.*, 2015). The clout variable is developed on the basis of research suggesting that those with high social power tend to use more first-person plural pronouns and social words (Jordan *et al.*, 2010; Kacewicz *et al.*, 2013). Use of social words (i.e. we) indicates that high-status individuals tend to be more collectively focused, thereby focusing their attention outward, toward others (Kacewicz *et al.*, 2013). Research suggests that high-rated reviews tended to present a higher level of confidence and expertise (i.e. high ratings for clout), whereas low-rated reviews typically presented a tentative, less confident style (Robertson *et al.*, 2021). Examples of the critical empirical research that underpins the algorithmic computation of the clout summary language variable include Kacewicz *et al.* (2013) and Fox and Royne Stafford (2021).
- (3) *Authenticity*: Pennebaker *et al.* (2015) suggest that higher values for authenticity are associated with honest, personal and disclosing text, whereas lower values seemingly present a guarded and distanced narrative. The summary variable for authenticity was developed following the research of Newman *et al.* (2003) that investigated the linguistic features that differentiate true and false stories. Their research identified that those misrepresenting the truth (i.e. being inauthentic) typically showed a lower cognitive complexity, used less self-references and made use of more words presenting negative emotions. Robertson *et al.* (2021) further identified that low-rated customer reviews tended to present higher levels of authenticity suggesting that authors presented a more open, honest and personal style in their review. Examples of the critical empirical research that underpins the algorithmic computation of the authenticity summary language variable is Newman *et al.* (2003) and Kalichman and Smyth (2021).
- (4) *Emotional tone*: The emotional tone summary language variable combines both positive and negative emotions into a single variable (Cohn *et al.*, 2004). Values above

50 indicate a positive tone whereas values below 50 are associated with a negative tone, representing feelings of anxiety, sadness and hostility (Pennebaker *et al.*, 2015; Robertson *et al.*, 2021). Values close to 50 indicate a lack of emotionality. As expected, Robertson *et al.* (2021) established that low-rated reviews tended to present negative emotions such as anxiety, sadness and hostility whereas high-rated reviews typically presented a more positive and upbeat style. An example of the critical empirical research that underpins the algorithmic computation of the authenticity summary language variable is Cohn *et al.* (2004).

Through the use of algorithms, values for the above-mentioned four summary language variables are created for each review in the dataset. Thereafter, a linear regression that seeks to regress the four summary language variables on the overall rating of the service experience addresses the first objective of the research by determining whether the text analysis allows for a quantifiable evaluation of customers' service experiences. The second objective is then addressed by examining which components of the computerized text analysis are the best predictors of a customer's service experience.

In order to exemplify the methodology in practice, three empirical studies are presented. Three distinct datasets comprising online customer reviews for three companies were used to provide a contextualized application of the method. In particular, the healthcare insurance industry was used as the service context, given that the analysis of unstructured data has already served as an important source of information within this industry in recent years (Appold, 2017; Balducci and Marinova, 2018). Traditionally, marketing has not focused on the healthcare industry, yet, as a professional services industry, the healthcare sector has a long tradition of leveraging and integrating data from multiple sources, as well as using analytics to inform product offerings, recommendations and the personalization of patient experiences to further drive patient value (Grewal *et al.*, 2020). As the healthcare landscape is increasingly fueled by technological advances and regulatory shifts, a more marketing-focused approach is needed to transform as an industry – from one that has provided patient care using episodic and reactive measures to one that relies on continuous and proactive assessments. This means that as an industry, the healthcare sector moves from rewarding volume to rewarding value, heralding a shift from being provider-centered to being more patient (customer)-centered (Burwell, 2015; Grewal *et al.*, 2020). As an industry which has exhibited intense competition (Goddard, 2015) and varied levels of customer satisfaction regarding service quality among different providers (Allahham, 2013), the healthcare insurance industry in South Africa provided a fitting context for analysis.

In order to address the research purpose through a descriptive research design, secondary data, in the form of online customer reviews freely available on an online review platform, was used as the data source. Using the Python program, data was scraped from an online review platform, namely Hellopeter, in February 2021. Hellopeter, a platform that seeks to connect South African customers and businesses (Hellopeter, 2018), was originally launched in 2000 in a pre-social media landscape and quickly earned a reputation as a complaint's platform, with users commonly airing their grievances with brands (Forbes, 2022). Today, the platform is a hub of review activity offering a host of both positive and negative reviews for a wide-ranging number of companies spanning multiple industries (Forbes, 2022). Hellopeter offers the platform for users to express their opinions without moderation, playing no further role in mediating communication or action between the customer and the business (Hellopeter, 2018). Reviews on the platform are publicly available without requiring registration, thus allowing the reviews to be readily scraped for analysis purposes. Given that all data was available as public data online, no individuals were able to be identified. In light of the fact that all data used in the study is available in the public domain, having been voluntarily offered by individuals, there were no potential ethical concerns noted.

The scraped data consisted of a single name or pseudonym attached to each review, a title for the review, the full, written review in English and an accompanying star rating that ranges from one to five stars. A total of 2,867 complete reviews from three different healthcare insurance providers in South Africa were scraped from the online review platform. The three healthcare insurance providers were Discovery ($n = 1,067$), Bonitas ($n = 857$) and Affinity ($n = 954$). Discovery and Bonitas are two of South Africa’s largest open healthcare insurance providers and had the highest number of reviews on Hellopeter.com. A smaller healthcare insurance provider, Affinity, had the third highest number of reviews. South Africa’s Customer Satisfaction Index found that Discovery and Bonitas experience similar customer satisfaction and customer loyalty levels (BusinessTech, 2020). Affinity was selected as the third medical insurance provider to offer a better market representation of healthcare insurance provider customer experience ratings.

Following the data scraping processes, the data was analyzed using the LIWC software in order to complete the computerized text analysis (Pennebaker and King, 1999; Pennebaker et al., 2003, 2011, 2015). LIWC is already widely deployed and simple to use, having already begun to gain traction in the medical field (Gong et al., 2018; Swol et al., 2020; Creten et al., 2022). The LIWC analysis was used to create descriptive statistics and the four summary language variables previously discussed. Following the completion of the LIWC analysis, the results were transferred to the SPSS software (version 27) to conduct a linear regression. The results obtained using these data analysis procedures are outlined below.

Results

A review of the descriptive statistics associated with the dataset indicates that Affinity holds the highest average star rating ($\mu = 3.07$; s.d. 1.69), followed by Discovery ($\mu = 2.78$; s.d. 1.60) and then Bonitas ($\mu = 2.45$; s.d. 1.55). As per Table 2, reviews for Affinity tended to present a much lower word count than the reviews for both Discovery and Bonitas. Beyond the word count, all remaining metrics are presented as a percentage of the total words used (Pennebaker et al., 2015).

The simplicity of the language is evaluated using a proxy measure that counts the number of words with six or more letters used throughout the reviews (Ferreira et al., 2022). As a percentage of total words, reviews for Affinity tended to have the highest level of language complexity (i.e. 23.35% of all words used were at least six letters in length). The reviews of both Discovery and Bonitas tended to present simpler language. LIWC assesses the presence of both positive and negative emotion by identifying linked words that exemplify the differing emotional states (Duncan et al., 2019). Linked words for positive emotions include words such as “love”, “nice” and “sweet”, whereas linked words for negative emotions include words such as “hurt”, “ugly” and “nasty” (Robertson et al., 2021). Given that Affinity held the highest average star rating, it is understandable that the reviews on average exhibited the highest level of positive emotions and the lowest level of negative emotions (Table 2), while reviews for Bonitas exhibited the lowest level of positive emotion and the highest level of negative emotions aligned with their relatively low average star rating.

Table 2. Basic sentence structure variables and associated inferential statistics

Linguistic variables	Affinity ($n = 954$)	Discovery ($n = 1,056$)	Bonitas ($n = 857$)
Star rating	3.07	2.78	2.45
Average word count	82.46	124.69	125.35
Language complexity	23.35	20.99	20.24
Positive emotions	7.19	3.98	2.84
Negative emotions	1.09	1.51	1.68

Table 3 presents the results of the four summary language variables created by LIWC, while Table 4 provides further context to these values by including examples of online reviews that received either a high or low rating for each of the summary language variables. The four summary variables have been rescaled to reflect a 100-point scale ranging from 0 to 100.

Summary language variable	Affinity (<i>n</i> = 954)	Discovery (<i>n</i> = 1,056)	Bonitas (<i>n</i> = 857)
Analytical thinking	56.82	59.91	58.41
Clout	62.73	51.78	54.01
Authenticity	35.05	60.40	57.15
Emotional tone	63.96	51.94	41.32

Table 3.
Comparisons of language summary variables

Summary language variable	High	Low
Analytical thinking	<p>Rating: 99</p> <p>"I had a pleasant call with *** (a consultant from Affinity Health this morning). I had made up my mind that I was canceling my health insurance because I was struggling with the payments. She was very empathetic and she made me realize I was making a rash decision – she outlined the advantages of downgrading rather than cancelling my insurance. It is because of [redacted] excellent customer care skills and knowledge of her company’s policies, that I have referred my sister to join Affinity Health. Thanks a lot [redacted] -you have represented your company very well! God bless you!"</p>	<p>Rating: 1.92</p> <p>"*** was very helpful..she even called me back . . . you are a star..thank you"</p>
Clout	<p>Rating: 91.35</p> <p>"A big thank you to *** at your Customer Care Department. Thank you for resolving my query and going the extra mile for me. It’s hard to find good customer service but your service is top-notch. Keep it up and I salute you."</p>	<p>Rating: 1.00</p> <p>"Affordable but too basic"</p>
Authenticity	<p>Rating: 99</p> <p>"I am very happy with Affinity Health. I had excellent service from *** earlier today. She answered all my questions i had on my policy. I understand everything clearly on my day to day cover now. And i will definitely recommend Affinity Health to my family and friends."</p>	<p>Rating: 4.97</p> <p>"Thank you *** for your help today, it is much appreciated. Very friendly and very helpful."</p>
Emotional tone	<p>Rating: 99</p> <p>"Hello! *** was absolutely amazing. You are calm, the information was clear enough for me to understand and I could hear you properly. You [are] definitely a great service provider! Thank you so much for helping me. You are highly recommended!!"</p>	<p>Rating: 25.77</p> <p>"These THIEVES ran an illegal debit order on my bank account out of the blue long after I cancelled the policy in writing well within the cooling off period! Do not trust them and do not do business with them! They are just as I said . . . a bunch of thieves!!"</p>

Table 4.
Contrasting high and low summary language variables

Reviews for Discovery present the highest levels of analytical thinking and authenticity, while reviews for Affinity present the lowest level of analytical thinking and authenticity together with the highest level of clout and the most positive emotional tone. Again, the positive tone is readily understood given that the highest star rating was awarded to Affinity. On the opposing end of the spectrum, we find that the reviews for Bonitas, who received the lowest star rating, present the most negative emotional tone.

In order to exemplify the use of a computerized text analysis to predict the evaluations of customers' service experiences, linear regressions were used on the LIWC output variables for each of the three healthcare insurance providers. The four summary language variables described above were used as predictors in the model, with the star rating being a proxy measure for an evaluation of a customer's service experience. The results of the regression analyses are presented in Table 5.

The results indicate that the overall models for each of the three healthcare insurance providers are significant. The R^2 values suggest that 36%, 40% and 60% of the variation in the star ratings for Discovery, Bonitas and Affinity respectively, can be explained through the four summary language variables. In the dataset for Discovery: clout, authenticity and emotional tone were identified as significant predictors of the overall rating for customers' service experience. In contrast with Bonitas, we found that analytical thinking, authenticity and emotional tone were significant predictors. Finally, only authenticity and emotional tone were significant predictors of the overall rating for customers' service experience for Affinity. It should be noted that authenticity and emotional tone were significant predictors across the three datasets, with emotional tone consistently being positively related to ratings of customers' service experience, whereas authenticity was consistently negatively related to ratings of customers' service experience. The negative standardized beta coefficients for authenticity suggest that as the level of authenticity presented in an online review increases, the rating for a customer's service experience tends to decrease. Emotional tone was consistently the strongest predictor of a customer's service experience, suggesting that as the level of emotional tone in an online review increases (i.e. becomes positive), the customer's service experience rating increases.

Discussion and implications

This methodological paper sought to examine how service firms can analyze unstructured data by using a digital technology in order to predict customers' service experience evaluations of a service firm. In particular, using a computerized text analysis tool to analyze online customer reviews to predict the evaluations of their customer-firm interactions. To exemplify the methodology and identify which linguistic components are the best predictors of a customer's service interaction evaluation, we presented three empirical studies within the healthcare insurance industry. The results suggest that a computerized text analysis of

Model summary Variable	Affinity $F_{(4, 949)} = 357.06$ ($p < 0.00$)			Discovery $F_{(4, 1051)} = 146.70$ ($p < 0.00$)			Bonitas $F_{(4, 852)} = 141.59$ ($p < 0.00$)		
	β	t	$R^2 = 0.60$ p	β	t	$R^2 = 0.36$ p	β	t	$R^2 = 0.40$ p
Analytical thinking	0.03	1.44	0.15	0.16	0.63	0.53	0.08	2.89	0.00*
Clout	0.04	1.64	0.10	0.13	4.23	<0.00**	0.02	0.74	0.46
Authenticity	-0.14	-5.52	<0.00**	-0.06	-1.97	0.04*	-0.06	-2.04	0.04*
Emotional tone	0.68	29.35	<0.00**	0.52	20.07	<0.00**	0.59	21.29	<0.00**

Note(s): * significant at a 5% level
** significant at a 1% level

Table 5.
Model summary and coefficients for all language summary variables

online customer reviews can predict a customer's service interaction rating of healthcare insurance providers, across all three empirical studies.

The implication thereof is that even in the absence of a star rating, content provided on social media platforms, written interactions with a company, or transcripts of spoken text from call centers can be used to provide a quantitative measure of the customer's evaluation of the customer-firm interaction, using LIWC. Using the theoretical lens of SAT and based on customers' qualitative online reviews (as a form of a speech act, e.g. praise, criticize, complain, commend, etc.), our results suggest that LIWC is able to predict the linguistic intent of these online customer reviews as a quantified measure. This aligns with previous research which have used different text analysis tools to analyze lexical and semantic content to predict customer satisfaction ratings specifically (Büschken and Allenby, 2016; Chatterjee *et al.*, 2021; Oh *et al.*, 2022).

The results further suggest that the four summary language variables available through LIWC, can explain between 36% and 60% of variation present in a customer's evaluation of their customer-firm interactions with various healthcare insurance providers. The implication thereof is that the variation within these evaluations that can be explained by the manner in which consumers talk is non-trivial. The ability to use language to, in some instances, explain more than half of the variation in customer evaluations of their customer-firm interactions is a powerful tool to be leveraged within the customer service industry. As customers heavily rely on the use of language to perform specific speech acts in online reviews, they do so within a network of interpersonal structures, motives and feelings (Holtgraves, 2021). Linguistically, these elements most likely play a critical role in the communication process as well as how firms would be able to interpret and process their evaluations of the customer-firm interaction. As such, the results provide evidence that the analysis of the four summary variables of LIWC, as micro-level speech act categories, are suitable to theoretically and meaningfully capture the intention of the speech act performance in an easy to understand quantifiable measure.

This result bears great significance for the market research industry, an industry with an estimated global revenue exceeding 73.4 billion U.S Dollars in 2019 alone (Statista, 2021). An examination of readily available online customer reviews, by means of inexpensive and easy to use computerized text analysis software, could provide a valid alternative, or an effective complementor, to traditional market research efforts when trying to evaluate customer-firm service interactions. As identified, traditional survey methods of data collection are commonly plagued by respondent fatigue (O'Reilly-Shah, 2017; Viswanathan and Kayande, 2012) and low survey response rates (Luo, 2009). An overall lack of motivation for survey participation, whether face-to-face or online (Ljepava, 2017), is able to dramatically impact the reliability of traditional survey data together with implications for target population representation. Reduced motivation to engage in survey data collection has resulted in a proliferation of paid panels of research participants that may also impact the representation of the sample and reliability of data. In addition, this further increases the costs of conducting traditional survey research (Ljepava, 2017).

This research therefore presents a different means of assessing customers' service experiences, one that would not face similar constraints, given the wealth of readily available unstructured data. As such, this research supports and builds on the notion that the rise and prevalence of unstructured data, such as online reviews, are a valuable source of insights for marketers (Robertson *et al.*, 2021). The implication thereof, is that the method can be expanded for use in the absence of a numerical star rating, expanding the availability of online content for text analysis from platforms that do not explicitly incorporate quantitative ratings.

Customer feedback and other sources of unstructured data are vitally important to firms (de Haan and Menichelli, 2020) – customers willingly provide the text, and in the case of many

service firms they do so in great volume and detail (e.g. TripAdvisor). For marketers and service providers, the focus could come from partnering with customers who could be viewed as co-creators of value for the firm, pushing service providers to use such collaboration for market research purposes (Singh *et al.*, 2022). Service firms should therefore seek to encourage customers to provide online reviews as this serves a dual purpose. First, as suggested by the results, as a form of unstructured data that is readily available, online customer reviews present real-time insight into service evaluations, particularly in the absence of numerical service quality ratings. Second, online reviews have become an increasingly important source of information, with many customers using online reviews to express their authentic and personal experiences with a service firm. This is of particular relevance in the healthcare insurance sector that is considered to be a credence good (Huck *et al.*, 2016), where social bonds and personal ties pertaining to service dimensions that offer interpersonal interactions and friendships, have been shown to lead to higher levels of customer commitment and loyalty (Hsieh *et al.*, 2005).

In aligning with the importance of how unstructured data can be used to aid decision-making at a firm level, the results provide further implications for customer service agents and those often tasked with managing and responding to online customer reviews. While this research made use of the LIWC software, suitable for use in the service marketing industry due to its price accessibility and ease of use requiring no coding or development skills, there are a multitude of other tools readily available for computerized text analysis (e.g. IBM Watson and DICTION). Establishing an automated process that could conduct the content analysis of online reviews and prioritize those with a low predicted service evaluation rating, could allow customer service teams to more effectively prioritize reviews that need a swift response. The importance of a timely response to negative reviews cannot be understated given that these responses have emerged as a vital component in managing firm reputation (Sparks and Bradley, 2017). In addition, positive feedback mechanisms could also provide helpful guidance to further enhance the customer experience. This triage process could reduce the amount of time that customer service agents expend on working through online reviews and create a more streamlined and accurate process of response coordination. In the event that a service firm has an already established partnership with their preferred research agency, they could incorporate continual reporting of service evaluations based on the computerized text analysis of online customer reviews as a key component of online social listening reporting requirements.

Limitations and recommendations for future research

The limitations of this research are threefold. First, the research is limited in scope as a result of specific methodological decisions. The research focused on healthcare insurance providers within a particular context, making use of three competitors within the industry, therefore limiting the ability for the findings to be generalized. However, the chosen industry and competitors provided a fitting service-oriented frame of reference to illustrate the potential of the methodology. Second, the data source consisted of reviews from a single online review platform. While this platform provided the opportunity to assess a relatively large dataset and examine the predictive ability of the summary variables, it is possible that other datasets sourced from other online review platforms could have resulted in potentially different insights. Third, as the online review platform used relies on voluntary contributions from customers, the reviews analyzed may be skewed towards an extreme view.

The research offers several avenues for further research. First, future researchers are encouraged to broaden the scope of the research by gathering data from additional sources and contrasting the ability to predict customers' service experiences from different data sources. Second, given that this research offers a potential alternative to traditional market

research surveys, future research should consider an amalgamation of additional sources of readily available information that could be used to offer a more nuanced, holistic understanding of service experience evaluations from a customer's perspective. Third, when assessing specific dimensions of customers' service experience, existing customer evaluation models can be compared to assess which is most suited to accurately capture specific topics as evidenced in the online reviews that customers post. Fourth, while LIWC provides a mechanical analysis of the four summated language variables, the richness of the data provided by online reviews could be complemented by a comparison of the results with a traditional content analysis.

References

- Al Awadhi, J., Obeidat, B. and Alshurideh, M. (2021), "The impact of customer service digitalization on customer satisfaction: evidence from telecommunication industry", *International Journal of Data and Network Science*, Vol. 5 No. 4, pp. 815-830.
- Allahham, A. (2013), "Determinants of customer satisfaction in healthcare services", *International Journal of Business and Management Invention*, Vol. 2 No. 12, pp. 59-64.
- Anderson, E. and Mittal, V. (2000), "Strengthening the satisfaction-profit chain", *Journal of Service Research*, Vol. 3 No. 2, pp. 107-120.
- Appold, K. (2017), "Turn data into insight – how predictive analytics can capture revenue", *Managed Healthcare Executive*, Vol. 27 No. 7, pp. 16-21.
- Argyris, Y., Monu, K., Kim, Y., Zhou, Y., Wang, Z. and Yin, Z. (2021), "Using speech acts to elicit positive emotions for complainants on social media", *Journal of Interactive Marketing*, Vol. 55 No. 1, pp. 67-80.
- Austin, J. (1975), *How to Do Things with Words*, Oxford University Press, London.
- Bagozzi, R., Dholakia, U. and Pearo, L. (2007), "Antecedents and consequences of online social interactions", *Media Psychology*, Vol. 9 No. 1, pp. 77-114.
- Balducci, B. and Marinova, D. (2018), "Unstructured data in marketing", *Journal of the Academy of Marketing Science*, Vol. 46, pp. 557-590.
- Berger, J., Humphreys, A., Ludwig, S., Moe, W., Netzer, O. and Schweidel, D. (2020), "Uniting the tribes: using text for marketing insight", *Journal of Marketing*, Vol. 84 No. 1, pp. 1-25.
- Boon, E., Bonera, M. and Bigi, A. (2013), "Measuring hotel service quality from online consumer reviews- A proposed method", *Information and Communication Technologies in Tourism 2014*, Springer, Cham, pp. 367-379.
- Burwell, S. (2015), "Setting value-based payment goals — HHS efforts to improve U.S. health care", *The New England Journal of Medicine*, Vol. 372 No. 10, pp. 897-899.
- Büschken, J. and Allenby, G. (2016), "Sentence-based text analysis for customer reviews", *Marketing Science*, Vol. 35 No. 6, pp. 953-975.
- BusinessTech (2020), "Best and worst major medical aids in South Africa", available at: <https://businesstech.co.za/news/lifestyle/444098/best-and-worst-major-medical-aids-in-south-africa/> (accessed 7 June 2021).
- Carr, C., Schrock, D. and Dauterman, P. (2012), "Speech acts within Facebook status messages", *Journal of Language and Social Psychology*, Vol. 31 No. 2, pp. 176-196.
- Charan, K. (2019), "The top 12 customer experience experts you need to follow in 2021", available at: <https://freshdesk.com/general/top-12-customer-experience-experts-you-need-to-follow-blog/> (accessed 8 June 2021).
- Chatterjee, S., Goyal, D., Prakash, A. and Sharma, J. (2021), "Exploring healthcare/health-product ecommerce satisfaction- A text mining and machine learning application", *Journal of Business Research*, Vol. 131, pp. 815-825.

- Cohn, M., Mehl, M. and Pennebaker, J. (2004), "Linguistic indicators of psychological change after September 11, 2001", *Psychological Science*, Vol. 15 No. 10, pp. 687-693.
- Creten, S., Heynderickx, P. and Dieltjens, S. (2022), "The stigma toward dementia on twitter: a sentiment analysis of Dutch language tweets", *Journal of Health Communication*, Vol. 27 No. 10, pp. 697-705.
- de Haan, E. and Menichelli, E. (2020), "The incremental value of unstructured data in predicting customer churn", Working Paper Series 2020, Report No. 20-105, Marketing Science Institute, pp. 1-49.
- Duncan, S., Chohan, R. and Ferreira, J. (2019), "What makes the difference? Employee social media brand engagement", *Journal of Business and Industrial Marketing*, Vol. 34 No. 7, pp. 1459-1467.
- Felbermayr, A. and Nanopoulos, A. (2016), "The role of emotions for the perceived usefulness in online customer reviews", *Journal of Interactive Marketing*, Vol. 36, pp. 60-76.
- Ferreira, C., Lam, J., Pitt, L., Caruana, A. and Brown, T. (2022), "Contrasting compulsive behavior: computerized text analysis of compulsion narratives", *Journal of Health Psychology*, Vol. 27 No. 8, pp. 1942-1958.
- Forbes (2022), "How South Africa's 'complaints platform' became its #1 online review platform for customers and businesses alike", available at: <https://www.forbesafrica.com/technology/2022/05/10/how-south-africas-complaints-platform-became-its-1-online-review-platform-for-customers-and-businesses-alike/> (accessed 26 September 2022).
- Fox, A. and Roynce Stafford, M. (2021), "Olympians on Twitter: a linguistic perspective of the role of authenticity, clout, and expertise in social media advertising", *Journal of Current Issues and Research in Advertising*, Vol. 42 No. 3, pp. 294-309.
- Goddard, M. (2015), "Competition in healthcare: good, bad or ugly?", *International Journal of Health Policy and Management*, Vol. 4 No. 9, pp. 567-569.
- Goncalves, H., Silva, G. and Martings, T. (2018), "Motivations for posting online reviews in the hotel industry", *Psychology and Marketing*, Vol. 35, pp. 807-817.
- Gong, Y., Shin, K. and Poellabauer, C. (2018), "Improving LIWC using soft word matching", in Shehu, A. and Wi, C. (Eds), *The 9th ACM Conference on Bioinformatics, Computational Biology and Health Informatics*, Association for Computing Machinery, New York.
- Grewal, D., Hulland, J., Kopalle, P. and Karahanna, E. (2020), "The future of technology and marketing: a multidisciplinary perspective", *Journal of the Academy of Marketing Science*, Vol. 48 No. 1, pp. 1-8.
- Grewal, D., Herhausen, D., Ludwig, S. and Ordenes, F. (2022), "The future of digital communication research: considering dynamics and multimodality", *Journal of Retailing*, Vol. 98 No. 2, pp. 224-240.
- He, W., Tian, X., Tao, R., Zhang, W., Yan, G. and Akula, V. (2017), "Application of social media analytics: a case of analyzing online hotel reviews", *Online Information Review*, Vol. 41 No. 7, pp. 921-935.
- Hellopeter (2018), "How does Hellopeter work?", available at: <https://blog.hellopeter.com/faqs/how-does-hellopeter-work> (accessed 3 June 2021).
- Holtgraves, T. (2021), "Understanding miscommunication: speech act recognition in digital contexts", *Cognitive Science*, Vol. 45 No. 10, e13023.
- Hsieh, Y., Chiu, H. and Chiang, M. (2005), "Maintaining a committed online customer: a study across search-experience-credence products", *Journal of Retailing*, Vol. 81 No. 1, pp. 75-82.
- Huck, S., Lünser, G., Spitzer, F. and Tyran, J. (2016), "Medical insurance and free choice of physician shape patient overtreatment: a laboratory experiment", *Journal of Economic Behavior and Organization*, Vol. 131, pp. 78-105.
- Humphreys, A. and Wang, R. (2018), "Automated text analysis for consumer research", *Journal of Consumer Research*, Vol. 44 No. 6, pp. 1274-1306.

- Jain, R., Aagja, J. and Bagdare, S. (2017), "Customer experience—a review and research agenda", *Journal of Service Theory and Practice*, Vol. 27 No. 3, pp. 642-662.
- Jordan, K., Sterling, J., Pennebaker, J. and Boyd, R. (2010), "Examining long-term trends in politics and culture through language of political leaders and cultural institutions", *Psychological and Cognitive Sciences*, Vol. 116 No. 9, pp. 3476-3481.
- Kacewicz, E., Pennebaker, J., Davis, M., Jeon, M. and Graesser, A. (2013), "Pronoun use reflects standings in social hierarchies", *Journal of Language and Social Psychology*, Vol. 33 No. 2, pp. 125-143.
- Kalichman, S. and Smyth, J. (2021), "'And you don't like, don't like the way I talk': authenticity in the language of Bruce Springsteen", *Psychology of Aesthetics, Creativity, and the Arts*, published online ahead of print.
- Keiningham, T., Ball, J., Benoit, S., Bruce, H., Buoye, A., Dzenkovska, J., . . . and Zaki, M. (2017), "The interplay of customer experience and commitment", *Journal of Services Marketing*, Vol. 31 No. 2, pp. 148-160.
- Key, T. and Keel, A. (2020), "How executives talk Exploring marketing executive value articulation with computerized text analysis", *European Journal of Marketing*, Vol. 54 No. 3, pp. 546-569.
- Korfiatis, N., Stamolampros, P., Kourouthanassis, P. and Sagiadinos, V. (2019), "Measuring service quality from unstructured data: a topic modeling application on airline passengers' online reviews", *Expert Systems with Applications*, Vol. 116, pp. 472-486.
- Krippendorff, K. (2018), *Content Analysis: An Introduction to its Methodology*, SAGE Publications.
- Labrecque, L., Swani, K. and Stephen, A. (2020), "The impact of pronoun choices on consumer engagement actions: exploring top global brands' social media communications", *Psychology and Marketing*, Vol. 37 No. 6, pp. 796-814.
- Li, M., Yin, D., Qiu, H. and Bai, B. (2021), "A systematic review of AI technology-based service encounters: implications for hospitality and tourism operations", *International Journal of Hospitality Management*, Vol. 95, pp. 1-10.
- LIWC (2023), "LIWC analysis", available at: <https://www.liwc.app/help/liwc#Summary-Measures> (accessed 24 January 2023).
- Ljepava, N. (2017), "Challenges of contemporary survey research: a review", *The 5th Global Conference on Business and Social Sciences 2017 Proceedings*, Kuala Lumpur, Malaysia.
- Lord Ferguson, S., Pitt, C. and Pitt, L. (2021), "Using artificial intelligence to examine online patient reviews", *Journal of Health Psychology*, Vol. 26 No. 13, pp. 2424-2434.
- Ludwig, S. and De Ruyter, K. (2016), "Decoding social media speak: developing a speech act theory research agenda", *Journal of Consumer Marketing*, Vol. 33 No. 2, pp. 124-134.
- Ludwig, S., De Ruyter, K., Friedman, M., Brüggem, E., Wetzels, M. and Pfann, G. (2013), "More than words: the influence of affective content and linguistic style matches in online reviews on conversion rates", *Journal of Marketing*, Vol. 77 No. 1, pp. 87-103.
- Luo, Y. (2009), "Using internet data collection in marketing research", *International Business Research*, Vol. 2 No. 1, pp. 196-202.
- Manchaiah, V., Swanepoel, D. and Bennett, R. (2021), "Online consumer reviews on hearing health care services: a textual analysis approach to examine psychologically meaningful language dimensions", *American Journal of Audiology*, Vol. 30 No. 3, pp. 669-675.
- Marketing Science Institute (2022), "MSI announces 2022-24 research priorities", available at: <https://www.msi.org/article/msi-announces-2022-24-research-priorities/> (accessed 13 February 2023).
- Newman, M., Pennebaker, J., Berry, D. and Richards, J. (2003), "Lying words: predicting deception from linguistic styles", *Personality and Social Psychology Bulletin*, Vol. 29 No. 5, pp. 665-675.
- Nguyen, N., Nguyen, D., Suseno, Y. and Bui Quang, T. (2020), "The flipped side of customer perceived value and digital technology in B2B professional service context", *Journal of Strategic Marketing*, pp. 1-21.

- Niu, R. and Fan, Y. (2018), "An exploratory study of online review management in hospitality services", *Journal of Service Theory and Practice*, Vol. 28 No. 1, pp. 79-98.
- Oh, S., Ji, H., Kim, J., Park, E. and del Pobil, A.P. (2022), "Deep learning model based on expectation-confirmation theory to predict customer satisfaction in hospitality service", *Information Technology and Tourism*, Vol. 24 No. 1, pp. 109-126.
- O'Reilly-Shah, V. (2017), "Factors influencing healthcare provider respondent fatigue by answering a globally administered in-app survey", *PeerJ*, Vol. 5, e3785.
- Palese, B. and Piccoli, G. (2016), "Online reviews as a measure of service quality", *Pre-ICIS SIGDSA/IFIP WG8.3 Symposium*, Dublin.
- Park, Y. (2018), "Predicting the helpfulness of online customer reviews across different product types", *Sustainability*, Vol. 10 No. 6, p. 1735.
- Paschen, J., Kietzmann, J. and Kietzmann, T. (2019), "Artificial intelligence (AI) and its implications for market knowledge in B2B marketing", *Journal of Business and Industrial Marketing*, Vol. 34 No. 7, pp. 1410-1419.
- Pemer, F. (2021), "Enacting professional service work in times of digitalization and potential disruption", *Journal of Service Research*, Vol. 24 No. 22, pp. 249-268.
- Pennebaker, J. (2011), *The Secret Life of Pronouns: What Our Words Say about Us*, Bloomsbury Publishing, New York.
- Pennebaker, J., Booth, R., Boyd, R. and Francis, M. (2015), *Linguistic Inquiry and Word Count (LIWC): LIWC2015*, Pennebaker Conglomerates (LIWC.net), Austin, TX.
- Pennebaker, J., Chung, C., Frazee, J., Lavergne, G. and Beaver, D. (2014), "When small words foretell academic success: the case of college admissions essays", *PloS One*, Vol. 9, p. 110.
- Pennebaker, J. and King, L. (1999), "Linguistic styles: language use as an individual difference", *Journal of Personality and Social Psychology*, Vol. 77 No. 6, pp. 1296-1312.
- Pennebaker, J., Mehl, M. and Niederhoffer, K. (2003), "Psychological aspects of natural language use: our words, our selves", *Annual Review of Psychology*, Vol. 54, pp. 547-577.
- Pitt, C., Mulvey, M. and Kietzmann, J. (2018), "Quantitative insights from online qualitative data: an example from the healthcare sector", *Psychology and Marketing*, Vol. 35 No. 12, pp. 1010-1017.
- Pollach, I. (2012), "Taming textual data: the contribution of corpus linguistics to computer-aided text analysis", *Organizational Research Methods*, Vol. 15 No. 2, pp. 263-287.
- Robertson, J., Ferreira, C. and Paschen, J. (2021), "Reading between the lines: understanding customer experience with disruptive technology through online reviews", *Australasian Marketing Journal*, Vol. 29 No. 3, pp. 215-224.
- Rooney, T., Krolikowska, E. and Bruce, H.L. (2021), "Rethinking relationship marketing as consumer led and technology driven: propositions for research and practice", *Journal of Relationship Marketing*, Vol. 20 No. 1, pp. 42-61.
- Rossetti, M., Stella, F. and Zanker, M. (2016), "Analyzing user reviews in tourism with topic models", *Information Technology and Tourism*, Vol. 16 No. 1, pp. 5-21.
- Searle, J. (1979), *Expression and Meaning*, Cambridge University Press, Cambridge.
- Singh, U., Singh, N., Gulati, K., Bhasin, N., Kumar, H. and Streejth, P. (2022), "A study on the revolution of consumer relationships as a combination of human interactions and digital transformations", *Materials Today: Proceedings*, Vol. 51, pp. 460-464.
- Song, B., Lee, C., Yoon, B. and Park, Y. (2016), "Diagnosing service quality using customer reviews: an index approach based on sentiment and gap analyses", *Service Business*, Vol. 10 No. 4, pp. 775-798.
- Sparks, B. and Bradley, G. (2017), "A "Triple A" typology of responding to negative consumer generated online reviews", *Journal of Hospitality and Tourism Research*, Vol. 41 No. 6, pp. 719-745.

-
- Statista (2021), "Market research industry – statistics and facts", available at: <https://www.statista.com/topics/1293/market-research/> (accessed 8 June 2021).
- Swol, L., Chang, C., Kerr, B. and Moreno, M. (2020), "Linguistic predictors of problematic drinking in alcohol-related Facebook posts", *Journal of Health Communication*, Vol. 25, pp. 214-222.
- Tausczik, Y. and Pennebaker, J. (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.
- Tsao, W. (2019), "Building the long bridge between visitors and customers through online general reviews", *Online Information Review*, Vol. 43 No. 2, pp. 201-218.
- Viswanathan, M. and Kayande, U. (2012), "Commentary on "common method bias in marketing: causes, mechanisms, and procedural remedies"", *Journal of Retailing*, Vol. 88 No. 4, pp. 556-562.
- Wang, F., Xu, H., Hou, R. and Zhu, Z. (2023), "Designing marketing content for social commerce to drive consumer purchase behaviors: a perspective from speech act theory", *Journal of Retailing and Consumer Services*, Vol. 70, 103156.
- Wirtz, J., Patterson, P., Kunz, W., Gruber, T., Lu, V., Paluch, S. and Martins, A. (2018), "Brave new world: service robots in the frontline", *Journal of Service Management*, Vol. 29 No. 5, pp. 907-931.
- Wu, L., Mattila, A., Wang, C. and Hanks, L. (2016), "The impact of power on service customers' willingness to post online reviews", *Journal of Service Research*, Vol. 19 No. 2, pp. 224-238.
- Yang, Z. and Fang, X. (2004), "Online service quality dimensions and their relationships with satisfaction: a content analysis of customer reviews of securities brokerage services", *International Journal of Service Industry Management*, Vol. 15 No. 3, pp. 302-326.
- Zhu, J., Chang, Y., Ku, C., Li, S. and Chen, C. (2021), "Online critical review classification in response strategy and service provider rating: algorithms from heuristic processing, sentiment analysis to deep learning", *Journal of Business Research*, Vol. 129, pp. 860-877.
- Zouari, G. and Abdelhedi, M. (2021), "Customer satisfaction in the digital era: evidence from Islamic banking", *Journal of Innovation and Entrepreneurship*, Vol. 10 No. 1, pp. 1-18.

Corresponding author

Caitlin Ferreira can be contacted at: caitlin.ferreira@uct.ac.za

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com