American Psychologist

© 2021 American Psychological Association
ISSN: 0003-066X

2022, Vol. 77, No. 4, 525–537
https://doi.org/10.1037/amp0000882

Using Natural Language Processing to Understand People and Culture

Jonah Berger[1] and Grant Packard[2]
[1] Wharton School, University of Pennsylvania
[2] Schulich School of Business, York University

Language can provide important insights into people, and culture more generally. Further, the digitization of information has made more and more textual data available. But by itself, all that data are just that: data. Realizing its potential requires turning that data into insight. We suggest that automated text analysis can help. Recent advances have provided novel and increasingly accessible ways to extract insight from text. While some psychologists may be familiar with dictionary methods, fewer may be aware of approaches like topic modeling, word embeddings, and more advanced neural network language models. This article provides an overview of natural language processing and how it can be used to deepen understanding of people and culture. We outline the dual role of language (i.e., reflecting things about producers and impacting audiences), review some useful text analysis methods, and discuss how these approaches can help unlock a range of interesting questions.

Public Significance Statement
This article offers an integrative discussion of how automated text analysis can be used to shed light on people and culture. It reviews recent methods and explains how they can be applied by nonspecialists to answer a range of research questions.

Keywords: natural language processing, automated text analysis, language, cultural success, culture

Supplemental materials: https://doi.org/10.1037/amp0000882.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Language is everywhere. It is how people express thoughts, communicate with others, and consume news, stories, and information. It is how parents parent, leaders lead, and salespeople sell. Language is how doctors communicate with patients, researchers communicate with study participants, and policymakers persuade the public.

Not surprisingly then, language has the potential to tell us a lot about people and culture. It can provide insight into who people are (e.g., personality), how they are feeling, and their attitudes, opinions, and reactions. Further, when aggregated across individuals, language can shed light on differences between groups or sociocultural contexts, and why some things (e.g., products or ideas) catch on.

But realizing that potential requires the right tools. The digitization of information has made more and more textual (i.e., language) data available. People write reviews, post online, and chat with friends, all of which provide information on people and the prevalence of stereotypes, innovations, and ideas. Books, songs, news articles, and movies provide a wealth of information on various cultural phenomena (e.g., Berger & Packard, 2018; Michel et al., 2011; Reagan et al., 2016).

By itself, though, all this data is just that—data. For this data to be useful, researchers must be able to parse it to extract insight.

This is where natural language processing (NLP) comes in. Automated text analysis is a computer-assisted NLP approach to quantify the information contained in text. These methods allow researchers to not only track the presence or prevalence of particular terms and ideas, but also to measure relationships between them, and how those relationships change over time. In short, natural language processing provides a powerful tool to help understand people and culture.

This article provides an overview of natural language processing and how it can be used to deepen understanding. First, we delineate the dual role of language and how it both reflects and impacts. Second, we review automated methods for extracting insight from text. While some psychologists

This article was published Online First December 16, 2021.
Jonah Berger https://orcid.org/0000-0002-2669-8211
Grant Packard https://orcid.org/0000-0002-8765-1994
Correspondence concerning this article should be addressed to Jonah Berger, Wharton School, University of Pennsylvania, 700 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104, United States. Email: jberger@wharton.upenn.edu

Jonah Berger

may be familiar with dictionary-based methods (e.g., Linguistic Inquiry and Word Count; Pennebaker et al., 2015), newer methods from computer science and statistics (e.g., topic modeling, embeddings, and neural network language models) have received less attention.[1] We explain these methods and how they might be used by nonspecialists. Along the way we discuss how these approaches can be applied to address a range of interesting questions.

Note that complete detail on how to do textual analysis is beyond the scope of this article. Many important topics should be carefully considered when acquiring textual data (e.g., how to scrape data from websites), cleaning and organizing it (e.g., tokenization, stemming, and removing stop words; Kern et al., 2016), and analyzing it (e.g., multiple hypothesis testing). While these topics are increasingly accessible to nonspecialists, they require more space than is available here, though see the online supplemental materials for some relevant resources.

The Dual Role of Language

Language serves a dual role: It both (a) reflects things about the person or people that produced it and (b) impacts or influences the audience that consumes it.

Language Reflects the Producer

Language can be viewed like a signature or fingerprint (Pennebaker, 2011). Different people use words differently and language can provide insight into the attentional focus of language producers (Boyd & Schwartz, 2021) as well as their states, traits, values, and personality (Boyd et al., 2020; Pennebaker et al., 2003). Language use differs by gender (Mehl & Pennebaker, 2003), age (Sap et al., 2014), and political affiliation (Sterling et al., 2020), for example, and can even signal things like an impending breakup (Seraj et al., 2021).

As a result, the language produced by a group or sociocultural context reflects or indicates things about the group or context that produced it (e.g., Holtgraves & Kashima, 2008; Morling & Lamoreaux, 2008).[2] Consistent with cross-cultural variation in the value of harmony and differentiation, for example, Korean advertisements emphasize conformity, while American ads emphasize uniqueness (Kim & Markus, 1999). Consistent with social class differences in the meaning of choice, the words used in car advertisements targeting working class individuals emphasize connecting with others, while ads targeting middle class individuals emphasize differentiation (Stephens et al., 2007).

Because people and cultures change over time, language also provides dynamic insight. At the individual level, this manifests through psychological states, such as how someone is feeling (Schwartz et al., 2014), but aggregated across people, analyzing language over time can provide insight into if, and how, cultures are changing. When referencing the two sexes, for example, news articles and books tend to put men before women (e.g., “men and women” rather than “women and men”), though this has reduced over time (Kesebir, 2017). Other work has used language to examine misogyny in music (Boghrati & Berger, 2020) and social class (Kozlowski et al., 2019).

In summary, because language reflects things about the people and groups that produce it, analyzing language can not only provide insight into individual differences and psychological states, but also sociocultural differences and how culture and cultures change over time (i.e., cultural analytics).[3]

Language Impacts the Audience

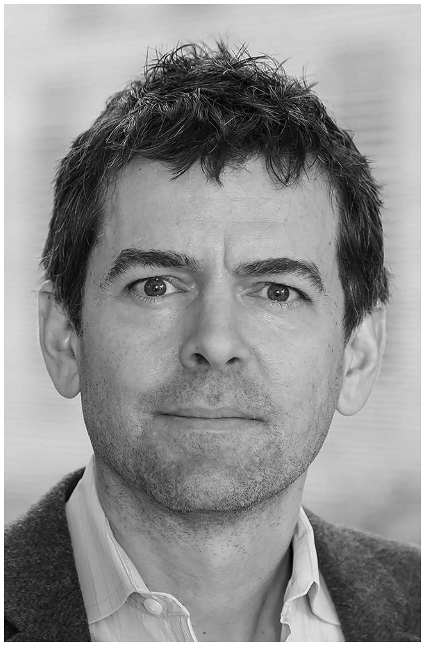
Beyond reflecting things about its producers, language also impacts the audience that is exposed to it. At the

[1] While some cognitive scientists have used these tools (e.g., Huth et al., 2016; Polyn et al., 2009; Steyvers et al., 2006) they have seen less widespread attention across areas of psychology.

[2] Cultural items play a critical role in the mutual constitution of culture and self (Markus & Hamedani, 2019). People learn norms, practices, and ways of being through social ties, but also through books, songs, and other cultural items. Because such items are shaped by the sociocultural context in which they are created, they carry meaning, reinforcing and propagating ways of being. American children’s books, for example, tend to reflect American values and biases. Consequently, American children are more likely to be exposed to, and adopt, sociocultural consistent values and stereotypes, and pass them on to others.

[3] Research in the digital humanities has also applied methods from computational linguistics to study cultural artifacts (e.g., Moretti, 2013). For surveys of the literature, see Jänicke et al. (2015), Underwood (2015), Gold (2012), Berry (2012). Also see Bail (2014) and Kozlowski et al. (2019) for recent discussions of extracting cultural insight from text.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.



Grant Packard

individual level, words, phrases, and other linguistic aspects can shape everything from memory and sustained attention to evaluation and social transmission. Phrases that relate to senses in metaphoric ways (e.g., cold person) are easier to remember (Akpinar & Berger, 2015), for example, and uncertain language increases attention (Berger et al., 2021). Questions shape interpersonal evaluations (Huang et al., 2017) and disgust and other high arousal emotions increase sharing (Berger, 2011; Berger & Milkman, 2012; Heath et al., 2001).

Consequently, at the collective level, language can shape whether cultural items (e.g., books or ideas) succeed or fail. Just as natural selection shapes evolution and success in the biological world, some (e.g., Dawkins, 1976; Heath et al., 2001) have suggested that processes of variation and selection shape cultural success. Some stories are longer while others are shorter. Some narratives are more circuitous while others are more direct. Some songs are thematically typical while others are more atypical.

The success of these different variants depends, in part, on how they fit with people. Sociocultural background shapes individual-level psychological process (e.g., cognition & attribution; Markus & Kitayama, 1991), but the reverse is also true. When shared across individuals, psychological processes can act as a selection mechanism, shaping the content of collective culture (Heath et al., 2001; Kashima, 2008; Schaller & Crandall, 2004). Processes of evaluation, memory, and transmission shape which items are liked, retained, and shared, and as a result, which items become popular and how culture evolves. More communicable traits, for example, are more likely to persist in ethnic stereotypes (Schaller et al., 2002) and minimally counterintuitive narratives are more popular (Norenzayan et al., 2006). Ideas are more successful in times when they are cued more frequently by the environment, and ideas with more prevalent habitats (i.e., more frequent cues), are more successful overall (Berger & Heath, 2005). More disgusting urban legends are more widely distributed (Heath et al., 2001) and news articles that evoke high arousal emotion are more likely to go viral (Berger & Milkman, 2012).

Taken together, research on language's impact suggests ways to encourage attention, persuasion, or memory. Further, it sheds light on cultural success. Because language impacts the audience that consumes it, items that include certain types of language may be more successful.

Unlocking the Potential of Text

Whether studying what language reflects, or how it impacts, two key shifts have greatly facilitated work in this space. The first is access to data. Technological changes have made it faster and easier to access a wealth of language-related data (see online supplemental materials). Millions of reviews, messages, and other content is posted online (Kern et al., 2016). Movie scripts, song lyrics, books, newspaper articles, and other content have been digitized, enabling researchers to comb hundreds of millions of words from millions of texts. Even everyday conversations can be recorded and turned into data.[4]

But parsing this data can be challenging. Manually reading content and coding the presence of words or themes used to be the main approach. But manual methods are difficult to scale. Having individuals read stories takes time and twice the number of stories takes twice as long. Further, because manual coding relies on human judgment, it is susceptible to bias.

Automated textual analysis, however, can help address these challenges. Below we review several useful techniques including dictionaries, topic modeling, embeddings, and more advanced neural network language models.

Dictionaries

Dictionaries are a simple way to begin to extract features from language data. This approach takes a predefined lexicon, or list of words, and searches texts for their presence.[5]

Some dictionaries count words in a category. Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2015), one of the most well-known dictionaries, scores texts on

[4] This shift has been particularly beneficial for cultural analytics (e.g., why some things catch on). Surveys are difficult to collect over decades. Further, it can be challenging to measure key features of cultural items (i.e., how much emotion they evoke). Natural language processing tools, combined with the digitization of content, have made it easier to measure the prevalence of cultural items (e.g., stereotypes) and how they change over time. Further, it has facilitated feature extraction from those items to help understand why some items may be more successful.

[5] Links to several dictionaries used for automated text analysis are provided in the online supplemental material.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

over 50 dimensions (e.g., pronouns and cognitive processes). The sentence "He is hungry now" would receive a 25% score on the pronoun dimension, for example, because one of the four words in the sentence is a pronoun (i.e., "he"). Each dimension is made up of a list of relevant words determined by judges and prior research. LIWC has been used to explore subjects like interracial feedback (Harber et al., 2019), self and other focus (Barasch & Berger, 2014), and racial disparities in police language (Voigt et al., 2017). Other word count-based dictionaries have been used to measure things like warmth and competence (Nicolas et al., 2021), gender stereotypes (Gaucher et al., 2011), references to nature (Kesebir & Kesebir, 2017), and liberal and conservative language (Neiman et al., 2016).

Other dictionaries use a more continuous approach. Rather than focusing on the count or percentage of words in a category, words are continuously scored on a particular dimension (e.g., love is a more positive word than hate). Brysbaert et al. (2014), for example, had participants rate 40,000 English words based on how abstract or concrete they were. Words like pitbull were rated as more concrete than words like essentialness. Similar approaches have been used to measure features such as dominance (Warriner et al., 2013) and attitude extremity (Rocklage et al., 2018).

Researchers can also create custom dictionaries. Given a particular concept and context, one can specify words or phrases that seem relevant, find synonyms, and use off the shelf packages, simple coding, or even spreadsheet software to measure their presence (Humphreys & Wang, 2018). Previously validated dictionaries increase validity, but they may miss some important words in a particular context, so some customization may be beneficial.

Dictionaries can be used to study both what language reflects and how it impacts. On the reflection side, dictionaries have been used to understand whether someone might be depressed (Eichstaedt et al., 2018), going through a breakup (Seraj et al., 2021), or whether cultural items from different sociocultural contexts vary on key dimensions (e.g., Kim & Markus, 1999; Snibbe & Markus, 2005). On the impact side, dictionaries have been used to understand gender differences in entrepreneurial fund raising (Huang et al., 2020) and how psychological processes (e.g., arousal) shape what content goes viral (Berger & Milkman, 2012).

Dictionaries can also be used as inputs to other methods. Once particular words have been quantified, they can be used to help calculate the similarity between texts (or parts of them). Linguistic style matching, for example, measures verbal coordination by analyzing how frequently different people use different types of function words (e.g., pronouns). Conversation partners tend to stylistically match one another (Niederhoffer & Pennebaker, 2002) and linguistic style matching predicts group cohesion (Gonzales et al., 2010), relationship stability (Ireland et al., 2011), and negotiation success (Taylor & Thomas, 2008).

Similar approaches can be applied more broadly. Danescu-Niculescu-Mizil et al. (2013) examined user lifecycles and linguistic change in online communities. They examined language use over time, and by measuring linguistic distance (i.e., similarity between new users' and community's language), investigated patterns of enculturation. While users initially adopted community language, eventually most stopped doing so as the community and its norms kept evolving. Further, linguistic distance predicted future engagement (i.e., whether the user stayed engaged in the community or left).

Srivastava et al. (2018) used a similar approach to examine enculturation in organizations. Analyzing over 10 million internal company emails showed a link between cultural adaptation and occupational outcomes. Employees with better cultural fit (i.e., more similar linguistic style to others in the firm) were more likely to be promoted, while those who were slower to enculturate (i.e., adapt others' linguistic style) were more likely to be fired. Employees who adapted initially but diverged later on were more likely to end up quitting.

These examples highlight some intriguing aspects of what words reflect. First, just as groups have behavioral norms, they have linguistic norms as well. Aggregated together, the words used by group members provides insight into how that group tends to communicate. Second, norm adherence can be measured through linguistic similarity, or the distance between any given individual and the group. Third, such distance provides insight into one's integration into the group, valuation by the group, and likely engagement with the group in the future.

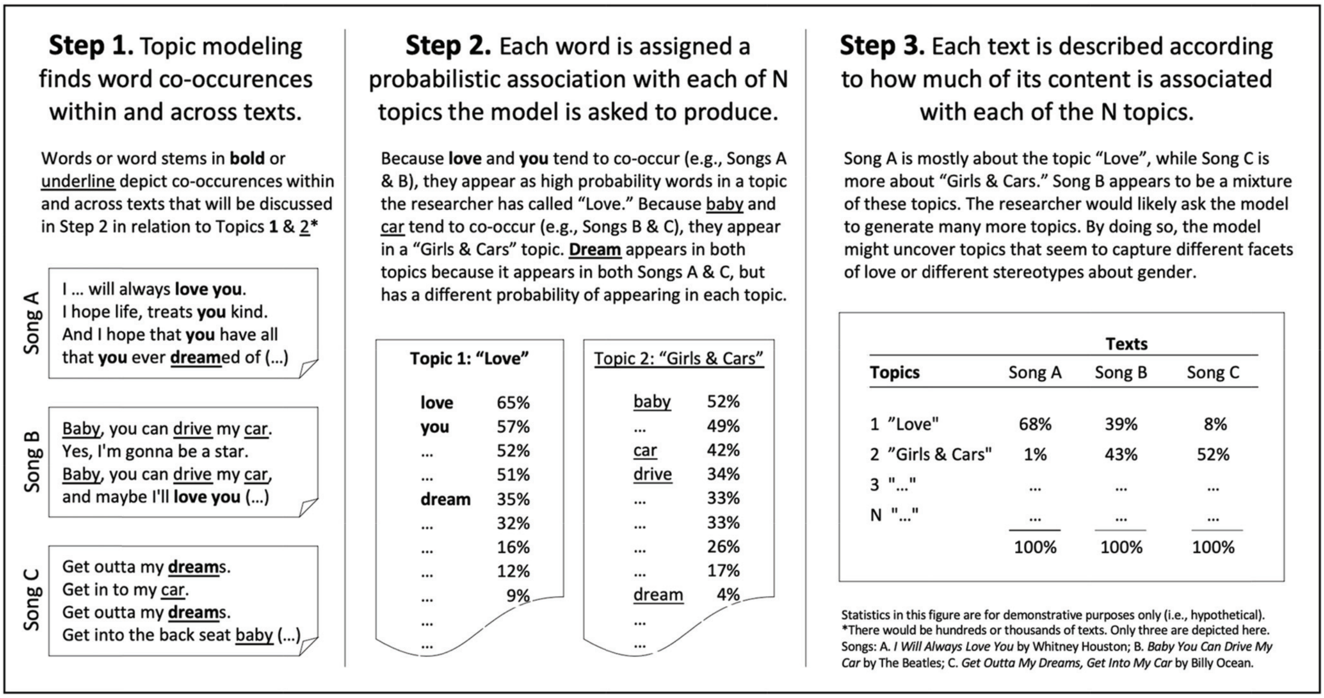
Measuring similarity opens up a range of avenues for future research. Researchers can compare the distance between people, groups, or sociocultural contexts, and how they change over time (see Bail et al., 2019). Linguistic similarity predicts friendship, for example, and friends also exhibit linguistic convergence (Kovacs & Kleinbaum, 2020). Comparing the language used in children's books from different countries, or the themes discussed in different literary genres, may provide insight into which are more similar and why.

Topic Modeling

Dictionaries can be useful, but topic modeling, or topic extraction, takes a more bottom up approach. Rather than relying on a prespecified list of words or phrases (i.e., closed vocabulary), topic modeling uses the structure of the data to identify the main themes in a body of text (i.e., open vocabulary, Eichstaedt et al., 2021; Kern et al., 2016). Similar to how factor analysis identifies underlying groupings among survey items by measuring the co-occurrence of words within and across texts, topic modeling identifies the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 1
Simplified Illustration of How Topic Modeling Works



latent themes or topics being discussed, and the words that make up each theme or topic (e.g., Wilson et al., 2016).[6]

Take song lyrics. While one could use dictionaries to measure how concrete a song's lyrics are, or whether the song uses many social words, those may or may not be the most relevant features. Instead, topic modeling starts with the data to discover the latent themes.

See Figure 1 for a simplified illustration of how topic modeling works. Given a set of texts (e.g., songs), Latent Dirichlet Allocation (LDA; Blei et al., 2003), a common topic modeling approach, uses machine learning to identify words that co-occur (i.e., appear together) both within and across texts. Songs that contain the word love, for example, may also tend to include words like you. Songs that contain the word car may also tend to contain words like baby. Based on co-occurrences, different clusters of words would be probabilistically assigned to different topics, with words more strongly associated with that topic receiving a higher weight. Finally, each text (i.e., song) is scored based on how much of each topic it contains.

Once topics have been identified, several questions can be asked. Researchers interested in what language reflects could compare what topics different types of people are talking or writing about, or analyze cultural items from different sociocultural contexts to shed light on differences between them. Analyzing song lyrics, for example, shows that while Dance and Rock songs talk a lot about fiery love, Pop songs talk more about uncertain love (Berger & Packard, 2018). Alternatively, one could examine variation over time (e.g., whether certain topics wax or wane in song lyrics and what that may indicate about cultural change). By allowing the themes to emerge from the data, rather than predetermining categories (i.e., using dictionaries), one may identify differences or changes that may not have been anticipated in advance.

Researchers can also use these tools to understand what becomes popular and why. Given American values of independence, for example, songs that talk more about independence may be more popular. Similarly, social media content is more likely to generate comments if it combines conversational themes that are not usually discussed together (Bail, 2016).

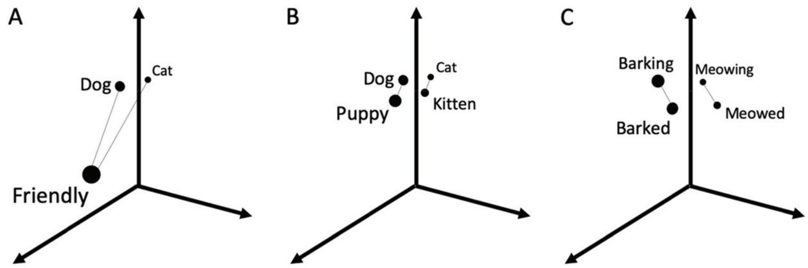
Once calculated, topics can also be used as inputs to measure similarity. Similar to groups, genres or categories have norms. Consequently, norm adherence can be used to understand why some things succeed or fail. Berger and Packard (2018) used topic modeling to identify the main themes in song lyrics, and how similar a given song is to its genre. Even controlling for a host of other things, more atypical songs ranked higher on the Billboard charts.

Similar approaches could be applied to other domains. Do successful movies tend to be similar or different from their genres? Are academic articles cited more if their mix

[6] While topic models are commonly run using off-the-shelf packages in R or Python, free online resources (e.g., http://textanalyzer.org/) permit less technical users to simply upload a text file and generate LDA results.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 2
Finding Relationships in Semantic Space With Word Embeddings



Note. Distances depicted are for illustrative purposes only.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

of themes is more similar to or different from other articles published in the same journal?

This raises the broader question of when similarity or difference should be beneficial. The value of novelty (Berlyne, 1970), and drive for stimulation (Zuckerman, 1979), suggests that atypicality should boost success. Exposure can lead to liking (Zajonc, 1968), however, and there are also benefits of familiarity (Kunst-Wilson & Zajonc, 1980). So which should dominate?

Some of this may depend on the domain. While atypical songs were generally more successful, this effect disappeared among Pop songs, a genre that is almost by definition more about mainstreaming than differentiation (Berger & Packard, 2018). One could imagine similar effects among movies or books. While atypicality may increase success overall, adhering to conventions may be more beneficial in some genres (e.g., action movies or romance films).

Successful cultural items may also mix and match, simultaneously similar in some respects (e.g., musical chords) and different in others (e.g., lyrics; see Berger et al., 2012, for related ideas). Being similar enough to evoke the warm glow of familiarity but differentiated enough to feel new and exciting.

Embeddings

Advances in computer science have provided even richer ways to measure similarity. Rather than comparing the prevalence of words or themes across documents, these approaches capture the relationship or distance between contextually related words or larger chunks of text.

Which is more related to grapefruit: kiwi, orange, or tiger? While this question is easy for people to answer, dictionaries or topic models may struggle. If there was a dictionary for fruit, it would be clear that tiger is not a member, while orange and kiwi are, but that would not suggest which of these two fruits are more related to grapefruit. Topic modeling also tends to be rather binary. If grapefruit and orange or kiwi often co-occur across documents, they may be included in the same topic. But if none of them appear in the same topic, it is hard to know which are more related. Further, simpler measures of word co-occurrence within a large document (e.g., Latent Semantic Analysis, LSA) may miss some nuance.[7] Two words that appear in close proximity (e.g., the same clause or paragraph) are probably more related than two words that appear in completely different sections of a document.

A computational linguistics approach called word embeddings addresses some of these issues (e.g., Word2vec, Mikolov et al., 2013 and GloVe, Pennington et al., 2014). Similar to Firth's (1957) famous suggestion that "you shall know a word by the company it keeps" (i.e., the distributional hypothesis), this neural network framework takes a corpus of text (e.g., books), and uses the local context in which words appear to determine how semantically related they are (see Eichstaedt et al., 2021, for a recent discussion in psychology). Just as people who hang out more often tend to live closer in geographic space, and social network analysis places people who have the same friends closer together in network space, word embeddings use word co-occurrence, distance between words, and word appearance in similar contexts across different texts to continuously position words in relation to one another in a multidimensional semantic space. By incorporating the distance between words within documents, word embeddings are able to achieve greater relational insight.

If the phrase "dogs are friendly" appears more frequently in a corpus of news articles than "cats are friendly," for example, the words "dog" and "friendly" would be seen as more related and would be located slightly closer together in space (Figure 2A). Beyond just co-occurrence, though, how closely words appear to one another also matters. The phrases dogs are friendly and "dogs love running and are also friendly" both contain the words dog and friendly, but

[7] Latent Semantic Analysis (i.e., LSA; Foltz, 2007; Graesser et al., 2004) analyzes the relationship between sentences, paragraphs, or whole texts by counting how many times different words appear in each. By constructing a matrix in which rows are unique words (or word stems) and columns are each document, it measures the similarity between documents by calculating the angle between their vectors. LSA has been used to measure the distance between adjoining paragraphs of books (e.g., Foltz, 1998) and the semantic evolution of thoughts to predict creativity (Gray et al., 2019).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

the first phrase has them closer together, which would lead the two words to be placed closer together. Words do not even have to necessarily co-occur; a word's presence across similar contexts helps shape that word's positioning in space. If the phrases "dogs are animals" and "animals are friendly" both appear frequently, the words dog and friendly would be placed closer together (than would, e.g., "cat" and "friendly"), even if the two words never directly co-occurred.

Because semantically related words are plotted closer together, one can measure the relationship between words by the distance between them. Consequently, embeddings can be a powerful tool to study people and culture. Examining the distance between different occupations and male and female words, for example, found that news articles exhibit strong gender stereotypes (Bolukbasi et al., 2016). Women were more likely to be associated with occupations like homemaker and receptionist, while men were more likely to be associated with occupations like captain or boss.

The same approach can be used to compare sociocultural contexts. Consistent with the principle of linguistic relativity, for example, women are more likely to be more associated with negatively valanced words in gendered languages (DeFranza et al., 2020).

Embeddings can also be used to help explore linguistic analogues by uncovering parallels in semantic space. For example, one can apply algebraic equations to embeddings to discover analogues for how people talk about a particular idea (e.g., dogs or cats) moderated by some other construct of interest (e.g., age or time; "dog minus old" may return "puppy," Figure 2B; "barking minus is" may return "barked," Figure 2C), although accuracy for predicting semantic and morphological analogues varies (Gladkova et al., 2016). Similar embedding equations for cat-related ideas may produce results that are mathematically, and semantically, parallel to those for the dog-related terms (Figure 2B and 2C).

Word embeddings are particularly useful for examining such shifts, and consequently, they can be used to study change. Garg et al. (2018), for example, analyzed the evolution of gender stereotypes and attitudes toward ethnic minorities over the past 100 years by tracking changes in distance between different adjectives, occupations, and words related to gender and ethnic groups. They found that linguistic shifts closely tracked demographic and occupational shifts. Analyzing 100 years of books shed light on how discussions of social class have shifted over time (Kozlowski et al., 2019).

More Advanced Neural Network Language Models

Dictionaries, topic modeling, and word embeddings can be quite useful for studying individual words, how culture influences their meanings, and the associations they elicit. But for more complex expressions of language, more accurate models of language may be valuable. Books, songs, movies, and other cultural artifacts can be distinguished from other forms of natural language by their structure, long-distance relationships between constituent parts (i.e., the meaning of a word or sentence often depends on the language that came before it), and other higher-order statistical features that word-level models do not capture well.

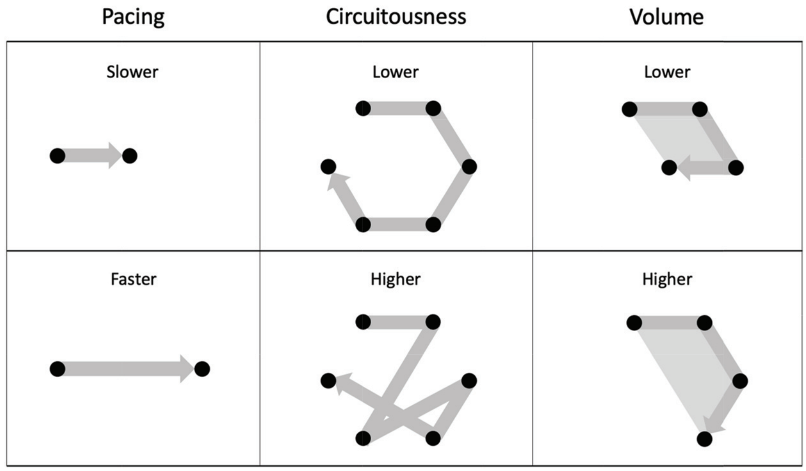
Fortunately, the last several years have seen the rapid development of neural network language models (e.g., BERT, Devlin et al., 2019; ELMo, Peters et al., 2018; and GPT-3, Brown et al., 2020) that vastly outperform prior approaches on a range of benchmarks. Language models are statistical tools that can predict language by mapping the probability with which particular words follow other words. If the word "wild" is often followed by the word "roses," for example, the model might use the incidence of the former in a text to predict that the latter is more likely to appear. This same type of analysis can then be applied to sentences or even entire paragraphs, allowing the model to predict how likely a given sentence is to follow a previous one. These models can help autocomplete sentences, answer questions, summarize documents, correct audio recordings, improve machine translation, and even generate novel content.

These approaches also have distinguishing features that make them particularly useful for analyzing complex language. First, they generate state-of-the-art results approaching human-level performance. These models have an unprecedented capacity to model the statistical structure of things like conversations, news articles, poems, and songs. When given a reading comprehension task, for example, GPT-3 performed almost as well as people (Brown et al., 2020). Similarly, when given a short prompt, GPT-3 was able to write high-quality news articles from scratch that were difficult to distinguish from those written by humans (Brown et al., 2020).

Second, they can do so even in settings that involve zero- or few-shot learning. Prior approaches were often pretrained on a large corpus (e.g., all Wikipedia articles) and then fine-tuned on whichever specific task they were being applied to (e.g., question answering). Such fine-tuning often required thousands or hundreds of thousands of task-specific examples, however, and requiring such large supervised training made these approaches difficult to apply. But by including many more machine learning parameters (e.g., 175 billion), and training on billions of words or tokens, recent models have been able to achieve so called "metalearning" such that they do not need to be retrained to perform a more specific task. Rather than requiring thousands of labeled examples, GPT-3 is able to perform quite well on many tasks with just a task description and prompt (and in some cases, a few examples). No fine-tuning required.

Third, these approaches take context into account. Simpler approaches often treat words as the unit of analysis.

Figure 3
Simplified Illustrations of Features of Semantic Progression



Note. Dots represent different points or "chunks" of discourse. Pacing captures how quickly things move in a single period. Circuitousness captures how directly or indirectly a narrative travels between the same points. Volume captures how much ground is covered.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

The word "bank," for example, might appear in a dictionary of financial terms, and topic modeling might group it with words like money, teller, or check. But the same word can mean different things in different contexts. The word bank, can refer to a financial institution, for example, or the side of a river. Consequently, ignoring the broader context in which individual words are situated can lead to imprecise inferences. By taking context into account, however, and doing so across longer spans of text, these newer neural network language models more accurately represent the meaning of language. Because they capture long-distance dependencies, these models may be particularly useful for understanding things like narrative or discourse structure and the meaning of longer cultural artifacts (e.g., novels or movies) in their entirety. Given the breadth of data on which these models are trained, one could even argue that the semantic space they delineate captures the culture of the current world.

Researchers are just beginning to leverage these tools, but one useful application may be quantifying discourse structure. Literary theorists have long argued that stories have common shapes (Freytag & MacEwan, 1900), but little empirical work has tested this possibility (Boyd et al., 2020). Further, while some work has begun to examine narratives using manual coding (McLean et al., 2020) or dictionary-based methods (Berger et al., 2021; Boyd et al., 2020; Reagan et al., 2016), less is known about how aspects of semantic progression may shape success (i.e., evaluations or sales).

Neural network language models may be able to provide insight. Just as embeddings can be used to measure similarity between words, larger chunks of text (i.e., sentences or paragraphs) can be represented as points in a multidimensional space, and more advanced tools can estimate how likely one chunk is to follow another. Then, by characterizing the relationship between adjoining points, or the set of points as a whole, one can extract features of narrative progression.

Take pacing, or how fast a narrative or discourse is moving. Adjacent chunks of discourse (i.e., adjoining paragraphs of a book or scenes of a movie) tend to be more semantically related than chunks that are further apart (Foltz, 2007; Toubia et al., 2021). A scene about two people getting married, for example, tends to be followed by a related scene (e.g., the afterparty) rather than a completely unrelated one (e.g., different characters robbing a bank). But while adjacent chunks tend to be more semantically related, how related varies across narratives (Foltz, 1998). Some narratives stick on a particular idea or theme for a while, while others move more quickly from one idea to the next (i.e., faster semantic progression, Figure 3). By measuring the average semantic similarity between adjoining passages of text, one can measure the speed of semantic progression (Laurino Dos Santos & Berger, 2020).

Similarly, looking at a sequence of chunks of a text can shed light on whether it takes the most direct path between different points or a more circuitous route (see Figure 3; Toubia et al., 2021). A story that goes from a wedding to the afterparty to different characters robbing a bank, for example, takes a more direct route than one that goes from the wedding to different characters robbing a bank, and then back to the afterparty.

Finally, by wrapping a shape around the set of points, one can measure volume or how much ground a discourse covers (Toubia et al., 2021). Does it focus only on a small set of things that are closely related (i.e., lower volume), or does it cover a wider set of ideas that may be less closely related (i.e., higher volume; Figure 3)?

Pacing, volume, and circuitousness are just a few features of semantic progression, but the ability to quantify such aspects opens up a range of interesting questions. Researchers could test whether narratives or texts that move more quickly are more successful, for example, because they provide more stimulation. Other work could test whether the link between volume and success depends on the length of the narrative. Covering a lot of ground may be good for a longer narrative (e.g., movie), for example, but detrimental for shorter ones (e.g., TV show). Similarly, while circuitousness may be bad in some cases, it may be good in others. More circuitous academic papers may be cited more, for example, because they make it easier to integrate disparate information (Toubia et al., 2021).

Similar ideas could be applied to characters. Are books and movies more successful, for example, when the main characters undergo significant changes throughout the course of the narrative? And if so, are certain types of shifts more well received? Tracking the language used by, and around, different characters over the course of a story may provide insight.

Discussion

Language is pervasive. Hardly a waking hour goes by where people are not creating or consuming language in some shape or form. Language reflects things about the people and socioeconomic context that produce it and impacts the audiences that consume it. Consequently, language has the potential to shed light on both people, and culture more generally.

Realizing that potential though, requires the right tools. That is where natural language processing comes in. Not only can these approaches parse features of language, and do so in a relatively objective way, but they can do so at scale. Consequently, these methods can shed light on a range of interesting questions.

Directions for Future Work

Given pronouns shift people's perceptions of relatedness (Fitzsimons & Kay, 2004), self and identity researchers might analyze pronoun use to understand variation in self-concepts and how people manage their self ("I") versus social ("we") identity. Motivation researchers could examine how employees talk about affiliation or achievement to understand what drives career aspirations and success. And construal level researchers could use dictionaries related to concreteness or specificity (e.g., Coltheart, 1981; Li & Nenkova, 2015) to better understand psychological distance (e.g., Snefjella & Kuperman, 2015).

Researchers studying attitudes and persuasion could use topic modeling to understand the main themes used in persuasive speech, and whether certain themes are more impactful. Researchers studying motivation could use topic modeling to examine the themes or approaches people use to discuss self-regulation, and how they differ across people or over time.

Embeddings offer similar possibilities. Emotion researchers could use embeddings to test the structure of emotion (see Jackson et al., 2019) and whether or how appraisals of discrete emotions shift across situations. Cognitive psychologists could use embeddings to capture perceptions of risk (Bhatia, 2019) and judgments of meaning, relatedness, and probability (see Bhatia et al., 2019, for a review). Because embeddings represent how information is organized and retrieved, memory and information processing researchers could use this approach to explore how certain ideas become linked in memory (Bhatia & Walasek, 2019).

The methods discussed here should also be useful for assessing temporal trends and historical roots of psychological phenomena. While surveys and experiments can capture what people think and feel now, they cannot be used retroactively. It is impossible to go back in time and survey people 10 years ago, for example, or before and after a major social or political event. But like insects preserved in amber, by analyzing language created at previous time points, it is possible to get some sense of attitudes, cognitions, and other aspects, and how they vary over time. Psychologists are just beginning to tap historical databases, and the new field of "historical psychology" (e.g., Muthukrishna et al., 2021) uses historical texts and artifacts to understand and explain changes in the drivers of cognition and behavior. Natural language processing allows researchers to parse this relatively unstructured data in a quantitative manner to extract psychological insights.

Language may be particularly useful when trying to measure attitudes, stereotypes, and biases that are otherwise challenging to capture. Researchers have long been interested in implicit attitude measures because they capture things that traditional self-report measures might not. Using automated text analysis to examine produced language may provide an alternative approach. Even if people do not make explicitly racist statements, word associations can provide evidence of biases (Caliskan et al., 2017).

This discussion begs the broader question of when language-based measures are more versus less useful. Language may be particularly useful in cases where other measures are difficult to collect (e.g., capturing things that have occurred in the past), people have less insight into their attitudes or preferences, or response bias may lead to inaccurate or false answers in undisguised measures (e.g., trait or state scales). That said, using the right language measures remains important. People who are depressed, for example, may not directly say so, but analyzing their

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

language may provide useful signals of this condition (Eichstaedt et al., 2018).

Limitations

While automated text analysis has many valuable aspects, it also has limitations. As with most methods, age-old questions of sampling, validity, and reliability arise. Survivor bias is a form of selection bias that arises when failures are ignored, and successes are oversampled. If researchers focus on the lyrics of culturally successful songs because they are easier to access, this could generate misleading inferences. They may end up reaching conclusions that do not apply to the general population (all song lyrics). Similarly, different word embedding methods can yield different vectors for complete words versus word stems, and many methods involve multiple choice points that may shape the results (e.g., number of topics to choose in LDA).

Using field data alone can also make it challenging to get at causality. A textual feature may be correlated with an outcome, but is it truly causing that outcome? One solution is to pair text analysis of field data with experimental methods to identify causality. When examining the meaning of happiness, Mogilner et al. (2011) analyzed millions of blogs, finding that the language linked to happiness shifts over one's lifetime. They then turned to experiments to test causality and shed light on what was driving the effect.

Natural experiments, or other approaches to causal inference in field data, can also be useful. Berger and Packard (2018) found that atypical songs are more popular, but to test causation, they looked at songs that charted in multiple genres (e.g., a song that appeared on both the country and hip-hop charts). This ensured that all other aspects of the song (e.g., music artist, lyrics, and release date) were identical, and provided a stronger causal test of whether songs are more popular when they are more differentiated from their genre.

Another limitation is the degree to which language fully represents people and culture. Beyond words, people also communicate through paralanguage (e.g., pitch, tone, or body language). Similarly, images reflect the cultures in which they are produced (Masuda et al., 2008). Visual and auditory modalities may convey the same information as text, or something different. Tools like Praat (Boersma & Weenink, 2018) can be used to extract pitch and tone from audio files (e.g., Van Zant & Berger, 2020) and research has started to use computer vision to extract features from images (Li & Xie, 2020; Zhang et al., 2017). While there has been less work in these areas than in text analysis, emerging approaches will hopefully enable better analysis of these important information channels.

Finally, it is important to recognize how the type of data analyzed may shape the language it contains. Books are longer than articles, which are longer than text messages. Online reviews are (mostly) created by individuals, but movies and news articles are created by groups or larger institutions. Social media posts broadcast online are more likely to be driven by self-presentation than everyday conversations between friends (Barasch & Berger, 2014). The motivations behind content creation, affordances of a given medium, and other features combine to shape the nature of the text. Just as language reflects things and the person or group that produced it, the content of different cultural items should reflect aspects of their creation.

Conclusion

In conclusion, technological advances have opened exciting new avenues to study people and culture. Natural language processing allows researchers to ask new questions and study age-old topics in new ways. Hopefully, more psychologists will adopt these tools and begin to extract wisdom from words.

References

Akpinar, E., & Berger, J. (2015). Drivers of cultural success: The case of sensory metaphors. *Journal of Personality and Social Psychology*, *109*(1), 20–34. https://doi.org/10.1037/pspa0000025

Bail, C. A. (2014). The cultural environment: Measuring culture with big data. *Theory and Society*, *43*(3-4), 465–482. https://doi.org/10.1007/s11186-014-9216-5

Bail, C. A. (2016). Combining natural language processing and network analysis to examine how advocacy organizations stimulate conversation on social media. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(42), 11823–11828. https://doi.org/10.1073/pnas.1607151113

Bail, C. A., Brown, T. W., & Wimmer, A. (2019). Prestige, proximity, and prejudice: How Google search terms diffuse across the world. *American Journal of Sociology*, *124*(5), 1496–1548. https://doi.org/10.1086/702007

Barasch, A., & Berger, J. (2014). Broadcasting and narrowcasting: How audience size affects what people share *Journal of Marketing Research*, *51*(3), 286–299. https://doi.org/10.1509/jmr.13.0238

Berger, J. (2011). Arousal increases social transmission of information. *Psychological Science*, *22*(7), 891–893. https://doi.org/10.1177/0956797611413294

Berger, J., Bradlow, E. T., Braunstein, A., & Zhang, Y. (2012). From Karen to Katie: Using baby names to understand cultural evolution. *Psychological Science*, *23*(10), 1067–1073. https://doi.org/10.1177/0956797612443371

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

Berger, J., Kim, Y. D., & Meyer, R. (2021). What makes content engaging? How emotional dynamics shape success. *Journal of Consumer Research*. Advance online publicaton. 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., & Packard, G. (2018). Are atypical things more popular? *Psychological Science*, *29*(7), 1178–1184. https://doi.org/10.1177/0956797618759465

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics*, *8*(5), 279–286. https://doi.org/10.3758/BF03212593

Berry, D. M. (2012). *Understanding digital humanities*. Palgrave Macmillan. https://doi.org/10.1057/9780230371934

Bhatia, S. (2019). Predicting risk perception: New insights from data science. *Management Science*, *65*(8), 3800–3823. https://doi.org/10.1287/mnsc.2018.3121

Bhatia, S., Richie, R., & Zou, W. (2019). Distributed semantic representations for modeling human judgment. *Current Opinion in Behavioral Sciences*, *29*, 31–36. https://doi.org/10.1016/j.cobeha.2019.01.020

Bhatia, S., & Walasek, L. (2019). Association and response accuracy in the wild. *Memory & Cognition*, *47*(2), 292–298. https://doi.org/10.3758/s13421-018-0869-6

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Boersma, P., & Weenink, D. (2018). PRAAT: Doing phonetics by computer [Computer program]. http://www.praat.org

Boghrati, R., & Berger, J. (2020). *Quantifying cultural change: An application to misogyny in music* (Working paper).

Bolukbasi, T., Chang, K., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *ArXiv*. https://arxiv.org/abs/1607.06520

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., Pasca, P., & Lanning, K. (2020). The personality panorama: Conceptualizing personality through big behavioral data. *European Journal of Personality*, *34*(5), 599–612. https://doi.org/10.1002/per.2254

Boyd, R. L., & Schwartz, H. A. (2021). Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field. *Journal of Language and Social Psychology*, *40*(1), 21–41. https://doi.org/10.1177/0261927X20967028

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., . . . Amodei, D. (2020). Language models are few-shot learners. *ArXiv*. https://arxiv.org/abs/2005.14165

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

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186. https://doi.org/10.1126/science.aal4230

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *33*(4), 497–505. https://doi.org/10.1080/14640748108400805

Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., & Potts, C. (2013). No country for old members: User lifecycle and linguistic change in online communities. *Proceedings of the 22nd international conference on World Wide Web* (pp. 307–318). Association for Computing Machinery. https://doi.org/10.1145/2488388.2488416

Dawkins, R. (1976). Hierarchical organisation: A candidate principle for ethology. *Growing Points in Ethology*, *7*, 54.

DeFranza, D., Mishra, H., & Mishra, A. (2020). How language shapes prejudice against women: An examination across 45 world languages. *Journal of Personality and Social Psychology*, *119*(1), 7–22. https://doi.org/10.1037/pspa0000188

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of deep bidirectional transformers for language understanding*. https://www.aclweb.org/anthology/N19-1423

Eichstaedt, J. C., Kern, M. L., Yaden, D. B., Schwartz, H. A., Giorgi, S., Park, G., Hagan, C. A., Tobolsky, V., 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*. Advance online publication. https://doi.org/10.31234/osf.io/t52c6

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preoţiuc-Pietro, D., Asch, D. A., & Schwartz, H. A. (2018). Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(44), 11203–11208. https://doi.org/10.1073/pnas.1802331115

Firth, J. R. (Ed.). (1957). A synopsis of linguistic theory, 1930–1955. *Studies in linguistic analysis* (pp. 1–32). Philological Society.

Fitzsimons, G. M., & Kay, A. C. (2004). Language and interpersonal cognition: Causal effects of variations in pronoun usage on perceptions of closeness. *Personality and Social Psychology Bulletin*, *30*(5), 547–557. https://doi.org/10.1177/0146167203262852

Foltz, P. W. (1998). Quantitative approaches to semantic knowledge representations. *Discourse Processes*, *25*(2–3), 127–130. https://doi.org/10.1080/01638539809545024

Foltz, P. W. (2007). Discourse coherence and LSA. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 167–184). Erlbaum Publishers.

Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes*, *25*(2-3), 285–307. https://doi.org/10.1080/01638539809545029

Freytag, G., & MacEwan, E. J. (1900). *Freytag's technique of the drama*. Scott Foresman.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(16), E3635–E3644. https://doi.org/10.1073/pnas.1720347115

Gaucher, D., Friesen, J., & Kay, A. C. (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. *Journal of Personality and Social Psychology*, *101*(1), 109–128. https://doi.org/10.1037/a0022530

Gladkova, A., Drozd, A., & Matsuoka, S. (2016). Analogy-based detection of morphological and semantic relations with word embeddings: What works and what doesn't. *Proceedings of the NAACL Student Research Workshop* (pp. 8–15). Association for Computational Linguistics. https://doi.org/10.18653/v1/N16-2002

Gold, M. K. (2012). *Debates in the digital humanities*. University of Minnesota Press.

Gonzales, A. L., Hancock, J. T., & Pennebaker, J. W. (2010). Language style matching as a predictor of social dynamics in small groups. *Communication Research*, *37*(1), 3–19. https://doi.org/10.1177/0093650209351468

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*, 193–202. https://doi.org/10.3758/BF03195564

Gray, K., Anderson, S., Chen, E. E., Kelly, J. M., Christian, M. S., Patrick, J., Huang, L., Kenett, Y. N., & Lewis, K. (2019). Forward flow": A new measure to quantify free thought and predict creativity. *American Psychologist*, *74*(5), 539–554. https://doi.org/10.1037/amp0000391

Harber, K. D., Reeves, S., Gorman, J. L., Williams, C. H., Malin, J., & Pennebaker, J. W. (2019). The conflicted language of interracial feedback. *Journal of Educational Psychology*, *111*(7), 1220–1242. https://doi.org/10.1037/edu0000326

Heath, C., Bell, C., & Sternberg, E. (2001). Emotional selection in memes: The case of urban legends. *Journal of Personality and Social Psychology*, *81*(6), 1028–1041. https://doi.org/10.1037/0022-3514.81.6.1028

Holtgraves, T. M., & Kashima, Y. (2008). Language, meaning, and social cognition. *Personality and Social Psychology Review*, *12*(1), 73–94. https://doi.org/10.1177/1088868307309605

Huang, K., Yeomans, M., Brooks, A. W., Minson, J., & Gino, F. (2017). It doesn't hurt to ask: Question-asking increases liking. *Journal of Personality and Social Psychology*, *113*(3), 430–452. https://doi.org/10.1037/pspi0000097

Huang, L., Joshi, P. D., Walslak, C. J., & Wu, A. (2020). Sizing up entrepreneurial potential: Gender differences in communication and investor perceptions of long-term growth and scalability. *Academy of Management Journal*. Advance online publication. https://doi.org/10.5465/amj.2018.1417

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

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*, 453–458. https://doi.org/10.1038/nature17637

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science*, *22*(1), 39–44. https://doi.org/10.1177/0956797610392928

Jackson, J. C., Watts, J., Henry, T. R., List, J. M., Forkel, R., Mucha, P. J., Greenhill, S. J., Gray, R. D., & Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure. *Science*, *366*(6472), 1517–1522. https://doi.org/10.1126/science.aaw8160

Jänicke, S., Franzini, G., Cheema, M. F., & Scheuermann, G. (2015). On close and distant reading in digital humanities: A survey and future challenges. In R. Borgo, F. Ganovelli, & I. Viola (Eds.), *Eurographics Conference on Visualization (EuroVis-STARs)* (pp. 83–103). The Eurographics Association.

Kashima, Y. (2008). A social psychology of cultural dynamics: Examining how cultures are formed, maintained, and transformed. *Social and Personality Psychology Compass*, *2*(1), 107–120. https://doi.org/10.1111/j.1751-9004.2007.00063.x

Kern, M. L., Park, G., Eichstaedt, J. C., Schwartz, H. A., Sap, M., Smith, L. K., & Ungar, L. H. (2016). Gaining insights from social media language: Methodologies and challenges. *Psychological Methods*, *21*, 507–525. https://doi.org/10.1037/met0000091

Kesebir, S. (2017). Word order denotes relevance differences: The case of conjoined phrases with lexical gender. *Journal of Personality and Social Psychology*, *113*(2), 262–279. https://doi.org/10.1037/pspi0000094

Kesebir, S., & Kesebir, P. (2017). A growing disconnection from nature is evident in cultural products. *Perspectives on Psychological Science*, *12*(2), 258–269. https://doi.org/10.1177/1745691616662473

Kim, H., & Markus, H. R. (1999). Deviance or uniqueness, harmony or conformity? A cultural analysis. *Journal of Personality and Social Psychology*, *77*(4), 785–800. https://doi.org/10.1037/0022-3514.77.4.785

Kovacs, B., & Kleinbaum, A. M. (2020). Language-style similarity and social networks. *Psychological Science*, *31*(2), 202–213. https://doi.org/10.1177/0956797619894557

Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, *84*(5), 905–949. https://doi.org/10.1177/0003122419877135

Kunst-Wilson, W. R., & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science*, *207*(4430), 557–558. https://doi.org/10.1126/science.7352271

Laurino Dos Santos, H., & Berger, J. (2020). *Pacing and audience response: Using deep learning sentence embeddings to analyze movie scripts* (Working paper).

Li, J. J., & Nenkova, A. (2015). Fast and accurate prediction of sentence specificity. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (pp. 2281–2287). AAAI Press.

Li, Y., & Xie, Y. (2020). Is a picture worth a thousand words? An empirical study of image content and social media engagement. *Journal of Marketing Research*, *57*(1), 1–19. https://doi.org/10.1177/0022243719881113

Markus, H. R., & Hamedani, M. G. (2019). People are culturally shaped shapers: The psychological science of culture and culture change. In D. Cohen & S. Kitayama (Eds.), *Handbook of cultural psychology* (pp. 11–52). Guilford Press.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Masuda, T., Gonzalez, R., Kwan, L., & Nisbett, R. E. (2008). Culture and aesthetic preference: Comparing the attention to context of East Asians and Americans. *Personality and Social Psychology Bulletin*, *34*(9), 1260–1275. https://doi.org/10.1177/0146167208320555

McLean, K. C., Syed, M., Pasupathi, M., Adler, J. M., Dunlop, W. L., Drustrup, D., Fivush, R., Graci, M. E., Lilgendahl, J. P., Lodi-Smith, J., McAdams, D. P., & McCoy, T. P. (2020). The empirical structure of narrative identity: The initial Big Three. *Journal of Personality and Social Psychology*, *119*(4), 920–944. https://doi.org/10.1037/pspp0000247

Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology*, *84*, 857–870. https://doi.org/10.1037/0022-3514.84.4.857

Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*, 176–182. https://doi.org/10.1126/science.1199644

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv*. https://arxiv.org/abs/1301.3781

Mogilner, C., Kamvar, S. D., & Aaker, J. (2011). The shifting meaning of happiness. *Social Psychological & Personality Science*, *2*(4), 395–402. https://doi.org/10.1177/1948550610393987

Moretti, F. (2013). *Distant reading*. Verso.

Morling, B., & Lamoreaux, M. (2008). Measuring culture outside the head: A meta-analysis of individualism-collectivism in cultural products. *Personality and Social Psychology Review*, *12*(3), 199–221. https://doi.org/10.1177/1088868308318260

Muthukrishna, M., Henrich, J., & Slingerland, E. (2021). Psychology as a historical science. *Annual Review of Psychology*, *72*, 717–749. https://doi.org/10.1146/annurev-psych-082820-111436

Neiman, J. L., Gonzalez, F. J., Wilkinson, K., Smith, K. B., & Hibbing, J. R. (2016). Speaking different languages or reading from the same script? Word usage of Democratic and Republican politicians. *Political Communication*, *33*(2), 212–240. https://doi.org/10.1080/10584609.2014.969465

Nicolas, G., Bai, X., & Fiske, S. T. (2021). Comprehensive stereotype content dictionaries using a semi-automated method. *European Journal of Social Psychology*. Advance online publication. https://doi.org/10.1002/ejsp.2724

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, *21*(4), 337–360. https://doi.org/10.1177/026192702237953

Norenzayan, A., Atran, S., Faulkner, J., & Schaller, M. (2006). Memory and mystery: The cultural selection of minimally counterintuitive narratives. *Cognitive Science*, *30*, 531–553. https://doi.org/10.1207/s15516709cog0000_68

Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us*. Bloomsbury Press/Bloomsbury Publishing. https://doi.org/10.1016/S0262-4079(11)62167-2

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. University of Texas at Austin.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

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*, 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. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532–1543). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1162

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 (Long Papers)* (pp. 2227–2237). Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-1202

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156. https://doi.org/10.1037/a0014420

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), 12. https://doi.org/10.1140/epjds/s13688-016-0093-1

Rocklage, M. D., Rucker, D. D., & Nordgren, L. F. (2018). The Evaluative Lexicon 2.0: The measurement of emotionality, extremity, and valence in language. *Behavior Research Methods*, *50*, 1327–1344. https://doi.org/10.3758/s13428-017-0975-6

Sap, M., Park, G. J., Eichstaedt, J. C., Kern, M. L., Stillwell, D., Kosinski, M., Ungar, L., & Schwartz, H. A. (2014). Developing age and gender predictive lexica over social media. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1146–1151). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1121

Schaller, M., Conway, L. G., III,., & Tanchuk, T. L. (2002). Selective pressures on the once and future contents of ethnic stereotypes: Effects of the communicability of traits. *Journal of Personality and Social Psychology*, *82*(6), 861–877. https://doi.org/10.1037/0022-3514.82.6.861

Schaller, M., & Crandall, C. S. (2004). *The psychological foundations of culture*. Erlbaum.

Schwartz, H. A., Eichstaedt, J., Kern, M. L., Park, G., Sap, M., Stillwell, D., Kosinski, M., & Ungar, L. (2014). Towards assessing changes in degree of depression through Facebook. *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* (pp. 118–125). Association for Computational Linguistics. https://doi.org/10.3115/v1/W14-3214

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 of the United States of America*, *118*(7), e2017154118. https://doi.org/10.1073/pnas.2017154118

Snefjella, B., & Kuperman, V. (2015). Concreteness and psychological distance in natural language use. *Psychological Science*, *26*(9), 1449–1460. https://doi.org/10.1177/0956797615591771

Snibbe, A. C., & Markus, H. R. (2005). You can't always get what you want: Educational attainment, agency, and choice. *Journal of Personality and Social Psychology*, *88*(4), 703–720. https://doi.org/10.1037/0022-3514.88.4.703

Srivastava, S. B., Goldberg, A., Manian, V. G., & Potts, C. (2018). Enculturation trajectories: Language, cultural adaptation, and outcomes in organizations. *Management Science*, *64*(3), 1348–1364. https://doi.org/10.1287/mnsc.2016.2671

Stephens, N. M., Markus, H. R., & Townsend, S. S. M. (2007). Choice as an act of meaning: The case of social class. *Journal of Personality and Social Psychology*, *93*(5), 814–830. https://doi.org/10.1037/0022-3514.93.5.814

Sterling, J., Jost, J. T., & Bonneau, R. (2020). Political psycholinguistics: A comprehensive analysis of the language habits of liberal and conservative social media users. *Journal of Personality and Social Psychology*, *118*(4), 805–834. https://doi.org/10.1037/pspp0000275

Steyvers, M., Griffiths, T. L., & Dennis, S. (2006). Probabilistic inference in human semantic memory. *Trends in Cognitive Sciences*, *10*, 327–334. https://doi.org/10.1016/j.tics.2006.05.005

Taylor, P. J., & Thomas, S. (2008). Linguistic style matching and negotiation outcome. *Negotiation and Conflict Management Research*, *1*(3), 263–281. https://doi.org/10.1111/j.1750-4716.2008.00016.x

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

Underwood, T. (2015). The literary uses of high-dimensional space. *Big Data & Society*. Advance online publication. https://doi.org/10.1177/2053951715602494

Van Zant, A. B., & Berger, J. (2020). How the voice persuades. *Journal of Personality and Social Psychology*, *118*(4), 661–682. https://doi.org/10.1037/pspi0000193

Voigt, R., Camp, N. P., Prabhakaran, V., Hamilton, W. L., Hetey, R. C., Griffiths, C. M., Jurgens, D., Jurafsky, D., & Eberhardt, J. L. (2017). Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(25), 6521–6526. https://doi.org/10.1073/pnas.1702413114

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

Wilson, S. R., Mihalcea, R., Boyd, R. L., & Pennebaker, J. W. (2016). Cultural influences on the measurement of personal values through words. *Proceedings of the 2016 AAAI Spring Symposium Series* (pp. 314–317). Association for the Advancement of Artificial Intelligence.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2), 1–27. https://doi.org/10.1037/h0025848

Zhang, S., Lee, D., Singh, P. V., & Srinivasan, K. (2017). How much is an image worth? Airbnb property demand estimation leveraging large scale image analytics. *SSRN Electronic Journal*. Advance online publication. https://doi.org/10.2139/ssrn.2976021

Zuckerman, M. (1979). *Sensation seeking: Beyond the optimum level of arousal*. Lawrence Erlbaum.

Received January 30, 2021
Revision received May 14, 2021
Accepted May 15, 2021 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.