

REVIEW



Wisdom from words: The psychology of consumer language

Jonah Berger¹ | Grant Packard²

¹Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania, USA

²Schulich School of Business, York University, Toronto, Ontario, Canada

Correspondence

Grant Packard, Schulich School of Business, York University, Toronto, Ontario, Canada.
Email: gpackard@schulich.yorku.ca

Abstract

Language plays a fundamental role in every aspect of life. But only recently has research begun to understand the role of language in consumer behavior. This paper offers an integrative discussion of research on the language of consumer psychology. We review some of the main areas of inquiry and discuss some key methodological approaches (e.g., automated textual analysis) that have been crucial to the area's development. Further, we outline some broad issues and opportunities in the space and highlight potential directions for future research. We hope to encourage more consumer psychologists to consider the great potential in producing new conceptual and substantive wisdom from words.

KEYWORDS

automated textual analysis, language, linguistics, natural language processing

1 | INTRODUCTION

Language is an integral part of both communication and consumer behavior more generally. It is how consumers and customers communicate with peers, colleagues, and friends and how public health officials, politicians, and other marketplace actors communicate with audiences. Marketers use language in their advertisements, salespeople use language in their pitches, and service representatives use language to help solve customer issues. Even people's thoughts are expressed through language.

Further, subtle variations in language can have important outcomes. For just a few examples, the pronouns and prepositions consumers use to share their opinions can shape purchases (Ludwig et al., 2013), concrete language shapes social perceptions and customer satisfaction (Packard & Berger, 2021), and the sounds different words make can impact brand attitudes and intentions (Lowrey & Shrum, 2007).

But while it is clear that language is both pervasive and important, research is only beginning to uncover some of the exciting insights in this area. This paper provides an integrative overview of research examining consumer and marketing language. First, we review recent research in the area across the main topics of inquiry. Second, we

outline some of the important approaches that have supported the area's explosion in recent years (e.g., automated textual analysis). While these approaches are not necessary to study consumer language, they can help unlock a range of interesting insights. Third, we discuss some key distinctions in this area and highlight potential directions for future research. Figure 1 provides a chapter overview to help readers interested in particular topics.

Our goal is simple: we hope to encourage more scholars to explore this emerging and exciting space. For researchers interested in the psychology of language, we hope that this paper provides a helpful review of relevant articles and resources. For researchers who are not interested in language per se, we hope this discussion will inspire them to see how language can help them explore interesting questions in their own areas of study.

2 | MAIN THEMES OF RESEARCH ON CONSUMER LANGUAGE

Language is a fundamental part of being human. It allows us to organize and share emotions, intentions, information, and aspirations, and it enables cooperation and debate, helping construct shared reality. But while language has been around for millennia, it is only recently that studying language as a source of psychological insight has

Jonah Berger and Grant Packard contributed equally and are listed in alphabetical order.

- 1 INTRODUCTION
- 2 MAIN THEMES OF RESEARCH ON CONSUMER LANGUAGE
 - 2.1 Individual Words
 - 2.2 Content Words
 - 2.3 Function Words
 - 2.4 Combining Words
 - 2.5 Conversation
 - 2.6 How Words Sound
 - 2.7 Cross Cultural Linguistics
- 3 ANALYZING LANGUAGE
 - 3.1 Experiments
 - 3.2 Manual Coding
 - 3.3 Dictionaries
 - 3.4 Topic Modeling
 - 3.5 Embeddings
 - 3.6 Broader Points
- 4 FUTURE RESEARCH
 - 4.1 Understanding Consumers
 - 4.2 Language as a Dependent Variable
 - 4.3 Acoustic Features
 - 4.4 Limitations
- 5 CONCLUSION

FIGURE 1 Chapter overview

become more common. Further, consumer research has been a noteworthy contributor to this growth.

Like the field of linguistics in which it is grounded, research on language in consumer psychology is diverse in terms of the nature and psychological functions of the aspects investigated. Some research examines “small” features such as those that operate at the level of individual words (e.g., syllable stresses or phonetics) or word-level choices across different parts of speech. Other work considers clause or phrase level aspects of verbal communication such as figurative language or speech, and the relative complexity of a set of words used to convey ideas. More macro research considers language as discourse, treating an entire text (e.g., an advertisement or a conversation) as the unit of analysis, or considers variations in how words are used and processed across cultures and different languages.

In this section, we discuss the major areas of research on consumer language, highlighting some of the key insights, and offering suggestions about the kind of questions future research might consider.

2.1 | Individual words

The most active area of research on the psychology of consumer language examines the semantics (i.e., meanings) of words, punctuation marks, letters, and paralinguistic symbols (e.g., emojis; Luangrath et al., 2016). Overall, this literature reveals the many ways that basic, and often subtle, differences in the choice of words and symbols reflect things about communicators and impact the audience they communicate with.

2.2 | Content words

The nouns, adjectives, and verbs communicators use are the primary signals of what it is they are talking about, or communication's *content*. Nouns tell us about the relevant actors in communication (e.g., consumers, marketing agents, or products), adjectives help describe those actors (e.g., smart, capable, or cold), and verbs tell us what those actors are doing (e.g., shopping, buying, or solving). Seemingly minor variations in the content words used to share information or persuade can shape audience inferences and attitudes. Customers are more satisfied with frontline employees who use more concrete content words (i.e., more vivid or specific nouns and verbs), for example, to reference the consumer's topic of interest (e.g., “I'll grab the shoes” versus “I'll get the order”). This occurs because concrete language signals that employees are paying more personal attention to the customer's specific needs (“shoes”) rather than just going through the motions (Packard & Berger, 2021).

One topic in this area that has received lots of attention is the use of forceful or assertive verb phrases (e.g., “Just do it” or “Visit us!”). Such language tends to be beneficial when its use is consistent with communicative expectations and norms. Because forceful language is linked to personal control and positive affect, communications about hedonic (vs. utilitarian) goods benefit from assertive content words and phrases, and accordingly, such language can make products seem even more hedonic (Kronrod et al., 2012). Assertive language is more effective for encouraging positive behavior than discouraging negative behavior (Grinstein & Kronrod, 2016), and for encouraging compliance from consumers who are less, rather than more, committed to a brand (Zemack-Ruger et al., 2017). Its impact is

moderated by various social and situated factors including effort (Baek et al., 2015), power (Wang & Zhang, 2020), and culture (Terlutter et al., 2010). One of the few examples of when less assertive or “softer” language might be beneficial (Hamilton et al., 2014) found that consumer reviewers often use verb phrases or hedge words that “soften the blow” of negative information (e.g., “I don’t want to be mean, but ...”). Doing so makes the persuasion source seem more credible and likeable, which encourages audiences to accept their product evaluations.

Another rich area of consumer language research considers emotional content words. People automatically use more emotional language when they want to persuade (Rocklage et al., 2018), or when they are speaking rather than writing (Berger, Rocklage, & Packard, 2022), and emotional language often increases impact. Words that convey stronger emotion (e.g., hate rather than dislike) attract attention, making content more viral (Akpınar & Berger, 2017; Berger & Milkman, 2012) and more persuasive (Ordenes et al., 2019). This may be in part because emotional language helps focus distracted audiences (Nielsen et al., 2018) or because certain types of emotional language (e.g., anxious words such as uncertain or worried) increase perceptions the speaker’s persuasion act is reasonable and based on genuine effort (Yin et al., 2017).

The signaling benefit of emotional language, however, has its limits. Too much or the wrong kind of emotionality can make communicators seem beholden to their feelings, making others less likely to follow their advice (Yin et al., 2017). Ludwig et al (2013), for example, found that positive affect words (e.g., thrilled or love) have diminishing returns on persuasion, as more extreme emotional language signals that opinions may be based on more subjective, rather than objective, experiences. This backfiring effect is particularly strong for utilitarian products, for which audience expectations of more rational decision making may be violated (Rocklage & Fazio, 2020). The impact of emotional language also applies to the words brands use in social media, where more emotional posts cause consumers to relate the brand to lower status reference groups (Lee, 2021).

Research that considers how context shapes the semantic meaning of linguistic content is known as pragmatics. Grewal et al. (1996) demonstrate that content words’ effects can depend on the setting in which they are used. Whether a pricing cue of a “comparison price” or a “regular price” was more effective, for instance, depended on whether participants imagined they saw this language at home versus in a store. Suspicions about the firm’s motivation were greater in the store context, causing participants to evaluate the meaning of the words “comparison” versus “regular” more carefully. Extending pragmatics to paralanguage, work on emoticons in customer service found that an emoticon’s effectiveness depended on the kind of relationship the customer wanted with the firm and situated factors related to the nature of the service offered (Li et al., 2018). Overall, emoticons tend to help when warmth is prioritized and can hurt customer evaluations when competence should come first.

2.3 | Function words

One of the most notable findings in language research is that rather than what is talked about (e.g., content words like nouns and verbs), the small, often overlooked words that link content words together (e.g., articles and prepositions like a, the, or in) can be surprisingly meaningful. These words that serve as the connective tissue of a clause or sentence are called function words, and hold particular importance in the psychology of language (Pennebaker, 2011).

Personal pronouns have proven to be an especially meaningful kind of function word because they signal the communicator’s attention or proximity to the self or others. As such, they have received considerable attention by consumer psychologists. Brands, for example, can signal a relationship with customers by using either the second person pronoun “You” (e.g., “You and Wells Fargo do X”) or the plural first person “We” (e.g., “We do X”). Sela et al. (2012) find that using the plural first person signals greater closeness between the customer and the brand, which can cause reactance among consumers who do not already feel a close affiliation.

Relatedly, using first person pronouns in advertisements makes consumers feel a personal sense of social belonging with a warm brand, increasing ad effectiveness. But third person pronouns (e.g., she, he, or they) are more impactful for competent brands because they stimulate comparison rather than belonging (Chang et al., 2019). In customer service, employees who use “I” rather than “we” to refer to who is helping the customer (I’m happy to help vs. We’re happy to help) seem more personally agentic and empathetic, increasing customer satisfaction and purchase (Packard et al., 2018). Even the personal pronouns contained within a product can impact liking and consumption. Songs that use more second person pronouns in their lyrics are more successful because they help listeners imagine a “you” in their own lives (Packard & Berger, 2020).

Another function word category that has seen some attention is conjunctions. Conjunctions provide important information about how people, things, and actions relate to one another (e.g., after, but, or still). In a consumer context, the conjunction “with,” for example, sends a stronger signal of togetherness for two products than “and” (e.g., Eggs with ham vs. Eggs and ham; Patrick & Haws, 2014).

2.4 | Combining words

Beyond research that looks at individual content or function words, other work considers how words can be combined to create new words, phrases, or grammatical clauses. Starting at the simplest level, something as simple as using contractions to collapse two words together (e.g., don’t vs. do not) can influence consumer judgments. Contractions suggest a reviewer is warmer, increasing positive perceptions towards them and the product they endorse (Kim et al., 2021).

Figurative (versus literal) language is another example of phrase or clause level language constructions that have received attention.

Artful figures of speech are expected and accepted in advertising copy (Phillips & McQuarrie, 2009) but are not always effective in consumer word of mouth. In that context, conversational norms for different kinds of consumption drive whether figurative speech is effective. Consumers are not used to seeing figures of speech in utilitarian contexts, for example, so telling business travelers that the “The food is yummy” is less persuasive than just saying “The food is good” (Kronrod & Danziger, 2013).

Analysis of the impact of syntax (i.e., how one combines words to convey a particular meaning) considers word combinations at the level of a clause, sentence, or even across an entire text. More complex syntax involves longer clauses or sentences, and less familiar words. These more complex language constructions often produce disfluency in information processing. As a result, complex syntax tends to hurt audience engagement and impede ad recall and recognition, because it reduces motivation to process (Berger, Moe, & Schweidel, 2022; Lowrey, 1998; Pancer et al., 2019). Consequently, while high involvement consumers are willing to engage with complex sentence structure (Lowrey, 2006), less interested consumers are more likely to tune out.

One way to engage low involvement consumers involves another simple structural change to a sentence: turning them from assertions into questions (Hagtvedt, 2015). Doing so increases arousal, which increases engagement for less involved audiences. Surprisingly, however, questions can turn off high involvement consumers, who feel more anxiety when faced with the uncertainty a question introduces.

A common thread across syntactic complexity research is that the nature and depth of one's involvement with the message moderates the willingness to process language carefully. Pogacar et al. (2018) synthesize this idea, offering a framework suggesting that language effects can be organized according to the complexity of the language and the consumer's processing mode (automatic vs. controlled). Rhetorical devices like metaphor, analogy, and careful argument structures (via syntax) are complex language features that demand more careful processing. Phonetics, which we discuss next, are relatively simple language features (low complexity) and are processed automatically. Overall, linguistic effects seem to be stronger when they are congruent with the kind of information processing required.

2.5 | Conversation

While most consumer language research has focused on individuals producing or consuming words, phrases, or entire texts, more recent work has started to explore dialogues across two or more people. The flood of social media conversations has been one rich setting (e.g., Chen & Berger, 2013). In addition to experiments, for example, Moore and McFerran (2017) studied the language of conversation in online forums. They found evidence for both affiliation and achievement motives in responding to others posts. Further, people mimicked positive emotion and social word use when they wanted to feel similar to a prior conversant, but used cognitive and descriptive words when they were motivated by status differentiation. Other research

finds that consumers tend to adopt others' linguistic style when they feel like they share the group's interests. Linguistic style matching captures the extent to which people tend to use similar types and amounts of function words. For example, someone might say either “I love Lady Gaga” or “Lady Gaga is lovely” when describing a performer. One uses a personal pronoun to do so, while the other uses impersonal voice. The extent to which consumers have similar linguistic styles can be used to predict which products consumers are interested in. Liu et al. (2019), for example, find that online reviews are rated as more useful when the language style used matches that of the intended customer. Similarly, Ludwig et al. (2013) report that increasing congruence between online reviews and the styles usually used by that group of reviews increases conversion rates.

Some work has also started to explore service conversations between consumers and employees. Employee use of singular first-person pronouns (e.g., I am happy to help vs. Happy to help vs. We are happy to help) boosts customer satisfaction and purchase, for example, because it signals that the employee is emotionally and behaviorally present (Packard et al., 2018). Other conversation research has examined the warmth/competence trade-off. Some research suggests that agents who use more competence-centered language are seen as more helpful, while warm language is costly (Güntürkün et al., 2020; Marinova et al., 2018). But work considering the moment-to-moment language dynamics of customer and agent discourse reveals that both warm and competent language matter, just at different conversational moments (Packard et al., 2022). Customers are more satisfied when agents use warm, affective language at the beginning and end of the conversation, but cognitive, competent language in the middle.

This work, though, only scratches the surface of what is possible. Research has begun to explore things like linguistic signals of conversational receptiveness (Yeomans et al., 2020), or openness to opposing viewpoints, linguistic features that lead to longer conversations (Boghrati & Berger, 2022), and even how paralinguistics features like pausing (Van Zant et al., 2022) and response speed (Templeton et al., 2022) impact social inferences and social connection. Even controlling for how interesting a particular product or brand is to discuss, the way conversation partners talk about something (i.e., the linguistic features used) shapes how long it is discussed. Speaking more concretely, asking narrow questions, matching linguistic styles, and allowing both partners to contribute, all seem to encourage longer discussion. See (Yeomans et al., 2022) for a review and discussion of motives in conversation.

2.6 | How words sound

Work on phonetics looks at how the acoustic signals that syllables and words produce when spoken convey meaning independent of the word's definition. Most consumer research in this space examines brand names. This work finds that brands with consonants that produce higher-frequency sounds (e.g., f or s) are seen as smaller, lighter, or closer, for example, while those with consonants that produce

lower frequency sounds (e.g., b or p), which seem bigger, heavier, and farther away (Klink, 2000; Maglio et al., 2014; Yorkston & Menon, 2004). These phoneme-based perceptions enhance consumer attitudes and intentions towards a brand if they match the brand's attributes or positioning.

Repeating word sounds is also important. Rhyme within brand names (e.g., Lean Cuisine or Rantifanti) can produce positive affect, improve product evaluation, and enhance brand recall (Argo et al., 2010; Carr & Miles, 1997), and the benefit of rhyming sounds persists across words in a slogan (Filkuková & Klempe, 2013). Other phonetics research finds that words with related sounds but disparate meanings and spellings (i.e., homophones) can activate the other word's meaning. Saying "Bye bye!" for example, primes consumption-related cognitions and behaviors because of its homophony with "buy" (Davis & Herr, 2014).

More recently, consumer phonetics research has examined the relationship between silent consonants in brand names and environmentalism (Joshi & Kronrod, 2020), how consonant voicing (i.e., air flow in pronunciation) impacts perceived harshness or softness (Pathak et al., 2020), and how the timing of syllable stresses signals brand gender and warmth (Pogacar et al., 2021).

Phonetic effects have also been found to transcend individual consonant, vowel, or syllable sounds to a speaker's own control of how word sounds are produced. A lower voice leads consumers to imagine the product discussed is larger (Lowe & Haws, 2017), for example, and low pitch vocalizations produce fewer negative cognitive responses and more positive brand attitudes (Chattopadhyay et al., 2003).

2.7 | Cross-cultural linguistics

Some work has also examined consumer language across cultures. Yorkston and De Mello (2005), for example, reveal how languages with gender marked nouns (e.g., la cerveza or il vino) can activate gender stereotypes and boost brand recall if the product's prototypical consumer gender is congruent with the gender marker. Other work (Schmitt & Zhang, 1998) finds that languages that use classifiers shift how consumers categorize products and their attributes, suggesting that language's grammatical structure can shape consumer behavior. In Mandarin Chinese, for example, rather than just saying "yuan" (umbrella in English), a speaker would first use the classifier "ba" to signal that the word coming next is an object that can be held in a hand. This research is just one example of how different languages may shape how its users categorize people and objects.

Turning to people who speak more than one language, bilinguals tend to prefer brands that use their primary language (Carroll & Luna, 2011) and are less likely to prefer brands that use their secondary language because it is harder to process affective aspects of the language (Karataş, 2020). Bilingual ads that "code-switch" from a secondary to primary language within a single utterance are more persuasive when the primary language is viewed more positively in a given culture (Luna & Peracchio, 2005) because it signals an appeal to the higher

status culture. Similarly, code-switching can help transfer meaning associated with a second language to a brand. Using some English in Hindi ads helped boost attitude towards an ad for luxury goods because English conveys sophistication, but didn't help for everyday goods (Krishna & Ahluwalia, 2008). For bilinguals, using a less native language makes lying feel less self-diagnostic, which allows the consumer to lie more without harming the self-concept (Gai & Puntoni, 2021).

3 | ANALYZING LANGUAGE

Now that we have reviewed the main themes in consumer language research, and some trends over time, we turn to how language can be analyzed to produce psychological insight.

3.1 | Experiments

The simplest way language can be examined is experimentally. A good deal of research has used experiments to manipulate language and examine the causal impact on attitudes, persuasion, and other dependent variables. Sela et al. (2012), for example, manipulated the pronouns used in marketing communications and measured the impact on consumers attitudes towards the brand. Similarly, Berger and Heath (2005) manipulated the topics of conversations and measured the influence on usage of different catchphrases, proverbs, and slang.

In addition to being used as an independent variable, language can also be used in experiments as a dependent variable or mediator. Melumad et al. (2019), for example, examined how using a smartphone or computer changed the language participants produced. Similarly, in studying how audience size changes word of mouth, Barasch and Berger (2014) measured the process of self versus other focus through counting the pronouns used in what participants wrote.

As in any other method, care needs to be given to ensure correct operationalization. When using language as an independent variable, for example, it is critical to be sure that only the focal language features are manipulated. Similarly, when using language as a dependent variable or mediator, ensuring that the features measured actually capture the constructs of interest.

3.2 | Manual coding

Beyond experiments, researchers can also manually code language in field data. In studying what makes online content viral, for example, Berger and Milkman (2012) had research assistant code how much anger or sadness different news articles evoked. Similarly, in research on personal pronouns in customer service, the authors used judges to code whether instances of "we", "our", and "ours" excluded or included the customer given their focus was on "we" uses that employees used to describe themselves as part of the company (Packard et al., 2018).

But while manual coding can be useful, it has some challenges. First, it is difficult to scale. Coding 10 articles often takes 10 times as long as coding one. Second, one may be concerned about subjectivity or bias. Different things may make different research assistants feel sad, and political leanings, gender, and other factors may shape how people perceive content.

As discussed, though, shifts in technology have enabled a huge amount of spoken and written language data to be digitized. Further, new tools have emerged to allow researchers to analyze this wealth of data at scale (Boyd & Pennebaker, 2015).

Consequently, it is important to understand these new methods for extracting psychological insight from language. There are two general classes of approaches. The first is relatively top down, involving a set of key, pre-defined constructs and variables that are related to them. The second is more bottom up, allowing the data to reveal potential patterns in language. First, we discuss the most common approach, dictionaries, which is relatively top down, and then discuss topic modeling and word embeddings, two more bottom-up approaches.

3.3 | Dictionaries

The most common approach for automated textual analysis, particularly among consumer psychologists, is dictionaries (for recent reviews see Berger et al., 2020; Berger & Packard, 2022; Humphreys & Wang, 2018). Dictionaries usually involve a mapping between words or phrases and key psychological constructs. Someone interested in measuring how much participants are attending to the past, for example, might utilize a dictionary including words like “was,” “had,” and “been.” By counting the number of these words that appear in different passages of text, the dictionary could provide a sense of which participants were focused more on the past.

The dictionary that has received the most attention is Linguistic Inquiry and Word Count (LIWC). The most recent iteration of this dictionary analyzes over 100 dimensions of text, including categories like pronouns, function words, social processes, and cognitive processes. A conservative estimate would suggest this dictionary has been cited over 10,000 times, but the actual number is likely much larger. In consumer psychology, it has been used to study everything from attention to the self versus others (Barasch & Berger, 2014), social acceptance in news media (Humphreys, 2010), emotional dynamics of movies (Berger et al., 2021), and the virality of emotional content (Berger & Milkman, 2012).

While LIWC is useful in many applications, a variety of other dictionaries also exist. They measure everything from authenticity (Kovacs et al., 2013) and brand personality (Opoku et al., 2006) to linguistic concreteness (Brysbaert et al., 2014) and values (Lasswell & Namenwirth, 1969), among other features. While some of these dictionaries just count the number of words that appear in different categories, others use a more continuous approach scoring words on different dimensions. The Evaluative Lexicon (Rocklage et al., 2018), for example, scores hundreds of words based on their emotionality,

and extremity. Other work has scored over 10,000 words based on valence, arousal, and dominance (Warriner et al., 2013). And the textual paralinguistic classifier (Luangrath et al., 2022) uses both dictionaries and rule-based algorithms to identify different types of nonverbal audible, tactile, and visual elements in text (e.g., emojis).

Dictionaries are particularly good when researchers know what they are looking for. If someone knows they want to measure sentiment, for example, then selecting validated, off-the-shelf dictionary will help them achieve that goal. Further, dictionaries can be used to measure aspects of *what* consumers are attending to or talking about (i.e., particular goals and motivations) and also *how* consumers are talking (i.e., whether they are using more or less concrete language).

There are only so many prevalidated dictionaries, however, so measures do not exist for everything researchers want to measure. Further, while researchers can create their own custom dictionaries, to ensure that those dictionaries actually measure the intended aspect, validation is often necessary.

3.4 | Topic modeling

Dictionaries have a number of valuable aspects, but in some cases, researchers may be less interested in a top-down approach and more interested in one that is bottom-up. They may not know exactly what feature they want to measure or may want to characterize language on more dimensions than there are dictionaries available.

In situations like these, a class of approaches called topic modeling can be particularly useful. Rather than starting with a particular pre-defined construct or constructs, topic modeling starts with the data itself and identifies the main topics or themes being discussed. If a researcher was interested in the main topics or themes talked about in restaurant reviews (e.g., food, service, and price), for example, they could use topic modeling to identify those themes and their relative prevalence in each review.

Topic modeling works by identifying co-occurrences among words. If words like price, cost, and expensive often show up together in restaurant reviews, for example, the approach may identify them all as belonging to the same topic or theme. Similarly, if words and phrases like tasty, delicious, and mouthwatering often show up together in reviews, the approach may identify them as belonging to the same topic or theme. When given a set of documents (e.g., online reviews, advertisements, or thought protocols), the approach identifies the main topics of themes that appear in those documents, the words or phrases that make up each topic or theme, and how strongly associated each word or phrase is with each topic or theme.

While this approach has not been used as frequently as dictionaries, it has begun to appear in a variety of papers by both consumer psychologists and marketing scholars more generally. Tirunillai and Tellis (2014), for example, used topic modeling to explore dimensions linked to quality, how they shift over time, and how that relates to brand positioning. Berger and Packard et al. (2018) used the approach to identify the main themes in song lyrics, and Li and Ma (2020) illustrate how marketers can use topic modeling and consumer search

data to identify where consumers are in the decision-making journey. Topic modeling can also be used as control variables, identifying different topics that might arise in things like customer service conversations (Packard & Berger, 2021) or news articles (Berger, Moe, & Schweidel, 2022).

Latent Dirichlet allocation (LDA; Blei et al., 2003) is the most commonly used topic modeling approach. Latent semantic analysis (LSA; Landauer et al., 1998) is also sometimes used, but while it is simpler and faster to implement, it often requires larger amounts of text and may achieve lower accuracy levels. Some work has also used Poisson factorization (Gopalan et al., 2013; Toubia, 2021) which has the benefit of not forcing topic probabilities to sum to one.

Note that topic modeling focuses on linguistic content rather than linguistic style. As discussed previously, researchers distinguish between words that indicate something about the content being discussed and words that indicate more about the linguistic style of the speaker (Pennebaker et al., 2003; Tausczik & Pennebaker, 2010). While the requests “would it be possible to get me a glass of water” and “grab me a glass of water” both indicate the speaker’s desire for water and directs the listener towards action, they go about those tasks in quite different ways. Style or function words include pronouns, prepositions, articles, conjunctions, auxiliary verbs, and a few other categories that appear among the nouns, regular verbs, adjectives, and adverbs that make up the content of language. Not surprisingly, then, style or function words often co-occur with multiple content words.

Topic modeling tends to ignore style and focus on content. In fact, as one of the standard text preprocessing steps, standard topic modeling packages often remove so-called common stop words (e.g., “a” or “the”) before assigning words to groups, removing many style words from the analysis. Consequently, this approach may be ideal for looking at content changes between groups or over time but may be less ideal for looking at variation in style.

3.5 | Embeddings

Advances in computer science have provided even more interesting ways to characterize language and text. Rather than just comparing the relative prevalence of a type of words or theme across documents, these approaches quantify the relationship or similarity between words or documents more generally.

Imagine trying to quantify how similar two words are to one another. While dictionaries might identify whether two words are both members of the same dictionary, for example, or score similarly on a particular dimension, this would only account for a single dimension out of many that the words could vary on. Similarly, while topic modeling could help identify whether the words co-occur, but if they do not appear in the same topic, it would be hard to know how similar they are. Further, such broad measures of coincidence within a large document miss a great deal of nuance. Two words that appear in the same sentence or paragraph are probably more related than two words that appear further away from one another.

A computational linguistics approach known as word embeddings (e.g., Word2vec, Mikolov et al., 2013 and GloVe, Pennington et al., 2014) helps address some of these issues. Similar to Firth’s (1957) suggestion that “you shall know a word by the company it keeps,” embeddings use a neural network framework to take a corpus of texts (e.g., online reviews) and the local context in which words appear to determine their semantic relatedness (see Eichstaedt et al., 2021, for a recent discussion). Just as people who see each other more often tend to live closer in geographic space, embeddings use word co-occurrence, distance between words, and appearance in similar contexts across different texts to position words in relation to one another in a continuous multidimensional semantic space. By taking into account where words appear in relation to one another within a given document, word embeddings are able to achieve greater insight into semantic relatedness.

If “Nikes are cool” shows up frequently in online reviews, for example, that would shrink the relative distance between the vector for Nike and the vector for cool. Beyond just incidence, though, the distance between occurrences also matters. Both “Nikes are cool” and “Nikes aren’t cheap, but they are cool” contain the words Nike and cool, but the first example has them closer together, which would decrease the vectors’ distance in word2vec space. Note that two words do not have to necessarily co-occur; occurring in similar contexts also shapes similarity. If “Nikes are high-tech” and “high-tech shoes are cool” both appear frequently, it decreases the distance between the vectors for “Nike” and “cool” even if the two words never appear near each other.

Embeddings have some useful properties. Words with similar semantic meanings appear closer together in the multidimensional space, and consequently, the relationship between words can be measured by the difference between their vectors. Words that mean similar things (e.g., fun and exciting), for example, are placed closer in word embedding space than words that mean different things (e.g., fun and dependable).

Consequently, embeddings can be a powerful tool to study consumer behavior. Bhatia and Olivola (2018, 2021), for example, demonstrate that word embeddings can be used to predict brand personality traits, and Aka et al. (2022) used embeddings to demonstrate that consumers prefer brands that “fit” their own psychological tendencies (see Nave et al., 2022, for a similar approach). Zhang et al. (2018) illustrated that embedding models can predict brand recall, and Timoshenko and Hauser (2019) demonstrate how embeddings can identify customer needs from product reviews.

Embeddings can also be used to quantify the distance between larger bodies of text. Just as word embeddings identifies the similarity between words, tools like Google’s Universal Sentence Encoder (Cer et al., 2018), BERT (Devlin et al., 2018), and ELMo (Peters et al., 2018) can determine the similarity of sentences, paragraphs, or even whole documents (e.g., how similar two online reviews are to one another). Depending on the approach, the similarity may be more semantic or may take syntax and other features into account.

Researchers are only starting to leverage these exciting tools, but one useful application may be the structure of discourse. While

theorists have long theorized that stories have common shapes (e.g., Campbell, 1949; Freytag & MacEwan, 1900), little empirical work has tested this possibility or examined whether different shapes or structures might be more effective or engaging.

Some work has started to examine narratives using manual coding or dictionary-based approaches (Berger et al., 2021; Boyd et al., 2020; Reagan et al., 2016), but embeddings can add additional insight into both the structure of discourse and its effects. Take stories, for example: first, each chunk or portion of the story (e.g., a paragraph) can be represented as a point in a high-dimensional space. Then, by looking at the relationship between adjoining portions of the story, or the set of points as a whole, one can extract features of the semantic progression. By computing the distance between adjoining passages, for example, one can get a sense of pacing or how fast a story is moving (Laurino Dos Santos & Berger, 2022). Similarly, by looking at how the story moves in semantic space, or the particular sequence between the set of points, one can quantify whether the story takes the most direct path between the different points or a more circuitous route. Finally, by looking at the set of points as a whole, one can measure the narrative's volume or how much ground it covers (Toubia et al., 2021).

Speed, circuitousness, and volume are only some of the features one could imagine extracting (e.g., see Berger et al., 2021 for work on emotional volatility), but the ability to quantify discourse features opens up a range of interesting research questions. TV shows and movies that move more quickly, for example, tend to be liked more (Toubia et al., 2021). But is the same true for online reviews, for example? Or might reviews that move faster be liked less because they require more effort to follow? What about conversations? Are they more enjoyable when the conversation moves faster or slower and why?

Similar questions could be asked with the concept of volume. While academic papers that covered more ground are cited more, TV shows that covered more ground are liked less (Toubia et al., 2021). What might account for this difference and how might this effect vary in different domains? Might online reviews that cover more ground be more useful because they integrate more information? Might conversations that cover more ground be more enjoyable or less?

And while these questions focus on the impact of linguistic features, or their effect on the audiences or peers that consume them, might these features also reflect things about content producers? Might the speed or volume of online reviews, for example, provide insight into how likely a given customer is to return the product? Or how loyal they will be in the future?

More generally, embeddings could be useful for a range of other topics. Embeddings have been used to capture risk perceptions (Bhatia, 2019) and judgments of relatedness, meaning, and probability (see Bhatia et al., 2019 for a review). Because embeddings represent how information is organized, memory and information processing researchers could use them to explore how ideas become linked in memory (Bhatia et al., 2018). Self and identity researchers might use these techniques to examine the malleability of self concepts over time, and emotion researchers could use them to test the structure of emotion, and whether or how appraisals of discrete emotions shift across situations.

3.6 | Broader points

This discussion of automated tools brings up a couple of larger points that are worth to note. First, note that studying language does not always require automated approaches. In situations where the number of things that need to be coded is relatively small, and the features to be coded are relatively simple and small in number, manual coding may be just fine. Researchers or research assistants can count the easily noted features or generate coding rubrics to manually code things. Automated approaches are particularly useful when the quantity of text is large, the features to be coded are relatively complex or difficult to perceive, and manual coding may introduce subjectivity.

Second, these approaches can be used to study *what* people are talking about, or *how* and *why* they are talking. Topic modeling, for example, focuses heavily on the *what*. When considering customer service calls, for example, topic modeling might identify that some calls focus more on the incorrect product being delivered while others focus more on an issue with the website. Other approaches, however, focus more on *how* a given topic is being talked about the psychological motivations behind *why*. A concreteness dictionary, for example, can help identify which customer service representatives are speaking more or less concretely, regardless of which topic they are talking about. Similar representatives using different pronouns may talk about the same issues, but by using different pronouns (e.g., I vs. we) may shape customer satisfaction by influencing how caring they seem.

Third, these techniques can both be used to study both what language reflects about its producer(s) and how it impacts audiences. Language serves a dual role. It both *reflects* things about the people, companies, and societies that produce it, and it *impacts* the audiences that consume it.

On the reflect side, for example, one can learn things about consumers, companies, and other marketplace actors based on the language they produce. Different people use language differently, and people use language differently based on transitory states, and as a result one can learn a lot about people based on the language they use. Language provides insight into people's personality and identity (Preotjuc-Pietro & Ungar, 2018; Pennebaker & King, 1999; Park et al., 2015) and also how they are feeling or what they are thinking about (Seraj et al., 2021). Consumers that use passive voice when complaining about a negative service experience, for example, are less likely to see themselves at fault for the issues, and thus more likely to share negative word of mouth (Sepehri & Berger, 2022). Similarly, the language companies use can provide insight into their attitudes towards women (Lawson et al., 2022), and the language used in ads can provide insight into cross-cultural differences (Han & Shavitt, 1994) in things like individualism and collectivism.

On the impact side, language also impacts the individuals, organizations, and societies that consume it. The language used in word of mouth not only reflects things about the consumer that shared it (e.g., Packard & Wooten, 2013), for example, it also impacts the consumers that listen to or read it, and their likelihood of purchasing that product (Berger, Rocklage, & Packard, 2022). Novices are more likely to say they "recommend" something, for example, and hearing

someone recommends (vs. likes) something is more likely to encourage action (Packard & Berger, 2017). Similarly, the language used by customer service representatives impacts customer satisfaction (Packard et al., 2018; Packard et al., 2022; Packard & Berger, 2021), the language used in song lyrics shapes whether songs become popular (Berger & Packard, 2018; Packard & Berger, 2020), the language used in social media posts shapes engagement (Cascio-Rizzo et al., 2022; Pezzuti et al., 2021), and the language used in news articles shapes whether they get read and shared (Berger & Milkman, 2012; Berger, Moe, & Schweidel, 2022).

Each of experiments, dictionaries, topic modeling, and embeddings offer benefits and limitations to help understand what language reflects about language producers and how language impacts language consumers.

4 | FUTURE RESEARCH

While the past few years have seen burgeoning interest in consumer language, much more remains to be done. Language research has only started to leak into other areas of study and there are lots of open opportunities. Further, the advent of new tools and greater availability of field data promises new chances to uncover consumer insights.

4.1 | Understanding consumers

One area in particular that deserves further study is using language to understand consumers. As noted previously, most consumer language research has focused on language's impact. How the use of words, phrases, and topics impacts the audiences that consume such language. There has been less attention, however, to what language reflects about consumers and their psychological processes. What can the language consumers use tell us about their traits, states, attitudes, cognitions, needs, and motives?

Consumer language research has barely scratched the surface of language as an indicator of one's inner mental life. Psychological research has shown that the language people use on social media can predict things like their personality (Park et al., 2015), or physical and mental health (Eichstaedt et al., 2015, 2018) with great accuracy, but similar approaches have, for the most part, yet to be applied in the consumer space. Consumer researchers could use language to better identify and understand consumer-relevant psychological factors like self-focus or construal level, and consumption-specific traits like materialism or frugality (Rick et al., 2008).

Further, while some of this work might focus on understanding, work on prediction may be just as useful. Consumer psychologists who study language tend to focus on understanding. Trying to understand why particular words, phrases, or other linguistic features, for example, have the impact that they do.

But more quantitative work has also used language for prediction. Rather than focus on any one particular linguistic feature (e.g., pronouns or concrete language), this type of work often extracts

as many linguistic features as possible from relevant text and uses them to predict an important outcome. Netzer et al. (2019), for example, examined whether the text that consumers wrote when applying for a loan could help predict whether borrowers would pay the loan back. They found that above and beyond the traditional financial and demographic variables often used in models to predict default, textual features added additional predictive power. Similarly, Dore and Berger (2022) examined why certain content is shared more on social media, using embeddings to represent posts in a multidimensional space and identifying a linguistic signature of sharing that predicted shares above and beyond traditional audience factors.¹

Similar approaches could be applied to a range of relevant outcomes. Topic modeling, for example, could be used to identify different motivations in a variety of context. Some work has applied this to hosting on Airbnb (Chung et al., 2022) but the same notion could be used to understand why people buy certain products or fail to achieve their goals. Along these lines, work could try to predict where consumers are in the customer journey (e.g., Humphreys et al., 2021), relevant cognitions, or variety of other aspects.

4.2 | Language as a dependent variable

Another area that might deserve more attention is text as a dependent variable. In most experimental research, researchers use scale measures to collect outcome variables. They select a small set of key variables of interest and include one or multiple measures of each of these constructs. But while this approach has many upsides, it is relatively restrictive. Researchers have to know in advance what they are looking for and can often only ask so many measures at once.

Language, however, can potentially be more open-ended. Rather than asking a small set of predetermined questions, researchers can ask more open-ended questions where participants write (or speak) their responses. Natural language techniques can then be used to extract insight from this data. By exploring correlations between language features and independent variables, or other relevant outcomes, researchers can begin to get a better sense of potentially important relationships and where to dig further.

Natural language may be particularly valuable as a dependent variable because it might help researchers study things that consumers may not want to express more explicitly. If asked, most people would not say that they are lying, yet their language often provides telltale signs that they are not telling the truth. Similarly, most individuals would not own up to being racist or sexist, but the language they use when describing different groups or individuals may provide evidence of such biases. Consequently, language may provide a way to gain insight into things that might not otherwise be revealed.

This is another situation where the distinction between content and function words becomes important. Most people are probably at least somewhat aware of *what* they are talking about: the topics or themes they are discussing. But they are likely less aware, and less in conscious control, of the way they express those things (Pennebaker, 2015). Consequently, function words may be particularly useful in

picking up things that consumers may not want to explicitly reveal, either because of response biases or lack of conscious awareness.

4.3 | Acoustic features

Acoustic features are another area of opportunity. While a decent amount of work has begun to explore language, or the words phrases and topics that people use, there has been less attention to acoustic paralinguistic or the vocal features used in verbal communication. The pitch, tone, volume, and other features consumers use when communicating with one another vary across individuals situations and other aspects. Further, these features can provide important insight into what people are thinking or feeling, or the likely impact of their communications (e.g., Van Zant & Berger, 2020; Wang et al., 2021). People shift their voice when trying to persuade (Van Zant & Berger, 2020), for example, and customer service agents who pause more frequently during conversational turns increase customer satisfaction, because they encourage customers to assent more often (i.e., say “yeah” or “ok”; Van Zant et al., 2022). While these are just a couple of examples, spoken communication always includes vocal features, and thus, this is a rich area for future work. And though written communication does not include acoustic features, it does often include textual paralinguistic, which is a rich opportunity for future study as well (e.g., Luangrath et al., 2016, 2022).

4.4 | Limitations

While there are lots of opportunities in consumer language research, that is not to say that there are no limitations. As with any research area, or any research method, there are some important challenges and caveats to keep in mind.

First, language is heavily context dependent. The language consumers use when broadcasting on social media is different than the language they use when speaking to their best friends, which is different than the language they use when speaking to their boss. Similarly, pragmatic factors, like modality (i.e., speaking vs. writing), channel (i.e., text vs. email), and devices (i.e., smartphone vs. PC), shape the language consumers use. Consequently, when analyzing language data in the field, or conducting experiments in the lab, it is important to keep context in mind. Context affects the language consumers create, and the language they think is appropriate to use.

Second, words does not capture all aspects of communication. Body language in face to face interactions or images in social media posts also convey information. Thus, while language often conveys a great deal, it is worth considering other ways to communicate as well.

5 | CONCLUSION

The words consumers produce and consume play an important role in their lives. A growing body of research has shed considerable light on

how language persuades and reflects the mind of the consumer. But much more remains to be done. The growing foundation in the psychology of language and wider accessibility of natural language processing tools and data have opened up a new era of possibility. This paper has tried to provide a glimpse into the diverse range of insights, approaches, and opportunities in this field to date. Hopefully, this will encourage more researchers to study the psychology of consumer or marketing language, or any other psychological theories of interest that can be elucidated, in part, through the wisdom on offer in words.

ORCID

Grant Packard  <https://orcid.org/0000-0002-8765-1994>

ENDNOTE

¹ In both these examples, the researchers were interested in more than just prediction. Beyond identifying the linguistic features added predictive power, they worked to identify which specific features seem to link to relevant outcomes.

REFERENCES

- Aka, A., Olivola, C., Nave, G. & Bhatia, S. (2022). Leveraging social media, digitized language data, and machine learning to identify consumer personality segments, facets of brand image, and predictors of brand liking, Working Paper.
- Akpinar, E., & Berger, J. (2017). Valuable virality. *Journal of Marketing Research*, 54(2), 318–330. <https://doi.org/10.1509/jmr.13.0350>
- Argo, J. J., Popa, M., & Smith, M. C. (2010). The sound of brands. *Journal of Marketing*, 74(4), 97–109. <https://doi.org/10.1509/jmkg.74.4.097>
- Baek, T. H., Yoon, S., & Kim, S. (2015). When environmental messages should be assertive: Examining the moderating role of effort investment. *International Journal of Advertising*, 34(1), 135–157. <https://doi.org/10.1080/02650487.2014.993513>
- Barasch, A., & Berger, J. (2014). Broadcasting and narrowcasting: How audience size affects what people share. *Journal of Marketing Research*, 51(3), 286–299.
- Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. *Journal of Marketing*, 84(1), 1–25. <https://doi.org/10.1177/0022242919873106>
- Berger, J., Kim, Y. D., & Meyer, R. (2021). What makes content engaging? How emotional dynamics shape success. *Journal of Consumer Research*, 48(2), 235–250. <https://doi.org/10.1093/jcr/ucab010>
- Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2), 192–205. <https://doi.org/10.1509/jmr.10.0353>
- Berger, J., Moe, W., & Schweidel, D. (2022). What leads to longer reads? Working Paper. Accessible at https://www.msi.org/wp-content/uploads/2020/11/MSI_Report_20-126.pdf
- Berger, J., & Packard, G. (2018). Are atypical things more popular? *Psychological Science*, 29(7), 1178–1184. <https://doi.org/10.1177/0956797618759465>
- Berger, J., & Packard, G. (2022). Using natural language processing to understand people and culture. *American Psychologist*, 77(4), 525–537.
- Berger, J., Rocklage, M. D., & Packard, G. (2022). Expression modalities: How speaking versus writing shapes word of mouth. *Journal of Consumer Research*, 49, 389–408. <https://doi.org/10.1093/jcr/ucab076>
- Berger, J. A., & Heath, C. (2005). Idea habitats: How the prevalence of environmental cues influences the success of ideas. *Cognitive Science*, 29(2), 195–221. https://doi.org/10.1207/s15516709cog0000_10

- Bhatia, S. (2019). Predicting risk perception: New insights from data science. *Management Science*, 65(8), 3800–3823.
- Bhatia, S., Goodwin, G. P., & Walasek, L. (2018). Trait associations for Hillary Clinton and Donald Trump in news media: A computational analysis. *Social Psychological and Personality Science*, 9(2), 123–130.
- Bhatia, S., & Olivola, C. (2018). Data-driven computational brand perception. *ACR North American Advances*, 46, 204–208.
- Bhatia, S., & Olivola, C. Y. (2021). Computational brand perception: Fine-tuned word embedding techniques for predicting consumer brand-trait associations. Working Paper.
- Bhatia, S., Richie, R., & Zou, W. (2019). Distributed semantic representations for modeling human judgment. *Current Opinion in Behavioral Sciences*, 29, 31–36.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993–1022.
- Boghtrati, R., & Berger, J. (2022). Quantifying gender bias in consumer culture. *arXiv preprint arXiv:2201.03173*.
- Boyd, R. L., Blackburn, K. G., & Pennebaker, J. W. (2020). The narrative arc: Revealing core narrative structures through text analysis. *Science Advances*, 6(32), eaba2196. <https://doi.org/10.1126/sciadv.aba2196>
- Boyd, R. L., & Pennebaker, J. W. (2015). Did Shakespeare write double falsehood? Identifying individuals by creating psychological signatures with text analysis. *Psychological Science*, 26(5), 570–582. <https://doi.org/10.1177/0956797614566658>
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3), 904–911. <https://doi.org/10.3758/s13428-013-0403-5>
- Campbell, J. (1949). *The hero with a thousand faces*. New World Library.
- Carr, D., & Miles, C. (1997). Rhyme attenuates the auditory suffix effect: Alliteration does not. *Quarterly Journal of Experimental Psychology Section a*, 50(3), 518–527. <https://doi.org/10.1080/713755722>
- Carroll, R., & Luna, D. (2011). The other meaning of fluency. *Journal of Advertising*, 40(3), 73–84. <https://doi.org/10.2753/JOA0091-3367400306>
- Cascio-Rizzo, G.L., Berger, J., De Angelis, M., & Pozharliev, R. (2022). How sensory language shapes consumer responses to influencer-sponsored content. Under Review.
- Cer, D., Yang, Y., Kong, S. Y., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y. H., & Kurzweil, R. (2018). Universal sentence encoder. *arXiv preprint arXiv:1803.11175*.
- Chang, Y., Li, Y., Yan, J., & Kumar, V. (2019). Getting more likes: The impact of narrative person and brand image on customer–brand interactions. *Journal of the Academy of Marketing Science*, 47, 1027–1045. <https://doi.org/10.1007/s11747-019-00632-2>
- Chattopadhyay, A., Dahl, D. W., Ritchie, R. J., & Shahin, K. N. (2003). Hearing voices: The impact of announcer speech characteristics on consumer response to broadcast advertising. *Journal of Consumer Psychology*, 13(3), 198–204. https://doi.org/10.1207/S15327663JCP1303_02
- Chen, Z., & Berger, J. (2013). When, Why, and How Controversy Causes Conversation. *Journal of Consumer Research*, 40(3), 580–593.
- Chung, J., Johar, G. V., Li, Y., Netzer, O., & Pearson, M. (2022). Mining Consumer Minds: Downstream Consequences of Host Motivations for Home-Sharing Platforms. *Journal of Consumer Research*, 48(5), 817–838.
- Davis, D. F., & Herr, P. M. (2014). From bye to buy: Homophones as a phonological route to priming. *Journal of Consumer Research*, 40(6), 1063–1077. <https://doi.org/10.1086/673960>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Dore, B., & Berger, J. (2022). o. Working Paper
- Eichstaedt, J. C., Kern, M. L., Yaden, D. B., Schwartz, H. A., Giorgi, S., Park, G., Hagan, C. A., Tobolsky, V. A., Smith, L. K., Buffone, A., Iwry, J., Seligman, M. E. P., & Ungar, L. H. (2021). Closed-and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychological Methods*, 26(4), 398–427. <https://doi.org/10.1037/met0000349>
- Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., Jha, S., Agrawal, M., Dziurzynski, L. A., Sap, M., Weeg, C., Larson, E. E., Ungar, L. H., & Seligman, M. E. (2015). Psychological language on Twitter predicts county-level heart disease mortality. *Psychological Science*, 26(2), 159–169.
- Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preoțiu-Pietro, D., Asch, D. A., & Schwartz, H. A. (2018). Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences*, 115(44), 11203–11208.
- Filkuková, P., & Klempe, S. H. (2013). Rhyme as reason in commercial and social advertising. *Scandinavian Journal of Psychology*, 54(5), 423–431. <https://doi.org/10.1111/sjop.12069>
- Firth, J. (1957). A Synopsis of Linguistic Theory 1930-1955. *Studies in Linguistic Analysis*.
- Freytag, G., & MacEwan, E. J. (1900). *Freytag's technique of the drama*. Scott Foresman.
- Gai, P. J., & Puntoni, S. (2021). Language and consumer dishonesty: A self-diagnostics theory. *Journal of Consumer Research*, 48(2), 333–351. <https://doi.org/10.1093/jcr/ucab001>
- Gopalan, P., Hofman, J. M., & Blei, D. M. (2013). Scalable recommendation with Poisson factorization. *arXiv preprint arXiv:1311.1704*.
- Grewal, D., Marmorstein, H., & Sharma, A. (1996). Communicating price information through semantic cues: The moderating effects of situation and discount size. *Journal of Consumer Research*, 23(2), 148–155. <https://doi.org/10.1086/209473>
- Grinstein, A., & Kronrod, A. (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), 423–441. <https://doi.org/10.1509/jmr.14.0224>
- Güntürkün, P., Haumann, T., & Mikolon, S. (2020). Disentangling the differential roles of warmth and competence judgments in customer-service provider relationships. *Journal of Service Research*, 23(4), 476–503. <https://doi.org/10.1177/1094670520920354>
- Hagtvedt, H. (2015). Promotional phrases as questions versus statements: An influence of phrase style on product evaluation. *Journal of Consumer Psychology*, 25(4), 635–641. <https://doi.org/10.1016/j.jcps.2014.12.005>
- Hamilton, R., Vohs, K. D., & McGill, A. L. (2014). We'll be honest, this won't be the best article you'll ever read: The use of dispreferred markers in word-of-mouth communication. *Journal of Consumer Research*, 41(1), 197–212. <https://doi.org/10.1086/675926>
- Han, S. P., & Shavitt, S. (1994). Persuasion and culture: Advertising appeals in individualistic and collectivistic societies. *Journal of Experimental Social Psychology*, 30(4), 326–350. <https://doi.org/10.1006/jesp.1994.1016>
- Humphreys, A. (2010). Megamarketing: The creation of markets as a social process. *Journal of Marketing*, 74(2), 1–19. <https://doi.org/10.1509/jm.74.2.1>
- Humphreys, A., Isaac, M. S., & Wang, R. J. H. (2021). Construal matching in online search: applying text analysis to illuminate the consumer decision journey. *Journal of Marketing Research*, 58(6), 1101–1119.
- Humphreys, A., & Wang, R. J. H. (2018). Automated text analysis for consumer research. *Journal of Consumer Research*, 44(6), 1274–1306. <https://doi.org/10.1093/jcr/ucx104>
- Joshi, P., & Kronrod, A. (2020). Sounds of green: How brand name sounds metaphorically convey environmental friendliness. *Journal of Advertising*, 49(1), 61–77. <https://doi.org/10.1080/00913367.2019.1696720>
- Karatas, M. (2020). Making decisions in foreign languages: Weaker senses of ownership attenuate the endowment effect. *Journal of Consumer Psychology*, 30(2), 296–303. <https://doi.org/10.1002/jcpy.1138>

- Kim, S., Moore, S. G., & Murray, K. (2021). "Don't buy" or "do not buy"? How negation style in online reviews influences consumer product evaluations. *Journal of Marketing Theory and Practice*, 29(3), 308–322. <https://doi.org/10.1080/10696679.2020.1839349>
- Klink, R. R. (2000). Creating brand names with meaning: The use of sound symbolism. *Marketing Letters*, 11(1), 5–20. <https://doi.org/10.1023/A:1008184423824>
- Kovacs, B., Carroll, G. R., & Lehman, D. W. (2013). Authenticity and consumer value ratings: Empirical tests from the restaurant domain. *Organization Science*, 25(2), 321–652.
- Krishna, A., & Ahluwalia, R. (2008). Language choice in advertising to bilinguals: Asymmetric effects for multinationals versus local firms. *Journal of Consumer Research*, 35(4), 692–705. <https://doi.org/10.1086/592130>
- Kronrod, A., & Danziger, S. (2013). "Wii will rock you!" The use and effect of figurative language in consumer reviews of hedonic and utilitarian consumption. *Journal of Consumer Research*, 40(4), 726–739. <https://doi.org/10.1086/671998>
- Kronrod, A., Grinstein, A., & Wathieu, L. (2012). Enjoy! Hedonic consumption and compliance with assertive messages. *Journal of Consumer Research*, 39(1), 51–61. <https://doi.org/10.1086/661933>
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25(2–3), 259–284. <https://doi.org/10.1080/01638539809545028>
- Lasswell, H. D., & Namenwirth, J. Z. (1969). The Lasswell value dictionary. New Haven.
- Laurino Dos Santos, H., & Berger, J. (2022). The speed of stories: Semantic progression and narrative success. *Journal of Experimental Psychology: General*, 151, 1833–1842. <https://doi.org/10.1037/xge0001171>
- Lawson, M. A., Martin, A. E., Huda, I., & Matz, S. C. (2022). Hiring women into senior leadership positions is associated with a reduction in gender stereotypes in organizational language. *Proceedings of the National Academy of Sciences*, 119(9), e2026443119. <https://doi.org/10.1073/pnas.2026443119>
- Lee, J. K. (2021). Emotional expressions and brand status. *Journal of Marketing Research*, 58(6), 1178–1196. <https://doi.org/10.1177/00222437211037340>
- Li, H., & Ma, L. (2020). Charting the path to purchase using topic models. *Journal of Marketing Research*, 57(6), 1019–1036. <https://doi.org/10.1177/0022243720954376>
- Li, X., Chan, K. W., & Kim, S. (2018). Service with emoticons: How customers interpret employee use of emoticons in online service encounters. *Journal of Consumer Research*, 45(5), 973–987. <https://doi.org/10.1093/jcr/ucy016>
- Liu, A. X., Xie, Y., & Zhang, J. (2019). It's not just what you say, but how you say it: The effect of language style matching on perceived quality of consumer reviews. *Journal of Interactive Marketing*, 46, 70–86. <https://doi.org/10.1016/j.intmar.2018.11.001>
- Lowe, M. L., & Haws, K. L. (2017). Sounds big: The effects of acoustic pitch on product perceptions. *Journal of Marketing Research*, 54(2), 331–346. <https://doi.org/10.1509/jmr.14.0300>
- Lowrey, T. M. (1998). The effects of syntactic complexity on advertising persuasiveness. *Journal of Consumer Psychology*, 7(2), 187–206. https://doi.org/10.1207/s15327663jcp0702_04
- Lowrey, T. M. (2006). The relation between script complexity and commercial memorability. *Journal of Advertising*, 35(3), 7–15. <https://doi.org/10.2753/JOA0091-3367350301>
- Lowrey, T. M., & Shrum, L. J. (2007). Phonetic symbolism and brand name preference. *Journal of Consumer Research*, 34(3), 406–414. <https://doi.org/10.1086/518530>
- Luangrath, A. W., Peck, J., & Barger, V. A. (2016). Textual paralinguistic and its implications for marketing communications. *Journal of Consumer Psychology*, 27(1), 98–107. <https://doi.org/10.1016/j.jcps.2016.05.002>
- Luangrath, A. W., Xu, Y., & Wang, T. (2022). Paralinguistic classifier (PARA): An algorithm for automatic coding of paralinguistic nonverbal parts of speech in text. *Journal of Marketing Research*. <https://doi.org/10.1177/00222437221116058>
- Ludwig, S., De Ruyter, K., Friedman, M., Brüggem, E. C., Wetzels, M., & Pfann, G. (2013). More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. *Journal of Marketing*, 77(1), 87–103. <https://doi.org/10.1509/jm.11.0560>
- Luna, D., & Peracchio, L. A. (2005). Advertising to bilingual consumers: The impact of code-switching on persuasion. *Journal of Consumer Research*, 31(4), 760–765. <https://doi.org/10.1086/426609>
- Maglio, S. J., Rabaglia, C. D., Feder, M. A., Krehm, M., & Trope, Y. (2014). Vowel sounds in words affect mental construal and shift preferences for targets. *Journal of Experimental Psychology: General*, 143(3), 1082–1096. <https://doi.org/10.1037/a0035543>
- Marinova, D., Singh, S. K., & Singh, J. (2018). Frontline problem-solving effectiveness: A dynamic analysis of verbal and nonverbal cues. *Journal of Marketing Research*, 55(2), 178–192. <https://doi.org/10.1509/jmr.15.0243>
- Melumad, S., Inman, J. J., & Pham, M. T. (2019). Selectively emotional: How smartphone use changes user-generated content. *Journal of Marketing Research*, 56(2), 259–275. <https://doi.org/10.1177/0022243718815429>
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26. <https://proceedings.neurips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html>
- Moore, S. G., & McFerran, B. (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), 229–245. <https://doi.org/10.1086/690942>
- Nave, G., Rentfrow, J., Bhatia, S. (2022). We are what we watch: Movie plots predict the personalities of their fans. Working Paper.
- Netzer, O., Lemaire, A., & Herzstein, M. (2019). When words sweat: Identifying signals for loan default in the text of loan applications. *Journal of Marketing Research*, 56(6), 960–980.
- Nielsen, J. H., Shapiro, S. A., & Mason, C. H. (2018). Emotionality and semantic onsets: Exploring orienting attention responses in advertising. *Journal of Marketing Research*, 47(6), 1138–1150. <https://doi.org/10.1509/jmkr.47.6.1138>
- Opoku, R., Abratt, R., & Pitt, L. (2006). Communicating brand personality: Are the websites doing the talking for the top South African business schools? *Journal of Brand Management*, 14(1), 20–39. <https://doi.org/10.1057/palgrave.bm.2550052>
- Ordenes, F., Grewal, D., Ludwig, S., Ruyter, K. D., Mahr, D., & Wetzels, M. (2019). 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. <https://doi.org/10.1093/jcr/ucy032>
- Packard, G., & Berger, J. (2017). How language shapes word of mouth's impact. *Journal of Marketing Research*, 54(4), 572–588. <https://doi.org/10.1509/jmr.15.0248>
- Packard, G., & Berger, J. (2020). Thinking of you: How second-person pronouns shape cultural success. *Psychological Science*, 31(4), 397–407. <https://doi.org/10.1177/0956797620902380>
- Packard, G., & Berger, J. (2021). How concrete language shapes customer satisfaction. *Journal of Consumer Research*, 47(5), 787–806.
- Packard, G., Li, Y., & Berger, J. (2022). When language matters. Working Paper.
- Packard, G., Moore, S. G., & McFerran, B. (2018). (I'm) happy to help (you): The impact of personal pronoun use in customer–firm interactions. *Journal of Marketing Research*, 55(4), 541–555. <https://doi.org/10.1509/jmr.16.0118>

- Packard, G., & Wooten, D. B. (2013). Compensatory knowledge signaling in consumer word-of-mouth. *Journal of Consumer Psychology*, 23(4), 434–450.
- Pancer, E., Chandler, V., Poole, M., & Noseworthy, T. J. (2019). How readability shapes social media engagement. *Journal of Consumer Psychology*, 29(2), 262–270. <https://doi.org/10.1002/jcpy.1073>
- Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. P. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, 108(6), 934–952. <https://doi.org/10.1037/pspp0000020>
- Pathak, A., Calvert, G. A., & Lim, L. K. (2020). Harsh voices, sound branding: How voiced consonants in a brand's name can alter its perceived attributes. *Psychology & Marketing*, 37(6), 837–847. <https://doi.org/10.1002/mar.21346>
- Patrick, V., & Haws, K. L. (2014). “And” bridges, “with” bonds: A lexical inferencing-based framework for influencing perceptions of product combinations. *Advances in Consumer Research*, 42, 136–140.
- Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us* (Vol. 211, pp. 42–45). Bloomsbury. [https://doi.org/10.1016/S0262-4079\(11\)62167-2](https://doi.org/10.1016/S0262-4079(11)62167-2)
- Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77(6), 1296–1312. <https://doi.org/10.1037/0022-3514.77.6.1296>
- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54(1), 547–577. <https://doi.org/10.1146/annurev.psych.54.101601.145041>
- Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1, Association for Computational Linguistics* (Long Papers), 2227–2237. <https://doi.org/10.18653/v1/N18-1202>
- Pezzuti, T., Leonhardt, J. M., & Warren, C. (2021). Certainty in language increases consumer engagement on social media. *Journal of Interactive Marketing*, 53, 32–46. <https://doi.org/10.1016/j.intmar.2020.06.005>
- Phillips, B. J., & McQuarrie, E. F. (2009). Impact of advertising metaphor on consumer belief: Delineating the contribution of comparison versus deviation factors. *Journal of Advertising*, 38(1), 49–62. <https://doi.org/10.2753/JOA0091-3367380104>
- Pogacar, R., Angle, J., Lowrey, T. M., Shrum, L. J., & Kardes, F. R. (2021). Is Nestlé a lady? The feminine brand name advantage. *Journal of Marketing*, 85(6), 101–117. <https://doi.org/10.1177/0022242921993060>
- Pogacar, R., Shrum, L. J., & Lowrey, T. M. (2018). The effects of linguistic devices on consumer information processing and persuasion: A language complexity × processing mode framework. *Journal of Consumer Psychology*, 28(4), 689–711. <https://doi.org/10.1002/jcpy.1052>
- Preoțiuc-Pietro, D., & Ungar, L. (2018). User-level race and ethnicity predictors from twitter text. In *Proceedings of the 27th International Conference on Computational Linguistics*, 1534–1545.
- Reagan, A. J., Mitchell, L., Kiley, D., Danforth, C. M., & Dodds, P. S. (2016). The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, 5(1), 1, 31–12. <https://doi.org/10.1140/epjds/s13688-016-0093-1>
- Rick, S. I., Cryder, C. E., & Loewenstein, G. (2008). Tightwads and spend-thrifts. *Journal of Consumer Research*, 34(6), 767–782.
- Rocklage, M. D., & Fazio, R. H. (2020). The enhancing versus backfiring effects of positive emotion in consumer reviews. *Journal of Marketing Research*, 57(2), 332–352. <https://doi.org/10.1177/0022243719892594>
- Rocklage, M. D., Rucker, D. D., & Nordgren, L. F. (2018). The intent to persuade transforms language via emotionality. *Psychological Science*, 29(5), 749–760. <https://doi.org/10.1177/0956797617744797>
- Schmitt, B., & Zhang, S. (1998). Language structure and categorization: A study of classifiers in consumer cognition, judgment, and choice. *Journal of Consumer Research*, 25(2), 108–122. <https://doi.org/10.1086/209530>
- Sela, A., Wheeler, S. C., & Sarial-Abi, G. (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), 644–661. <https://doi.org/10.1086/664972>
- Septhri, A., & Berger J. (2022), The consequences of passive voice, Working Paper.
- Seraj, S., Blackburn, K. G., & Pennebaker, J. W. (2021). Language left behind on social media exposes the emotional and cognitive costs of a romantic breakup. *Proceedings of the National Academy of Sciences*, 118(7), e2017154118. <https://doi.org/10.1073/pnas.2017154118>
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54. <https://doi.org/10.1177/0261927X09351676>
- Templeton, E. M., Chang, L. J., Reynolds, E. A., Cone LeBeaumont, M. D., & Wheatley, T. (2022). Fast response times signal social connection in conversation. *Proceedings of the National Academy of Sciences*, 119(4), e2116915119. <https://doi.org/10.1073/pnas.2116915119>
- Terlutter, R., Diehl, S., & Mueller, B. (2010). The cultural dimension of assertiveness in cross-cultural advertising: The perception and evaluation of assertive advertising appeals. *International Journal of Advertising: The Quarterly Review of Marketing Communications*, 29(3), 369–399. <https://doi.org/10.2501/S0265048710201233>
- Timoshenko, A., & Hauser, J. R. (2019). Identifying customer needs from user-generated content. *Marketing Science*, 38(1), 1–20. <https://doi.org/10.1287/mksc.2018.1123>
- Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *Journal of Marketing Research*, 51(4), 463–479. <https://doi.org/10.1509/jmr.12.0106>
- Toubia, O. (2021). A Poisson factorization topic model for the study of creative documents (and their summaries). *Journal of Marketing Research*, 58(6), 1142–1158.
- Toubia, O., Berger, J., & Eliashberg, J. (2021). How quantifying the shape of stories predicts their success. *Proceedings of the National Academy of Sciences*, 118(26), e2011695118. <https://doi.org/10.1073/pnas.2011695118>
- Van Zant, A., Berger, J., Packard, G., & Wang, H. (2022). Getting to “yeah.” How pausing shapes helpfulness perceptions in conversation. Working Paper.
- Van Zant, A. B., & Berger, J. (2020). How the voice persuades. *Journal of Personality and Social Psychology*, 118(4), 661.
- Wang, C. X., & Zhang, J. (2020). Assertive ads for want or should? It depends on consumers' power. *Journal of Consumer Psychology*, 30(3), 466–485. <https://doi.org/10.1002/jcpy.1165>
- Wang, X., Lu, S., Li, X. I., Khamitov, M., & Bendle, N. (2021). Audio mining: The role of vocal tone in persuasion. *Journal of Consumer Research*, 48(2), 189–211.
- Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45(4), 1191–1207. <https://doi.org/10.3758/s13428-012-0314-x>
- Yeomans, M., Minson, J., Collins, H., Chen, F., & Gino, F. (2020). Conversational receptiveness: Improving engagement with opposing views. *Organizational Behavior and Human Decision Processes*, 160, 131–148. <https://doi.org/10.1016/j.obhdp.2020.03.011>

- Yeomans, M., Schweitzer, M., & Brooks, A. W. (2022). The conversational circumplex: Identifying, prioritizing, and pursuing informational and relational motives in conversation. *Current Opinion in Psychology*, 44, 293–302.
- Yin, D., Bond, S., & Zhang, H. (2017). Keep your cool or let it out: Non-linear effects of expressed arousal on perceptions of consumer reviews. *Journal of Marketing Research*, 54(3), 885–900. <https://doi.org/10.1509/jmr.13.0379>
- Yorkston, E., & De Mello, G. E. (2005). Linguistic gender marking and categorization. *Journal of Consumer Research*, 32(2), 224–234. <https://doi.org/10.1086/432232>
- Yorkston, E., & Menon, G. (2004). A sound idea: Phonetic effects of brand names on consumer judgments. *Journal of Consumer Research*, 31(1), 43–51. <https://doi.org/10.1086/383422>
- Zemack-Ruger, Y., Moore, S. G., & Fitzsimons, G. J. (2017). Just do it! Why committed consumers react negatively to assertive ads. *Journal of Consumer Psychology*, 27(3), 287–301. <https://doi.org/10.1016/j.jcps.2017.01.002>
- Zhang, Z., Li, J., Zhu, P., Zhao, H., & Liu, G. (2018). Modeling multi-turn conversation with deep utterance aggregation. *arXiv preprint arXiv:1806.09102*.

How to cite this article: Berger, J., & Packard, G. (2022). Wisdom from words: The psychology of consumer language. *Consumer Psychology Review*, 1–14. <https://doi.org/10.1002/arcp.1085>