From Text to Thought
How Analyzing Language Can Advance Psychological Science

Joshua Conrad Jackson[1*], Joseph Watts[2,3,4], Johann-Mattis List[2], Ryan Drabble[1], Kristen Lindquist[1]

[1]Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill

[2]Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History

[3]Center for Research on Evolution, Belief, and Behaviour, University of Otago

[4]Religion Programme, University of Otago

*Corresponding author

Main Text Word Count, Excluding References: 8,116

Reference Count: 123

Acknowledgements: JCJ is supported by the National Science Foundation, the Royster Society of Fellows, and the John Templeton Foundation. JW is supported by the John Templeton Foundation and the Marsden Foundation of New Zealand (19-VUW-188). JML is supported by the European Research Council. The views in this paper do not necessarily reflect the views of these funding agencies.

Abstract: Humans have been using language for thousands of years, but psychologists seldom consider what natural language can tell us about the mind. Here we propose that language offers a unique window into human cognition. 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—comparative linguistics and natural language processing—are already contributing to how we understand emotion, creativity, and religion, and overcoming methodological 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 techniques with behavioral 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

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

1. Introduction

Humans have been using language in some form for millennia, and compiling written records for at least the last 5,000 years (Walker & Chadwick, 1990). In that time, our species has written nearly 130 million books containing over half a trillion words, we have produced 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 under-explored trove of data on the human experience.

Human language—be it spoken, written, or signed—has the power to reveal how we organize thoughts into categories and view associations between these categories. It shows how we view the salience of difference concepts, and how our understanding of these concepts may change over time. On a broader level, language can reveal variation in thought processes across different cultural and ideological groups, and illuminate universal and variable patterns in how humans view concepts as central to human experience as *God*, *emotion*, and *the self*. Language is thus a rich and dynamic window into human psychology that promises to shed light across the branches of psychological science.

Yet the promise of language analysis for psychological science has been largely unrealized, since most records of language have historically been inaccessible. Books have gathered dust on shelves, sacred texts have lay stored in museums, and songs have been stored either in human memory, cassette tapes, or albums. These vast stores of natural linguistic data lay out of reach over the 20th and early 21st centuries, while psychologists developed increasingly sophisticated measures of brain behavior (Nichols & Holmes, 2002), physiological activity

(Kagan, Reznick, & Snidman, 1987), self-report scales (Likert, 1932), and implicit categorization (Greenwald, McGhee, & Shwartz, 1998). But this is beginning to change.

Just as the printing press made language accessible to the masses, digital 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 digitalizing written text, but also of efficiently processing, storing, and recognizing patterns in this text. As a result of these innovations, records of language are no longer hidden away, but freely and easily accessible. We can now retrieve vast stores of written language 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 data to make predictions about cognition and culture by explaining how popular new methods of language analysis work and explaining which psychological predictions are most suitable for these methods. Our main goal is that, as the trickle of text analysis in psychology becomes a flood, we will be prepared to analyze language rigorously, accurately, and in a manner that takes full advantage of these methods' promise.

We also highlight practical advantages of language analysis. For example, we suggest that comparative linguistic paradigms are uniquely suited to resolve problems of representation and diversity in psychology by incorporating traditionally underrepresented cultures (Chandler et al., 2019; Henrich, Heine, & Norenzayan, 2010; Rad, Martingano, & Ginges, 2018), whereas computational paradigms such as natural language processing are uniquely suited to resolve problems with low power in psychological science by incorporating millions of datapoints (Bakker, Hartgerink, Wicherts, & van der Maas, 2016; Cohen, 1992).

Because of its theoretical and practical advantages, we suggest that rigorous language analysis is at least as valuable as Likert-scale responses, neuroimaging, psychophysiological readings, and other more common methods in psychological science. By complementing more traditional experimental and correlational methods with language analysis, we can gain a more complete understanding of human cognition and behavior.

2. What Does It Mean to Analyze Language?

Humans are intuitive language analysts. We process words, search for hidden meanings or innuendos, and react to sentiment and affect that is embedded in sentences. However, formal language analysis requires going beyond this intuition to quantitatively deconstruct the meaning of language. People may feel inspired when they perceive a rousing speech, but how can we capture this inspiration by quantifying the length, content, and format of a sentence? People may feel angry when someone expresses a differing opinion, but how can we predict this anger based on the linguistic content of two people engaged in conversation?

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 (1901), Murray's (1943) Thematic Apperception Test (TAT) analyzed the linguistic content of stories that people told in response to pictures, and Allport counted words in a dictionary to identify the structure of personality (Allport & Vernon, 1930). Many of these early methods had important limitations, and are rarely used in contemporary quantitative research, but they foreshadowed the potential impact of language analysis on psychological science.

The impact of language analysis for psychological theorizing was not realized until the development of computational methods of language analysis, the most popular of which may be the Linguistic Inquiry and Word Count (LIWC) technique (Pennebaker, Francis, & Booth, 2001; Tauscik & Pennebaker, 2010). LIWC uses word frequency to yield insight into the meaning of language. For example, excerpts of language with more communal word use (e.g. “we,” “us”) are probably expressing more affiliative meaning than language with more tribal language (“they,” “them”). LIWC uses these word-count methods with pre-programmed “dictionaries” which represent semantic categories. A “negative emotion” dictionary counts a predetermined set of words with negative affect whereas a “pronouns” dictionary counts pronouns. LIWC gives the percentage of words in a corpus that represents each dictionary. This method has been generative in psychology, and studies have applied LIWC to understand the psychological effects of ageing (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, Kufner, & Egloff, 2010; Cohn, Mehl, & Pennebaker, 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, nor can it determine the source vs. the target of this positivity. Consider, for example, an excerpt from Martin Luther King Jr.'s 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 papers have pointed out the limitations of these "bag of words" approaches which simply count the number of words rather than examining how these words are used in context (Enriquez, Troyano, & Lopez-Solaz, 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 (the words "up" and "down" describe both physical placement and psychological mood; Crawford et al., 2006, Meier & Robinson, 2006; Landau, Meier, & Keefer, 2010). However, a drawback of these methods is that they qualitatively analyze language, and therefore are not suitable for analyzing large-scale or cross-cultural datasets.

Another limitation of word-count methods is that they are almost exclusively focused 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 they may account for more than 99% of language analysis research published in psychology journals. Some efforts have been made to translate LIWC to other languages, but these efforts are very recent and focus more on replication than comparison (Windsor, Cupit, & Windsor, 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, comparative linguistics, involves the comparison of languages in order to distinguish patterns of linguistic descent and borrowing. Another tradition, natural language processing, uses methods developed in computer science to analyze semantic patterns in language. Both methods were developed outside of psychology, but have great potential for research into cognition and culture.

Comparative Linguistics as a way to Understand Cultural Diversity and Universality

Comparative linguistics has its roots in the early 19th century, when linguists like the Danish scholar Rasmus Rask (1787-1832) and the German scholar Jacob Grimm (1785-1863) first pointed to striking similarities between geographically dispersed languages like Sanskrit, Gothic, Latin, and Greek (Geisler and List 2013). As early psychoanalysts were trying to use spontaneous language to peer into the human unconscious, a different school of researchers had been focusing on a very different goal for decades: trying to estimate the historical relationship of languages and to determine how they had evolved into their current form. The methods of historical and typological language comparison were originally qualitative—systematically comparing words across several languages and estimating the likelihood that they had a common origin. Recent computational advances have further expanded the scale and ambition of the comparative enterprise, and allow researchers to repurpose methods of reconstructing evolutionary ancestry that were originally developed in biology. While these are not necessarily psychological discoveries, the models and histories represented in language phylogenies have a surprisingly wide range of applications to psychological questions.

One application of quantitative comparative linguistics is in cross-cultural psychology. Most cross-cultural studies survey the distribution of some variable across societies, but rarely model the interdependence between societies. For example, studies often treat Italy and Spain as independent units, even though 80% of their lexicons overlap and the two societies share many features due to their recent common ancestry and ongoing interactions (Campbell, 2013). From a statistical standpoint, this is problematic because it violates the assumption of independence in most regression models and can lead to inaccurate inferences—a violation known as "Galton's problem." For example, the highly cited link between cultures' pathogen prevalence and political conservatism is rendered non-significant when controlling for cultural and linguistic interdependence (Bromham et al., 2018).

Modeling cultural interdependence not only adds statistical rigor, it also allows for cross-cultural psychologists to ask new questions about evolution and human behavior. Most comparative linguists identify linguistic relatedness by examining "cognates"— words or parts of words in different languages which trace back to common ancestral forms (Crystal, 2011). The word for the number "one" is a cognate that shares its basic form across English ("one"), French ("une") and German ("eins"). By identifying and mapping overlapping cognates, it is possible 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. For example, in the tree depicted in Figure 1, Portugal and Brazil are sister cultures, as are Austria and Germany, and these two pairs of sister cultures share a common ancestor. These ancestral trees provide insights into the development of cultural differences. For example, if cultures such as Austria and Germany share a cultural psychological feature (e.g. individualism), this could suggest that this feature might have stemmed from common ancestry.

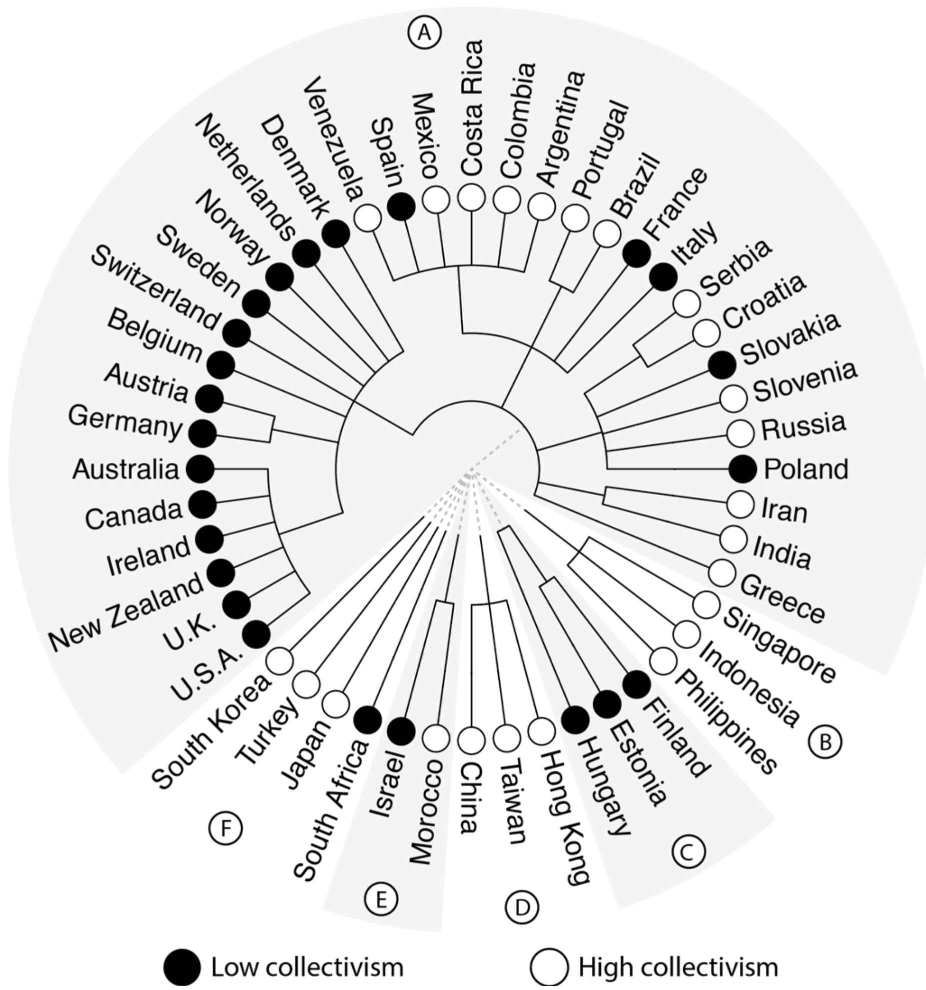

Figure 1. The global distribution of individualism-collectivism, colored such that dark nodes represent individualist cultures and light nodes represent collectivist cultures. “High” and “low” is determined by whether scores fall above or beyond the midpoint of the 1-100 scale from https://www.hofstede-insights.com/product/compare-countries/. This distribution is represented on a language-based phylogeny. Cultures connected by solid lines are part of the same language family (language family data retrieved from Bromham et al., 2018). The letters represent the following language families: A = Indo-European, B = Austronesian, C = Uralic, D = Sino-Tibetan, E = Afro-Asiatic, F = Other.

While phylogenies represent the vertical inheritance of language and culture, 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 “boulevard” in English is borrowed from the French language. Many comparative language databases flag suspected borrowings, and the WOLD database is specifically designed to catalogue borrowings between languages. In principle, data on borrowings between languages could be used to build large-scale networks representing histories of contact and horizontal transmission between societies. In a similar way to how language phylogenies can be used to account for and model the common ancestry of cultures, language borrowing networks could be used to account for and model the diffusion of cultural traits such as monogamy. We are not aware of research using language-based diffusion networks to answer psychological questions, but as we note below, such an approach could be a promising avenue for future research.

Psychological science has recently begun grappling with the tremendous diversity in human culture and psychology, as well as the issues associated with focusing on WEIRD cultures (Henrich, Norenzayan, & Heine, 2010). Modeling the evolutionary history of cognitive and behavioral traits also makes it possible to speculate about the origins of these differences. For example, surveys published in *Science* and *Science Advances* have argued that rice (vs. wheat) farming is responsible for current-day cultural differences in collectivism (Talhelm et al., 2014; Talhelm, Zhang, & Oishi, 2018), but these correlational surveys have not been able to estimate whether variation in farming predated variation in collectivism or vice versa. Using analyses that incorporate both phylogenetic trees and borrowing networks could help establish causal direction by testing between different models of coevolution between rice faming and collectivism (Gray & Watts, 2018).

Phylogenetic language trees can also yield insights about universal tendencies, such as the tendency to transmit some cultural artifacts more reliably than others. Consider how the words for *one* are remarkably similar across Indo-European languages (e.g. "eins" in German, "uno" in Spanish), whereas the words for *twenty* are much more diverse ("zwanzig" in German, "veinte" in Spanish). Comparative linguistics have used these patterns to claim that lower numbers are transmitted more reliably than higher numbers (Pagel & Meade, 2018), perhaps because lower numbers are used more frequently and there is more of a change cost to replacing them (Pagel, Atkinson, & Meade, 2007). But these studies have not yet considered the psychological factors that may also drive cultural transmission. For example, psychological studies suggest that high-arousal words are transmitted more reliably than low-arousal words, and that social words are transmitted more reliably than asocial words. Most of this psychological research studies transmission via brief experiments using the "Bartlett method," where messages or concepts are passed through a set of participants like a game of telephone (Bartlett, 1932/1995; Mesoudi & Whiten, 2008). Comparing results from this paradigm to real rates of lexical evolution could allow us to test whether the same concepts that evolve quickly in social interactions over a few minutes also rapidly evolve over thousands of years.

Colexification represents another phenomenon from comparative linguistics with psychological implications. Colexification occurs when two concepts are expressed with a single word (Francois, 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 (Francois, 2008). By extension, universal patterns of colexification should illustrate universal patterns of semantic associations. Youn and colleagues (2016) studied these associations by showing that people around the world perceived concepts such as “moon” and “sun” as similar, and concepts such as “sea” and “lake” as similar. Yet colexification can also demonstrate cross-cultural variation, especially if there are systematically different patterns of colexification across languages (Jackson et al., 2019). Given this potential, colexification is a promising paradigm for testing whether taxonomies of personality (e.g. “the big five”), emotion (“basic emotions”), and morality (“moral foundations”) are universal, or whether they have a varying semantic structure across cultures.

These methods are not just theoretically relevant to psychological questions; they are also publicly accessible and free to download. For example, the “D-Place” database contains language phylogenies representing the historical relationships between over 1,000 human societies from around the world (Kirby et al., 2016), and the “Cross-Linguistic Colexifications” (CLICS) database contains colexifications from over 2,000 languages (Rymski et al., 2020). Other databases contain information on cross-cultural variation in grammar (Dryer & Haspelmath, 2013), word borrowing (Haspelmath & Tadmore, 2009), and vocabulary (Dellert et al., 2020) from a range of large and small languages. These databases allow psychologists to collect rigorously vetted stimulus sets from a high-powered sample of cultures, while also extending psychological theories to small-scale cultural groups that are frequently underrepresented in psychological research. Table 1 summarizes several of these resources and provides links to their publicly available data.

Table 1. 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 meta-data on sources of information.
Cross-Linguistic Colexification Database	https://clics.clld.org/	Contains data on concept colexification from over 2000 languages.
World Loanword Database	https://wold.clld.org/	Contains vocabularies of 1000-2000 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 7000 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.

Natural Language Processing as a Tool for Studying Large-Scale Patterns of Cognition

Whereas comparative linguistics is more than two centuries old, natural language processing (NLP)—the interdisciplinary study of computer interaction with human language—is considerably younger. NLP's earliest notable paradigm was the "Turing Test": the hypothetical test wherein a computer's speech must appear indistinguishable from a human's (Turing, 1950/2009). Other early NLP developments involved ELIZA—a computer therapist that could respond to human complaints ("I have a stomach ache") with realistic therapist comments ("and why does your stomach ache?"), and Jabberwacky—a computer program designed in the 1980s to simulate entertaining but realistic human conversations, which is still running under the new name of "Cleverbot."

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 often involves building and testing models about the cognitive processes underlying human language. Early machine learning models of language translation and production were built using constrained statistical (Weaver, 1955), rule-based (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. Growing recognition of the flexibility and complexity of language (Evans & Levinson, 2009) and the failure of these methods to advance past relatively rudimentary tasks suggests that such approaches are unlikely to ever pass the Turning Test.

Today, artificial neural nets are now at the forefront of research in machine learning and NLP. These networks are designed to simulate the neurological mechanism of organic brains. Like human brains, artificial neural networks go through an extensive “training phase” that involves providing them with vast bodies of data and evaluating the way they classify or process that data. Once trained, these models excel at rapidly, accurately and efficiently preforming many classification and recognition tasks. Like the human brain, the way they process language can be complex and difficult to understand. But unlike the human brain, researchers can easily and 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).

It is impossible to exhaustively describe the many ways that NLP algorithms process language, but some specific NLP models can help illustrate how these methods recognize structure, diversity, and overlap in how humans use language. For example, consider that the words “happy” and “birthday” often appear together. A simple approach named topic modeling may observe this textual grouping to be semantically meaningful, and may create a semantic category that includes these words along with other co-occurring words such as “cake,” “candle,” and “gift” (Hong & Davison, 2010; Wallach, 2006). A more advanced approach such as word2vec may use word embeddings (in which words or phrases are mapped to vectors of numbers) to create numerical distances that represent semantic distances between these words (Goldberg & Levy, 2014; Mikolov et al., 2013). These estimates make NLP algorithms better suited than a simple word-count approach (Wallach, 2006). In Martin Luther King Jr.’s speech, for example, a more advanced NLP model could recognize that the negative words frequently precede positive words and could use this information to evaluate the tone of the text. NLP algorithms could also recognize speeches that are similar to King's "I have a dream" speech.

A multitude of other preprocessing tools can be used in conjunction with NLP models. For example, the method of lemmatization will remove inflectional endings to create a single form for words like "walk," "walking," and "walked" (Plisson, Lavrac, & Mladenic, 2004). "Sentence breaking" will identify symbols such as periods or semi-colons that demarcate semantic chunks (Pringle, Swerdlow, Wysoker, 2002). An emerging field of "word sense disambiguation" uses context to disambiguate the true meaning of words that can be interpreted in different ways (e.g. the English word "funny"; Navigli, 2009). These preprocessing tools can help cut through the ideocracies of language to maximize the ratio of signal to noise.

One distinct advantage of NLP machine learning 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 digitalized corpus of books published in several languages over the last 400 years totaling more than 150 billion words (Michel et al., 2010). 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 over 100 million words of digitalized Time magazine articles from 1923 to 2006 (Davies, 2007). 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.

Several papers have already applied NLP approaches to these corpora to gain valuable insights into processes of cultural and psychological change. For example, Jackson and colleagues (2019) used word2vec to identify clusters of words in the Google News Corpus related to “freedom” and “constraint” and then tracked the frequency of these words over time, finding that words related to freedom have risen over time whereas words related to constraint have fallen over time. Others have applied similar techniques to track the rise of harm-related concepts (Vylomova, Murphy, & Haslam, 2019), complementing word-count studies which have tracked the rise of individualism (Grossmann & Varnum, 2015) and historical changes in sexism (Varnum & Grossmann, 2017).

NLP approaches are also useful for studying contemporaneous processes. When they are paired with word count techniques, these algorithms can simultaneously identify semantic categories, and track the frequency of these categories across a wide range of behavioral contexts. These approaches have been used to identify symptoms of depression on personal blogs (Wang et al., 2013), predict the outcome of elections using political sentiment (Bermingham & Smeaton, 2011), and track emotional reactions to natural disasters and terrorist attacks (Vo et al., 2013; Garg, Garg, & Ranga, 2017; Garcia & Rime, 2019). These early results show the promise of these methods for identifying the stability of human emotion across context, the relationship between negative affect expression and harmful behavior, the potential barriers to political reconciliation, and a wide range of other social and affective processes.

Each of these studies has operated on a scale previously unimaginable in psychological science. Rather than recruiting dozens of subjects to participate in an experiment over several months, NLP algorithms can efficiently analyze millions of datapoints in seconds. These algorithms can also 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 like Facebook or Twitter.

NLP analyses may have been historically rare in psychology because they require advanced coding abilities, and most training in psychology has not included coding. However, these barriers are both now falling away. Psychologists are increasingly using R as a software to design and analyze studies, and methods such as word2vec, sentiment analysis, and topic modeling have now been translated from Javascript and Python to easy-to-use R packages (see Table 2). These innovations make it more feasible than ever to learn NLP machine learning approaches to complement existing psychological methods.

Table 2. R Text Analysis Packages

Package Name	**Link**	**Description**
OpenNLP	https://cran.r-project.org/web/packages/openNLP/index.html	Provides functions for chunking, parsing, annotation, part-of-speech tagging, and sentence segmentation
Rweka	https://cran.r-project.org/web/packages/RWeka/index.html	Provides a wide range of data mining functions, including tokenization and stemming functions, n-

		gram string detection, and simple work tokenization.
tm	https://cran.r-project.org/web/packages/tm/index.html	Text mining functions including stop-word removal, whitespace removal, stemming, and removal of sparse terms.
languageR	https://cran.r-project.org/web/packages/languageR/index.html	Provides functions to detect vocabulary richness, and other statistical analyses of text.
RKEA	https://cran.r-project.org/web/packages/RKEA/index.html	Package specializing in keyword detection and extraction from documents.

Language Analysis as a Complementary Tool in Psychological Science

Language analysis has many advantages over traditional psychological methods, but it also comes with important limitations. The scale and efficiency of NLP algorithms comes at the expense of precision (a human coder will often categorize language semantics better than a computer program), and all methods are limited by the fact that language is only a rough approximate of human experience. These limitations 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.

Different forms of language analysis can also be used *together,* as well as in tandem with other methods. Comparative linguistics and NLP were developed for different goals and in very different fields, and as such have different 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. It would be an important testament to internal validity, for example, if the same concept associations derived by analyzing colexifications replicated using word embeddings to map out the semantic relationship between concepts. Yet to date, research teams have rarely combined methods (see Figure 2). As we illustrate below, studies using comparative linguistics and NLP are already revealing major insights into cognition and culture on a scale unprecedented in psychological science.

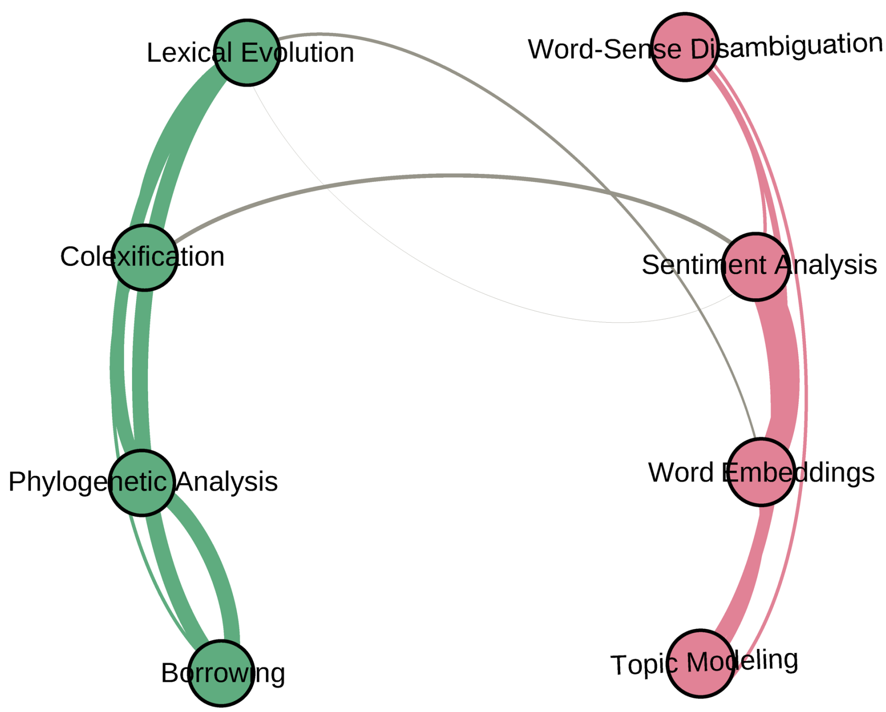

Figure 2. A bibliometric analysis of eight forms of language analysis. Each node is a method, and links between nodes represent first-authors who have published using more than one method.

Colors represent communities of nodes that group together, derived from an infomap community detection algorithm. This algorithm automatically separated comparative linguistics methods (in red) and NLP methods (in blue). This figure illustrates little cross between these two traditions, whereas there is high interconnectedness within these traditions (i.e., researchers who use phylogenetic mapping also study borrowing but do not study word embeddings). Table S1 in our supplementary materials includes the data from Figure 2, which may provide helpful background reading for those interested in these methods.

Psychological science still has a long road ahead, before our discipline can master these methods of language analysis and begin to use them together. Fortunately, comparative linguistics and NLP models are already making a major impact in three fields of psychological science—religious belief, emotion, and creativity—that foreshadows their ability to impact all fields of psychology research.

3. Three Case Studies

Methods of language analysis based on NLP and comparative linguistics are just beginning to find their way into psychological research. Here, focus on the insights NLP and comparative linguistics have provided in the study of emotion, creativity, and religion, three major areas of psychological research. We show how these analyses can be harnessed, and how they can complement behavioral and experimental research.

Emotion

Human emotion is one of the great paradigms of psychological science. Several early psychologists and biologists, including Darwin (1872/1998), James (1884), Spencer (1894), and Wundt (1897) advanced major theories of emotion that evolved over time into two broad perspectives. On the one hand, a universalist thesis stemming from Darwin proposed that emotions are universal and discrete properties that arose in the distant past through a process of biological evolution (Plutchik, 1991; Izard, 2013; Ekman & Friesen, 2003). On the other hand, a constructionist hypothesis proposed that emotions are not modular or biologically discrete, but emerge from more basic physiological and sociocultural mechanisms (Russell, 2003; Lindquist et al., 2012). Some “appraisal” perspectives have also fallen between these two camps, arguing that emotions can vary in their interpretation and frequency across cultures, but that they are nevertheless the function of basic “appraisals” that trigger universal emotional reactions (Scherer, 1984; Smith & Ellsworth, 1985; Siemer, Mauss, & Gross, 2007; Brosch, Pourtois, & Sander, 2010).

Early research into human emotion yielded no clear support for either side of this debate. A series of field studies in Papua New Guinea found that people in both cultures recognized posed facial expressions at rates better than chance (Ekman & Friesen, 1971). However, other studies among the Shuar of Ecuador (Bryant & Barrett, 2008), the Himba of Namibia (Gendron et al., 2014), and the Hadza of Tanzania (Gendron et al., 2020) found more variability in how people recognized and classified emotion. Many of these inconsistent results stem from methodological factors and demand characteristics. For example, nearly all studies sampled only two cultures, which severely limits the conclusions that can be drawn about universality versus cultural specificity. Furthermore, studies finding evidence for universality tend to train participants to recognize the meaning of certain emotions before the task, whereas studies that find evidence for cultural specificity tend to use more open-ended tasks (Gendron et al., 2015).

Language analysis is a useful method for overcoming the limitations of prior cross-cultural research on emotion by leveraging datasets of unprecedented scale—sampling the world's languages rather than recruiting subjects from two cultures—and by studying how people naturally view the meaning of emotion concepts outside of behavioral experiments. Early applications of language analysis to emotion used qualitative comparative linguistics methods. Wierzbicka (1992) analyzed dictionaries and wrote about which English-language emotion words had equivalents in other languages. Russell (1994) used a similar approach, studying the ethnographic record to find emotion words in small-scale societies which had no direct translation in English.

In a recent approach using quantitative methods of comparative linguistics, we performed a global analysis of colexifications involving emotion concepts (Jackson et al., 2019). Emotion colexifications such as when the Persian word *aenduh* colexifies the concepts of "grief" and "regret" are insightful because they show how language speakers view the semantic similarity between two emotion concepts (Francois, 2008). If the same colexifications occur across most of the world's languages in a universal pattern, this would suggest that people around the world view the relationships and semantic properties of emotion categories in the same way (Youn et al., 2016). A variable pattern of colexifications, however, would be evidence for the constructionist hypothesis that emotion semantics show important cross-cultural differences.

We carried out this analysis across 2474 languages and focused specifically on 24 emotion concepts. This study involved constructing networks in which nodes represented emotion concepts and edges represented colexifications, and then comparing emotion colexification networks across language families. The results showed wide cross-cultural variability; emotions' concept clustering varied nearly four times more than the clustering structure of color across the same set of language families. Figure 3 displays the variation in these clustering patterns. For example, “anxiety” was related to “fear” among Tai-Kadai languages, but was more related to “grief” among Austroasiatic languages.

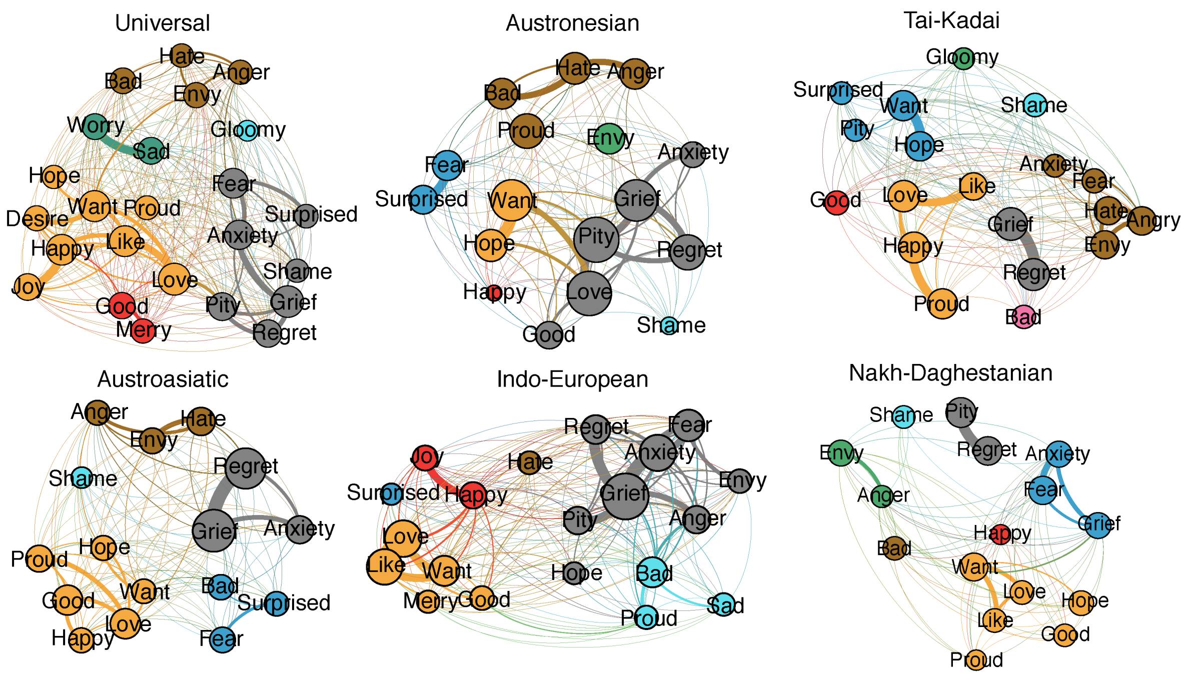

Figure 3. The colexification structure of emotion concepts for all languages (top left) and for six individual language families in Jackson and colleagues (2019) analysis of emotion. Nodes are emotion concepts and links between concepts represent colexifications. Color indicates semantic community. Figure reproduced with permission from AAAS.

Although there was wide linguistic variability, there were also systematic patterns within this cross-cultural variation. For example, language families in close geographic proximity had the most similar colexification networks, suggesting that the meaning of emotion can be transmitted through historical patterns of contact (e.g. warfare, trade) and common ancestry. All language families also appeared to organize emotions in terms of hedonic valence (the hedonic pleasantness versus unpleasantness of emotions) and to a lesser extent, activation (the physiological arousal associated with experiencing an emotion). This analysis represented the largest-scale investigation of human emotion to date, and illustrated the relevance of cross-cultural variation and universal structure to emotion semantics across cultures.

Whereas Jackson and colleagues' (2019) analysis addressed old questions in the study of emotion, NLP-derived sentiment analyses have begun to ask new questions that were not testable with behavioral paradigms. For example, Morin and Acerbi (2015) used sentiment analysis to examine English fiction from 1800 to 2000, finding a decrease in positive emotion (but not negative emotion) 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. non-fiction), posing an open question about why emotion expression in literature has changed so dramatically over time. Other studies have used sentiment analysis of social media (Roberts et al., 2012; Yu & Wang, 2015), showing that affective sentiment conveyed by language on social media websites like Facebook is contagious (Kramer, Guillory, & Hancock, 2014), and 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 emotion expression might have changed over history, and how it spreads in newly developed online means of social interaction.

Religion

Like emotion, religion is an enduring puzzle of human behavior. Many evolutionary anthropologists and cultural psychologists have approached this puzzle from a historical perspective, studying how religions have changed over time and how religion has changed human culture (e.g. Norenzaynan et al., 2016; Johnson, 2016). Psychologists of religion and spirituality have approached religion from a phenomenological perspective, studying how religion relates to well-being, meaning-making, and prosociality (Spilka, 2002; Hood, Hill, & Spilka, 2018), and how religious experience may have changed over time (Caluori et al., 2019). Language analysis promises to meet both goals.

Advances in comparative linguistics have already yielded insights into the historical development and impact of religious belief. Most of these insights have focused on the development of religion in the Pacific Islands, where phylogenetic analyses have mapped out cultural family trees using patterns of linguistic variation (Gray, Drummond, & Greenhill, 2009). These language trees provide a proxy for the ancestral relationships between different Pacific cultures, including how recently different cultures shared a common ancestor. Using these estimates, analyses can infer the probability that some form of religious belief was practiced earlier in history, and the probability that it caused some other sociocultural development.

The first study to apply this method to religion tested the “supernatural monitoring hypothesis,” the thesis that watchful and punitive gods contributed to human social complexity by increasing in-group cooperation (Johnson, 2016; Norenzayan et al., 2016). Over a century ago, Durkheim (1912/2008) suggested that religion’s major social function was to build cohesion and cooperation within communities, and this hypothesis was refined by a series of more recent high-profile papers in evolutionary anthropology (Johnson, 2005) and cultural psychology (Norenzayan & Shariff, 2008). Yet these papers focused on cross-cultural correlations, and could not infer whether social complexity preceded or followed the emergence of moralizing religion. By mapping social complexity and supernatural punishment beliefs to phylogenetic trees, Watts and colleagues (2015) showed that both hypotheses could be true. This research used a method known as “Pagel’s Discrete” that not only tests whether two traits are related, but also models how they coevolve over time (e.g. rice farming; Pagel, 1999). The Pagel’s Discrete method estimates direction of causality by inferring the temporal order of the emergence of traits and the effect that they have on one another (Watts et al., 2015; Watts et al., 2017). The results of this study suggested that forms of broad supernatural punishment (e.g. punishment for violating religious taboos) tended to precede and facilitate social complexity across Austronesia. However, watchful and punitive high gods (e.g. the Christian God) tended to follow social complexity, and analyses suggested that these beliefs were more likely a result of social complexity than a cause of social complexity.

Other phylogenetic analyses shed light on the darker side of religious evolution. As well as increasing parochial cooperation, religion can also be used to justify and legitimize a broad range of other behaviors and actions. This includes ritualized human sacrifice, which was practiced in early human societies throughout the 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. Evidence in support of this theory is largely based on individual case studies showing that higher classes often orchestrated ritual sacrifices (Turner & Turner, 1999; Carrasco, 1999). Watts and colleagues (2017) 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 (Gray, Drummond, & Greenhill, 2009). Using Pagel's Discrete, they found evidence that ritual human sacrifice often preceded, facilitated and helped to sustain social inequalities which provides general support for the social control hypothesis.

Psychological research has used NLP analyses to study the experience of religious belief, rather than its evolutionary history. Some social theorists view religion as a primarily positive force, since it reinforces social connections and promotes well-being (Bloom, 2012; Brooks, 2006). 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 (Hitchens, 2007; Dawkins & Ward, 2006). Since these predictions involve large scale societal patterns, they have been difficult to test in the laboratory, and most evidence for each perspective was either anecdotal or arose from surveys with design limitations. 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-reported responses to these questions unreliable.

NLP analyses are able to overcome these social desirability limitations, and have begun to yield novel insights into the association between religion and psychological processes. Some of these studies have provided evidence that religiosity is linked to higher well-being. For example, Ritter and Preston (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 non-religious users. Wallace et al (2018) conducted a creative analysis of obituaries, finding that people with obituaries mentioning religion lived significantly longer than obituaries that did not mention religion, even controlling for demographic information. While this analysis is correlational, it provides some early support for the possibility that religion’s effect on social connectedness and well-being can stave off the grave.

Other language 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 non-religious 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 non-religious people's explanations. Using a text analytic approach based on Jaccard distances between world explanations, Watts and colleagues (2020) found a more nuanced pattern. Religious people's explanations of supernatural concepts were more homogenous than non-religious people's explanations, but their explanations of natural phenomena (e.g. the prevalence of parasites) were more *diverse* than non-religious people's explanations, apparently because they drew on supernatural as well as scientific concepts in their explanations.

Taken together, these studies show how language analyses can address longstanding questions about the historical and cultural impact of religion, but also about how religion affects the way we make meaning and find well-being in everyday life.

Creativity

As with emotion and religion, creativity is a human experience that is both theoretically contentious and difficult to measure. Creativity contributes to well-being and self-fulfillment at an individual level as well as innovation at a societal level (Pratt & Jeffcutt, 2009; Wright & Walton, 2003). However, measuring creativity can be challenging, and a dozen paradigms have emerged in psychology that attempt to capture creativity with varied success. For example, one popular measure asks participants to name multiple uses for common household items such as paper clips and bricks (Guilford, 1950), whereas others require participants to think of creative marketing schemes (Lucas & Norgren, 2015) or draw an alien from another planet (Ward, 1994). In each paradigm, participants' responses are qualitatively scored on creativity by trained research assistants. While 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. Since these measures require custom tasks and research assistant instruction, they are also rarely suitable for analyzing spontaneous behavior in the real world.

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" (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, since "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 (Gray et al., 2019). Forward flow in celebrities' social media posts can even predict their creative achievement (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, Chen, & Choi, 2013). Much like an unorthodox-looking alien, unorthodox patterns of language can signal creativity. However, it can be difficult to distinguish non-normative and creative language (e.g. "metal to the petal") from non-normative and non-sensical language (e.g. the metal petal 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 basic questions about whether we prefer creative cultural products. 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 2014-2016 on its “typicality” based on 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.

These recent 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. While 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 et al., 2019). But this research was done on American culture and it is not clear whether these findings would generalize around the world.

4. Conclusion

Language provides a powerful window into the human mind. Humans use language to express our thoughts, emotions, and biases, and we now have the tools to analyze and interpret this language on an unprecedented scale. Here we summarized emerging research using language analysis to make psychological discoveries in three key areas of psychological inquiry that have been historically difficult to study. While this research is still young, it has already yielded major insights into emotion, religious belief, creativity, and many other psychological processes. We have focused on social and cultural psychology in this paper given our own interests and areas of expertise, but the techniques we describe here are just as suitable for testing longstanding and new predictions about clinical, developmental, and cognitive psychology.

Our goal is not only to summarize the theoretical potential of language analysis, but also to provide background reading and methodological tools for psychological scientists who are interested in adopting language analysis within their own research. To this end, we encourage interested readers to browse Table S1, which contains 150 papers employing the methods we have summarized here. We also encourage readers to browse the resources in Tables 1 & 2, which are all publicly and freely accessible.

With the proper rigor and training, the use of language analysis has the power to transform psychological science. It allows us to analyze human behavior on a previously unimaginable scale, conduct truly global studies of human culture, and survey indigenous and historical groups that have been underrepresented in past psychological research. Language analysis is not perfect, and it seldom offers the same level of causal inference as experimentation. In tandem with these other methods, however, language analysis promises an enriched and more globally representative study of behavior and cognition.

References

Allport, G. W., & Vernon, P. E. (1930). The field of personality. *Psychological bulletin*, *27*(10), 677-730.

Back, M. D., Küfner, A. C., & Egloff, B. (2010). The emotional timeline of September 11, 2001. *Psychological Science*, *21*(10), 1417-1419.

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.

Bartlett, F. C., & Bartlett, F. C. (1995). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Berger, J., & Packard, G. (2018). Are atypical things more popular? *Psychological Science*, *29*(7), 1178-1184.

Bermingham, A., & Smeaton, A. (2011, November). On using Twitter to monitor political sentiment and predict election results. In *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)* (pp. 2-10).

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*, *114*(28), 7313-7318.

Brosch, T., Pourtois, G., & Sander, D. (2010) The perception and categorisation of emotional stimuli: A review. *Cognition and Emotion* 24, 377-400.

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), 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 (AZ).

Bryant, G. A., & Barrett, H. C. (2008) Vocal emotion recognition across disparate cultures. *Journal of Cognition and Culture* 8, 135-148.

Caluori, N., Jackson, J. C., Gray, K., Gelfand, M. G. (2020). Conflict changes how people view God. *Psychological Science, 31*(3), 280-292.

Campbell, L. (2013). *Historical linguistics*. Edinburgh University Press.

Carrasco, D. *City of Sacrifice* (Beacon Press, 1999).

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.

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.

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., & Prodger, P. (1998). *The expression of the emotions in man and animals*. Oxford University Press, USA.

Davies, M. (2007). TIME Magazine Corpus: 100 million words, 1920s–2000s. *Retrieved September*, *1*, 2008.

Dawkins, R., & Ward, L. (2006). *The god delusion*. Boston: Houghton Mifflin Company.

Durkheim, E., & Swain, J. W. (2008). *The elementary forms of the religious life*. Courier Corporation.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124-129.

Ekman, P., & Friesen, W. V. *Unmasking the face: A guide to recognizing emotions from facial clues*. (Ishk, 2003).

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.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioural and Brain Sciences*, *32*(5), 429-448.

François, A. (2008). Semantic maps and the typology of colexification. *From polysemy to semantic change: Towards a typology of lexical semantic associations*, *106*, 163.

Freud, S. (1901). Psychopathology of everyday life. New York: Basic Books.

Garcia, D., & Rimé, B. (2019). Collective emotions and social resilience in the digital traces after a terrorist attack. *Psychological Science*, *30*(4), 617-628.

Garg, P., Garg, H., & Ranga, V. (2017, May). Sentiment analysis of the uri terror attack using twitter. In *2017 International Conference on Computing, Communication and Automation (ICCCA)* (pp. 17-20). IEEE.

Geisler, H. and J.-M. List (2013): Do languages grow on trees? The tree metaphor in the history of linguistics. In: Fangerau, H., H. Geisler, T. Halling, and W. Martin (eds.): Classification and evolution in biology, linguistics and the history of science. Concepts – methods – visualization. Franz Steiner Verlag: Stuttgart. 111-124.

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. Cultural relativity in perceiving emotion from vocalizations. *Psychological Science* 25, 911–920 (2014).

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.

Gendron, M. et al. Perception in Hadza Hunter Gatherers. Preprint at https://psyarxiv.com/pf2q3/ (2019).

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 preprint arXiv:1402.3722*.

Gray, R. D., & Watts, J. (2017). Cultural macroevolution matters. *Proceedings of the National Academy of Sciences*, *114*(30), 7846-7852.

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, K., Anderson, S., Chen, E. E., Kelly, J. M., Christian, M. S., Patrick, J., ... & Lewis, K. (2019). "Forward flow": A new measure to quantify free thought and predict creativity. *American Psychologist*, 74(5), 539-554.

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.

Grossmann, I., & Varnum, M. E. (2015). Social structure, infectious diseases, disasters, secularism, and cultural change in America. *Psychological Science*, *26*(3), 311-324.

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.

Hammarström, Harald & Forkel, Robert & Haspelmath, Martin & Bank, Sebastian. 2020. Glottolog 4.2.1. Jena: Max Planck Institute for the Science of Human History. (Available online at http://glottolog.org, Accessed on 2020-06-27.)

Haspelmath, M., & Tadmor, U. (2009). World Loanword Database (WOLD). Retrieved from 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.

Holmes, D., Alpers, G. W., Ismailji, T., Classen, C., Wales, T., Cheasty, V., ... & 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, July). Empirical study of topic modeling in twitter. In *Proceedings of the first workshop on social media analytics* (pp. 80-88).

Hoffer, B. L. (2002). Language borrowing and language diffusion: An overview. *Intercultural communication studies*, *11*(4), 1-37.

Hood Jr, R. W., Hill, P. C., & Spilka, B. (2018). *The psychology of religion: An empirical approach*. Guilford Publications.

Izard, C. E. *Human emotions*. (Springer Science & Business Media, 2013).

Jackson, J. C., Watts, J., Henry, T. R., List, J. M., Forkel, R., Mucha, P. J., ... & Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure, *Science, 366*(6472), 1517-1522.

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 Behavior*, *3*(3), 244.

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, USA.

Johnson, D. D. (2005). God's punishment and public goods. *Human Nature*, *16*(4), 410-446.

Kagan, J., Reznick, J. S., & Snidman, N. (1987). The physiology and psychology of behavioral inhibition in children. *Child Development*, 1459-1473.

Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H. J., ... & Leehr, D. (2016). D-PLACE: A global database of cultural, linguistic and environmental diversity. *PLoS One*, *11*(7), e0158391.

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*, *111*(24), 8788-8790.

Kuznetsova, P., Chen, J., & Choi, Y. (2013, October). Understanding and quantifying creativity in lexical composition. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 1246-1258).

Landau, M. J., Meier, B. P., & Keefer, L. A. (2010). A metaphor-enriched social cognition. *Psychological bulletin*, *136*(6), 1045.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *The Behavioral and Brain Sciences*, *35*(3), 121.

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.

List, Johann Mattis & Rzymski, Christoph & Greenhill, Simon & Schweikhard, Nathanael & Pianykh, Kristina & Tjuka, Annika & Wu, Mei-Shin & Forkel, Robert (eds.) 2020. Concepticon 2.3.0. Jena: Max Planck Institute for the Science of Human History. (Available online at http://concepticon.clld.org, Accessed on 2020-06-27.

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.

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.

Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., ... & Pinker, S.

(2011). Quantitative analysis of culture using millions of digitized books. *Science, 331*(6014), 176-182.

Michaelis, Susanne Maria & Maurer, Philippe & Haspelmath, Martin & Huber, Magnus (eds.) 2013. Atlas of Pidgin and Creole Language Structures Online. (Available online at http://apics-online.info, Accessed on 2020-06-27.)

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).

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.

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. *Artificial and Human Intelligence*, 351-354.

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.

"The Oxford English Corpus". Sketch Engine. Lexical Computing CZ s.r.o. Retrieved 27 October 2016.

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*.

Norenzayan, A., & Shariff, A. F. (2008). The origin and evolution of religious prosociality, *Science, 322*(5898), 58-62.

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), 20160517.

Packard, G., & Berger, J. (2020). Thinking of You: How Second-Person Pronouns Shape Cultural Success. *Psychological Science*, *31*(4), 397-407.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). Linguistic inquiry and word count: LIWC [Computer software]. *Austin, TX: liwc. net*, *135*.

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.

Plisson, J., Lavrac, N., & Mladenic, D. (2004). A rule-based approach to word lemmatization. In *Proceedings of IS* (Vol. 3, pp. 83-86).

Plutchik, R. *The emotions*. (University Press of America. 1991).

Pratt, A. C., & Jeffcutt, P. (2009). *Creativity, innovation and the cultural economy*. Routledge.

Pringle, L. G., Swerdlow, R. W., & Wysoker, A. (2002). *U.S. Patent No. 6,470,306*. Washington, DC: U.S. Patent and Trademark Office.

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*, *115*(45), 11401-11405.

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 *Lrec* (Vol. 12, pp. 3806-3813).

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145-172.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological bulletin*, *115*(1), 102-141.

Scherer, K. R. in *Approaches to Emotion* (Eds. Scherer, K., & Ekman, P.) 293-317 (Psychology Press, 1984).

Siemer, M., Mauss, I., & Gross, J. J. Same situation--different emotions: how appraisals shape our emotions. *Emotion*, 7, 592-600 (2007).

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology 48*, 813-838.

Spencer, H. *Principals of psychology*. (D. Appelton & Company, 1894).

Spilka, B. (2002). Psychology of religion: Empirical approaches. In *Religion and Psychology* (pp. 42-54). Routledge.

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), eaap8469.

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.

Turing, A. M. (2009). Computing machinery and intelligence. In *Parsing the Turing Test* (pp. 23-65). Springer, Dordrecht.

Turner, C. G. & Turner, J. A. *Man Corn: Cannibalism and Violence in the Prehistroic American Southwest* (Univ. Utah Press, 1999)

Varnum, M. E., & Grossmann, I. (2016). Pathogen prevalence is associated with cultural changes in gender equality. *Nature Human Behaviour*, *1*(1), 1-4.

Vo, B. K. H., & Collier, N. I. G. E. L. (2013). Twitter emotion analysis in earthquake situations. *International Journal of Computational Linguistics and Applications*, *4*(1), 159-173.

Vylomova, E., Murphy, S., & Haslam, N. (2019, August). Evaluation of semantic change of harm-related concepts in psychology. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change* (pp. 29-34).

Walker, C. B., & Chadwick, J. (1990). *Reading the past: Ancient writing from cuneiform to the alphabet*. Univ 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, June). Topic modeling: beyond bag-of-words. In *Proceedings of the 23rd international conference on Machine learning* (pp. 977-984).

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 *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 201-213). Springer, Berlin, Heidelberg.

Ward, T. B. (1994). Structured imagination: The role of category structure in exemplar generation. Cognitive Psychology, 27, 1–40. http://dx.doi .org/10.1006/cogp.1994.1010

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), 20142556.

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.

Watts, J., Passmore, S., Jackson, J. C., Rzymski, C., & Dunbar, R. I. (2020). Text analysis shows conceptual overlap as well as domain-specific differences in Christian and secular worldviews. *Cognition*, *201*, 104290.

W. Weaver (1955). Translation (1949). In: *Machine Translation of Languages*, MIT Press, Cambridge, MA.

Wierzbicka, A. (1994). Emotion, language, and cultural scripts. In S. Kitayama & H. R. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (p. 133–196). American Psychological Association.

Windsor, L. C., Cupit, J. G., & Windsor, A. J. (2019). Automated content analysis across six languages. *PloS one*, *14*(11), e0224425.

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. *Outlines of psychology* (C. H. Judd, trans.). (Thoemmes Press, 1897)

Youn, H., Sutton, L., Smith, E., Moore, C., Wilkins, J. F., Maddieson, I., ... & Bhattacharya, T. (2016). On the universal structure of human lexical semantics. *Proceedings of the National Academy of Sciences*, *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.