Article

AMERICAN MARKETING ASSOCIATION

Journal of Marketing Research
1-21

DOI: 10.1177/00222437221116058
journals.sagepub.com/home/mrj
SAGE

Paralanguage Classifier (PARA): An Algorithm for Automatic Coding of Paralinguistic Nonverbal Parts of Speech in Text

Andrea Webb Luangrath, Yixiang Xu, and Tong Wang

Abstract

Brands and consumers alike have become creators and distributors of digital words, thus generating increasing interest in insights to be gained from text-based content. This work develops an algorithm to identify textual paralanguage, defined as nonverbal parts of speech expressed in online communication. The authors develop and validate a paralanguage classifier (called PARA) using social media data from Twitter, YouTube, and Instagram ($N = 1{,}241{,}489$ posts). Using auditory, tactile, and visual properties of text, PARA detects nonverbal communication cues, aspects of text often neglected by other word-based sentiment lexica. This work is the first to reveal the importance of textual paralanguage as a critical indicator of sentiment valence and intensity. The authors further demonstrate that automatically detected textual paralanguage can predict consumer engagement above and beyond existing text analytics tools. The algorithm is designed for researchers, scholars, and practitioners seeking to optimize marketing communications and offers a methodological advancement to quantify the importance of not only *what* is said verbally but *how* it is said nonverbally.

Keywords

textual paralanguage, text analytics, consumer linguistics, social media, emoji, nonverbal communication, online word of mouth, consumer engagement

Online supplement: https://doi.org/10.1177/00222437221116058

Although the sheer volume of text-rich data offers exciting opportunities for marketing managers, the task of deciding what to consider is often overwhelming. Recently, the field of marketing has made great strides in understanding linguistic aspects of marketing messages predictive of relevant consumer attitudes and behaviors (e.g., Berger et al. 2020; Berger, Rocklage, and Packard 2021; Moore 2015; Netzer et al. 2012; Villarroel Ordenes et al. 2018). Recent research has called for analytical tools to aid researchers and practitioners in identifying and analyzing text-based content (Humphreys and Wang 2018). The purpose of our work is to develop a comprehensive automatic classifier for nonverbal communication expressed in text-based messages: the textual paralanguage classifier (PARA; www.textualparalanguage.com).

Currently, text analytics researchers mostly consider actual words themselves, forming inferences and making predictions from nuances in the meaning of words. In this research, we take a different approach: we focus on the extratextual elements in online written communication, termed "textual paralanguage" (TPL; Luangrath, Peck, and Barger 2017). "Nonverbal communication" refers to communication that is effected by means other than words (Knapp, Hall, and Horgan 2014). We focus on the subtleties in *how* something is written rather than the verbal aspects of *what* is written.

Online text differs dramatically from traditional English prose; the use of symbols, text-based images, and unique qualifiers of speech has given rise to fragmentation of linguistic norms. Extratextual features are often thought to be trivial, and many market researchers and firms employing text analytics begin cleaning data sets by normalizing spelling, extracting extraneous punctuation, or eliminating other symbols thought to be generating "noise." We propose here that these traditionally neglected aspects of text speech are actually quite meaningful.

The expression of nonverbal cues in text is a classic demonstration of the complexity and heterogeneity found in

Andrea Webb Luangrath is Associate Professor of Marketing, Tippie College of Business, University of Iowa, USA (email: andrea-luangrath@uiowa.edu). Yixiang Xu is a doctoral candidate in marketing, University of California, Berkeley, USA (email: yixiang-xu@berkeley.edu). Tong Wang is Assistant Professor of Business Analytics, Tippie College of Business, University of Iowa, USA (email: tong-wang@uiowa.edu).

unstructured data (Gandomi and Haider 2015) in the sense that it is nonnumeric and multifaceted and maintains concurrent representation (Balducci and Marinova 2018). Meaningful insights from unstructured data can be gleaned by giving theoretically relevant structure to an otherwise unstructured space. By developing a methodological approach to automatically detect textual paralanguage in online content, PARA gives structure to text data through the lens of nonverbal communication. To the best of our knowledge, this is the first and most comprehensive tool dedicated to detecting features of nonverbal cues in text.

In this work, we demonstrate the functionality of PARA by revealing the theoretical and empirical relevance of textual paralanguage to the detection of sentiment valence and intensity. We show that PARA is particularly adept at identifying sentiment intensity relative to other text tools. We also reveal the usefulness of this tool in a marketing-relevant context via the prediction of consumer engagement. Scholars and practitioners can use PARA to detect paralinguistic cues to glean insights from and optimize social media communication. We begin with a conceptual discussion of textual paralanguage, situate this topic within a broader body of work on consumer linguistics in marketing, detail the development and validation of PARA, and conclude by discussing relevant future research directions.

Theoretical Foundation of TPL

TPL is defined as written manifestations of nonverbal audible, tactile, and visual elements that supplement or replace written language and that can be expressed through words, symbols, images, punctuation, demarcations, or any combination of these elements (Luangrath, Peck, and Barger 2017). Face-to-face communication is rich with nonverbal behavior, including body language, eye contact, and tone of voice; these elements of communication are often referred to as "paralanguage" (Poyatos 2002; Trager 1958). When people exchange written messages electronically, these elements need to be translated into text to be communicated.

Text is imbued with nonverbal cues in myriad ways (Table 1). Auditory aspects of speech are generally indicative of how words would be spoken. For example, vocal aspects can convey tempo (e.g., "amazingggggg," denoted with "stretchable words"; Gray, Danforth, and Dodds 2020). A nonverbal expression can be suggestive of physical contact with another person, with examples including emojis that convey touch (e.g., the hug emoji) or even indications of touch using alphanumeric letters and symbols (e.g., *high-five*). Visual forms of nonverbal communication can merely correspond to presentational elements of a message (e.g., ★彡) or can relate to different forms of body language (e.g., the thumbs-up emoji). Any form of nonverbal communication has the potential to be translated into text.

Humphreys and Wang (2018, p. 1295) acknowledge that "while text analysis provides ample information, giving meaning to it requires theory." We use an established conceptual model of textual paralanguage (Luangrath, Peck, and Barger 2017), such that our detection of nonverbal communication cues in text is grounded in nonverbal and sensory marketing theory. Analogous to the identification of properties of speech such as nouns, verbs, and prepositions in verbal content, we categorize the "properties of nonverbal speech" denoted in text. Indeed, scholars have acknowledged that this approach (i.e., to identify parts of speech) can reveal a great deal about human personality, emotions, and behavior (e.g., Packard and Berger 2020; Packard, Moore, and McFerran 2018; Pennebaker 2011; Sela, Wheeler, and Sarial-Abi 2012). With an aim of being as comprehensive as possible in the identification of TPL, PARA intentionally classifies these nonverbal cues at a rudimentary sensory level.

Our process aligns with modern approaches to the study of nonverbal expressions in that these cues are not viewed as deterministic of specific functions or emotions (Barrett 2017). For example, artificial intelligence systems can detect the presence of actual facial expressions (e.g., Emotient 2015), and yet the mere existence of a facial expression does not presuppose emotional inference (Barrett 2017); that is, there is a difference between the presence of a facial expression (e.g., a smile) and the inference that one is experiencing an emotion (e.g., joy). Likewise with TPL, the presence of a smiling emoji does not definitively imply emotion felt by the sender or emotional inference by the receiver. Thus, PARA is built with the intention to detect these nonverbal parts of speech rather than to deterministically categorize these elements according to emotion or any other higher-order function. The typology laid forth by Luangrath, Peck, and Barger (2017) provides the scaffolding on which the classifier is built and aids in the identification of the basic nonverbal features in text.

Sentiment Valence and Intensity Derived from TPL

One of the most prominent ways to distill meaningful inferences from text is through sentiment analysis, which has been used to identify brand and consumer sentiment (e.g., Homburg, Ehm, and Artz 2015; Schweidel and Moe 2014). Measurement of opinions and affective states have generally relied on two aspects of sentiment analysis: valence (whether a message is positive, negative, or neutral) and intensity (the degree of positivity or negativity of a message) (Liu 2012; Pang and Lee 2008; Villarroel Ordenes et al. 2017). We argue that the detection of TPL is crucial for a complete and robust analysis of sentiment valence and intensity.

Instances of TPL may have more than one denotative or connotative meaning, as is true of many words (Akkaya, Wiebe, and Mihalcea 2009); TPL elements vary in their direct mapping to sentiment valence. Research has demonstrated that some emojis, for example, have strong positive or negative associations while the valence of others is more ambiguous (Miller et al. 2016). Similarly, instances of TPL can vary in conveyance of negativity and positivity. For example, heart emojis

Table 1. Output Variables Generated by PARA and Types of Textual Paralanguage.

	Superordinate Sensory Level	PARA Output Variables	Sample Text
	—	—	This coffee is amazing.
Textual Paralinguistic Nonverbal Part of Speech	Auditory TPL	Stress	This coffee is AMAZING.
		Tempo	This coffee is amazingggggg.
		Rhythm	This. Coffee. Is. Amazing.
		Emphasis	This coffee is amazing!!!!!!
		Pitch	THiS CoFfEe iS aMaZiNg.
		Volume	This coffee is amazing. *peaceful silence*
		Censorship	This coffee is amazing. Holy S#!T.
		Spelling	This coffee is a-m-a-z-i-n-g.
		Alternant	This coffee is amazing. Mmm.
		Differentiator	This coffee is amazing. *gulp*
	Tactile TPL	Tactile Emoji	This coffee is amazing.🤗
		Tactile Emoticon	This coffee is amazing. :-*
		Alphahaptics	This coffee is amazing. *high-five*
	Visual TPL	Bodily Emoji	This coffee is amazing.👍
		Bodily Emoticon	This coffee is amazing. :)
		Alphakinesics	This coffee is amazing. *grinning*
		Nonbodily Emoji	This coffee is amazing.☕
		Nonbodily Emoticon	This coffee is amazing. ★彡
		Formatting	1. Get coffee. 2. Drink. 3. It's amazing.
Aggregate Variables		Emoji Count	Raw count of the number of emojis
		Emoji Index	Summation of tactile emojis, bodily emojis, and nonbodily emojis
		Emoticon Index	Summation of tactile emoticons, bodily emoticons, and nonbodily emoticons
		TPL Index	Summation of all TPL nonverbal parts of speech (e.g., stress, tempo)

Notes: In addition to the parts of speech, PARA generates other aggregate variables and indices, for a total of 23 output variables, which are written as one line of data to an output file. For a full descriptive typology of TPL, see Luangrath, Peck, and Barger (2017).

tend to be associated with positive sentiment and thus are more apt to contribute information to valence assessments. Conversely, stress (e.g., using all capital letters), could be applied to text content that is either positive or negative (e.g., "EXCELLENT," "TERRIBLE"); thus, the mere existence of capital letters is not informative about the valence of a message.

In a similar fashion to valence, TPL is an indicator of sentiment intensity in text. Indeed, "many applications would benefit from being able to determine not just the binary polarity (positive versus negative), but also the *strength* of the sentiment expressed in text" (Hutto and Gilbert 2014, p. 218, emphasis in original). As such, this has been acknowledged and captured by existing text tools (e.g., Hutto and Gilbert's [2014] VADER; Rocklage, Rucker, and Nordgren's [2018] Evaluative Lexicon). Prior work has also connected in-person nonverbal cues to expression intensity. For example, louder voices and faster speech have been shown to affect perceptions of emotional intensity (Juslin and Laukka 2001). We expect that TPL will facilitate the amplification, or intensity, of a message (e.g., via variations of alphanumeric keystrokes like "!!" vs. "!!!!!!"). As with valence, certain TPL features more readily convey intensity. The use of capital letters on certain words, while less capable of revealing sentiment valence, likely plays a larger role in affecting sentiment intensity. Thus, the intention of our methodological approach is to identify and classify nonverbal cues in text and use the detection of these nonverbal parts of speech to inform the measurement of sentiment valence and intensity.

TPL as a Facilitator of Social Engagement

Linguistic aspects of communication can provide insight into consumer- and brand-relevant outcomes and the power of sharing thoughts, feelings, and emotions with others. Extant research shows that consumer-generated text, or online word of mouth, shapes preferences for products (Packard and Berger 2017), affects what is shared (Barasch and Berger 2014), and influences buying behavior (Berger 2014; Chen and Berger 2016; Chevalier and Mayzlin 2006). Increasingly, researchers are investigating nonverbal cues as a source of linguistic insight. For example, textual paralinguistic cues, in the form of emoticons, have been shown to positively influence warmth perceptions of customer service employees but negatively influence competence perceptions (Li, Chan, and Kim 2019).

Given the proposed relevance of TPL to sentiment valence and intensity, we expect that it will also possess predictive power for an important marketing-related outcome: consumer engagement. Consumer engagement, broadly defined as action taken by a consumer to interact with content on social media, is of great concern to marketers (Brodie et al. 2013; Calder, Malthouse, and Maslowska 2016; Hollebeek, Glynn, and Brodie 2014). Prior work shows that virality of online content is not driven by emotion alone; rather, arousal levels, or the intensity of that emotion, motivate sharing behaviors (Berger and Milkman 2012). Likewise, as nonverbal expression influences sentiment valence and intensity, it is likely to be predictive of consumer engagement with social media content. Further, appealing to the sensory system by adding sensory-laden stimuli has been shown to create more engaging content (e.g., Sanchez-Vives and Slater 2005) and thus, we expect a sensory-oriented view of language to improve prediction of consumer engagement.

In the remainder of this article, we detail the development and validation of PARA. We present the results of three studies (Table 2). Holistically, Study 1 documents construct validity. Using Twitter data from large corporate brands, we detail dictionary development and ensure comprehensive coverage of paralinguistic elements via a process of theoretical saturation (Study 1a). We illustrate the workflow and rule-based criteria (Study 1b) and assess the accuracy of PARA using comparisons with human-coded instances of textual paralanguage (Study 1c). In Study 2, we demonstrate causal validity by showing that PARA influences human ratings of sentiment valence and intensity as well as incremental validity by showing that PARA aids in the assessment of sentiment above and beyond existing linguistic tools. Finally, in Study 3, we show PARA's predictive validity by demonstrating that it significantly improves prediction of consumer engagement relative to using an extensive list of word and topic-based linguistic comparisons. We conclude with a broad theoretical and managerial discussion of the applications of PARA.

Methodology

Study 1a: Keyword Dictionary Development for PARA

Identification of TPL relies on both dictionary-based and rule-based approaches. No dictionary for TPL exists; therefore, we began by developing a panel of dictionaries to guide classification. We discovered or created a total of five subdictionaries that identify whether words, symbols, or text-based images are indicative of TPL (the first three are standardized, and we constructed the latter two): an emoji dictionary, an emoticon dictionary, a symbol dictionary, a doppelgänger dictionary, and a keyword dictionary. The purpose of Study 1a is to detail the methods and results for the development of the constructed keyword and doppelgänger dictionaries (see Web Appendix A for a description of the standardized dictionaries).

Data

We gathered data from official Twitter accounts of 69 national brands across a range of industries (N = 11,032). Note that some of the brand Twitter data (N = 22 brands) in the current study are also reported in Luangrath, Peck, and Barger (2017). To supplement these data and ensure broad coverage of TPL features, we collected additional tweets from other brands (N = 47), some of which were chosen because they are part of "*Forbes*' Most Valuable Brands" (Badenhausen 2018) and others because of their noteworthy presence on social media (Grossman 2014) (see Table 3 for sample tweets and Web Appendix A for a full list of brands). We gathered an additional small data set (N = 200) of tweets to test for theoretical saturation.

Keyword Dictionary Development Procedure and Results

Item generation. Using TAMS analyzer, an open-source software program for coding (Weinstein 2020), three research assistants coded the tweets for TPL. Coders were trained on the construct of TPL, given the definitions of the categories, and provided with examples. After independently coding the tweets, they included all identified instances of TPL in the keyword dictionary. They discussed any categorization discrepancies among themselves and achieved consensus. This resulted in the identification of 115 unique keywords included in the dictionary, which served as the basis for further human validation.

Construct validation of items generated. We gave the generated keyword dictionary to three new research assistants, who voted to either include or exclude keywords from the dictionary as well as suggest keywords they believed should be included. Following established procedures for dictionary refinement (Humphreys 2010), we established inclusion/exclusion criteria as follows: (1) if two of three coders voted for inclusion, the keyword was included; (2) if two of three coders voted for exclusion, it was excluded; and (3) if one of three coders offered a word to include, coders discussed it among themselves. Two or all three coders recommended the removal of 35 instances of the 115 keywords generated. Independent of the data, the new coders then generated an additional 223 instances of TPL for the dictionary, of which they agreed on 205 for inclusion. For example, the data revealed items such as "ugh" and then the coders identified additional instances such as "arg," "argh," and "humph," which could be "like-instances" or merely thought to represent the construct. In total, 285 keywords were included in the dictionary.

Keyword theoretical saturation. To ensure that we adequately captured the TPL construct, we followed a saturation procedure similar to prior work (Weber 2005). The logic behind this procedure is that a new set of data is coded until it yields no new information (Berger et al. 2020). Using the separate small data set of tweets, two new individuals coded for TPL. We captured no new instances of TPL through this process, which suggests theoretical saturation was achieved. The resulting keyword dictionary is a constructed dictionary that captures

Table 2. Overview of Textual Paralanguage Classifier (PARA) Validation Studies.

Study	TPL Validation	Validation Technique	Outcome	Data	N	Results
Study 1	Construct validity	Dictionary development/ rule-based algorithm	—	Brand Twitter data[a]	11,032	Instances of TPL are generated from brand data, and human coders evaluated fit of TPL instances and categorization. Details of the rule-based algorithm are developed.
		Saturation		General public, Twitter data	200	Assessment of theoretical saturation illustrates the robustness of dictionary creation.
		Hit rate (accuracy)	Human-coded Instances of TPL	General public, Twitter data	5,000	Construct validity is demonstrated with high accuracy of PARA with human-coded instances of TPL. Krippendorff's alpha across all TPL features is .896 (auditory TPL: .916, tactile TPL: .863, visual TPL: .910).
Study 2	Causal validity	Causal inference from text (selection on observables design)	Human ratings of sentiment intensity and valence	Twitter and YouTube data	9,200	Using publicly available Twitter data (Hutto and Gilbert 2014) as well as YouTube comments data, TPL significantly affects human-rated assessment of intensity and valence (establishing causal validity). We identify the relative influence of TPL features on human-coded sentiment intensity and valence. We also show the generalizability of PARA by incorporating data from a different social media platform.
	Incremental validity	Comparison with existing text analytic tools[a]				We demonstrate incremental validity via comparisons with other established linguistic tools. PARA adds explanatory/predictive power to detect human-coded valence and intensity above and beyond existing linguistic tools[a] by capturing nonverbal parts of speech.
Study 3	Predictive validity	Holdout sample	Consumer engagement (i.e., likes, retweets, comments)	Twitter data, Instagram data	1,149,913, 66,144	PARA has superior predictive performance relative to other established text analytic tools that rely predominantly on word-based assessments. Inclusion of TPL improves prediction on top of an extensive list of tweet-specific controls as well as numerous word and topic-based comparisons.

[a]LIWC (Pennebaker et al. 2015), VADER (Hutto and Gilbert 2014), Arousal-Dominance-Valence Lexicon (Warriner, Kuperman, and Brysbaert 2013), Evaluative Lexicon (Rocklage, Rucker, and Nordgren 2018), Hedonometer (Dodds et al. 2011), SenticNet (Cambria et al. 2018), SenticWordNet (Baccianella, Esuli, and Sebastiani 2010).

expressions of nonverbal communication conveyed in alphanumeric notation.

Doppelgänger Dictionary Development Procedure and Results

We also developed a subdictionary for the purposes of identifying elements that may *seem* to be TPL but in fact are not, aptly named the "doppelgänger dictionary." This dictionary is exclusionary (rather than inclusionary, like the keyword dictionary) and was born out of PARA errors. In total, we identified 256 doppelgängers. For example, words written in capital letters are not always indications of nonverbal cues. A retweet on Twitter is indicated by "RT," which is not TPL. Frequently used abbreviations like states (e.g., AL, AK), time zones (e.g., EST), or agencies/organizations (e.g., NASA, IRS) are also not TPL. In addition, it is also inappropriate to say that

Table 3. Sample Tweets Containing Textual Paralanguage from Brands (Study 1a).

Brand	Twitter Handle	Sample Tweet	TPL Subcategory
Arby's	@arbys	Brisssssket. Say it out loud. Let it marinate the air. #meatcraft <LINK>	Tempo: Elongation of "s"
Burger King	@burgerking	When someone says we're going to BK: (-__-)ノ	Bodily emoticon: (-__-)ノ
Coca-Cola	@cocalcola	Next time you're at the movies don't feel bad about taking that last "SLURRRP!" of Coke. Everyone's been there.	Stress: SLURRRP Tempo: RRR Differentiator: slurp
Digiorno Pizza	@digiornopizza	Always look at the bright side of life. *turns on oven light, watches pizza do its thing*	*asterisk action*
Frappuccino	@frappuccino	OoOoOoOo…SPARKLYYYYY! #HappyFourth #4thofJuly <LINK>	Pitch: OoOoOoOo Tempo: Ellipses … Stress: SPARKLYYYYY Tempo: YYYYY
General Electric	@generalelectric	A Scientist's Checklist: ✅ Research 📖 ✅ Tools 🔬 ✅ Safety glasses 👓 ✅ Coffee ☕	Artifact list formatting Check mark emoji (4) Open book emoji Microscope emoji Glasses emoji Hot beverage emoji
Hamburger Helper	@helper	Is the karate chop a legit MMA move? Cause if it is, I think I might have another career after cooking. ✋👊💪	Bodily emoji: Raised hand Tactile emoji: Oncoming fist Bodily emoji: Flexed biceps
Hot Pockets	@hotpockets	Mmmmmmmmmmmmmmmmmmmm…delicious. #HotPockets	Tempo: Elongation of "m" Alternants: Mmm Tempo: Ellipses …
McDonald's	@mcdonalds	It's like opening a present. Every. Single. Morning. <LINK>	Rhythm: Every. Single. Morning.
Nescafe	@nescafe	It's Wednesday already?! Time for a well deserved #NESCAFE	Emphasis: ?!
Progressive	@progressive	Americans drive over 3 TRILLION miles each year. That's halfway from the sun to the edge of the solar system! <LINK>	Stress: TRILLION
Smokey the Bear	@smokey_bear	Happy #NationalHugDay #SmokeyBearHug <|:') x < <LINK>	Tactile emoticon: <|:') x <
Taco Bell	@tacobell	When you drive past Taco Bell and don't stop 😭	Bodily emoji: Crying face

Notes: Tweets often contained links to other content, which are denoted here as <LINK>.

capital letters should apply only to English words; this would neglect any word not properly spelled (e.g., "This sale is HUUUUGE"). Thus, this doppelgänger dictionary aids PARA in distinguishing between capitalization for the purposes of nonverbal conveyance from that which is merely functional. All of the dictionaries work in tandem with rule sets to support TPL classification.

Study 1b: Rule-Based Algorithm for PARA

To detect paralanguage, we also developed a rule-based procedure (Abraham 2005; Clancey 1983). Rule-based systems encode and translate human knowledge into several hard-coded association rules that a system can then execute. Classification began by creating modules, one for each subcategory of textual paralanguage (Table 1). Each of the modules relied on three basic phases (Figure 1) (see Web Appendix B for full rule sets for all modules).

Phase I: Input, cleaning, and restructuring. The first processing block cleans and restructures the input text to remove items that frequently show up in social media but are not relevant to TPL classification. For example, the block eliminates all text content attached to and beginning with "#," "http:," "@," and "RT" from further processing, as are all tweets not in English. In addition, tweets are split into a bag of items chunked according to spaces.

Phase II: Pattern detection. Using the cleaned text, the next processing block identifies patterns of text and symbols necessary to determine a specific paralinguistic feature. We programmed a module for each subcategory using regular expression and strict if-then rules. For example, in the detection of tempo, letter repetition is often expressed in the elongation of a word, so a tweet that contains "hmmmmmm" would be identified in this phase for the elongation of the letter "m" as a potential candidate for indication of tempo.

Phase III: Stemming and screening. Phase III, which further screens potential instances of TPL, is critical for classification accuracy. Stemming, or reconstructing a word into its most

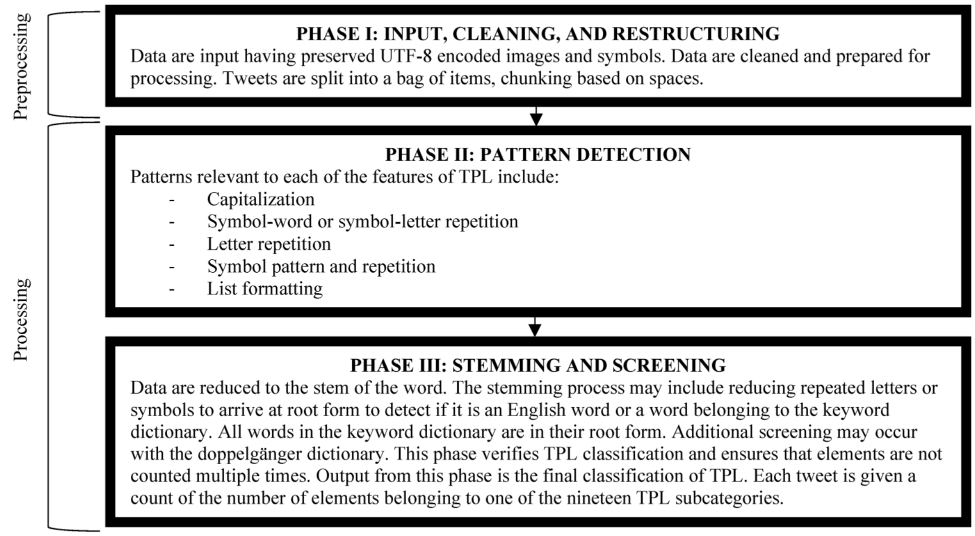

Figure 1. PARA Workflow for Classification of Textual Paralanguage.

basic form, is common in computational linguistics (Lovins 1968) to capture all forms of a word (e.g., lov- for love, loved, loves, loving, lover). In our case, stemming identifies the root of the word to determine whether it represents a paralinguistic element. For example, Phase II would detect "hmmmmmm" as containing letter repetition, and in Phase III, the word would be transformed to its root "hmm," thereby eliminating the elongation. This root belongs to the keyword dictionary, so it would be counted. To explicate the complexity of this phase, a single word could be elongated in a variety of ways (e.g., "really" could be expressed as "reeeeeally," "reallllly," or "reallyyyyyy"). PARA can detect these varieties and captures all possible stems for a word. In a sense, the algorithm has been programmed to engage in iterative processing by systematically bouncing between modules and expanding or contracting word forms to check against our dictionaries. After this phase, the detection and categorization of TPL is complete.

Study 1c: Construct Validation with Human Coders of TPL

The purpose of Study 1c is to test the construct validity of PARA by examining whether the classifier identifies what it purports to identify. We test whether the tool accurately classifies instances of TPL compared with a human coder.

Data and Method

We generated data from general Twitter users, not from brands, because tweets originating from national brands are vetted internally and curated with intention, often going through various approvals before posting. From a composition perspective, tweets from the public are arguably messier and more idiosyncratic. Because we are interested in demonstrating accuracy beyond formalized corporate tweets, we opted to test classification accuracy on a data set of tweets from the general public. We queried the Twitter streaming API at various times of day until 5,000 tweets were acquired. Following a similar process as in Study 1a, the tweets were human-coded for TPL.

We defined the accuracy of PARA as the proportion of correct classifications among the total number of observations. For every tweet, a correct classification occurred when the number of TPL instances in each category (i.e., auditory, tactile, or visual) detected by PARA exactly matched that identified by human coders. We calculated two accuracy rates: (1) accuracy across all tweets, which reflects the ability of PARA to detect TPL when there was, in fact, TPL, and (2) accuracy for TPL-containing tweets, which reflects how well PARA categorizes and counts instances of TPL.

Results

After Phase I of processing through PARA, 4,589 tweets remained, as posts that only contained a hyperlink, for example, did not proceed for further processing. We identified TPL in 1,873 (40.81%) of the remaining tweets. Then, we compared the performance of PARA with that of human coders and determined an accuracy rate on all tweets of 96.70%. PARA attained high accuracy across all of the TPL categories (auditory TPL: 93.51%, tactile TPL: 99.41%, and visual TPL: 97.19%). The second accuracy rate, which was determined using TPL-containing tweets only, was 92.94%. The categories of TPL varied in the level of accuracy but again resulted in high accuracy (auditory TPL: 85.90%, tactile TPL: 98.77%, and visual TPL: 94.13%). The intraclass correlation coefficient, a meta-reliability metric between the two accuracy rates, is .86, which indicates high reliability. Krippendorff's alpha across all TPL is .896 (auditory TPL: .916, tactile TPL: .863, and visual TPL: .910). PARA also achieved good performance for other metrics including sensitivity, specificity, precision, and the F1-score at not only the category level but also the subcategory level (Web Appendix A). Results demonstrated construct validity with high accuracy of PARA using the human coder as the benchmark.

Study 2: TPL Facilitates Assessments of Sentiment Valence and Intensity

The purpose of Study 2 is to investigate whether TPL facilitates the detection of sentiment. In this study, we sought to establish both causal and incremental validity. Causal validity, a form of internal validity, assesses the extent to which a construct, as operationalized in a data set, is actually the cause of another construct or outcome (Berger et al. 2020). Given the theoretical relevance of in-person nonverbal cues to expressed sentiment (e.g., Juslin and Laukka 2001), we expect TPL to causally affect sentiment valence and intensity in text.

In Study 2, we also investigated incremental validity, defined as the degree to which a measure explains or predicts a phenomenon of interest, above and beyond other measures (Haynes and Lench 2003). Thus, we expected the effect of TPL on text sentiment to hold even after we compared other sentiment-related text tools or lexica concurrently. Given that PARA is robust at capturing elements of nonverbal parts of speech, we would theoretically expect Study 2 to demonstrate incremental validity not only relative to tools that assess verbal content but also above and beyond the nonverbal components captured by existing text analytic tools.

Data

We compiled a data set of social media text. The data set contained publicly available data from the authors of VADER (Hutto and Gilbert 2014), which included 4,200 tweets that had been manually coded for sentiment. These data preserved alphanumeric indications of TPL (e.g., "!!!," and emoticons) but did not preserve emojis. To compensate for the lack of emojis and the limitation of data acquired from one social media platform, we enriched the data set with a sample of scraped YouTube comments. To this end, we used an open-source web client (Klostermann 2019) to query the YouTube comment API at several times of day until we acquired all comments from the top 100 most-viewed YouTube videos (YouTube 2019). We then randomly sampled YouTube comments written in English (N = 5,000).

Measurement

We asked four coders to manually evaluate all 5,000 YouTube comments on sentiment valence and intensity. We instructed these coders to rate each comment on a nine-point scale (1 = "extremely negative," and 9 = "extremely positive"); the procedure and measurement were consistent with prior work (Hutto and Gilbert 2014). Thus, these manually coded sentiment scores served as the outcome variables. For descriptive statistics, see Web Appendix C.

We processed all of the data using a variety of common text analytic tools or lexica including the Arousal-Dominance-Valence Lexicon (Warriner, Kuperman, and Brysbaert 2013), Evaluative Lexicon (Rocklage, Rucker and Nordgren 2018), Hedonometer (Dodds et al. 2011), Linguistic Inquiry and Word Count (LIWC; Pennebaker et al. 2015), SenticNet (Cambria et al. 2018), SenticWordNet (Baccianella, Esuli, and Sebastiani 2010), VADER (Hutto and Gilbert 2014), and PARA (for descriptive statistics for Studies 2 and 3, see Web Appendices C and E). Important to note is that VADER (Hutto and Gilbert 2014) is a tool that acknowledges some paralinguistic features (e.g., CAPS, "!!!") to aid in assessments of sentiment intensity and valence. While many tools strip away symbols and unique markers, this tool preserves certain nonverbal features. Similarly, LIWC (Pennebaker et al. 2015) generates an output called "netspeak," which captures some nonverbal components as well. The inclusion of both of these tools as benchmarks provided a conservative test of incremental validity.

Although VADER captured verbal and nonverbal components of a message, it did not delineate these two aspects of speech explicitly in the output, thus obfuscating the distinction between nonverbal and verbal indicators. To mitigate this issue, we removed all the nonverbal elements in the data and used VADER to process those verbal-exclusive text, which resulted in a VADER-generated *verbal* score. We further computed a VADER-generated *nonverbal* score by regressing the VADER intensity score (generated from the original data) on the VADER verbal score and taking the residual as the approximate measure of the VADER nonverbal score.

Methods

This study employed a quasi-experimental design using a causal inference from text approach following prior methods (Feder et al. 2021; Manzoor et al. 2020; Roberts, Stewart, and Nielsen 2020). In contrast to a lab experiment, in which the treatment is randomly assigned for full control over confounding variables, our data were partially generated in a controlled environment minimizing the potential for confounds while at the same time utilizing actual text data. Any potential confound would need to affect both TPL and the outcome variables (sentiment valence or intensity), with such confounds operating through channels observable to the human coders. These human-rated indicators of sentiment ensured that the only potential source of confounding was from the content of the text. In other words, causal inference was possible based on "selection on observables," meaning that the treated group (text with TPL) and the nontreated group (text without TPL) differed only in terms of a set of observable characteristics—namely, the content of the text itself. Thus, by controlling for the text, it blocked spurious pathways between TPL and sentiment, which enabled us to demonstrate observationally causal effects of TPL on valence and intensity.

We used a robust set of variables to control for text. First, it is plausible that text may vary depending on the social media platform; thus, we included platform fixed effects. Second, we controlled for the sentiment of the text by extracting 22 sentiment-related variables generated by a set of seven text tools. Third, to control for additional features of the text itself, we identified *words* and *topics* with which TPL and sentiment may co-occur. For example, people may be more likely to use an emoji with the word “happy” than with the word “okay.” In addition, more serious topics like political news might use less TPL, whereas more entertaining topics like celebrity gossip might use more. The general idea is that if we can extract and control for the word and topic confounds from text, then we will be able to identify the causal effect of TPL on sentiment valence and intensity. To do so, we used a structural topic model (STM). Unlike standard topic models (e.g., latent Dirichlet allocation; Blei, Ng, and Jordan 2003), STM allowed us to integrate the treatment information (i.e., existence of TPL) into the topic generation process and, therefore, derive topics that were associated with our treatment. We controlled for word-based confounds using an STM-based dimension reduction technique, which generated a text-based propensity score that captured how TPL is affected by words (Roberts, Stewart, and Nielsen 2020; see Web Appendix D for additional details).

TPL instances per post were often zero or one, which caused statistical inference to depend on unjustifiable extrapolation. To avoid such extrapolation, we binarized each subcategory indicating its presence or absence. We ran two sets of regression, one using the binary indicator of PARA and the other using the TPL subcategories (using 19 distinct TPL features; Table 1) as the regressors on both sentiment valence and intensity. We specified the regressions as follows:

$$\text{i.}\quad \mathrm{Y_i} = \alpha_1 + \beta_1\mathrm{PARA_i} + \gamma_1\mathrm{PS_i} + \theta_1\mathrm{Platform_i} + \sum_j \omega_j\mathrm{Tool_{ij}} + \sum_j \delta_j\mathrm{Topic_{ij}} + \varepsilon_i,$$

$$\text{ii.}\quad \mathrm{Y_i} = \alpha_1 + \sum_{j=1}^{19}\beta_j\mathrm{TPLsubcategory_{ij}} + \theta_1\mathrm{Platform_i} + \sum_j \omega_j\mathrm{Tool_{ij}} + \sum_j \delta_j\mathrm{Topic_{ij}} + \sum_{j=1}^{19}\gamma_j\mathrm{PS_{ij}} + \varepsilon_i, \text{ where}$$

$\mathrm{Y_i}$ is manually coded sentiment valence or sentiment intensity;

$\mathrm{PARA_i}$ indicates whether TPL of any kind is present;

$\mathrm{TPLsubcategory_{ij}}$ indicates whether a particular TPL subcategory is used;

$\mathrm{Platform_i}$ indicates whether the text is from Twitter or YouTube and controls for the fixed effects of data source;

$\mathrm{Tool_{ij}}$ includes all the variables generated by the set of sentiment-related text tools;

$\mathrm{Topic_{ij}}$ includes all topics extracted from the text based on the structural topic model (we determined the number of topics using Mimno and Lee’s (2014) algorithm); and

$\mathrm{PS_{ij}}$ is the propensity score. In the regressions using PARA, it measures the probability that a tweet contains TPL of any kind. In the regression using TPL subcategory, each $\mathrm{PS_{ij}}$ measures the probability of a tweet to contain a particular TPL subcategory.

Results

Causal validity results. After Phase I processing, 9,139 observations remained. TPL occurred in more than half of all posts (N = 4,667; 51.07%). Among the robust set of control variables, the topic modeling extracted 65 latent topics from the text data. We tested the main contention that the paralinguistic features captured by PARA significantly affected the sentiment valence and intensity perceived by a human. Results revealed that, on average, PARA significantly increased both sentiment valence (B = .092, SE = .018, $p < .001$; Table 4, Model 3) and sentiment intensity (B = .36, SE = .018, $p < .001$; Table 4, Model 6).

Table 4 presents estimates of PARA using regression i by gradually including control variables. Importantly, across all model specifications, the effect of PARA on both sentiment valence and intensity remained statistically significant. Results demonstrated that a simplistic model with no controls overestimated the effect of PARA on both sentiment valence (B = .221, SE = .019, $p < .001$; Table 4, Model 1) and intensity (B = .472, SE = .016, $p < .001$; Table 4, Model 4), likely because the model did not account for other words, topics, platforms, or lexical features with which the paralinguistic elements co-occur. Controlling for platform fixed effects and other text tools only partially mitigated these estimation biases (Table 4, Models 2 and 5). A specification curve analysis, with approximately 152 alternative specifications, demonstrates the robustness of our findings (Web Appendix D). These results provide evidence of observational causality; in other words, PARA significantly influenced sentiment perceptions.

Regressions using the TPL subcategories (i.e., regression ii) further revealed heterogeneous effects on sentiment (Figure 2). For example, the auditory TPL subcategories of stress (B = .393, SE = .026, $p < .001$), emphasis (B = .30, SE = .024, $p < .001$), and tempo (B = .097, SE = .022, $p < .001$) revealed significant effects on sentiment intensity but had no statistically significant effects on valence.

Incremental validity results. Our regression i results further revealed significant effects of PARA on sentiment valence and intensity over and above existing text analytic tools. First, the significant main effect of PARA remained after controlling for the series of comparison text tools (sentiment valence: B = .097, SE = .014, $p < .001$, Table 4, Model 2; sentiment intensity: B = .372, SE = .015, $p < .001$, Table 4, Model 5). To standardize the units for comparison between PARA and the other tools, we computed effect sizes by converting

Table 4. Results of PARA Effects on Sentiment Valence and Intensity (Regression i).

	Dependent Variable					
	Sentiment Valence			Sentiment Intensity		
	(1)	(2)	(3)	(4)	(5)	(6)
PARA	.221***	.097***	.092***	.472***	.372***	.360***
	(.019)	(.014)	(.018)	(.016)	(.015)	(.018)
Platform fixed effects	No	Yes	Yes	No	Yes	Yes
Text tools	No	Yes	Yes	No	Yes	Yes
Topics	No	No	Yes	No	No	Yes
Propensity score	No	No	Yes	No	No	Yes
Observations	9,139	9,139	9,139	9,139	9,139	9,139
R^2	.015	.495	.507	.088	.320	.358
Adjusted R^2	.015	.493	.502	.087	.318	.352
Residual standard error	.897	.643	.638	.762	.658	.642
	(d.f. = 9,137)	(d.f. = 9,114)	(d.f. = 9,049)	(d.f. = 9,137)	(d.f. = 9,114)	(d.f. = 9,049)
F-statistic	138.142***	371.931***	104.629***	876.732***	178.810***	56.809***
	(d.f. = 1; 9,137)	(d.f. = 24; 9,114)	(d.f. = 89; 9,049)	(d.f. = 1; 9,137)	(d.f. = 24; 9,114)	(d.f. = 89; 9,049)

*$p < .05$, **$p < .01$, ***$p < .001$.
Notes: For tools that generate multiple text dimensions, we included the dimensions related to sentiment. This includes sentiment scores from the SenticNet and SentiWordNet; negative, positive, neutral, compound sentiment scores from VADER; arousal, dominance, and valence scores from the Arousal-Dominance-Valence lexicon; average valence, extremity, and emotionality scores from Evaluative Lexicon; WC, analytic, clout, authentic, tone, affect, posemo, negemo, and netspeak from LIWC; and hedonometer score from Hedonometer. Scores were generated using SenticNet 5 lexicon, SentiWordNet using Python package nltk.corpus.sentiwordnet, VADER components using Python package nltk.sentiment.vader, Arousal-Dominance-Valence lexicon, Evaluative Lexicon software, LIWC2015 software, and hedonometer using Python package pyhmeter.

regression coefficients into Pearson correlation coefficients (denoted as γ; Cohen 2013).

Effects of PARA on valence and intensity were robust and competitive compared to other text tools (Figure 3). The effect size of PARA on sentiment intensity ($\gamma = .207$, 95% confidence interval = [.187, .226]) outperformed the other comparison tools, whereas the effect size of PARA on sentiment valence ($\gamma = .054$, 95% confidence interval = [.034, .075]) was similar to the comparison tools although still statistically significant. (We also assessed incremental validity of PARA via 10-fold cross validation and found that the inclusion of TPL subcategories on top of the comparison tools significantly improved prediction accuracy; see Web Appendix C, Table W9).

Study 2 Discussion

Using a causal inference from text approach, we show that the identification of TPL influenced human ratings of sentiment valence and intensity. By identifying nonverbal parts of speech using PARA, the results suggest that PARA captures features of text that a human coder uses to determine sentiment. That is, we are tapping into elements that people actually use in making these assessments. This occurred despite a robust set of controls not only from the social media data (e.g., duration since posting) but also after controlling for word- and topic-based variables via topic modeling and 22 sentiment-related variables generated from existing tools. This allowed us to establish causal validity using a causal inference from text approach.

Further, there were heterogeneous effects in how the TPL subcategories mapped onto valence and intensity. For example, several auditory-based TPL features, including stress, emphasis, tempo, and rhythm convey how a message should be spoken and inherently give a text more vocal variation. Often these elements, added with intention, help amplify the intensity of a message but are not informative of the valence of a message because they can be applied to text content that is either positive or negative. Not surprisingly, we found significant positive effects of those TPL features on sentiment intensity but insignificant effects on valence. Study 2 also demonstrated incremental validity above and beyond other text analytic tools. We showed the value of PARA, even compared with tools such as VADER and LIWC, that capture aspects of both verbal and nonverbal parts of speech.

Study 3: Establishing Predictive Validity of PARA on Consumer Engagement

The purpose of Study 3 was to establish predictive validity by showing that PARA could predict outcomes relevant for marketers. From Study 2, we learned that the detection of TPL significantly affected human-perceived sentiment valence and intensity and did so relative to other text analytic tools. Prior research has shown that arousal levels related to activation are what motivates sharing behaviors (Berger and Milkman 2012). Logically, we would expect PARA to have predictive power on consumer engagement. We operationalized engagement as the number of likes (Twitter and Instagram) and

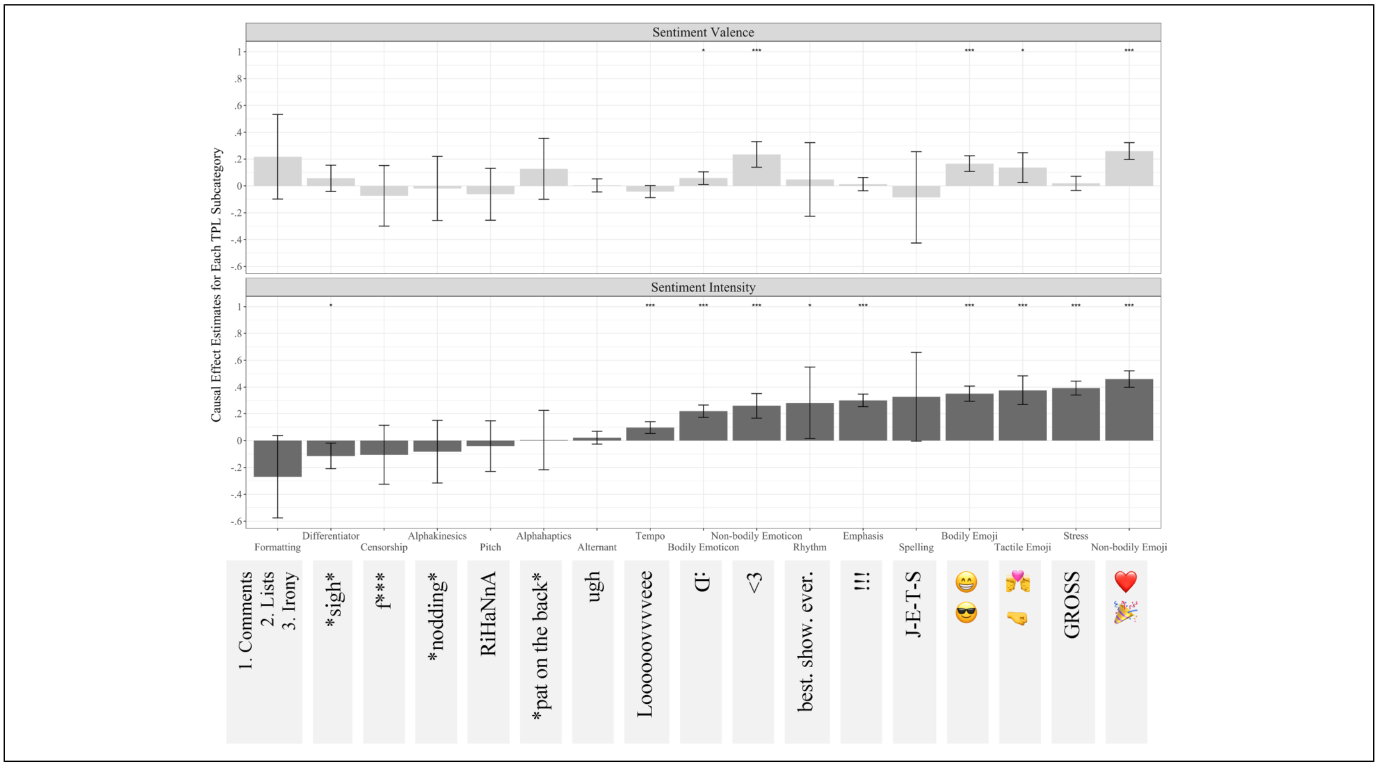

Figure 2. Heterogeneous Effects of TPL Subcategories on Valence and Intensity (Regression ii).
$*p < .05$, $**p < .01$, $***p < .001$.
Notes: Exemplars of the subcategories are derived from the data. Subcategories are ordered by effect on sentiment intensity. Estimates for two subcategories (volume and tactile emoticons) are not displayed as there were ten or fewer instances in the data. Error bars show 95% confidence interval.

retweets (Twitter only) of a post. In this study, the aim was to test the predictive power of PARA on consumer engagement and, specifically, (1) whether PARA predicted engagement relative to a baseline set of control variables, (2) the predictive value of TPL-derived sentiment, and (3) whether PARA provided additional predictive power on top of the same set of text tools and lexica as in Study 2.

Data and Method

Data for Study 3 were selected to ensure broad representation of both within-platform and across-platform content. Data included publicly available Twitter data associated with coronavirus-related content (N = 897,127; Smith 2020), publicly available Twitter data associated with Olympic content (N = 252,786; Preda 2021), and Instagram data from the top 50 social influencers[1] (N = 66,144). The Python package InstaLooter (Larralde 2021) was used to scrape Instagram posts.

For each outcome variable (i.e., likes/retweets), we conducted out-of-sample predictions based on feature sets: (1) Baseline (i.e., simple characteristics of the post and account, such as the number of followers, number of friends, year, month, weekday of the post, and topic- and word-based controls from topic modeling; see Web Appendix E for details), (2) Baseline + TPL-based Valence Scores, (3) Baseline + TPL-based Intensity Scores (i.e., the models of regression (ii) trained on the sentiment data in Study 2 generated valence and intensity scores for this study), (4) Baseline + PARA, (5) Baseline + Text Tools (i.e., 22 variables generated by the same set of text analytic tools as Study 2), (6) Baseline + Text Tools + PARA. In addition to feature sets 1–6, we conducted separate prediction analyses to assess the predictive power of PARA compared with each text tool independently. For both outcome variables (i.e., likes/retweets) and for each comparison text $tool_j$ (N = 7), we conducted an out-of-sample prediction based on (7) Baseline + Text $Tool_j$ and (8) Baseline + Text $Tool_j$ + PARA.

We used the following nested cross-validation procedure to carry out predictions: first, we randomly and evenly split the entire data set into ten groups of tweets; second, for each of the ten holdout groups, we trained a linear model to predict each of the outcome variables by fitting a Lasso regularized linear regression (Tibshirani 1996) to the remaining 90% of the data; and third, using that trained model, we conducted out-of-sample predictions for the remaining 10% of the data (i.e., the holdout group). Following procedures from previous work, we estimated the predictive accuracy by calculating the variance explained by predictive models based on cross-

[1] List of influencers: https://web.archive.org/web/20210517183940/https://en.wikipedia.org/wiki/List_of_most-followed_Instagram_accounts.

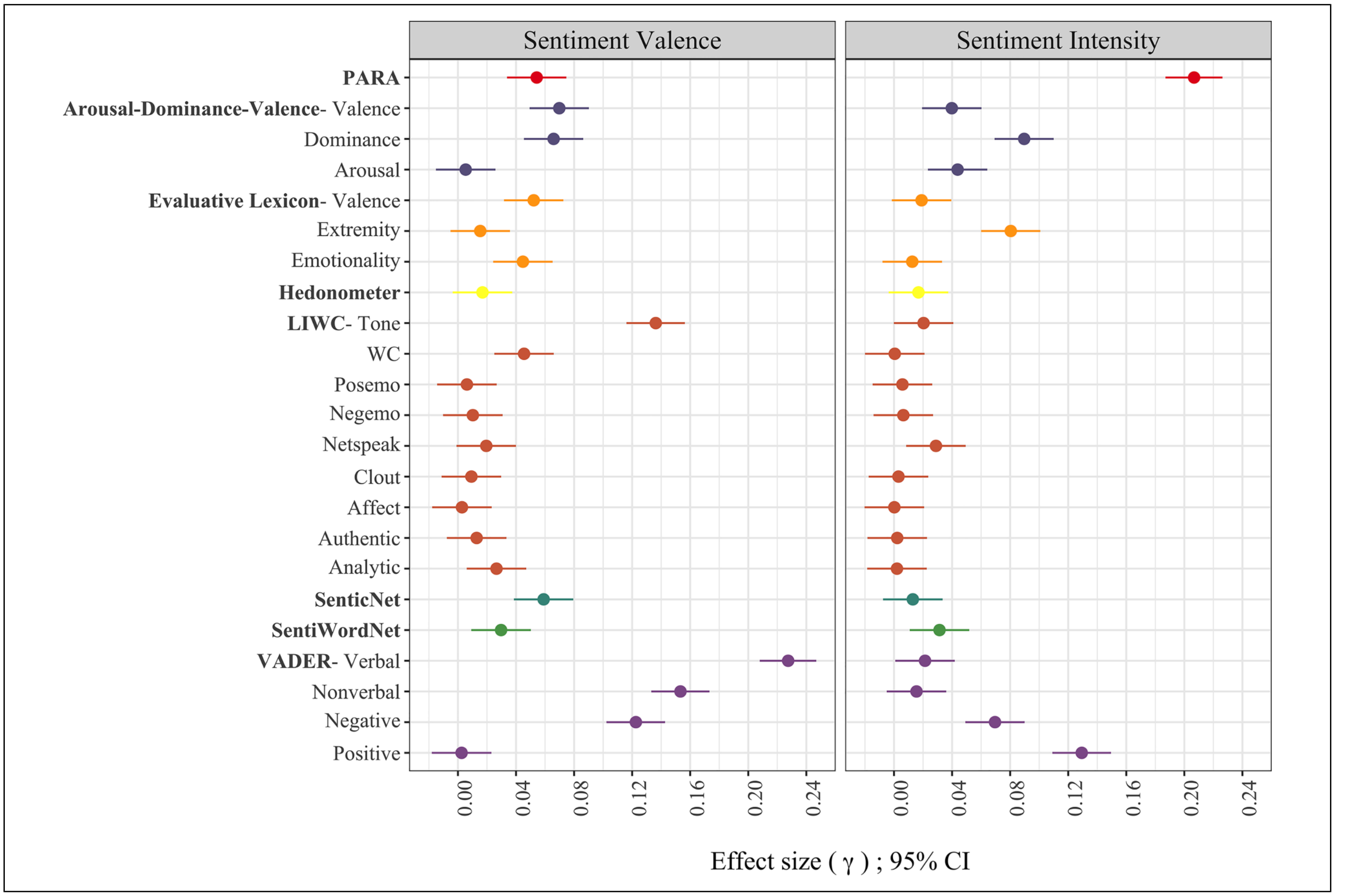

Figure 3. Relative Comparisons of PARA to Other Text Tools on Sentiment Valence and Intensity.

validation (VEcv; Li 2017), the Pearson correlation (γ) between the actual and predicted outcome variables (Nave et al. 2018), and the root mean square error (RMSE). As features sets were added to the analysis, we saw prediction improvement as demonstrated by higher values of VEcv and Pearson correlations (γ) and by lower values of RMSE.

Results

Prediction results of PARA relative to baseline. PARA identified instances of TPL in 304,255 (26.46%) Twitter posts and 37,287 (56.37%) Instagram posts. The prediction analysis revealed that the inclusion of PARA (feature set 4; Table 5) improved prediction accuracy across the outcome variables relative to the baseline (feature set 1). Using paired t-tests, key comparisons revealed that including PARA statistically significantly improved prediction relative to relying on the baseline features alone on Twitter likes (comparison of feature set 4 vs. 1, ΔVEcv = .269, SE = .054, t = 4.955, $p < .001$; Table 6), Twitter retweets (ΔVEcv = .040, SE = .009, t = 4.513, $p < .001$), and Instagram likes (ΔVEcv = .150, SE = .026, t = 5.677, $p < .001$). The inclusion of PARA consistently improved prediction accuracy across all performance metrics (see Table 5).

Prediction results of TPL-derived sentiment scores relative to baseline. Analyses revealed that TPL-derived sentiment valence significantly improved prediction relative to the baseline model on Twitter likes (comparison of feature set 2 vs. 1; ΔVEcv = .026, SE = .005, t = 5.603, $p < .001$; Table 6) and marginally significantly, with 6% confidence, on retweets (ΔVEcv = .009, SE = .004, t = 2.046, $p = .055$) but did not show significant improvement on Instagram likes (ΔVEcv = .000, SE = .003, t = .083, $p = .936$). We observed more improvement with the TPL-derived sentiment intensity score, which significantly improved prediction relative to the baseline on Twitter likes (comparison of feature set 3 vs. 1; ΔVEcv = .058, SE = .007, t = 7.965, $p < .001$) and retweets (ΔVEcv = .014, SE = .004, t = 3.271, $p = .004$) but not significantly on Instagram likes (ΔVEcv = .006, SE = .004, t = 1.635, $p = .137$). This suggests that TPL-derived sentiment, particularly sentiment intensity, contributes to the prediction of consumer engagement on Twitter.

Prediction results of PARA relative to other text tools. PARA significantly improved prediction relative to the joint set of text tools and baseline features on Twitter likes (comparison of feature set 6 vs. 5; ΔVEcv = .231, SE = .049, t = 4.706, $p < .001$; Table 6), retweets (ΔVEcv = .037, SE = .008, t = 4.329,

Table 5. Prediction Accuracy of PARA (Out of Sample) on Consumer Engagement.

	Performance Metric								
	VEcv			Pearson Correlation (γ)			RMSE		
Feature Set	Twitter Likes	Twitter Retweets	Instagram Likes	Twitter Likes	Twitter Retweets	Instagram Likes	Twitter Likes	Twitter Retweets	Instagram Likes
1. Baseline	4.458	1.395	51.237	20.907	11.901	71.619	14,899.557	97.351	873,826.691
2. Baseline + TPL valence scores	4.483	1.404	51.237	20.973	11.936	71.620	14,898.151	97.350	873,823.252
3. Baseline + TPL intensity scores	4.516	1.410	51.243	21.047	11.964	71.623	14,893.187	97.347	873,776.938
4. Baseline + PARA	4.727	1.435	51.387	21.490	12.071	71.725	14,860.726	97.330	872,474.616
5. Baseline + text tools	4.738	1.487	51.408	21.506	12.322	71.739	14,860.377	97.265	872,291.981
6. Baseline + text tools + PARA	4.969	1.523	51.564	21.986	12.468	71.848	14,826.126	97.249	870,884.399

Notes: Text tools include Arousal-Dominance-Valence Lexicon, Evaluative Lexicon, Hedonometer, LIWC, SenticNet, SenticWordNet, and VADER. TPL = textual paralanguage; RMSE = root mean square error.

$p < .001$), and Instagram likes (ΔVEcv = .156, SE = .026, t = 5.912, $p < .001$). The prediction improvement made by PARA was statistically significant and comparable to those made by the set of text tools. On Twitter likes, VEcv improved over the baseline with the inclusion of the text tools (comparison of feature set 5 vs. 1, ΔVEcv = .280) and improved further with the inclusion of PARA (comparison of feature set 6 vs. 5, ΔVEcv = .231). The same was true for both retweets (comparison of feature set 5 vs. 1, ΔVEcv = .092; comparison of feature set 6 vs. 5, ΔVEcv = .037) and Instagram likes (comparison of feature set 5 vs. 1, ΔVEcv = .171; comparison of feature set 6 vs. 5: ΔVEcv = .156).

We demonstrated TPL-associated prediction improvement by comparing Baseline + Text Tool_j + PARA relative to Baseline + Text Tool_j for each text tool independently (Figure 4). Across all text tools, the inclusion of PARA statistically significantly improved prediction. In Figure 4, zero represents no improvement, and adding PARA to the prediction analysis was positive and significant when added to each tool. For example, PARA showed the most improvement on top of Sentiwordnet on Twitter likes (ΔVEcv = .269, SE = .054, t = 4.948, $p < .001$) and the least improvement when included with LIWC on Twitter retweets (ΔVEcv = .030, SE = .008, t = 3.940, $p < .001$), though still significant. The improvement made by including PARA along with each text tool (the dark gray bars in Figure 4) was competitive with those made by each text tool alone relative to the baseline (the light gray bars in Figure 4). Adding PARA robustly improved prediction accuracy over and above the other comparison text tools (see Web Appendix E for details on other metrics).

Study 3 Discussion

These results address the three main aims of this study. First, the data reveal that PARA contributes significantly to the prediction of consumer engagement (i.e., likes and retweets) even after accounting for numerous post-specific factors and word- and topic-based controls. This study thereby establishes the predictive validity of PARA.

Second, the prediction results also suggest that TPL-derived sentiment contributes to the prediction of consumer engagement but does not fully account for the improvement attained by including PARA. The prediction accuracy of PARA improved over and above the TPL sentiment scores, and we rationalize that this may have occurred for two reasons. First, it could suggest that TPL serves functions other than affecting sentiment valence and intensity that have not been captured here. In other words, sentiment partially (but not fully) accounts for the effect of TPL on consumer engagement, such that it captures some of the variance, but there is variance left to explain. Second, the TPL-derived sentiment scores were generated from Twitter and YouTube data in Study 2 but applied to Twitter and Instagram data in Study 3. The finding that sentiment scores significantly improved prediction on the Twitter data but not the Instagram data could be indicative of domain shift, an issue that arises when the data are distributed differently between the training data set and the data on which a measure is used. In other words, it is possible that TPL could be used to communicate, or perceived to communicate, sentiment differently across platforms. For example, one may speculate that the sentiment derived from visual or graphical features of TPL (e.g., emojis) may be weakened on Instagram because they are competing in a more graphical platform (as opposed to a more textual platform on Twitter). Thus, the sentiment scores themselves could fall prey to issues of domain shift. The prediction results from PARA did not suffer from this same issue because the detection of the paralinguistic features was independent of the data.

Third, and importantly, the inclusion of PARA improved prediction relative to a robust set of text analytic tools that rely predominantly on word-based assessments. PARA extracted unique information from text that other existing tools did not detect, and these features improved prediction.

Table 6. Prediction Accuracy Comparisons on Consumer Engagement.

		Performance Metric											
		ΔVEcv				ΔPearson Correlation (γ)				ΔRMSE			
Comparisons	Outcome Variable	β	SE	t	p	β	SE	t	p	β	SE	t	p
PARA versus baseline, feature set 4 versus 1	Twitter likes	.269	.054	4.955	<.001	.583	.114	5.103	<.001	−38.830	9.387	−4.136	.001
	Twitter retweets	.040	.009	4.513	<.001	.171	.033	5.119	<.001	−.021	.006	−3.508	.002
	Instagram likes	.150	.026	5.677	<.001	.105	.019	5.481	<.001	−1,352.075	244.858	−5.522	<.001
TPL valence scores versus baseline, feature set 2 versus 1	Twitter likes	.026	.005	5.603	<.001	.065	.013	5.094	<.001	−1.406	.336	−4.187	<.001
	Twitter retweets	.009	.004	2.046	.055	.035	.013	2.680	.015	−.001	.000	−1.887	.074
	Instagram likes	.000	.003	.083	.936	.000	.002	.149	.885	−3.439	22.561	−.152	.882
TPL intensity scores versus baseline, feature set 3 versus 1	Twitter likes	.058	.007	7.965	<.001	.140	.015	9.318	<.001	−6.369	1.501	−4.244	<.001
	Twitter retweets	.014	.004	3.271	.004	.063	.014	4.527	<.001	−.004	.001	−3.347	.003
	Instagram likes	.006	.004	1.635	.137	.004	.002	1.529	.161	−49.753	30.630	−1.624	.139
Text tools versus baseline, feature set 5 versus 1	Twitter likes	.280	.052	5.337	<.001	.599	.106	5.645	<.001	−39.179	9.348	−4.191	<.001
	Twitter retweets	.092	.023	4.073	.001	.421	.103	4.075	.001	−.086	.020	−4.244	<.001
	Instagram likes	.171	.025	6.759	<.001	.120	.017	6.842	<.001	−1,534.710	229.220	−6.695	<.001
PARA versus text tools and baseline, feature set 6 versus 5	Twitter likes	.231	.049	4.706	<.001	.480	.100	4.791	<.001	−34.251	8.316	−4.119	.001
	Twitter retweets	.037	.008	4.329	<.001	.146	.029	5.053	<.001	−.016	.005	−3.349	.003
	Instagram likes	.156	.026	5.912	<.001	.109	.019	5.728	<.001	−1,407.582	244.885	−5.748	<.001

Notes: Comparisons are mean differences of the feature sets from Table 5 representing prediction improvement. The t-value is calculated using a paired t-test.

PARA could be used as a stand-alone text analytic tool to detect paralinguistic features, or it could be used in tandem with existing text tools to conduct more comprehensive text-based prediction analyses.

General Discussion

Bearden, Netemeyer, and Haws (2011, p. 1) note that "measurement is at the heart of virtually all scientific endeavors"; indeed, researchers have long acknowledged the need for sound measurement of constructs (Churchill 1979). Answering this call, we develop a tool to detect nonverbal parts of speech to facilitate the discovery of insights from text content. Prior research has developed a conceptual model of textual paralanguage (Luangrath, Peck and Barger 2017), and the current work translates this model into an automatic paralanguage classification tool called PARA. We further answer the call to develop more theoretically grounded text analytic tools (Humphreys and Wang 2018).

The study of TPL (Luangrath, Peck, and Barger 2017) is situated alongside a rapidly growing body of linguistic work in marketing (Berger, Rocklage, and Packard 2021; Berman et al. 2019; Chen 2017; Chakraborty, Kim, and Sudhir 2022; Lee and Kronrod 2020; Packard and Berger 2021; Zemack-Rugar, Moore, and Fitzsimons 2017). From text mining to understanding factors driving online posting behavior (Moe and Schweidel 2012; Toubia and Stephen 2013) to discourse analysis of institutional forces shaping the legitimation of consumption practices (Humphreys 2010), scholars have taken a variety of approaches to the study of language and text. This research contributes to a growing body of work within marketing to "unite the tribes" of those studying linguistic phenomena (Berger et al. 2020).

In contrast to tools that are predominantly word-based (e.g., Hovy, Melumad, and Inman 2021), PARA identifies nonverbal features. Analogous to properties of speech (e.g., nouns, verbs, prepositions) in the verbal content, we categorize the "properties of nonverbal speech" denoted in text. Ours is certainly not the first inquiry into extratextual elements (Das, Wiener, and Kareklas 2019; Eisner et al. 2016, Felbo et al. 2017). Prior research has shown that emoticons expressing positive emotions correlate negatively with stress levels (Settanni and Marengo 2015) and affect perceptions of online service encounters (Li, Chan, and Kim 2019). Research on product reviews finds that textual characteristics such as word count, average word length, occurrences of exclamation marks, and customer ratings predict whether a review is fake (Anderson and Simester 2014). However, most research that has considered extratextual features has focused on one or a few elements, and no work to our knowledge has taken a comprehensive approach to the automatic classification of textual paralanguage.

To do so, we employed a supervised learning approach in which humans coded social media content for the presence of TPL. We developed dictionaries and rule-based algorithms to automatically detect nonverbal communication in text. In contrast to an inductive approach, which may use latent Dirichlet allocation to generate topic discovery (Blei, Ng, and Jordan 2003; Netzer, Lemaire, and Herzenstein 2019), we used an existing framework of textual paralanguage to generate this classifier (Luangrath, Peck, and Barger 2017). We validated PARA by demonstrating construct, causal, incremental, and predictive validity. By detecting a robust set of nonverbal cues in text, PARA aids in the assessment of sentiment valence and

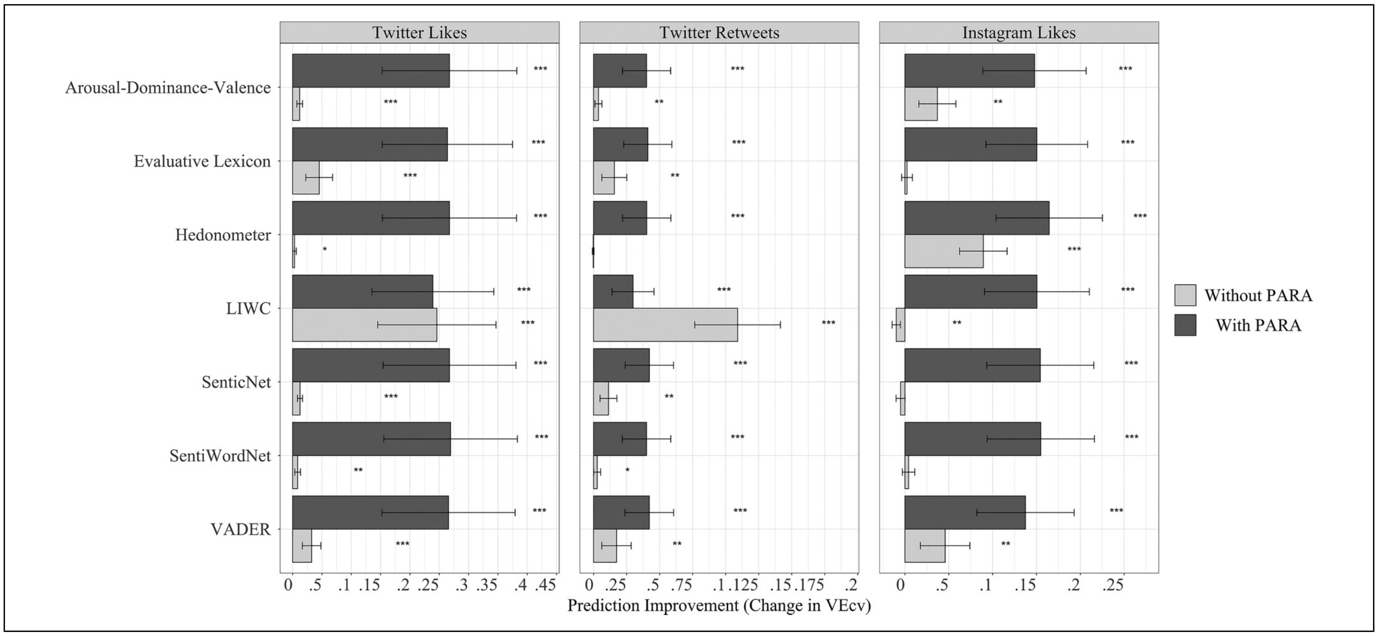

Figure 4. Consumer Engagement Prediction Accuracy Improvement with the Inclusion of PARA (Measured by Change in VEcv).
*$p < .05$, **$p < .01$, ***$p < .001$.
Notes: The light and dark gray bars represent mean differences in prediction accuracy improvement. The stars represent whether the prediction improvement is significantly greater than zero. Error bars represent the 95% confidence interval. Without PARA is the comparison Text Tool$_j$ vs. Baseline (or feature sets [7] vs. [1]). With PARA is the comparison PARA versus Text Tool$_j$ and Baseline (or feature sets [8] vs. [7]). The t-value is calculated using a paired t-test.

intensity and improves prediction accuracy on consumer engagement with social media content. Indeed, comparative analyses of text classification methods are commonplace (Hartmann et al. 2019), and we demonstrate that PARA has predictive power above and beyond both the nonverbal and verbal components of existing tools. Whether in isolation or in conjunction with other text tools, PARA enriches the study of text analytics through the detection of nonverbal parts of speech.

PARA is well-equipped to process social media data, as we have tested here, but it can also be applied to other sources of text data such as email, messaging, customer service chats, and text created in apps. Text content generated via mobile device is a likely modality by which TPL occurs, as individuals have easy access to emoji keyboards and are known to generate briefer, emotional, and more “gist-based” content on mobiles as opposed to computers (Melumad, Inman, and Pham 2019).

Theoretical Applications of PARA

Our work has implications that apply to a variety of theoretical areas within marketing. Given the proliferation of ways that nonverbal communication can be expressed in text-based messages, opportunities abound for theory-driven hypotheses and research questions (Table 7; additional applications in Web Appendix F). For example, studies of nonverbal communication often isolate specific nonverbal features of language (e.g., pitch). By identifying nonverbal features, PARA provides ample opportunity to form and test paralinguistic hypotheses within specific nonverbal domains. In other words, this algorithm breaks down nonverbal communication cues into their linguistically based categories, which then can be tested independently.

We foresee two main routes for theoretical application of PARA, including extensions from (1) in-person nonverbal communication theory and (2) text-based linguistic theory. First, research on in-person nonverbal communication documents interpersonal interactions in face-to-face contexts that have the potential to be translated into text. For example, Schroeder and Epley (2016) show that paralinguistic cues in human voice convey mental capacities. Future work could consider the theoretical applicability of a similar question with paralinguistic cues expressed in text. For example, do certain TPL features (perhaps those that are bodily) more readily convey humanlike minds? Thus, prior work coupled with PARA could give rise to new theoretical extensions.

Second, applications of nonverbal linguistic phenomena could also emerge from text-based linguistic research. For example, Grinstein and Kronrod (2016) demonstrate that urging consumers with an assertive tone (e.g., “You are doing a lot for your health. You must do more!” vs. “You are doing a lot for your health. You can do more!”) is effective in encouraging desired behaviors (e.g., handwashing, financial retirement planning), because the assertive tone intensifies the message. While extant studies explore these theoretical ideas in the context of text-based *language*, extensions of this literature stream could be viewed through the lens of text-based *paralanguage*. Comparisons of assertiveness and sentiment intensity between verbal and nonverbal components could be considered.

Table 7. Potential Applications of PARA to Paralinguistic Study in Consumer Contexts.

Linguistic Modality	Article	Main Finding	Sensory Feature	TPL Feature	Potential Application of PARA
In-person nonverbal communications	Chattopadhyay et al. (2003)	• Voices with faster-than-normal syllable speed and low pitch produce less negative advertisement-directed cognitive responses and more favorable ad/brand attitudes.	Auditory	Tempo/pitch	Are ad attitudes driven by TPL indications of pitch? Syllable speed could be operationalized with TPL by elongating certain vowels over others.
In-person nonverbal communications	Lowe and Haws (2017)	• Acoustic pitch influences consumers' interpretation of product size such that lower pitches increase size perceptions.	Auditory	Pitch	Pitch alterations via fluctuations in capitalizations or elongation mapping onto size perceptions.
In-person nonverbal communications	Luangrath, Peck, and Gustafsson (2020)	• Interpersonal touch initiators fear imposing intimacy on recipients of touch and underestimate how positively it will be received by the recipient.	Tactile	Tactile emoji/ alphahaptics	Is there similar trepidation around using tactile kinesics? Perhaps there are forecasting errors in how senders think tactile TPL will be perceived by recipients.
In-person nonverbal communications	Schroeder and Epley (2016)	• Human voices naturally convey humanlike mind with paralinguistic qualities (e.g., pace/ tempo). Yet, adding human visual cues to text (i.e., seeing a person perform in a subtitled video clip) did not increase inferences of human creator.	Auditory	Tempo/ differentiators	Perhaps human visual cues (e.g., bodily emojis) are not as effective at conveying humanlike minds than other auditory TPL indicators (e.g., haha, *yawn*). Opportunity to identify TPL features that convey basic mental capacities.
In-person nonverbal communications	Van Zant and Berger (2020)	• Louder voices signal dominance.	Auditory	Volume/stress	Status and dominance inferences from text. Do volume and stress conveyed via text communicate confidence and generate more persuasive advertising content?
Text-based language	Grinstein and Kronrod (2016)	• When marketers praise (scold) consumers, an assertive tone is more (less) effective because of its intensity, influencing socially responsible actions such as washing hands and financial decision making.	Auditory	Stress/emphasis	Test this theory by identifying nonverbals from text using the TPL features particularly adept at intensifying messages (e.g., Stress).
Text-based language	Humphreys, Isaac, and Wang (2020)	• Quantify consumers' construal levels from abstract (concrete) language in search queries, marketers can increase engagement with search-based advertising by matching ad content with consumers' mindsets.	Tactile	Tactile emojis/ emoticons/ alphahaptics	TPL could be investigated as another feature to match construal mindsets. Certain visual features (e.g., emojis) could prompt concrete construals while verbal processing of TPL could facilitate abstract construals.
Text-based language	Lafreniere, Moore, and Fisher (2022)	• Consumers judge Yelp reviews with swearwords as more useful. Swearwords increase the intensity of product attributes and of the reviewer's feelings.	Auditory	Censorship	How does censorship of swearwords (e.g., S#!T) affect perceived usefulness of reviews? Consider comparisons of sentiment intensity between actual and censored swearwords.

(continued)

Table 7. (continued)

Linguistic Modality	Article	Main Finding	Sensory Feature	TPL Feature	Potential Application of PARA
Text-based language	Lee and Kronrod (2020)	• In consumer WOM, consensus language is more persuasive for weak ties than strong ties.	Tactile	Tactile emojis	Perhaps tactile kinesics are less effective when coupled with consensus language if it communicates relational closeness thus facilitating strong (over weak) ties.
Text-based language	Li, Chan, and Kim (2019)	• Customers perceive online chat employees who use emoticons as warmer but less competent, affecting service satisfaction.	Visual	Bodily emojis/ emoticons	Testing warmth and competence ratings at scale by comparing resultant satisfaction ratings from bodily emoticons vs. emojis.
Text-based language	Ludwig et al. (2013)	• Affective content and linguistic style matching in online reviews jointly increase conversion rates. Conversion rates taper off for increases in positive (but not negative) affect.	TPL index	All	Use TPL features to quantify the nonverbal drivers of affective content and linguistic style matching in online reviews to examine effects on conversion rates.
Text-based language	Moore and McFerran (2017)	• Consumers mimic linguistic content in online WOM. Same genders mimic positive emotion and social word use; similar status individuals mimic cognitive and descriptive word use.	Visual	All	Do consumers mimic positive and negative emojis in consumer online WOM similar to word use?
Text-based language	Packard and Berger (2020)	• Customers are more satisfied, willing to purchase, and purchase more when employees speak to them concretely	Voice qualities	Alternants	Utterances (e.g., "um," "uh," "hmm") could dilute perceptions of concreteness, negatively affecting customer satisfaction.
Text-based language	Rocklage and Fazio (2020)	• Emotionality of Amazon reviews is expected, and thus viewed favorably, for hedonic products but not utilitarian products.	Visual	Emoji	Use nonverbal features as controls in statistical analyses (e.g., control for emoji use when testing effects of emotionality).

Notes: WOM = word of mouth. For articles on in-person nonverbals, paralinguistic effects have been demonstrated in interpersonal interactions (i.e., face-to-face) but have the potential to be examined in text-based contexts using PARA. Text-based language articles investigate linguistic effects by studying written communication. In all cases, extensions are plentiful by shifting an orientation toward TPL, or nonverbal cues in text.

We enumerate ways in which PARA provides fodder for future theoretical application and inquiry (Table 7).

Managerial Implications

Text analytics firms and practitioners often eliminate what they consider "extraneous" TPL features from their data, as these features have traditionally been viewed as irrelevant noise. However, our study shows that the linguistic phenomenon of TPL is relevant to marketing managers in both content creation and consumer listening. Whether replying to customers on Twitter, managing a brand's Facebook page, or composing an email to notify customers of a data breach, the way something is written affects how consumers interpret and interact with content. Managers monitoring customer feedback via text could benefit from this tool. Being able to recognize how inclusion of nonverbal cues changes sentiment valence and intensity can indicate when rhetoric is heating up or cooling down. Messages flagged as high in intensity could be given prioritized response times.

The way in which a brand communicates affects the level of engagement with content. Consumer engagement with brands on social media is considered both an outcome of online communication and an indicator of performance (Hollebeek, Glynn, and Brodie 2014) and carries the potential to impact brand awareness, loyalty, and brand personality (Barger, Peltier, and Schultz 2016). It can also affect the organic reach of branded content on social media because social media platforms often prioritize content with high engagement rates (Escobedo 2017).

In recent years, online industries have emerged around influencer marketing. Vetting of influencers is something that brands should do before forming partnerships. This vetting process could use PARA to understand how an influencer communicates online as well as give recommendations to influencers as to how they should be communicating with consumers. In a similar sense, TPL is a facet of one's overall voice online and could be considered in making hiring decisions for those who manage brands' social media accounts or for those in customer-facing roles such as online customer service representatives. The applications of this analytical tool could be viewed as a support system to facilitate managerial decisions (Chica and Rand 2017), thus encouraging the dissemination of theoretically grounded research in practice.

Limitations

We acknowledge some limitations with PARA and this research more generally. First, our goal has been to produce a paralinguistic tool that is as comprehensive as possible, and to do so, we survey the domain broadly to capture the varied ways in which TPL is revealed in text. To this end, the classifier is built on a scaffolding rooted at the intersection of sensory marketing and linguistics, identifying the nature of the paralinguistic element from a sensory perspective. Thus, TPL is based more in its sensory nature than in the higher-order functions it serves, and it is likely that certain TPL features are better than others at specific functions (e.g., sarcasm, playfulness). This work does not speak to these higher-order functions. Future work could investigate specific TPL features and map them onto an array of important linguistic functions.

Second, the subcategories of TPL produce heterogeneous effects, as Study 2 demonstrates. While this finding is logical and expected across many different features of nonverbal cues, it does present difficulties with making aggregate claims regarding overall TPL effects. The effectiveness of specific features of TPL could vary across topic, context, and so on, and future research should explore the boundary conditions across and within features of TPL.

Third, we demonstrate that the sentiment derived from paralanguage significantly predicts consumer engagement with social media posts. However, sentiment does not fully account for the prediction improvement of PARA. The data-driven approach to the construction of the sentiment scores in Study 3 could have introduced an issue of domain shift as they were generated on one social media platform but then applied to another. More information is needed to mitigate the distributional differences of the data across platforms. Finally, PARA operates on identification of nonverbals rooted in the English language. Certain paralinguistic features (e.g., emojis) are more standard across languages, while other alphabetic features (e.g., utterances) may be more language specific. While addressing this limitation is beyond the scope of this current work, we see ample opportunity for future work to generate non-English TPL dictionaries.

Conclusion

Textual data are ubiquitous. The intent of this work is to give structure to unstructured data and make the study of textual paralanguage accessible to researchers and practitioners with the development of the PARA text analytic tool. Studies of linguistic elements and communication styles are becoming increasingly important to investigate marketing-relevant outcomes (Berger et al. 2020). We contribute to the field by mapping a real-world phenomenon to a construct (MacInnis et al. 2020), and we do so by developing a methodological text tool to detect language that moves beyond *what* is said verbally to *how* it is said nonverbally.

Acknowledgments

The authors thank Neeraj Arora, Victor Barger, Robby Cortes, Sarah Francisco, Joann Peck, Bryce Pyrah, Naveen Ramaraju, Min Tian, participants in the University of Iowa behavioral research seminar, three anonymous reviewers, and the *JMR* editorial team for their suggestions and feedback. The authors gratefully acknowledge the University of Iowa and the Tippie College of Business Research Excellence Grant for funding this work.

Associate Editor

Jonah Berger

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Abraham, Ajith (2005), "Rule-Based Expert Systems," in Handbook of Measuring System Design, Peter H. Sydenham and Richard Thorn, eds. New York: John Wiley & Sons, 909–19.

Akkaya, Cem, Janyce Wiebe, and Rada Mihalcea (2009), "Subjectivity Word Sense Disambiguation," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA: Association for Computational Linguistics, 190–99.

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

Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani (2010), "Sentiwordnet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," *Lrec*, 10, 2200–2204.

Badenhausen, Kurt (2018), "The World's Most Valuable Brands 2018," *Forbes* (May 23), https://www.forbes.com/sites/kurtbadenhausen/2018/05/23/the-worlds-most-valuable-brands-2018/#3759476610c1.

Balducci, Bitty and Detelina Marinova (2018), "Unstructured Data in Marketing," *Journal of the Academy of Marketing Science*, 46 (4), 557–90.

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

Barger, Victor, James W. Peltier, and Don E. Schultz (2016), "Social Media and Consumer Engagement: A Review and Research

Agenda," *Journal of Research in Interactive Marketing*, 10 (4), 268–87.

Barrett, Lisa F. (2017), "The Theory of Constructed Emotion: An Active Inference Account of Interoception and Categorization," *Social Cognitive and Affective Neuroscience*, 12 (1), 1–23.

Bearden, William O., Robert G. Netemeyer, and Kelly L. Haws (2011), "Values and Goals," in *Handbook of Marketing Scales: Multi-Item Measures for Marketing and Consumer Behavior Research*. Thousand Oaks, CA: SAGE Publications, 151–236.

Berger, Jonah (2014), "Word of Mouth and Interpersonal Communication: A Review and Directions for Future Research," *Journal of Consumer Psychology*, 24 (4), 586–607.

Berger, Jonah, Ashlee Humphreys, Stephan Ludwig, Wendy W. Moe, Oded Netzer, and David A. Schweidel (2020), "Uniting the Tribes: Using Text for Marketing Insight," *Journal of Marketing*, 84 (1), 1–25.

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

Berger, Jonah, Matthew D. Rocklage, and Grant Packard (2021), "Expression Modalities: How Speaking Versus Writing Shape Word of Mouth," *Journal of Consumer Research*, 49 (2), 398–408.

Berman, Ron, Shiri Melumad, Colman Humphrey, and Robert Meyer (2019), "A Tale of Two Twitterspheres: Political Microblogging During and After the 2016 Primary and Presidential Debates," *Journal of Marketing Research*, 56 (6), 895–917.

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

Brodie, Roderick J., Ana Ilic, Biljana Juric, and Linda Hollebeek (2013), "Consumer Engagement in a Virtual Brand Community: An Exploratory Analysis," *Journal of Business Research*, 66 (1), 105–14.

Calder, Bobby J., Edward C. Malthouse, and Ewa Maslowska (2016), "Brand Marketing, Big Data and Social Innovation as Future Research Directions for Engagement," *Journal of Marketing Management*, 32 (5/6), 579–85.

Cambria, Erik, Soujanya Poria, Devamanyu Hazarika, and Kenneth Kwok (2018), "SenticNet 5: Discovering Conceptual Primitives for Sentiment Analysis by Means of Context Embeddings," in *Thirty-Second AAAI Conference on Artificial Intelligence*. Menlo Park, CA: Association for the Advancement of Artificial Intelligence.

Chakraborty, Ishita, Minkyung Kim, and K. Sudhir (2022), "Attribute Sentiment Scoring with Online Text Reviews: Accounting for Language Structure and Missing Attributes," *Journal of Marketing Research*, 59 (3), 600–622.

Chattopadhyay, Amitava, Darren W. Dahl, Robin J.B. Ritchie, and Kimary N. Shahin (2003), "Hearing Voices: The Impact of Announcer Speech Characteristics on Consumer Response to Broadcast Advertising," *Journal of Consumer Psychology*, 13 (3), 198–204.

Chen, Zoey (2017), "Social Acceptance and Word of Mouth: How the Motive to Belong Leads to Divergent WOM with Strangers and Friends," *Journal of Consumer Research*, 44 (3), 613–32.

Chen, Zoey and Jonah Berger (2016), "How Content Acquisition Method Affects Word of Mouth," *Journal of Consumer Research*, 43 (1), 86–102.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (3), 345–54.

Chica, Manuel and William Rand (2017), "Building Agent-Based Decision Support Systems for Word-of-Mouth Programs: A Freemium Application," *Journal of Marketing Research*, 54 (5), 752–67.

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

Clancey, William J. (1983), "The Epistemology of a Rule-Based Expert System—A Framework for Explanation," *Artificial Intelligence*, 20 (3), 215–51.

Cohen, Jacob (2013), *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.

Das, Gopal, Hillary J.D. Wiener, and Ioannis Kareklas (2019), "To Emoji or Not to Emoji? Examining the Influence of Emoji on Consumer Reactions to Advertising," *Journal of Business Research*, 96, 147–56.

Dodds, Peter Sheridan, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss, and Christopher M. Danforth (2011), "Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter," *PLOS One*, 6 (12), 1–26.

Eisner, Ben, Tim Rocktäschel, Isabelle Augenstein, Matko Bošnjak, and Sebastian Riedel (2016), "Emoji2vec: Learning Emoji Representations from Their Description," in *Proceedings of the Fourth International Workshop on Natural Language Processing for Social Media*. Stroudsburg, PA: Association for Computational Linguistics, 48–54.

Emotient (2015), *Emotient^{TM} Native SDK User Guide*. San Diego: Emotient Incorporated.

Escobedo, Joe (2017), "How to Sell the Importance of Engagement over Fan Growth on Social Media," *Forbes* (June 12), https://www.forbes.com/sites/joeescobedo/2017/06/12/social-media-engagement/#6f1721e82193.

Feder, Amir, Katherine A. Keith, Emaad Manzoor, Reid Pryzant, Dhanya Sridhar,, Zach Wood-Doughty, et al. (2021), "Causal Inference in Natural Language Processing: Estimation, Prediction, Interpretation and Beyond," arXiv preprint arXiv:2109.00725.

Felbo, Bjarke, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann (2017), "Using Millions of Emoji Occurrences to Learn Any-Domain Representations for Detecting Sentiment, Emotion and Sarcasm," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA: Association for Computational Linguistics, 1615–25.

Gandomi, Amir and Murtaza Haider (2015), "Beyond the Hype: Big Data Concepts, Methods, and Analytics," *International Journal of Information Management*, 35 (2), 137–44.

Gray, Tyler J., Christopher M. Danforth, and Peter Sheridan Dodds (2020), "Hahahahaha, Duuuuude, Yeeessss!: A Two-Parameter Characterization of Stretchable Words and the Dynamics of Mistypings and Misspellings," *PLOS One*, 15 (5), e0232938.

Grinstein, Amir and Ann Kronrod (2016), "Does Sparing the Rod Spoil the Child? How Praising, Scolding, and an Assertive Tone Can Encourage Desired Behaviors," *Journal of Marketing Research*, 53 (3), 433–41.

Grossman, Samantha (2014), "The 13 Sassiest Brands on Twitter," *Time* (February 7), http://time.com/5151/sassiest-brands-on-Twitter-ranked/.

Hartmann, Jochen, Juliana Huppertz, Christina Schamp, and Mark Heitmann (2019), "Comparing Automated Text Classification Methods," *International Journal of Research in Marketing*, 36 (1), 20–38.

Haynes, Stephen N. and Heather C. Lench (2003), "Incremental Validity of New Clinical Assessment Measures," *Psychological Assessment*, 15 (4), 456–66.

Hollebeek, Linda D., Mark S. Glynn, and Roderick J. Brodie (2014), "Consumer Brand Engagement in Social Media: Conceptualization, Scale Development and Validation," *Journal of Interactive Marketing*, 28 (2), 149–65.

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

Hovy, Dirk, Shiri Melumad, and J. Jeffrey Inman (2021), "Wordify: A Tool for Discovering and Differentiating Consumer Vocabularies," *Journal of Consumer Research*, 48 (3), 394–414.

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

Humphreys, Ashlee, Mathew S. Isaac, and Rebecca Jen-Hui Wang (2020), "Construal Matching in Online Search: Applying Text Analysis to Illuminate the Consumer Decision Journey," *Journal of Marketing Research*, 58 (6), 1101–19.

Humphreys, Ashlee and Rebecca Jen-Hui Wang (2018), "Automated Text Analysis for Consumer Research," *Journal of Consumer Research*, 44 (6), 1274–1306.

Hutto, Eric and C.J. Gilbert (2014), "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," in *Eighth International Conference on Weblogs and Social Media* (ICWSM-14). Palo Alto, CA: Association for the Advancement of Artificial Intelligence.

Juslin, Patrik N. and Petri Laukka (2001), "Impact of Intended Emotion Intensity on Cue Utilization and Decoding Accuracy in Vocal Expression of Emotion," *Emotion*, 1 (4), 381–412.

Klostermann, Philip (2019), "YouTube Comment Scraper," http://ytcomments.klostermann.ca.

Knapp, Mark L., Judith A. Hall, and Terrence G. Horgan (2014), *Nonverbal Communication in Human Interaction*, 8th ed. Boston: Wadsworth.

Lafreniere, Katherine C., Sarah G. Moore, and Robert J. Fisher (2022), "The Power of Profanity: The Meaning and Impact of Swear Words in Word of Mouth," *Journal of Marketing Research*, 59 (5), 908–25.

Larralde, Martin (2021), "InstaLooter (Version 2.4.4)," Python package (accessed December 16, 2021), https://instalooter.readthedocs.io/en/latest/usage.html.

Lee, Jeffrey K. and Ann Kronrod (2020), "The Strength of Weak-Tie Consensus Language," *Journal of Marketing Research*, 57 (2), 353–74.

Li, Jin (2017), "Assessing the Accuracy of Predictive Models for Numerical Data: Not R nor R^2, Why Not? Then What?" *PLOS One*, 12 (8), e0183250.

Li, Xueni (Shirley), Kimmy Wa Chan, and Sara Kim (2019), "Service with Emoticons: How Customers Interpret Employee Use of Emoticons in Online Service Encounters," *Journal of Consumer Research*, 45 (5), 973–87.

Liu, Bing (2012), "Sentiment Analysis and Opinion Mining," *Synthesis Lectures on Human Language Technologies*, 5 (1), 1–167.

Lovins, Julie Beth (1968), "Development of a Stemming Algorithm," *Mechanical Translation and Computational Linguistics*, 11 (1/2), 22–31.

Lowe, Michael L. and Kelly L. Haws (2017), "Sounds Big: The Effects of Acoustic Pitch on Product Perceptions," *Journal of Marketing Research*, 54 (2), 331–46.

Luangrath, Andrea Webb, Joann Peck, and Victor A. Barger (2017), "Textual Paralanguage and Its Implications for Marketing Communications," *Journal of Consumer Psychology*, 27 (1), 98–107.

Luangrath, Andrea Webb, Joann Peck, and Anders Gustafsson (2020), "Should I Touch the Customer? Rethinking Interpersonal Touch Effects from the Perspective of the Touch Initiator," *Journal of Consumer Research*, 47 (4), 588–607.

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

MacInnis, Deborah J., Vicki G. Morwitz, Simona Botti, Donna L. Hoffman, Robert V. Kozinets,, Donald R. Lehmann, et al. (2020), "Creating Boundary-Breaking, Marketing-Relevant Consumer Research," *Journal of Marketing*, 84 (2), 1–23.

Manzoor, Emaad, George H. Chen, Dokyun Lee, and Michael D. Smith (2020), "Influence via Ethos: On the Persuasive Power of Reputation in Deliberation Online," arXiv preprint arXiv:2006.00707.

Melumad, Shiri, J. Jeffrey Inman, and Michel Tuan Pham (2019), "Selectively Emotional: How Smartphone Use Changes User-Generated Content," *Journal of Marketing Research*, 56 (2), 259–75.

Miller, Hannah, Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson, Loren Terveen, and Brent Hecht (2016), "'Blissfully Happy' or 'Ready to Fight': Varying Interpretations of Emoji," in *Proceedings of ICWSM*. Palo Alto, CA: Association for the Advancement of Artificial Intelligence.

Mimno, David and Moontae Lee (2014), "Low-Dimensional Embeddings for Interpretable Anchor-Based Topic Inference," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Stroudsburg, PA: Association for Computational Linguistics, 1319–28.

Moe, Wendy W. and David A. Schweidel (2012), "Online Product Opinions: Incidence, Evaluation, and Evolution," *Marketing Science*, 31 (3), 372–86.

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

Moore, Sarah G. and Brent McFerran (2017), "She Said, She Said: Differential Interpersonal Similarities Predict Unique Linguistic Mimicry in Online Word of Mouth," *Journal of the Association for Consumer Research*, 2 (2), 229–45.

Nave, Gideon, Juri Minxha, David M. Greenberg, Michal Kosinski, David Stillwell, and Jason Rentfrow (2018), "Musical Preferences Predict Personality: Evidence from Active Listening and Facebook Likes," *Psychological Science*, 29 (7), 1145–58.

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

Netzer, Oded, Alain Lemaire, and Michal Herzenstein (2019), "When Words Sweat: Identifying Signals for Loan Default in the Text of Loan Applications," *Journal of Marketing Research*, 56 (6), 960–80.

Packard, Grant and Jonah Berger (2017), "How Language Shapes Word of Mouth's Impact," *Journal of Marketing Research*, 54 (4), 572–88.

Packard, Grant and Jonah Berger (2020), "Thinking of You: How Second-Person Pronouns Shape Cultural Success," *Psychological Science*, 31 (4), 397–407.

Packard, Grant and Jonah Berger (2021), "How Concrete Language Shapes Customer Satisfaction," *Journal of Consumer Research*, 47 (5), 787–806.

Packard, Grant, Sarah G. Moore, and Brent McFerran (2018), "(I'm) Happy to Help (You): The Impact of Personal Pronoun Use in Customer–Firm Interactions," *Journal of Marketing Research*, 55 (4), 541–55.

Pang, Bo and Lillian Lee (2008), "Opinion Mining and Sentiment Analysis," *Foundations and Trends® in Information Retrieval*, 2 (1–2), 1–135.

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

Pennebaker, James W., Roger J. Booth, Ryan L. Boyd, and Martha E. Francis (2015), *Linguistic Inquiry and Word Count: LIWC2015*. Austin, TX: Pennebaker Conglomerates.

Poyatos, Fernando (2002), *Nonverbal Communication Across Disciplines: Paralanguage, Kinesics, Silence, Personal and Environmental Interaction*. Philadelphia: John Benjamins Publishing.

Preda, Gabriel (2021), "Tokyo Olympics 2020 Tweets—July," Kaggle, https://www.kaggle.com/gpreda/tokyo-olympics-2020-tweets.

Roberts, Margaret, Brandon Stewart, and Richard Nielsen (2020), "Adjusting for Confounding with Text Matching," *American Journal of Political Science*, 64 (4), 887–903.

Rocklage, Matthew and Russell Fazio (2020), "The Enhancing Versus Backfiring Effects of Positive Emotion in Consumer Reviews," *Journal of Marketing Research*, 57 (2), 332–52.

Rocklage, Matthew, Derek Rucker, and Loran Nordgren (2018), "The Evaluative Lexicon 2.0: The Measurement of Emotionality, Extremity, and Valence in Language," *Behavior Research Methods*, 50 (4), 1327–44.

Sanchez-Vives, Maria V. and Mel Slater (2005), "From Presence to Consciousness Through Virtual Reality," *Nature Reviews Neuroscience*, 6 (4), 332–39.

Schroeder, Juliana and Nicholas Epley (2016), "Mistaking Minds and Machines: How Speech Affects Dehumanization and Anthropomorphism," *Journal of Experimental Psychology: General*, 145 (11), 1427–37.

Schweidel, David and Wendy Moe (2014), "Listening in on Social Media: A Joint Model of Sentiment and Venue Format Choice," *Journal of Marketing Research*, 51 (4), 387–402.

Sela, Aner, S. Christian Wheeler, and Gülen Sarial-Abi (2012), "We Are Not the Same as You and I: Causal Effects of Minor Language Variations on Consumers' Attitudes Toward Brands," *Journal of Consumer Research*, 39 (3), 644–61.

Settanni, Michele and Davide Marengo (2015), "Sharing Feelings Online: Studying Emotional Well-Being via Automated Text Analysis of Facebook Posts," *Frontiers in Psychology*, 6, 1–7.

Smith, Shane (2020), "Coronavirus (Covid 19) Tweets—Early April," Kaggle https://www.kaggle.com/smid80/coronavirus-covid19-tweets-early-april.

Tibshirani, Robert (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, 58 (1), 267–88.

Toubia, Olivier and Andrew Stephen (2013), "Intrinsic vs Image-Related Utility in Social Media: Why Do People Contribute Content to Twitter?" *Marketing Science*, 32 (3), 368–92.

Trager, G.L. (1958), "Paralanguage: A First Approximation," *Studies in Linguistics*, 13 (1), 1–12.

Van Zant, Alex B. and Jonah Berger (2020), "How the Voice Persuades," *Journal of Personality and Social Psychology*, 118 (4), 661–82.

Villarroel Ordenes, Francisco, Dhruv Grewal, Stephan Ludwig, Ko De Ruyter, Dominik Mahr, and Martin Wetzels (2018), "Cutting Through Content Clutter: How Speech and Image Acts Drive Consumer Sharing of Social Media Brand Messages," *Journal of Consumer Research*, 45 (5), 988–1012.

Villarroel Ordenes, Francisco, Stephan Ludwig, Ko De Ruyter, Dhruv Grewal, and Martin Wetzels (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), 875–94.

Warriner, Amy Beth, Victor Kuperman, and Marc Brysbaert (2013), "Norms of Valence, Arousal, and Dominance for 13,915 English Lemmas," *Behavior Research Methods*, 45 (4), 1191–207.

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

Weinstein, Matthew (2020), "TAMS: Text Analysis Markup System: An Open Source Qualitative Analysis System," http://tamsys.sourceforge.net/.

YouTube (2019), "Top 1,000 Most Viewed YouTube Videos of All Time," https://www.youtube.com/playlist?list=PL11E57E1166929B60.

Zemack-Rugar, Yael, Sarah Moore, and Gavan Fitzsimons (2017), "Just Do It! Why Committed Consumers React Negatively to Assertive Ads," *Journal of Consumer Psychology*, 27 (3), 287–301.