



From Text to Thought: How Analyzing Language Can Advance Psychological Science

Joshua Conrad Jackson¹ , Joseph Watts^{2,3,4} ,
Johann-Mattis List², Curtis Puryear¹, Ryan Drabble¹, and
Kristen A. Lindquist¹ 

¹Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill; ²Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History; ³Center for Research on Evolution, Belief, and Behaviour, University of Otago; and ⁴Religion Programme, University of Otago

Abstract

Humans have been using language for millennia but have only just begun to scratch the surface of what natural language can reveal about the mind. Here we propose that language offers a unique window into psychology. After briefly summarizing the legacy of language analyses in psychological science, we show how methodological advances have made these analyses more feasible and insightful than ever before. In particular, we describe how two forms of language analysis—natural-language processing and comparative linguistics—are contributing to how we understand topics as diverse as emotion, creativity, and religion and overcoming obstacles related to statistical power and culturally diverse samples. We summarize resources for learning both of these methods and highlight the best way to combine language analysis with more traditional psychological paradigms. Applying language analysis to large-scale and cross-cultural datasets promises to provide major breakthroughs in psychological science.

Keywords

natural-language processing, comparative linguistics, historical linguistics, psycholinguistics, cultural evolution, emotion, religion, creativity

Humans have been using language for millennia and compiling written records for at least the past 5,000 years (Walker & Chadwick, 1990). In that time, humans have written nearly 130 million books (Tachyer, 2010), producing sprawling religious scriptures, millions of songs, countless speeches, and expansive dictionaries that explain and translate entire lexicons. These records of human language represent a rich but underexplored trove of data on the human experience.

Human language—be it spoken, written, or signed—has the power to reveal how humans organize thoughts into categories, view associations between these categories, and use these categories in daily life for communication and social influence. It can be used to understand how humans view the salience of different ideas and how understanding of these ideas may change over time. On a broader level, language can reveal variation in thought processes and verbal behavior across different cultural and ideological groups and

illuminate universal and variable patterns in how humans understand constructs such as God, emotion, and the self. Language is thus a rich and dynamic window into human experience that promises to yield new insights in each branch of psychological science.

The promises of language analysis for psychological science were largely unrealized for most of the field's history because most records of language were inaccessible. Books gathered dust on shelves, sacred texts lay in museums, and songs were stored either in human memory, on cassette tapes, or in albums. These vast stores of natural linguistic data sat out of reach over the 20th and early 21st centuries, and psychologists developed increasingly sophisticated measures of

Corresponding Author:

Joshua Conrad Jackson, Kellogg School of Management, Northwestern University

Email: joshua.jackson@kellogg.northwestern.edu

explicit attitudes (Likert, 1932), implicit attitudes (Greenwald et al., 1998), brain activity (Nichols & Holmes, 2002), and physiology (Kagan et al., 1987). But this is beginning to change.

Just as the printing press made language accessible to the masses, computational innovations are now making language analyzable for the academic masses. A methodological arms race in computational linguistics and computer science is producing new techniques that are capable not only of digitizing written language but also of efficiently processing, storing, and quantifying patterns in this language. As a result of these innovations, records of language are no longer hidden away but are freely and easily accessible. Researchers can now retrieve vast stores of digitized written text from thousands of languages around the world and throughout history and finally begin realizing the potential of language analysis for psychological science.

With newly developed databases and analytic tools, language analysis is trickling into psychological science. Here we discuss how psychologists can best leverage these tools to make predictions about human experience by explaining popular new methods of language analysis and psychological predictions that are suitable for these methods. We focus primarily on topics central to social psychology, such as emotion, religion, and creativity, but we also give examples from clinical, developmental, and cognitive psychology.

The main goal of this article is to provide a “one-stop shop” for psychological scientists to read about the history and best practices associated with different methods of language analysis and to provide resources for easily learning these methods. Although there are existing reviews of specific language-analysis methods (e.g., Bittermann & Fischer, 2018; Pennebaker et al., 2007; Rudkowsky et al., 2018) and some broader reviews about the utility of language analysis for the social and organizational sciences (e.g., Berger et al., 2020; Boyd & Schwartz, 2020; Kjell et al., 2019; Short et al., 2018), few articles have discussed how multiple forms of linguistic analysis can be integrated to address a range of psychological questions. We provide this information so that, as the trickle of text analysis in psychology becomes a flood, psychologists will be prepared to analyze language rigorously, accurately, and in a manner that takes full advantage of each method’s promise.

We also highlight systemic advantages of language analysis, focusing on the promise of natural-language processing (NLP) and comparative linguistics. NLP paradigms may be uniquely suited to resolve problems associated with the generalizability of psychological findings because they sample from real-life conversations, speeches, and texts and are useful for solving the problems associated with low statistical power because they often incorporate millions of datapoints (Bakker et al.,

2016; Cohen, 1992). Comparative-linguistics paradigms may be uniquely suited to resolve problems of representation and diversity in psychology by incorporating traditionally underrepresented cultures (Chandler et al., 2019; Henrich et al., 2010; Rad et al., 2018). Language analysis is therefore well suited to address several of the largest current challenges in psychological science.

We suggest that language-analysis methods, because of their theoretical and practical advantages, are at least as valuable as Likert scales, measures of implicit bias, behavioral measures, neuroimaging, psychophysiology, and other paradigms in psychological science. We also review limitations of language analysis that make it well suited to complement (rather than replace) these existing methods. By complementing traditional methods with rigorous language analysis, we can gain a more complete understanding of the human mind.

What Does It Mean to Analyze Language?

Humans are intuitive language analysts. Just as psychologists use measurements to index latent constructs, humans infer the latent meaning being conveyed via language. Humans recognize words, react to sentiment and affect in sentences, and search for meaning in metaphors and innuendos. Formal language analysis requires going beyond this intuition to quantitatively deconstruct the meaning of language and measure the constructs that it conveys. People may feel inspired when they hear a rousing speech, but how can the construct of “inspiration” be quantified by examining the length, content, and format of a sentence? Translation dictionaries may equate two words and report that they have the same meaning, but how can researchers test whether language speakers actually use these words to communicate the same ideas?

The roots of language analysis in psychological science

Questions about how psychological meaning is embedded in language have deep roots in psychology, and many of the earliest psychologists were keenly aware of the promise of language analysis. Freud’s analytic techniques involved examining free associations and slips of the tongue (Freud, 1901). Murray’s Thematic Apperception Test analyzed the linguistic content of stories that people told in response to pictures (Murray, 1943), and Allport counted words in a dictionary to identify the structure of personality (Allport & Vernon, 1930). These early methods had substantial limitations and are rarely used in contemporary quantitative research, but they foreshadowed the impact of language analysis on psychological science.

The promise of language analysis for psychological theorizing was not fully realized until the development of computational methods of language analysis, the most popular of which may be the technique known as *linguistic inquiry and word count* (LIWC; Pennebaker et al., 2007; Tausczik & Pennebaker, 2010). LIWC uses word frequency to yield insight into the meaning of language. For example, words referencing social in-groups (e.g., “we,” “us”) are probably expressing more affiliative meaning than words referencing out-groups (“they,” “them”). LIWC uses these word-count methods with preprogrammed dictionaries that represent semantic categories and correspond to psychological constructs of interest. A negative-emotion dictionary counts a predetermined set of words that connote feelings of negative affect, whereas a pronouns dictionary counts instances of “she,” “I,” “they,” and other pronouns that can be used to assess whether someone is referring to the self or others. LIWC gives the percentage of words in a corpus that fall into each dictionary. This method has been generative in psychology, and studies have applied LIWC to understand the psychological effects of aging (Pennebaker & Stone, 2003), the content of lies (Newman et al., 2003), mental-health stressors such as bullying and domestic abuse (Holmes et al., 2007), political messaging (Gunsch et al., 2000; Pennebaker & Lay, 2002), the emotional toll of terrorist attacks (Back et al., 2010; Cohn et al., 2004), and the popularity of songs (Packard & Berger, 2020).

One of LIWC’s major strengths is its parsimony. The software takes corpora—stores of written text that have been structured in a way that makes them downloadable and analyzable by algorithms—and returns simple percentages summarizing the text’s content. But this strength is also a limitation. When analyzing a sentence with many positive words, counting alone cannot distinguish whether words are meant ironically or as part of a counterfactual statement, and it cannot determine the source or the target of this positivity. Consider, for example, an excerpt from Martin Luther King Jr.’s (1963) famous “I have a dream” speech:

We have also come to this hallowed spot to remind America of the fierce urgency of Now. This is no time to engage in the luxury of cooling off or to take the tranquilizing drug of gradualism. Now is the time to make real the promises of democracy. Now is the time to rise from the dark and desolate valley of segregation to the sunlit path of racial justice. Now is the time to lift our nation from the quicksands of racial injustice to the solid rock of brotherhood.

In just a few sentences, King’s speech uses the words “luxury,” “desolate,” “segregation,” and “justice.” A

counting approach could identify themes of positivity, negativity, morality, and inequity, yet it would not identify the nuanced way that King intended these words to signal perseverance and a fight for progress. Many articles have pointed out the limitations of these “bag of words” approaches that simply count the number of words rather than examining how these words are used in context (Enríquez et al., 2016; Wallach, 2006). Some psychological paradigms have sought to address these gaps. For example, research on conceptual metaphors explores how words take on multiple meanings and how these can reflect psychological associations (e.g., the concepts “up” and “down” describe both physical placement and psychological mood; Crawford et al., 2006; Landau et al., 2010; Meier & Robinson, 2006). However, a drawback of conceptual-metaphor methods is that they qualitatively analyze language, making them difficult to apply to large-scale or cross-cultural datasets.

Another limitation of word-count methods is that they are focused almost exclusively on the English language, which limits their historical and cross-cultural generalizability. The English language (including Old English and Middle English) has existed for a small fraction of human history, and approximately 5% of people today speak English as a first language, yet English speakers probably account for more than 99% of language-analysis research published in psychology journals (Lewis, 2009). Some efforts have been made to translate LIWC to other languages, but these efforts are very recent and focus more on replication than on comparison (Windsor et al., 2019). This leaves open many questions about how seemingly equivalent words have different meanings across languages and whether more closely related languages have more similar meaning structures than more distantly related languages.

These limitations notwithstanding, word-count methods such as LIWC have been tremendously useful in psychology, and their limitations can be addressed by supplementing them with other methods of language analysis that are currently rarer in psychology. One of these traditions, NLP, uses methods developed in computer science to analyze semantic patterns in language. Another tradition, comparative linguistics, involves the comparison of languages to determine how languages have evolved over time, how they may communicate meaning in unique ways. Both methods were developed outside of psychology but have great potential for psychological research.

NLP as a tool for studying large-scale patterns of cognition

Background. NLP—the interdisciplinary study of computer interaction with human language—is a relatively

young area of study. NLP's earliest notable paradigm was the "Turing Test": the hypothetical test wherein a computer mimics human language so well that an observer cannot differentiate the computer from a real person (Turing, 1950/2009). Other early NLP developments involved ELIZA (Weizenbaum, 1966)—a computer therapist that could respond to human complaints ("I feel sad") with realistic therapist comments ("and why do you feel sad?")—and Jabberwacky.com, now running as "Cleverbot," which was designed in the 1980s to simulate entertaining but realistic human conversations.

NLP was not necessarily designed with psychological insights in mind, but building algorithms to simulate human speech has obvious psychological implications. Many of these insights derive from the advancement of "machine learning"—computer algorithms that can improve automatically through experience. Machine learning approaches can either be *unsupervised*, in which algorithms such as topic models try to classify words without researchers providing feedback, or *supervised*, in which algorithms are trained on the evaluation and classification of data using feedback from researchers. For example, an unsupervised machine-learning algorithm could use a corpus of speeches to automatically identify major semantic themes on the basis of co-occurring words, whereas a supervised algorithm could be trained to recognize that negative words frequently precede positive words or even to recognize metaphors (Jacobs & Kinder, 2017). When applied to King's speech, this algorithm would be able to do far more than a simple word-counting technique by potentially revealing themes of justice and liberty and identifying that metaphors such as a "sunlit path" are referring to morally commendable action.

Although early machine-learning approaches were limited by statistical methods and computational power, machine learning has taken huge steps in the past several decades. Early machine-learning models of language translation and production were built using constrained statistical methods (Weaver, 1955), rule-based methods (Nirenburg, 1989), and example-based methods (Nagao, 1984). These methods made simplistic assumptions about the cognitive processes underlying the production of language, such as the existence of a universal structure to grammar across languages. Today, artificial neural nets are at the forefront of research in machine learning and have more promise for actually understanding psychological processes. These networks are loosely modeled after the structure of organic brains by modeling associative networks of co-occurrence across many variables. Like the human brain, the way they process language can be complex and difficult to understand. But unlike the human brain, researchers can often ethically gain access to, and modify, the

precise mechanisms underlying how these algorithms process language by delving into their code. This opens a new way of building and testing scientific theories within psychology (Cichy & Kaiser, 2019). In Figure S1 in the supplementary materials on OSF (<https://osf.io/xycbd>), for example, we describe a neural network that is designed to classify U.S. presidents' speeches as being from either the pre-Civil War period or post-Civil War period to show how one of these algorithms can use text to make complex classification judgments.

NLP approaches now have a wide range of applications to psychological questions. These methods allow researchers to quantify the meaning of constructs in text or speech, identify the presence and extent of certain attitudes and emotions, and distill the meaning of words on the basis of how they are used in context. These algorithms can efficiently analyze millions of datapoints in seconds and have the potential to analyze more representative samples of subjects than typical undergraduate research pools or Mechanical Turk experiments, especially when they are applied to online blogs, diaries, or social-media websites such as Facebook or Twitter.

Application 1: quantifying the meaning of constructs. One of the most fundamental applications of NLP involves identifying the meaning of constructs and finding sets of constructs that cluster together in meaning. Topic modeling is a classic unsupervised NLP method that accomplishes this goal by finding co-occurring words that may represent psychological categories of interest. For example, a topic model might observe a construct such as "birthday" on the basis of the co-occurrence of such words as "happy," "birthday" "cake," "candle," and "gift" (Hong & Davison, 2010; Wallach, 2006). Topic models can either match words to a predefined number of topics or freely extract the best-fitting number of topics from a set of texts using optimization.

Topic models each share a basic structure and output format, but they can be generated by different algorithms. One of these algorithms, latent semantic analysis, is arguably the most foundational method of generating topic models (Landauer et al., 1998), but it is not the only method. Probabilistic latent semantic analysis will include probabilities that words belong in topics (Steyvers & Griffiths, 2007); latent Dirichlet allocation is a Bayesian version of probabilistic latent semantic analysis (Blei et al., 2003), and structural topic models examine the relationships between variables and the prevalence of topics (Roberts et al., 2019). Researchers have used these kinds of topic models to estimate cross-cultural differences in people's personal values (Wilson et al., 2016); predict the likelihood of

clinical depression using people's social-media updates (De Choudhury et al., 2013; Eichstaedt et al., 2018); quantify differences in the meaning of language across gender, age, and personality style (Schwartz et al., 2013); and estimate why some requests for favors are more effective than others (Althoff et al., 2014). A related set of models that classify texts (e.g., newspaper articles) rather than topics have helped match students' reading level to their reading material (Graesser et al., 2011), identified differences in the thinking process of individuals with psychosis compared with control subjects (Elvevaag et al., 2010), and recognized different responses to a geopolitical event (Mishler et al., 2015).

Whereas topic models are focused on categorization, approaches involving word embedding quantify the meaning of concepts in a more continuous way; methods such as word2vec or GloVe (global vectors for word representation) map words or phrases to vectors of numbers using neural network models to create continuous numerical distances that represent differences in meaning (Goldberg & Levy, 2014; Mikolov et al., 2013). The semantic vectors produced by word embeddings allow researchers to map the meaning between any two concepts and to collect clusters of concepts that are the most similar to theoretically important "seed" concepts. For example, the seed concept of "freedom" might be closest in vector space to "autonomy" and relatively close to "choice" and "liberty." These comparisons can help psychologists to measure and quantify otherwise abstract psychological constructs such as "freedom." This approach has helped detect increasingly permissive culture in the United States via an increase in vocabulary related to "freedom" (Jackson, Gelfand, et al., 2019) and track the expanding concept of harm across the 20th and 21st centuries (Vylomova et al., 2019).

Application 2: tracking attitudes and emotions in unstructured data. Another NLP approach known as "sentiment analysis" goes beyond quantifying meaning and focuses on tracking attitudes and mood over time. Sentiment analysis is actually an umbrella term to capture a range of methods. "Knowledge-based" methods of sentiment analysis are similar to LIWC, insofar as they detect the frequency of different prespecified words and track how the frequency of these words changes over time (Caluori et al., 2020). For example, Cohn et al. (2004) tracked changes in affect after trauma, showing that positive-emotion language dropped sharply after the 9/11 terrorist attacks but then rebounded over time. Garcia and Rimé (2019) did a similar analysis of positive and negative collective emotions following the Paris terrorist attacks of 2015, and Vo and Collier (2013) used the approach to capture spikes in fear and anxiety following

earthquakes. Hutto and Gilbert (2014) recently developed VADER (valence-aware dictionary and sentiment reasoner), a knowledge-based form of sentiment analysis that builds on LIWC by quantifying the intensity as well as the prevalence of positive and negative sentiment in text and incorporating slang into its dictionaries. VADER also uses several grammatical rules to detect preferences and emotions in nuanced contexts, such as when preferences are expressed through negations ("I do not dislike my partner") or modifiers ("sometimes I really hate my friends").

Combining grammatical rules with a human-validated lexicon (as VADER does) is a powerful and easily interpreted approach to sentiment analyses. Because the researcher specifies the set of rules ahead of time, there is no "black box" obscuring how the algorithm scores a segment of text. However, this strength is also its weakness. More complex tasks often benefit from learning which rules help to understand and classify text. Machine-learning methods, such as random forests and neural networks, are often better equipped to mine opinions in context because they can flexibly learn how patterns in input text (e.g., a smiley face) relate to some output (e.g., positive affect). Supervised approaches will often use a set of hand-labeled texts to train a sentiment classifier. Over the course of training, the model can learn how the presence of negation, emojis, or information from previous sentences help to correctly classify the text without requiring the researcher to explicitly implement any of these rules (Kiritchenko et al., 2014). For example, Wang and colleagues (2013) used a machine-learning approach to detect depression using the textual content of personal blogs with 80% accuracy, whereas Oscar and colleagues (2017) used a supervised machine-learning approach to capture stigma toward individuals with dementia.

Application 3: distilling linguistic information. A third set of NLP techniques is focused on more practical tasks, such as distilling and disambiguating the meaning of language as part of "preprocessing" text before additional analyses. These methods allow researchers to increase the signal in their data and reduce noise before testing hypotheses. For example, the method of lemmatization will remove inflectional endings to create a single form for words such as "walk," "walking," and "walked." Sentence breaking will identify symbols such as periods or semicolons that demarcate semantic chunks. The emerging field of word-sense disambiguation uses context to disambiguate the true meaning of words that can be interpreted in different ways, such as the English word "funny" (Navigli, 2009). These preprocessing tools help distill language so that filler words are cut and words conveying important meaning are retained and made

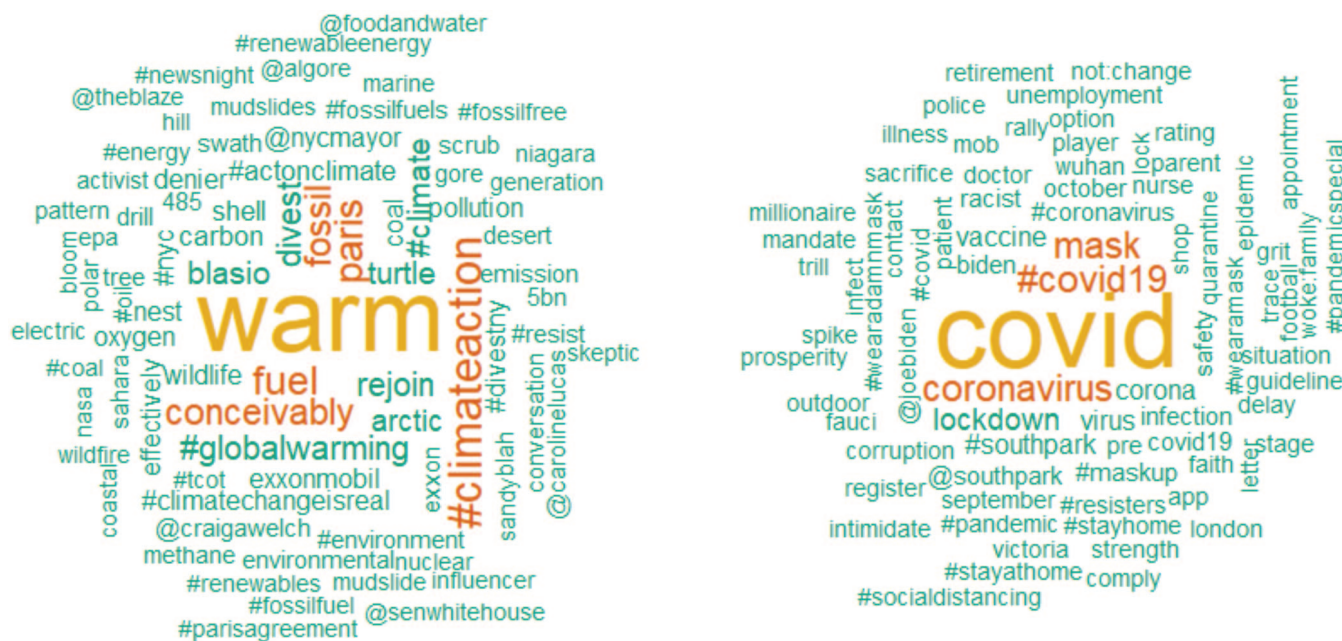


Fig. 1. Words from tweets about climate change (left) and COVID-19 (right). These word clouds come from an algorithm called *term frequency-inverse document frequency* (TF-IDF), which is designed to highlight words that best distinguish between two corpora. This text was preprocessed using lemmatization and stop-word removal before visualization. Code for generating these plots is available in the supplementary materials on OSF (<https://osf.io/hvcg3/>).

easier to detect. For example, Figure 1 shows a word-cloud of preprocessed keywords from tweets about climate change and tweets using COVID-19 hashtags. Note that there are no filler words such as “the” or “and” and that redundant forms of keywords (“ill” and “illness”) have been combined to minimize redundancy.

NLP resources. One distinct advantage of NLP algorithms is that they can operate over any sufficiently large digitally accessible corpora. In the early days of these algorithms, such corpora were difficult to find. But now there is a virtually limitless supply of digitalized text. As a case in point, the entire World Wide Web represents a digitalized corpus, and other corpora offer billions of words related to specific functions. The Google Books database contains a digitized corpus of books published in several languages over the past 400 years totaling more than 150 billion words (Michel et al., 2011). The Oxford English Corpus is the largest corpus of 21st century English, totaling more than 2.1 billion words across multiple English-language cultures (Oxford English Corpus, 2016). The *TIME Magazine* corpus of American English contains more than 100 million words of digitized *TIME Magazine* articles from 1923 to 2006 (Davies, 2007). The social-media sites Twitter (<https://developer.twitter.com>) and Reddit (<https://www.reddit.com/dev/api/>) both have easily accessible application programming interfaces (APIs), providing public access to millions of human

interactions. Training NLP models can be an arduous task, and this training process benefits from large sources of data, but once models are trained, they can be easily applied to datasets of any size. Table 1 contains a list of corpora that were built for text analysis.

NLP analyses may have been historically rare in psychology because they require advanced coding abilities. However, these barriers are now falling away as more psychologists develop proficiency with the R software environment (R Core Team, 2021). To help facilitate NLP proficiency in psychological science, we have created a five-part tutorial on NLP methods that covers (a) data acquisition and R packages, (b) preprocessing text data, (c) sentiment analysis using VADER, (d) word embeddings using GloVe, and (e) topic modeling. This R-based tutorial is available alongside our tutorial in comparative-linguistics methods in the supplementary materials at OSF (<https://osf.io/hvcg3/>).

Comparative linguistics as a way to understand cultural diversity and universality

Background. Research on comparative linguistics—the study of similarities and differences between languages and the evolution of these characteristics—is far older than NLP but has been applied to psychological questions only recently. In the earliest days of the field,

Table 1. Text Analysis Corpora

Corpus name	Link	Description
American National Corpus	http://www.anc.org/	Text corpus of American English containing 22 million words of spoken and written data since 1990. Mediums include email, tweet, and Web data, annotated for part of speech, lemma, and named entities.
British National Corpus	http://www.natcorp.ox.ac.uk/	Text corpus containing 100 million words of spoken and written language from the late 20th century from a variety of sources. Of the words, 90% are written and 10% are spoken. Tagged for parts of speech.
Corpus of Contemporary English	https://www.english-corpora.org/coca/	Text corpus containing 1 billion words of text from 1990 to 2019 from fiction, popular magazines, academic texts, TV and movie subtitles, blogs, and web pages. Allows searching by individual word. Tagged for parts of speech.
Google Books NGram Corpus	https://www.english-corpora.org/googlebooks/	Text corpus containing 200 billion words of written books. Subdivided into British English, American English, and Spanish. Mark Davies has made this corpus more accessible by allowing search by word, phrase, substring, lemma, part of speech, synonym, and collocates (nearby words). One strength of this corpus is its historical time span.
Oxford English Corpus	https://www.sketchengine.eu/oxford-english-corpus/	Text corpus of 21st-century English used by the makers of the Oxford English Dictionary, containing over 2 billion words. Includes language from many English-speaking countries and comprises many sources, including blogs, newspaper articles, emails, and social media. Tagged with extensive metadata. Users must apply for access through Oxford University Press.

linguists such as the Danish scholar Rasmus Rask (1787–1832) and the German scholar Jacob Grimm (1785–1863) pointed to striking similarities between such geographically dispersed languages as Sanskrit, Gothic, Latin, and Greek (Geisler & List, 2013). Many of these early insights relied on the qualitative classification of cognates, defined as words or parts of words in different languages that trace back to common ancestral forms (Crystal, 2011). The word for the number 1, for instance, is a cognate that shares its basic form and sound across Indo-European languages such as English (“one”), French (“une”), and German (“eins”), suggesting that these languages evolved from a parent language that had a similar word for this number.

Recent computational advances have expanded the scale and ambition of comparative linguistics. In particular, researchers have repurposed methods from biology to reconstruct language’s evolutionary ancestry. These approaches computationally aggregate many cognate classifications and use these classifications to develop language phylogenies (i.e., phylogenetic trees) that can be used to provide a proxy for cultural ancestry in the same way that biological phylogenetic trees display species’ ancestry. Figure 2 shows one such phylogenetic tree, in which modern countries are organized

on the basis of the historical relationships between their predominant languages. This map shows that countries such as Singapore and Indonesia are “sister cultures” that share a more common ancestor than do Singapore and the United States. The center of Figure 2 represents a hypothetical common ancestor for all languages, which diverged and diversified as humans spread around the world.

Comparative-linguistics insights are interesting in their own right, but they also have a surprisingly wide range of application to psychological questions involving culture and psychology. Many of these applications rely on modeling the relationship between cultures, analyzing patterns of coevolution between cultural and behavioral factors, and comparing the meaning of constructs across languages. Computational comparative-linguistics approaches have also allowed for the compilation of huge databases of words and their associated meanings, which allows for cross-cultural comparisons on an unprecedented scale.

Application 1: modeling cultural interdependence.

One of the most basic applications of comparative linguistics involves modeling interdependent datapoints in cross-cultural studies. Cross-cultural analyses will usually

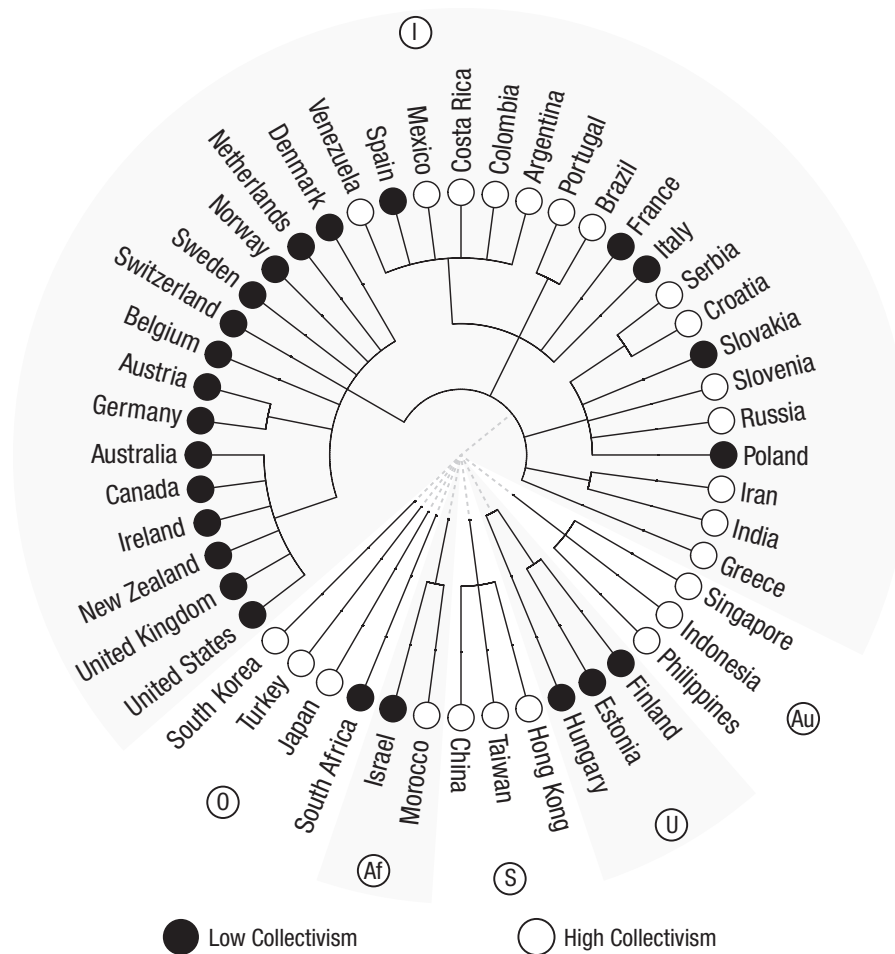


Fig. 2. The global distribution of individualism and collectivism. Filled nodes represent individualist cultures (low collectivism; scores fall below the midpoint of the 1-to-100 scale from <https://www.hofstede-insights.com/product/compare-countries/>) and open nodes represent collectivist cultures (high collectivism; scores fall above the midpoint of the 1-to-100 scale). This distribution is represented on a language-based phylogeny. Cultures connected by solid lines are part of the same language family (language family data are from Bromham et al., 2018). The circled letters represent the following language families: I = Indo-European, Au = Austronesian, U = Uralic, S = Sino-Tibetan, Af = Afro-Asiatic, O = other.

use regression to test for and explain patterns of variation across countries. These regressions assume that observations are independent, but comparative-linguistics research shows that many countries are interdependent because of their shared histories. Studies often treat Italy and Spain as independent units, for example, even though 80% of their lexicons overlap and the two societies share many features because of their recent common ancestry (Campbell, 2013). From a statistical standpoint, this is a case of “Galton’s problem”—interdependence between countries can lead to spurious correlations. For example, there is a highly cited link between cultures’ pathogen prevalence and political conservatism, which many scholars cite as evidence that disgust sensitivity makes people more conservative (Inbar et al., 2012). Yet this

link is rendered nonsignificant when controlling for cultural and linguistic interdependence via cultures’ shared language families and geographic regions, suggesting that pathogen prevalence and political conservatism do not have a causal relationship (Bromham et al., 2018).

Fortunately, concerns about Galton’s problem can be partially alleviated by nesting cultures within their language families (Jackson et al., 2020). Modeling Indo-European as a group-level variable in a multilevel regression makes it less likely that a spurious association arises because of similarities between countries such as Italy and Spain. This kind of nested analysis is slowly becoming more common in cross-cultural research (e.g., Jackson et al., 2020; Skogsgard et al., 2020) but it is still not standard practice in cross-cultural psychology.

Application 2: detecting patterns of cultural development. Cultural phylogenies also have the potential to yield important insights into the development of cultural differences because they track the relationship between linguistic and cultural groups over thousands of years. For example, consider worldwide variation in individualism–collectivism, which refers to cultures’ tendencies to either value individual rights and achievements (individualism) versus collective obligations and goals (collectivism). Most studies have observed that European countries are more individualistic than East Asian countries (Markus & Kitayama, 1991), but a cultural phylogeny can show that countries around the world with Germanic and Uralic languages are more consistently individualistic than countries with Latin and Slavic languages, suggesting that Northern and Central Europe may have historically been more individualistic than Western and Eastern Europe. In this way, phylogenetic trees can shed light on where and how cultural differences in human experience first emerged.

Whereas phylogenies represent the vertical inheritance of language and culture—where cultural information is passed down from one generation to another—it is also important to recognize that traits can be borrowed between groups, a process also known as horizontal transmission (Hoffer, 2002). For example, the word “honesty” in English is borrowed from the French language. Many comparative language databases flag suspected borrowings, and the World Loanword Database (WOLD; Haspelmath & Tadmor, 2009) is specifically designed to catalogue borrowings between languages. In principle, data on borrowings between languages could be represented in large-scale networks representing histories of contact and horizontal transmission between societies. Just as language phylogenies model the ancestry of cultures, language borrowing networks can model the diffusion of cultural constructs such as monogamy or psychological constructs such as intelligence. By tracking the diffusion of constructs through language, borrowing analyses have the potential to identify whether these factors are universal and, if they are not, why they have spread around the world over time. One plausible example could track whether the construct of self-esteem first emerged in individualist cultures in Western Europe and then was borrowed by collectivist cultures in South American and East Asia.

Modeling the evolutionary history of cultural variation also makes it possible to speculate about the causal origins of this variation. For the past decade, psychological science has begun grappling with the tremendous diversity in human culture and psychology, as well as the issues associated with focusing on WEIRD (Western, educated, industrialized, rich, and democratic) cultures (Henrich et al., 2010). Comparative-linguistics

methods can not only analyze diverse samples but also examine sources of cultural diversity. For example, surveys published in *Science* and *Science Advances* have argued that rice farming (vs. wheat farming) is responsible for current-day cultural differences in collectivism (Talhelm et al., 2014, 2018), but these correlational surveys have not been able to causally test this hypothesis or even establish whether agricultural changes predated cultural changes. Using analyses that incorporate both phylogenetic trees and borrowing networks could help establish causal direction by testing between different models of coevolution between rice farming and collectivism (R. D. Gray & Watts, 2017).

Phylogenetic language trees can also yield insights about universal tendencies in how people change and transmit words, concepts, and behaviors over time. Many articles show that words for lower numbers are transmitted more reliably than words for higher numbers during the formation of new languages, perhaps because lower numbers are used more frequently than higher numbers (Pagel et al., 2007; Pagel & Meade, 2018). For example, the Latin word for the number 2, “duo,” has a similar sound and spelling to the French word “deux” and the Italian word “due,” but the Latin word “undeviginti,” meaning “19,” looks and sounds less similar to the French word “dix-neuf” and the Italian word “diciannove.” However, these studies have not yet considered how psychological variables could influence such cultural transmissions. On the other hand, psychological studies using the “Bartlett method”—in which statements are transmitted from person to person like a game of “telephone”—have uncovered several psychological transmission biases (Bartlett & Bartlett, 1932/1995). For example, high-arousal concepts are transmitted more reliably than low-arousal concepts, and social concepts are transmitted more reliably than asocial concepts (Mesoudi & Whiten, 2008), illustrating the salience of high-arousal feelings (Kensinger, 2004) and sociality (Cacioppo & Cacioppo, 2018) to human experience. Comparing results from this paradigm with rates of lexical evolution (the evolution of words) could assess whether concepts that are reliably transmitted in minutes-long social interactions are also reliably transmitted over thousands of years of history.

Application 3: quantifying cross-cultural differences in meaning. Comparative-linguistics methods are also well-suited to examine the meaning of emotions, moral values, personality traits, or other psychological factors across cultures by examining how these factors are expressed as linguistic *concepts* (meanings attached to words; Jackendoff, 1992). Insofar as language represents the psychological categories that are relevant to its speakers, it is a useful tool for psychologists to measure the

extent to which a latent psychological *construct* (the latent meaning attached to clusters of observations; Fried, 2017) is shared within a culture over time or across cultures. For instance, researchers could examine how *concepts* such as “anger,” “disgust,” and “fear” are related to the psychological *construct* of emotion within or across languages.

One method for addressing this question examines a linguistic phenomenon called *colexification*, which occurs when two concepts are expressed with a single word (François, 2008; List et al., 2018). For example, the English word “funny” colexifies the concepts of “humorous” and “odd,” whereas the Russian word “ruka” colexifies “arm” and “hand.” As these examples illustrate, colexification often occurs when concepts are perceived as similar by speakers of a language (François, 2008), which makes frequency of colexification a useful measure of semantic closeness.

Studies are now beginning to build networks of colexifications to illustrate universality and cultural variation in semantic association across cultures. For example, Youn and colleagues (2016) showed that languages around the world had a similar meaning for physical entities such as “moon” and “sun” or “sea” and “lake,” suggesting that these concepts may have a universal meaning. Yet these colexification networks can also demonstrate cross-cultural variation if concepts show systematic variation in their colexifications across languages (Jackson, Watts, et al., 2019). For example, if “humorous” were colexified only with “odd” in European languages, this would suggest that strangeness is not a central aspect of humor across the world. Colexification is therefore a promising paradigm for testing whether Western theories about the universal structure of personality (e.g., “the big five”; Costa & McCrae, 2008), emotion (“basic emotions”; Ekman, 1999), morality (“moral foundations”; Graham et al., 2013), or psychopathology (American Psychiatric Association, 2013) generalize to non-Western cultures.

Comparative-linguistics resources. Comparative-linguistics resources are widely available, even though they are seldom used by psychologists. Many databases and datasets of comparative linguistics are publicly accessible and free to download. For example, the D-Place database contains language phylogenies representing the historical relationships among more than 1,000 human societies from around the world (Kirby et al., 2016), and the Database of Cross-Linguistic Colexifications (CLICS) contains colexifications from more than 2,000 languages (Rzymiski et al., 2020). Other databases contain information on cross-cultural variation in grammar (Dryer & Haspelmath, 2013), word borrowing (Haspelmath & Tadmor, 2009), and vocabulary (Dellert et al., 2020) from a range of large

and small languages. These databases provide rigorously vetted stimulus sets from enormous samples of cultures, and they often include data from small-scale cultural groups that are frequently underrepresented in psychological research. Table 2 summarizes several of these resources and provides links to their publicly available data.

Our supplementary materials at OSF (<https://osf.io/hvcg3/>) also contain tutorials for how to analyze phylogenetic trees (in R) and build colexification networks (in Python). These resources are intended for scholars with basic coding abilities but who have not yet used methods from comparative linguistics.

Limitations and opportunities for language analysis

Language analysis has many advantages over traditional psychological methods, but it also comes with important limitations. Although NLP approaches offer an unprecedented scale of analysis, they will seldom be more accurate than a human coder. NLP techniques also carry the same gender and racial biases as the language- and human-generated labels they are trained upon (Garg et al., 2018; Kiritchenko & Mohammad, 2018). Preprocessing methods in NLP analysis also have a trade-off between parsimony and accuracy. An algorithm that removes stop words and lemmatizes key words will help make text analysis simpler, but it can also neglect important information in context. Words such as “warm” and “warming” may be lemmatized even though they have different implications for climate-change belief.

Comparative-linguistics methods face different challenges. One challenge to using language to study cultural variation is that language groups do not always neatly correspond to cultural groups. Cultural groups can speak multiple languages, and languages can span many cultures. A language phylogeny therefore provides only an approximation of how societies developed and diverged from one another and may not be appropriate when large-scale language replacement has occurred in a sample. Language phylogenies may also be biased by word borrowings. Language phylogenies are built from datasets that exclude known borrowings, but undetected borrowings can make two languages seem more similar than they really are (Greenhill et al., 2009). Finally, all language-analysis methods are limited by the fact that language is only a rough approximate of human experience.

The limitations of NLP and comparative-linguistics are not insurmountable. Methods of separating the likelihood of horizontal and vertical inheritance are growing more advanced (Atkinson et al., 2016; Sookias et al.,

Table 2. Public Datasets of Historical and Cross-Cultural Language

Database	Link	Description
D-Place	https://d-place.org/	Aggregates data on cultures' evolutionary histories, ecologies, sociocultural structures, and geographic locations into one repository with rich metadata on sources of information, including previously established phylogenetic trees.
Cross-Linguistic Colexification Database	https://clics.clld.org/	Contains data on concept colexification from over 2,000 languages.
World Loanword Database	https://wold.clld.org/	Contains vocabularies of 1,000 to 2,000 entries for 41 languages around the world, as well as the likelihood that these words were borrowed from other languages.
Natural History of Song	https://osf.io/jmv3q/	Contains ethnographic descriptions of songs from 60 cultures. Also contains features of songs from 86 societies that were gathered through field recordings.
APiCS Online	https://apics-online.info/	A database of structural properties of creole and pidgin languages gathered from descriptive materials.
Glottolog	https://glottolog.org	A reference catalog of the worlds languages, providing expert classifications, geolocations, and references for more than 7,000 spoken and signed languages.
Concepticon	https://concepticon.clld.org	A reference catalog of concepts that are typically used in cross-linguistic studies, offering definitions, links to datasets in which the concepts were used, and additional metadata on psychological categories (norms, ratings, relations).
World Atlas of Language Structures	https://wals.info/	A large database of structural properties of language gathered from descriptive materials.

Note: Many of these databases are still in development, so their coverage will likely expand from these estimates.

2018), and subsets of machine-learning classifications can be vetted by human coders to confirm their accuracy before interpretation. However, these limitations are important to acknowledge, and they make language analysis well suited to complement (rather than replace) other methods in psychology, such as experimental design, correlational surveys, neuroimaging, psychophysiology, and computational modeling.

Using different forms of language analysis together also combines their relative strengths. NLP and comparative linguistics were developed for different goals and in very different fields, and thus have mostly distinct strengths and weaknesses. Whereas NLP can analyze data on the scale of millions and with high granularity across time and person, comparative linguistics operates on a truly global scale and can make inferences about human culture long before the advent of writing. For this reason, these methods are a perfect match, and some articles are showing the potential of combining these methods. For example, one recent article on cultural differences in word meaning showed that semantic vectors in word embeddings correlated highly with colexification (Thompson et al., 2020), validating the two approaches and suggesting that long-standing patterns of meaning in language persist today.

Unfortunately, researchers are rarely trained in both comparative linguistics and NLP. Figure 3 displays this dynamic in a network in which nodes represent methods and edge thickness represents the number of researchers who have been the first author on articles using different methods. The purpose of this figure, the data for which were drawn from a review of 200 different articles across NLP and comparative linguistics, is to underscore the lack of research that combines the scale of NLP with the cross-cultural and historical scope of comparative-linguistics methods. This network clearly shows that many researchers publish multiple methods within NLP and comparative linguistics, but few researchers publish methods that overlap both areas. Training in both sets of methods could foster interdisciplinary collaboration and increase the kinds of questions that scholars are able to answer.

Applying Language Analysis in Psychological Science: Three Case Studies

Psychological science still has work to do before researchers can master NLP and comparative linguistic methods. We dedicate the rest of this article to

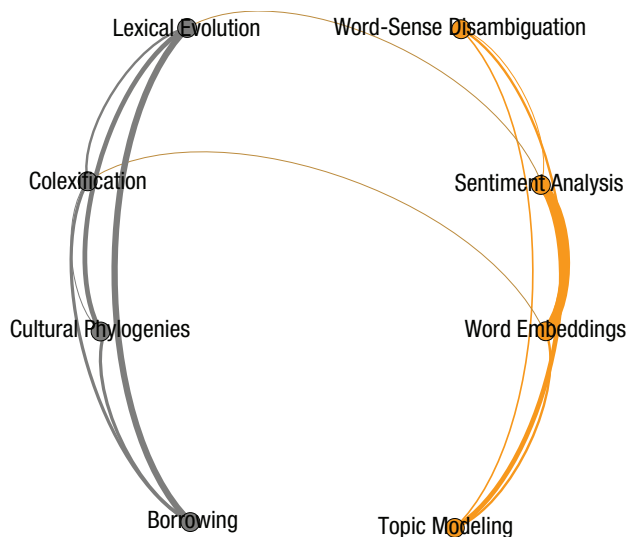


Fig. 3. A bibliometric analysis of eight forms of language analysis. Each node is a method, and links between nodes represent first authors who have published using both methods. Colors are communities of clustering nodes from the community-detection algorithm infomap. This algorithm separated comparative-linguistics methods (in gray) and NLP methods (in orange), which have little cross-over but high within-cluster interconnectedness (i.e., researchers who use phylogenetic mapping also study borrowing but do not study word embeddings). Data come from Table S1 in the supplementary materials on OSF (<https://osf.io/hvcg3/>).

illustrating how that might happen. First, we present Figure 4, which is a visual flowchart illustrating how the language-analysis methods discussed in this article can be employed to address psychological questions. We then summarize three case studies that demonstrate how NLP and comparative linguistics can yield new insights and increase the scale and diversity of study into three psychological constructs that have been notoriously difficult to study—emotion, religion, and creativity. In these sections, we highlight research that has used language analysis to address new questions or solve long-standing debates or that has used language-analysis methods to increase the scale or cultural diversity of research in these fields. This work illustrates the utility of language analysis for asking enduring psychological questions and foreshadows the potential of these tools to address psychological constructs across social, cultural, cognitive, clinical, and developmental psychology.

Emotion

Questions and debates about the nature of human emotion have existed since the earliest days of psychological science (Darwin, 1872/1998; James, 1884; Spencer, 1894; Wundt, 1897) and are relevant to psychological questions pertinent to social, clinical, and developmental

psychology. Language-analysis methods have already increased the scope of this long-standing field and generated original methods of addressing old debates.

One of the most enduring debates about emotions concern whether emotions are universal, inborn categories that possess little variation around the world or are socially learned categories that vary in their experience and conceptualization across cultures (Cowen & Keltner, 2020; Ekman & Friesen, 1971; Izard, 2013; Plutchik, 1991; Lindquist et al., 2012; Mesquita et al., 2016; Russell, 2003). We recently addressed this question by means of a comparative-linguistics approach using colexifications (Jackson, Watts, et al., 2019). This analysis allowed us to increase the scale and generalizability over previous field studies of cross-cultural differences in emotion that had relied on smaller sample sizes and two-culture comparisons (Bryant & Barrett, 2008; Ekman & Friesen, 1971; Gendron et al., 2014, 2015, 2020).

In our study, we computationally aggregated thousands of word-lists and translation dictionaries into a large database named “CLICS” (<https://clics.clld.org/>), and we used this database to examine colexification patterns of 24 emotion concepts across 2,474 languages. We constructed networks of colexification in which nodes represented concepts (e.g., “anger”) and edges represented colexifications (instances in which people had named two concepts with the same word), and then compared emotion colexification networks across language families. In contrast to Youn and colleagues (2016), who found universal colexification patterns involving concepts such as “sun” and “sky,” we found wide cultural variation in the colexification of emotion concepts such as “love” and “fear.” In fact, clusters of emotion colexification varied more than three times as much as the clustering patterns of colors—our set of control concepts—across language families (see Fig. 5). For example, “anxiety” was perceived as similar to “fear” among Tai-Kadai languages, but was more related to “grief” in Austroasiatic languages, suggesting that speakers of these language may conceptualize anxiety differently.

The variability in emotion meaning that we observed was associated with the geographic proximity of language families, suggesting that the meaning of emotion may be transmitted through historical patterns of contact (e.g., warfare, trade) and common ancestry. We also found that emotions universally clustered together on the basis of their hedonic valence (whether or not they were pleasant to experience) and to a lesser extent, by their physiological activation (whether or not they involved high levels of physiological arousal), suggesting valence and physiological activation might be biologically based factors that provide “minimal” universality to the meaning of emotion. In sum, this study

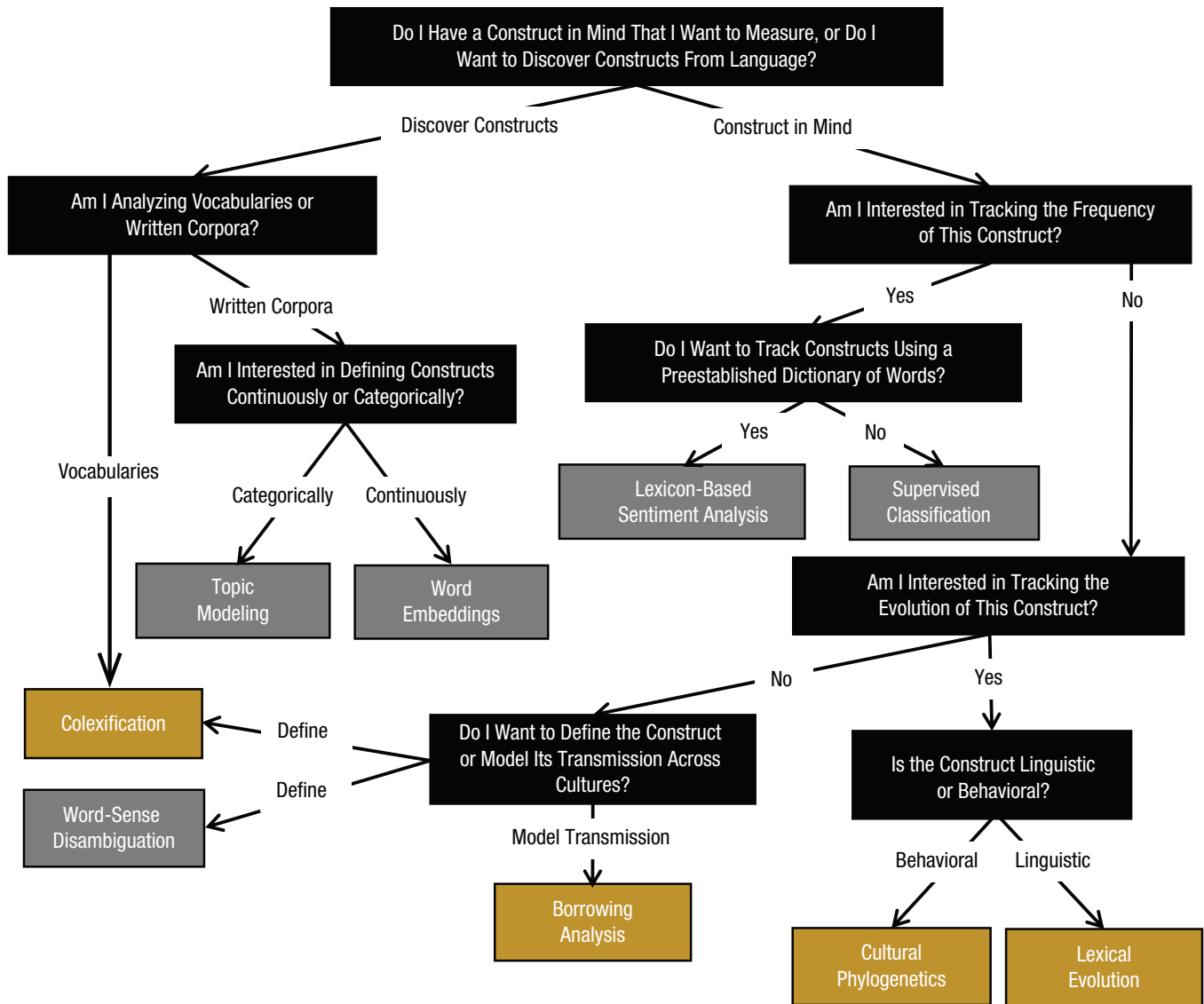


Fig. 4. A flowchart of different language-analysis methods and the kinds of questions they are best suited to answer. Orange boxes represent methods from comparative linguistics, and gray boxes represent methods from NLP. Black boxes approximate the questions that may guide researchers toward these methods. Concepts are defined here as the meaning associated with words. This is meant as a general guide for researchers interested in language analysis, and there is some overlap in classifications. For example, word embeddings can show how language conveys moods and attitudes, and collexification can sometimes uncover evolutionary dynamics.

used an unprecedented sample of cultures to yield new insights into the structure and cultural variation of human emotion.

A different set of language-analysis studies involving NLP are improving how psychologists measure emotion and track it over time and across social networks. For example, in a study of unprecedented historical scale, Morin and Acerbi (2017) used sentiment analysis to examine English fiction from 1800 to 2000 to assess whether the expression of emotion had changed systematically over time. They found a decrease in positive (but not negative) emotions conveyed in language over

history in three separate corpora of text. This change could not be explained by changing writer demographics (e.g., age and gender), vocabulary size, or genre (fiction vs. nonfiction), raising the possibility that something about emotion or its expression has itself changed over time.

Other studies have also used language analysis to track faster emotional dynamics, such as measuring the emotional qualities of social-media posts (Roberts et al., 2012; Yu & Wang, 2015) and testing whether the emotions of one person are likely to rapidly spread via language throughout that person’s social network. Such

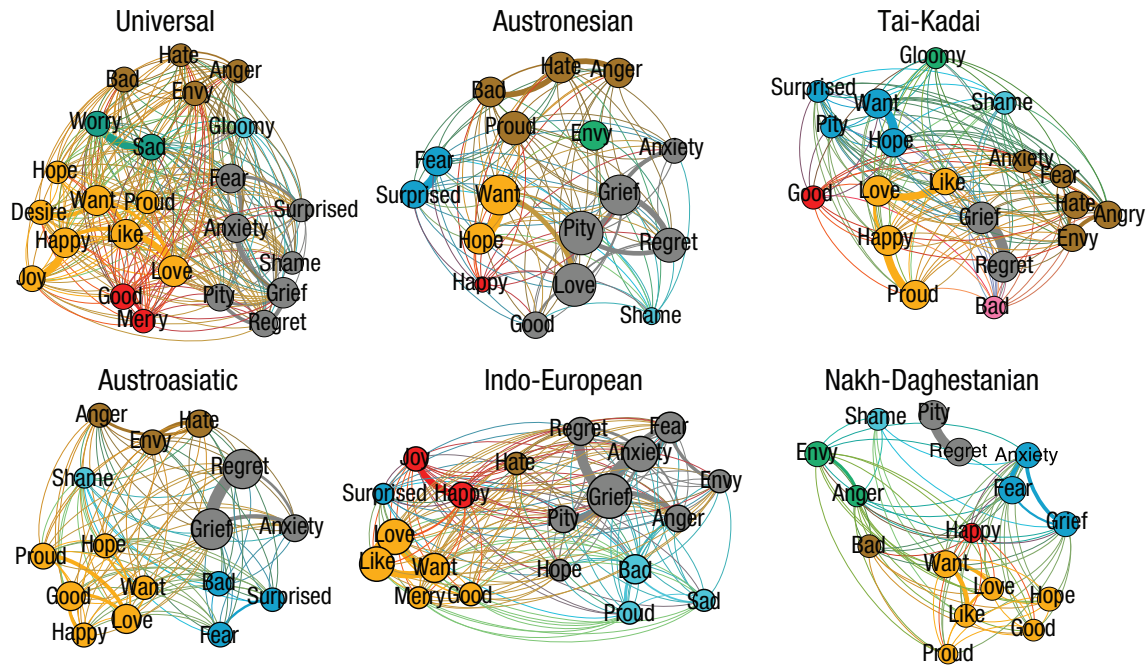


Fig. 5. The colexification structure of emotion concepts for all languages (top left) and for five individual language families in Jackson and colleagues (2019) analysis of emotion. Nodes are emotion concepts, and links between concepts represent the likelihood that these concepts will be colexified in a language. Color indicates semantic community, which refers to clusters of emotions that are similar in meaning. From Jackson, J. C., Watts, J., Henry, T. R., List, J. M., Forkel, R., Mucha, P. J., Greenhill, S., 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>. Reprinted with permission from AAAS.

studies have shown experimentally that emotional sentiment conveyed by language on social-media websites (e.g., Facebook) is more likely to make individuals who view that language express similar emotions (Kramer et al., 2014). Correlational studies find that social-media information with high emotional content is more likely to be shared than information with low emotional content (Brady et al., 2017). These studies show how affect can spread across many social-media users in a short period of time.

Religion

The science of religion has a rich legacy equal to that of the psychology of emotion; many psychological studies have addressed questions about the social value and historical development of religion. Language analysis has recently begun answering both kinds of questions with a scope and ecological validity that was not possible with traditional methods.

NLP analyses have shed light on the positive and negative ways that religion affects happiness and intergroup relations. Some social theorists view religion as a primarily positive force because it reinforces social connections and promotes well-being (Brooks, 2007).

On the other hand, “New Atheism” suggests that religion has a more negative effect on psychology by narrowing people’s worldviews and homogenizing the beliefs of religious adherents (Dawkins & Ward, 2006; Hitchens, 2008). Evidence for this debate has been mixed because of methodological challenges. For example, religious people frequently report more well-being than atheists in large national surveys, but they also show more social-desirability bias (Gervais & Norenzayan, 2012), which makes their self-reports less reliable.

NLP analyses are able to overcome these social-desirability limitations and have begun to show ecologically valid evidence that religion is linked to well-being. For example, Ritter et al. (2014) conducted a sentiment analysis of 16,000 users on Twitter and found that Christians expressed more positive emotion, less negative emotion, and more social connectedness than nonreligious users. Wallace et al. (2019) conducted a creative analysis of obituaries, finding that people whose obituaries mentioned religion had lived significantly longer than people whose obituaries did not mention religion, even controlling for demographic information.

Other NLP research has called the New Atheist proposition of religious worldview homogeneity into

question. For example, Watts and colleagues (2020) analyzed the explanations that Christian and nonreligious participants generated to explain a wide range of supernatural and natural phenomena and estimated the overlap of these explanations as a measure of worldview homogeneity. If religion does indeed homogenize adherents' worldviews, one would expect that religious people's explanations would share greater overlap than nonreligious people's explanations. Watts and colleagues (2020) used a text analysis approach known as Jaccard distances, which was able to estimate the similarity between participants' explanations of the world using overlapping key words, and test whether religious people offered more homogeneous explanations than did nonreligious people. Using this algorithm, the researchers found that religious people's explanations of supernatural phenomena were more homogeneous than nonreligious people's explanations, but their explanations of natural phenomena (e.g., the prevalence of parasites) were more *diverse* than were nonreligious explanations, probably because they drew on supernatural as well as scientific concepts when explaining the natural world.

Comparative linguistics has mostly contributed to questions about how religion has developed over time across cultures. Many of these analyses have focused on the "supernatural monitoring hypothesis": that watchful and punitive gods contributed to the evolution of social groups by increasing in-group prosociality and fostering large-scale cooperation (Johnson, 2016; Norenzayan et al., 2016). This idea is nearly a century old, arguably dating back to Durkheim (1912/2008), but most tests of the hypothesis have been correlational, and there is an ongoing debate about whether societies with large-scale cooperation tend to adopt moralistic religions or societies that adopt moralistic religions tend to be more cooperative (Whitehouse et al., 2019).

Researchers using comparative-linguistics methods recently addressed these debates by focusing on the development of religion in the Pacific Islands, where linguistic analyses have mapped out cultural phylogenies that can then be repurposed for cross-cultural research (R. D. Gray et al., 2009). Using these phylogenetic trees and implementing a method known as *Pagel's discrete* (Pagel, 1999), Watts and colleagues (2015) inferred the probability that ancestor cultures had high levels of political complexity (indicating large-scale cooperation), the probability that they believed in supernatural punishment, and the probability that they worshiped moralizing high gods. Their results showed partial support for both sides of the debate about religion and cooperation. Broad supernatural punishment (e.g., punishment for violating taboos)

tended to precede and facilitate political complexity. However, belief in watchful and punitive high gods (e.g., the Christian God) tended to occur only when societies were already politically complex.

Phylogenetic analyses have also shed light on the darker side of religious evolution, such as ritualized human sacrifice practices, which were common across the ancient world. According to the social-control hypothesis, ritual human sacrifice was used as a tool to help build and maintain social inequalities by demonstrating the power of leaders and instilling fear among subjugates. Yet evidence in support of this theory was based largely on individual case studies showing that higher classes often orchestrated ritual sacrifices (Carrasco, 1999; Turner & Turner, 1999). Watts and colleagues (2016) tested this prediction by examining patterns of ritual human sacrifice and social inequality across 93 Pacific societies that had been mapped onto an established language phylogeny (R. D. Gray et al., 2009). They found evidence that ritual human sacrifice often preceded, facilitated, and helped to sustain social inequalities, supporting the social-control hypothesis.

Creativity

Compared with the psychology of emotion and religion, that of creativity has a shorter history in psychology. Most psychologists agree that creativity contributes to personal feelings of self-fulfillment and societal innovation (Pratt & Jeffcutt, 2009; Wright & Walton, 2003), but the field is still exploring the best ways to measure creativity as a psychological construct. More than a dozen creativity-measurement paradigms exist in psychology. One such measure asks participants to name multiple uses for common household items such as article clips and bricks (Guilford, 1950), whereas others require participants to think of creative marketing schemes (Lucas & Nordgren, 2015) or draw an alien from another planet (Ward, 1994). In each paradigm, responses are qualitatively scored on creativity by trained research assistants. Although these tasks are themselves quite creative, the coding process can be onerous, and it can take months to obtain creativity ratings for a small behavioral study. Because these measures require custom tasks and laboratory settings, they are also rarely suitable for analyzing real-world creative behavior.

Language analysis has only recently been applied to study creativity, but NLP techniques are already advancing the measurement of creativity with paradigms that can be applied to both individuals in a small study as well as millions of people around the world. One such paradigm is "forward flow" (K. Gray et al., 2019). Forward flow asks people to free associate concepts, much

like classic psychoanalysis methods. But rather than qualitatively deconstructing these free associations, forward flow uses word embeddings to quantitatively analyze the extent that present thoughts diverge from past thoughts. For example, because “dog” and “cat” are frequently used together in large corpora, “dog” → “cat” would not represent as much divergence as “dog” → “fortress,” which are less frequently used together. Forward flow correlates with higher creativity scores on validated behavioral tasks such as the multiple uses task, and creative professionals such as actors, performance majors, and entrepreneurs score highly on forward flow (K. Gray et al., 2019). Forward flow in celebrities’ social-media posts can even predict their creative achievement (K. Gray et al., 2019). Forward flow may represent a rich and low-cost measure that could help capture creativity across people and societies.

Other NLP analyses have captured creativity in terms of divergences from normative language (e.g., Kuznetsova et al., 2013). Much like an unorthodox-looking alien, unorthodox patterns of language can signal creativity. However, it can be difficult to distinguish nonnormative and creative language (e.g., “metal to the pedal,” which is a reformulation of “pedal to the metal”) from nonnormative and nonsensical language (e.g., “the metal pedal to”). Berger and Packard (2018) developed a potential solution to this problem in a study of the music industry and used this method to test how creativity related to a product’s success. Their approach first used topic modeling to develop words that frequently appeared in different genres of music. For instance, words about bodies and movement were often featured in dance songs, whereas words about women and cars were often featured in country music songs. The study next quantified each song from the sample on its typicality according to how much it used language typical of its genre. Analyzing these trends found that songs that broke from tradition and featured atypical language performed better than songs featuring more typical language, offering some evidence that people prefer creative cultural products.

Recent language-analysis studies have already made a considerable impact on the study of creativity and show the potential of NLP for capturing and quantifying variability in creativity across people and products. Although no comparative-linguistics research has examined creativity, this subfield also has great potential for examining whether creativity varies in its structure across cultures and how creativity has evolved across history. Some historical analyses suggest that creativity has been highest during periods of societal looseness—periods with less rigid social norms and more openness (Jackson, Gelfand, et al., 2019). But this research was

done on American culture, and it is not clear whether these findings would generalize around the world.

Conclusion

Humans use language to express thoughts, convey emotions, and show biases. Researchers now have the tools to analyze and interpret this language, and here we encourage psychologists to use these tools to advance the field. Although research using language analysis is still young, it has already yielded major insights into emotion, religion, creativity, and many other processes. We have focused primarily on social, affective, and cultural psychology in this article given our own areas of expertise, but language-analysis methods are just as suitable for personality, clinical, developmental, and cognitive psychology. For example, many studies referenced in this article used language analysis to detect psychopathology or dementia and to help improve learning material in classrooms, which are core challenges in these other psychological subfields.

Our goal is not only to summarize the theoretical potential of language analysis but also to provide resources for psychological scientists who are interested in adopting language analysis. To this end, we encourage interested readers to browse Table S1, which contains 200 articles employing the methods we have summarized here. We also encourage readers to browse the resources in Tables 1 and 2, which are all publicly and freely accessible, and to visit our tutorials at <https://osf.io/hvcg3/> to see how language-analysis techniques are implemented in R.

With the proper rigor and training, the use of language analysis has the power to transform psychological science. It also allows our field to analyze data on a previously unimaginable scale and survey indigenous and historical groups that have been underrepresented in past psychological research. When used with more traditional methods, language analysis promises an enriched and more globally representative study of human cognition and behavior.

Transparency

Action Editor: June Gruber

Editor: Laura A. King

Declaration of Conflicting Interests


The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding


J. C. Jackson is supported by the National Science Foundation (NSF), the Royster Society of Fellows, and the John Templeton Foundation. J. Watts is supported by the John

Templeton Foundation and the Marsden Foundation of New Zealand (Grant 19-VUW-188). J.-M. List is supported by the European Research Council. K. A. Lindquist is funded by NSF Grant BCS 1941712 and National Institute on Drug Abuse Grant R01-DA051127. The views in this article do not necessarily reflect the views of these funding agencies.

ORCID iDs

Joshua Conrad Jackson  <https://orcid.org/0000-0002-2947-9815>

Joseph Watts  <https://orcid.org/0000-0002-7737-273X>

Kristen A. Lindquist  <https://orcid.org/0000-0002-5368-9302>

References

- Allport, G. W., & Vernon, P. E. (1930). The field of personality. *Psychological Bulletin*, *27*(10), 677–730.
- Althoff, T., Danescu-Niculescu-Mizil, C., & Jurafsky, D. (2014). *How to ask for a favor: A case study on the success of altruistic requests*. arXiv. <https://arxiv.org/abs/1405.3282>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). <https://doi.org/10.1176/appi.books.978089042559>
- Atkinson, Q. D., Coomber, T., Passmore, S., Greenhill, S. J., & Kushnick, G. (2016). Cultural and environmental predictors of pre-European deforestation on Pacific Islands. *PLOS ONE*, *11*(5), Article e0156340. <https://doi.org/10.1371/journal.pone.0156340>
- Back, M. D., Küfner, A. C., & Egloff, B. (2010). The emotional timeline of September 11, 2001. *Psychological Science*, *21*(10), 1417–1419. <https://doi.org/10.1177/0956797610382124>
- Bakker, M., Hartgerink, C. H., Wicherts, J. M., & van der Maas, H. L. (2016). Researchers' intuitions about power in psychological research. *Psychological Science*, *27*(8), 1069–1077. <https://doi.org/10.1177/0956797616647519>
- Bartlett, F. C., & Bartlett, F. C. (1995). *Remembering: A study in experimental and social psychology*. Cambridge University Press. (Original work published 1932)
- 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.
- Berger, J., & Packard, G. (2018). Are atypical things more popular? *Psychological Science*, *29*(7), 1178–1184. <https://doi.org/10.1177/0956797618759465>
- Bittermann, A., & Fischer, A. (2018). How to identify hot topics in psychology using topic modeling. *Zeitschrift für Psychologie*, *226*(1), 3–13. <https://doi.org/10.1027/2151-2604/a000318>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.
- Boyd, R. L., & Schwartz, H. A. (2020). 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.
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences, USA*, *114*(28), 7313–7318.
- Bromham, L., Hua, X., Cardillo, M., Schneemann, H., & Greenhill, S. J. (2018). Parasites and politics: Why cross-cultural studies must control for relatedness, proximity and covariation. *Royal Society Open Science*, *5*(8), Article 181100. <https://doi.org/10.1098/rsos.181100>
- Brooks, A. C. (2007). *Who really cares: The surprising truth about compassionate conservatism—America's charity divide: Who gives, who doesn't, and why it matters*. Basic Books.
- Bryant, G. A., & Barrett, H. C. (2008). Vocal emotion recognition across disparate cultures. *Journal of Cognition and Culture*, *8*, 135–148.
- Cacioppo, J. T., & Cacioppo, S. (2018). The growing problem of loneliness. *The Lancet*, *391*(10119), 426.
- Caluori, N., Jackson, J. C., Gray, K., & Gelfand, M. G. (2020). Conflict changes how people view God. *Psychological Science*, *31*(3), 280–292. <https://doi.org/10.1177/0956797619895286>
- Campbell, L. (2013). *Historical linguistics*. Edinburgh University Press.
- Carrasco, D. (1999). *City of sacrifice*. Beacon Press.
- Chandler, J., Rosenzweig, C., Moss, A. J., Robinson, J., & Litman, L. (2019). Online panels in social science research: Expanding sampling methods beyond Mechanical Turk. *Behavior Research Methods*, *51*(5), 2022–2038.
- Cichy, R. M., & Kaiser, D. (2019). Deep neural networks as scientific models. *Trends in Cognitive Sciences*, *24*(4), 305–317.
- Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, *1*(3), 98–101. <https://doi.org/10.1111%2F1467-8721.ep10768783>
- Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, *15*(10), 687–693. <https://doi.org/10.1111/j.0956-7976.2004.00741.x>
- Costa, P. T., Jr., & McCrae, R. R. (2008). *The Revised NEO Personality Inventory (NEO-PI-R)*. Sage.
- Cowen, A. S., & Keltner, D. (2020). Universal facial expressions uncovered in art of the ancient Americas: A computational approach. *Science Advances*, *6*(34), Article eabb1005. <https://doi.org/10.1126/sciadv.abb1005>
- Crawford, L., Margolies, S. M., Drake, J. T., & Murphy, M. E. (2006). Affect biases memory of location: Evidence for the spatial representation of affect. *Cognition and Emotion*, *20*(8), 1153–1169.
- Crystal, D. (2011). *A dictionary of linguistics and phonetics* (Vol. 30). John Wiley & Sons.
- Darwin, C. (1998). *The expression of the emotions in man and animals*. Oxford University Press. (Original work published 1872)
- Davies, M. (2007). *TIME Magazine corpus*. <https://www.english-corpora.org/time/>
- Dawkins, R., & Ward, L. (2006). *The god delusion*. Houghton Mifflin Company.
- De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. In *Proceedings of the 7th International AAAI Conference on Weblogs and*

- Social Media* (Vol. 7, pp. 128–137). Association for the Advancement of Artificial Intelligence. <https://ojs.aaai.org/index.php/ICWSM/article/view/14432/14281>
- Dellert, J., Daneyko, T., Münch, A., Ladygina, A., Buch, A., Clarius, N., & Mühlenbernd, R. (2020). NorthEuraLex: A wide-coverage lexical database of Northern Eurasia. *Language Resources and Evaluation*, 54(1), 273–301.
- Dryer, M., & Haspelmath, M. (2013). *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology. <http://wals.info/>
- Durkheim, E. (2008). *The elementary forms of the religious life*. Courier Corp. (Original work published 1912)
- Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preotiuc-Pietro, D., Asch, D. A., & Schwartz, H. A. (2018). Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences, USA*, 115(44), 11203–11208.
- Ekman, P. (1999). Basic emotions. In T. Dalgeish & T. Power (Eds.), *Handbook of cognition and emotion* (pp. 45–60). John Wiley & Sons.
- Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2), 124–129.
- Elvevaag, B., Foltz, P. W., Rosenstein, M., & DeLisi, L. E. (2010). An automated method to analyze language use in patients with schizophrenia and their first-degree relatives. *Journal of Neurolinguistics*, 23(3), 270–284.
- Enríquez, F., Troyano, J. A., & López-Solaz, T. (2016). An approach to the use of word embeddings in an opinion classification task. *Expert Systems With Applications*, 66, 1–6.
- François, A. (2008). Semantic maps and the typology of colexification. In M. Vanhove (Ed.), *From polysemy to semantic change: Towards a typology of lexical semantic associations* (No. 106, pp. 163). John Benjamins Publishing Company.
- Freud, S. (1901). *Psychopathology of everyday life*. Basic Books.
- Fried, E. I. (2017). What are psychological constructs? On the nature and statistical modelling of emotions, intelligence, personality traits and mental disorders. *Health Psychology Review*, 11(2), 130–134.
- Garcia, D., & Rimé, B. (2019). Collective emotions and social resilience in the digital traces after a terrorist attack. *Psychological Science*, 30(4), 617–628. <https://doi.org/10.1177/0956797619831964>
- 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, USA*, 115(16), 3635–3644.
- Geisler, H., & List, J.-M. (2013). Do languages grow on trees? The tree metaphor in the history of linguistics. In H. Fangerau, H. Geisler, T. Halling, & W. Martin (Eds.), *Classification and evolution in biology, linguistics and the history of science. Concepts – methods – visualization* (pp. 111–124). Franz Steiner Verlag.
- Gendron, M., Hoemann, K., Crittenden, A. N., Msafiri, S., Ruark, G., & Barrett, L. F. (2020). *Perception in Hadza hunter gatherers*. PsyArXiv. <https://psyarxiv.com/pf2q3/>
- Gendron, M., Roberson, D., & Barrett, L. F. (2015). Cultural variation in emotion perception is real: A response to Sauter, Eisner, Ekman, and Scott (2015). *Psychological Science*, 26(3), 357–359. <https://doi.org/10.1177/0956797614566659>
- Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014). Cultural relativity in perceiving emotion from vocalizations. *Psychological Science*, 25, 911–920. <https://doi.org/10.1177/0956797613517239>
- Gervais, W. M., & Norenzayan, A. (2012). Like a camera in the sky? Thinking about God increases public self-awareness and socially desirable responding. *Journal of Experimental Social Psychology*, 48(1), 298–302.
- Goldberg, Y., & Levy, O. (2014). *word2vec Explained: Deriving Mikolov et al.'s negative-sampling word-embedding method*. arXiv. <https://arxiv.org/abs/1402.3722v1>
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40(5), 223–234.
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In P. Devine & A. Plant (Eds.), *Advances in experimental social psychology* (Vol. 47, pp. 55–130). Academic Press.
- Gray, K., Anderson, S., Chen, E. E., Kelly, J. M., Christian, M. S., Patrick, J., Huang, L., Kennett, Y. N., & Lewis, K. (2019). “Forward flow”: A new measure to quantify free thought and predict creativity. *American Psychologist*, 74(5), 539–554.
- Gray, R. D., Drummond, A. J., & Greenhill, S. J. (2009). Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science*, 323(5913), 479–483.
- Gray, R. D., & Watts, J. (2017). Cultural macroevolution matters. *Proceedings of the National Academy of Sciences, USA*, 114(30), 7846–7852. <https://doi.org/10.1073/pnas.1620746114>
- Greenhill, S. J., Currie, T. E., & Gray, R. D. (2009). Does horizontal transmission invalidate cultural phylogenies? *Proceedings of the Royal Society B: Biological Sciences*, 276(1665), 2299–2306.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480.
- Guilford, J. P. (1950). Creativity. *American Psychologist*, 5, 444–454.
- Gunsch, M. A., Brownlow, S., Haynes, S. E., & Mabe, Z. (2000). Differential forms linguistic content of various of political advertising. *Journal of Broadcasting & Electronic Media*, 44(1), 27–42.
- Haspelmath, M., & Tadmor, U. (2009). *World Loanword Database (WOLD)*. <https://wold.clld.org/>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83.
- Hitchens, C. (2008). *God is not great: How religion poisons everything*. McClelland & Stewart.
- Hoffer, B. L. (2002). Language borrowing and language diffusion: An overview. *Intercultural Communication Studies*, 11(4), 1–37.

- Holmes, D., Alpers, G. W., Ismailji, T., Classen, C., Wales, T., Cheasty, V., Miller, A., & Koopman, C. (2007). Cognitive and emotional processing in narratives of women abused by intimate partners. *Violence Against Women, 13*(11), 1192–1205.
- Hong, L., & Davison, B. D. (2010). Empirical study of topic modeling in Twitter. In *SOMA '10: Proceedings of the First Workshop on Social Media Analytics* (pp. 80–88). Association for Computing Machinery. <https://doi.org/10.1145/1964858.1964870>
- Hutto, E., & Gilbert, C. H. E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media* (Vol. 8, pp. 216–225). Association for the Advancement of Artificial Intelligence. <https://ojs.aaai.org/index.php/ICWSM/article/view/14550/14399>
- Inbar, Y., Pizarro, D., Iyer, R., & Haidt, J. (2012). Disgust sensitivity, political conservatism, and voting. *Social Psychological and Personality Science, 3*(5), 537–544.
- Izard, C. E. (2013). *Human emotions*. Springer Science & Business Media.
- Jackendoff, R. (1992). What is a concept? In A. Lehrer, E. F. Kittay, & R. Lehrer (Eds.), *Frames, fields, and contrasts: New essays in semantics and lexical organization* (pp. 191–208). Routledge.
- Jackson, J. C., Gelfand, M., De, S., & Fox, A. (2019). The loosening of American culture over 200 years is associated with a creativity–order trade-off. *Nature Human Behaviour, 3*(3), 244–250. <https://doi.org/10.1038/s41562-018-0516-z>
- Jackson, J. C., Gelfand, M., & Ember, C. R. (2020). A global analysis of cultural tightness in non-industrial societies. *Proceedings of the Royal Society B: Biological Sciences, 287*(1930), Article 20201036. <https://doi.org/10.1098/rspb.2020.1036>
- Jackson, J. C., Watts, J., Henry, T. R., List, J. M., Forkel, R., Mucha, P. J., Greenhill, S., 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>
- Jacobs, A. M., & Kinder, A. (2017). “The brain is the prisoner of thought”: A machine-learning assisted quantitative narrative analysis of literary metaphors for use in neurocognitive poetics. *Metaphor and Symbol, 32*(3), 139–160.
- James, W. (1884). What is an emotion? *Mind, 9*, 188–205.
- Johnson, D. (2016). *God is watching you: How the fear of God makes us human*. Oxford University Press.
- Kagan, J., Reznick, J. S., & Snidman, N. (1987). The physiology and psychology of behavioral inhibition in children. *Child Development, 58*, 1459–1473.
- Kensinger, E. A. (2004). Remembering emotional experiences: The contribution of valence and arousal. *Reviews in the Neurosciences, 15*(4), 241–252.
- King, M. L., Jr. (1963, August 28). *I have a dream* [Speech audio recording]. American Rhetoric. <https://www.americanrhetoric.com/speeches/mlkihavedream.htm>
- Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H. J., Blasi, D., Botero, C., Bownen, C., Ember, C., Leehr, D., Low, B., McCarter, J., Divale, W., & Gavin, M. C. (2016). D-PLACE: A global database of cultural, linguistic and environmental diversity. *PLOS ONE, 11*(7), Article e0158391. <https://doi.org/10.1371/journal.pone.0158391>
- Kiritchenko, S., & Mohammad, S. (2018). Examining gender and race bias in two hundred sentiment analysis systems. In M. Nissim, J. Berant, & A. Lenci (Eds.), *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics* (pp. 43–53). Association for Computational Linguistics. <https://aclanthology.org/S18-2005/>
- Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research, 50*, 723–762.
- Kjell, O. N., Kjell, K., Garcia, D., & Sikström, S. (2019). Semantic measures: Using natural language processing to measure, differentiate, and describe psychological constructs. *Psychological Methods, 24*(1), 92–115. <https://doi.org/10.1037/met0000191>
- Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences, USA, 111*(24), 8788–8790.
- Kuznetsova, P., Chen, J., & Choi, Y. (2013, October). Understanding and quantifying creativity in lexical composition. In D. Yarowsky, T. Baldwin, A. Korhonen, K. Livescu, & S. Bethard. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 1246–1258). Association for Computational Linguistics. <https://aclanthology.org/D13-1124/>
- Landau, M. J., Meier, B. P., & Keefer, L. A. (2010). A metaphor-enriched social cognition. *Psychological Bulletin, 136*(6), 1045–1067.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*(2–3), 259–284.
- Lewis, M. P. (2009). *Ethnologue: Languages of the world* (16th ed.). SIL International.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 22*(140), 55.
- Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences, 35*(3), 121–143.
- List, J. M., Greenhill, S. J., Anderson, C., Mayer, T., Tresoldi, T., & Forkel, R. (2018). CLICS2: An improved database of cross-linguistic colexifications assembling lexical data with the help of cross-linguistic data formats. *Linguistic Typology, 22*(2), 277–306.
- Lucas, B. J., & Nordgren, L. F. (2015). People underestimate the value of persistence for creative performance. *Journal of Personality and Social Psychology, 109*, 232–243.
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224–253.
- Meier, B. P., & Robinson, M. D. (2006). Does “feeling down” mean seeing down? Depressive symptoms and vertical selective attention. *Journal of Research in Personality, 40*(4), 451–461.

- Mesoudi, A., & Whiten, A. (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1509), 3489–3501.
- Mesquita, B., Boiger, M., & De Leersnyder, J. (2016). The cultural construction of emotions. *Current Opinion in Psychology*, *8*, 31–36.
- Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., 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*(6014), 176–182.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (pp. 3111–3119). Neural Information Processing Systems. <https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf>
- Mishler, A., Crabb, E. S., Paletz, S., Hefright, B., & Golonka, E. (2015, August). Using structural topic modeling to detect events and cluster Twitter users in the Ukrainian crisis. In C. Stephanidis (Ed.), *International Conference on Human-Computer Interaction* (pp. 639–644). Springer.
- Morin, O., & Acerbi, A. (2017). Birth of the cool: A two-centuries decline in emotional expression in Anglophone fiction. *Cognition and Emotion*, *31*(8), 1663–1675. <https://doi.org/10.1080/02699931.2016.1260528>
- Murray, H. A. (1943). *Thematic apperception test*. Harvard University Press.
- Nagao, M. (1984). A framework of a mechanical translation between Japanese and English by analogy principle. In S. Nirenburg, H. L. Somers, & Y. Wilks (Eds.), *Artificial and human intelligence* (pp. 351–354). MIT Press.
- Navigli, R. (2009). Word sense disambiguation: A survey. *ACM Computing Surveys (CSUR)*, *41*(2), 1–69.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, *29*(5), 665–675.
- Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, *15*(1), 1–25.
- Nirenburg, S. (1989). Knowledge-based machine translation. *Machine Translation*, *4*(1), 5–24.
- Norenzayan, A., Shariff, A. F., Gervais, W. M., Willard, A. K., McNamara, R. A., Slingerland, E., & Henrich, J. (2016). The cultural evolution of prosocial religions. *Behavioral and Brain Sciences*, *39*, Article 1. <https://doi.org/10.1017/S0140525X14001356>
- Oscar, N., Fox, P. A., Croucher, R., Wernick, R., Keune, J., & Hooker, K. (2017). Machine learning, sentiment analysis, and tweets: An examination of Alzheimer's disease stigma on Twitter. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *72*(5), 742–751.
- The Oxford English Corpus. (2016). *Sketch engine*. Lexical Computing CZ s.r.o. <https://www.sketchengine.eu/oxford-english-corpus/>
- 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>
- Pagel, M. (1999). The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic Biology*, *48*(3), 612–622.
- Pagel, M., Atkinson, Q. D., & Meade, A. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, *449*(7163), 717–720.
- Pagel, M., & Meade, A. (2018). The deep history of the number words. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1740), Article 20160517. <https://doi.org/10.1098/rstb.2016.0517>
- Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic inquiry and word count: LIWC* (Version LIWC2015) [Computer software]. liwc.net.
- Pennebaker, J. W., & Lay, T. C. (2002). Language use and personality during crises: Analyses of Mayor Rudolph Giuliani's press conferences. *Journal of Research in Personality*, *36*(3), 271–282.
- Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: Language use over the life span. *Journal of Personality and Social Psychology*, *85*(2), 291–301.
- Plutchik, R. (1991). *The emotions*. University Press of America.
- Pratt, A. C., & Jeffcutt, P. (2009). *Creativity, innovation and the cultural economy*. Routledge.
- Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences, USA*, *115*(45), 11401–11405.
- R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. <http://www.R-project.org>
- Ritter, R. S., Preston, J. L., & Hernandez, I. (2014). Happy tweets: Christians are happier, more socially connected, and less analytical than atheists on Twitter. *Social Psychological and Personality Science*, *5*(2), 243–249.
- Roberts, K., Roach, M. A., Johnson, J., Guthrie, J., & Harabagiu, S. M. (2012, May). EmpaTweet: Annotating and detecting emotions on Twitter. In N. Calzolari, K. Choukri, T. Declerk, M. U. Dogan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *LREC'12: Proceedings of the Eighth International Conference on Language Resources and Evaluation* (pp. 3806–3813). European Language Resources Association. <https://aclanthology.org/L12-1059/>
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of Statistical Software*, *91*(2), 1–40. <http://doi.org/10.18637/jss.v091.i02>
- Rudkowsky, E., Haselmayer, M., Wastian, M., Jenny, M., Emrich, Š., & Sedlmair, M. (2018). More than bags of words: Sentiment analysis with word embeddings. *Communication Methods and Measures*, *12*(2–3), 140–157.
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172.
- Rzyski, C., Tresoldi, T., Greenhill, S. J., Wu, M.-S., Schweikhard, N. E., Koptjevskaja-Tamm, M., Gast, V.,

- Bodt, T. A., Hantgan, A., Kaiping, G., Chang, S., Lai, Y., Morozova, N., Arjava, H., Hubler, N., Koile, E., Pepper, S., Proos, M., Van Epps, B., . . . List, J. M. (2020). The Database of Cross-Linguistic Colexifications, reproducible analysis of cross-linguistic polysemies. *Scientific Data*, 7(1), Article 13. <https://doi.org/10.1038/s41597-019-0341-x>
- Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLOS ONE*, 8(9), Article e73791. <https://doi.org/10.1371/journal.pone.0073791>
- Short, J. C., McKenny, A. F., & Reid, S. W. (2018). More than words? Computer-aided text analysis in organizational behavior and psychology research. *Annual Review of Organizational Psychology and Organizational Behavior*, 5, 415–435.
- Skogsgard, I., Ember, C. R., Pitek, E., Jackson, J. C., & Carolus, C. (2020). Resource stress predicts changes in religious belief and increases in sharing behavior. *Human Nature*, 31(3), 249–271.
- Sookias, R. B., Passmore, S., & Atkinson, Q. D. (2018). Deep cultural ancestry and human development indicators across nation states. *Royal Society Open Science*, 5(4), Article 171411. <https://doi.org/10.1098/rsos.171411>
- Spencer, H. (1894). *Principals of psychology*. D. Appleton & Company.
- Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 427–448). Routledge.
- Tachyer, L. (2010, August 5). Books of the world, stand up and be counted! All 129,864,880 of you. *Google Books Search*. <http://booksearch.blogspot.com/2010/08/books-of-world-stand-up-and-be-counted.html>
- Talhelm, T., Zhang, X., & Oishi, S. (2018). Moving chairs in Starbucks: Observational studies find rice-wheat cultural differences in daily life in China. *Science Advances*, 4(4), Article eaap8469. <https://doi.org/10.1126/sciadv.aap8469>
- Talhelm, T., Zhang, X., Oishi, S., Shimin, C., Duan, D., Lan, X., & Kitayama, S. (2014). Large-scale psychological differences within China explained by rice versus wheat agriculture. *Science*, 344(6184), 603–608.
- 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.
- Thompson, B., Roberts, S. G., & Lupyan, G. (2020). Cultural influences on word meanings revealed through large-scale semantic alignment. *Nature Human Behaviour*, 4, 1029–1038. <https://doi.org/10.1038/s41562-020-0924-8>
- Turing, A. M. (2009). Computing machinery and intelligence. In R. Epstein, G. Roberts, & G. Bebet (Eds.), *Parsing the Turing test* (pp. 23–65). Springer. (Original work published 1950)
- Turner, C. G., & Turner, J. A. (1999). *Man corn: Cannibalism and violence in the prehistoric American Southwest*. University of Utah Press.
- Vo, B. K. H., & Collier, N. (2013). Twitter emotion analysis in earthquake situations. *International Journal of Computational Linguistics and Applications*, 4(1), 159–173. <https://www.gelbukh.com/ijcla/2013-1/IJCLA-2013-1-pp-159-173-09-Twitter.pdf>
- Vylomova, E., Murphy, S., & Haslam, N. (2019). Evaluation of semantic change of harm-related concepts in psychology. In N. Tahmasebi, L. Borin, A. Jatowt, & Y. Xu (Eds.), *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change* (pp. 29–34). Association for Computational Linguistics.
- Walker, C. B., & Chadwick, J. (1990). *Reading the past: Ancient writing from cuneiform to the alphabet*. University of California Press.
- Wallace, L. E., Anthony, R., End, C. M., & Way, B. M. (2019). Does religion stave off the grave? Religious affiliation in one's obituary and longevity. *Social Psychological and Personality Science*, 10(5), 662–670.
- Wallach, H. M. (2006). Topic modeling: Beyond bag-of-words. In W. Cohen & A. Moore (Eds.), *ICML '06: Proceedings of the 23rd International Conference on Machine Learning* (pp. 977–984). Association for Computing Machinery. <https://doi.org/10.1145/1143844.1143967>
- Wang, X., Zhang, C., Ji, Y., Sun, L., Wu, L., & Bao, Z. (2013, April). A depression detection model based on sentiment analysis in micro-blog social network. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Eds.), *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 201–213). Springer.
- Ward, T. B. (1994). Structured imagination: The role of category structure in exemplar generation. *Cognitive Psychology*, 27, 1–40.
- Watts, J., Greenhill, S. J., Atkinson, Q. D., Currie, T. E., Bulbulia, J., & Gray, R. D. (2015). Broad supernatural punishment but not moralizing high gods precede the evolution of political complexity in Austronesia. *Proceedings of the Royal Society B: Biological Sciences*, 282(1804), Article 20142556. <https://doi.org/10.1098/rspb.2014.2556>
- Watts, J., Passmore, S., Jackson, J. C., Rzymiski, C., & Dunbar, R. I. (2020). Text analysis shows conceptual overlap as well as domain-specific differences in Christian and secular worldviews. *Cognition*, 201, Article 104290.
- Watts, J., Sheehan, O., Atkinson, Q. D., Bulbulia, J., & Gray, R. D. (2016). Ritual human sacrifice promoted and sustained the evolution of stratified societies. *Nature*, 532(7598), 228–231. <https://doi.org/10.1038/nature17159>
- Weaver, W. (1955). Translation. In W. Locke & A. D. Booth (Eds.), *Machine translation of languages* (pp. 15–23). The MIT Press. (Original work published 1949)
- Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45. <https://doi.org/10.1145/365153.365168>
- Whitehouse, H., Francois, P., Savage, P. E., Currie, T. E., Feeney, K. C., Cioni, E., Purcell, R., Ross, R. M., Larson, J., Baines, J., ter Haar, B., Covey, A., & Turchin, P. (2019). Complex societies precede moralizing gods throughout world history. *Nature*, 568(7751), 226–229.

- Wilson, S., Mihalcea, R., Boyd, R., & Pennebaker, J. (2016, November). Disentangling topic models: A cross-cultural analysis of personal values through words. In D. Bamman, S. Dogruoz, J. Eisenstein, D. Hovy, D. Jurgens, B. O'Connor, A. Oh, O. Tsur, & S. Volkova (Eds.), *Proceedings of the First Workshop on NLP and Computational Social Science* (pp. 143–152). Association for Computational Linguistics. <https://aclanthology.org/W16-5619>
- Windsor, L. C., Cupit, J. G., & Windsor, A. J. (2019). Automated content analysis across six languages. *PLOS ONE*, *14*(11), Article e0224425. <https://doi.org/10.1371/journal.pone.0224425>
- Wright, T. A., & Walton, A. P. (2003). Affect, psychological well-being and creativity: Results of a field study. *Journal of Business & Management*, *9*(1), 21–33.
- Wundt, W. (1897). *Outlines of psychology* (C. H. Judd, Trans.). Thoemmes Press.
- Youn, H., Sutton, L., Smith, E., Moore, C., Wilkins, J. F., Maddieson, I., Croft, W., & Bhattacharya, T. (2016). On the universal structure of human lexical semantics. *Proceedings of the National Academy of Sciences, USA*, *113*(7), 1766–1771.
- Yu, Y., & Wang, X. (2015). World Cup 2014 in the Twitter World: A big data analysis of sentiments in US sports fans' tweets. *Computers in Human Behavior*, *48*, 392–400.