Title

Identifying Trends in the Behavior Analytic Literature via Natural Language Processing (I):

Proof of Concept with Simple Descriptions of Text

Abstract

Every day, people respond to and emit large amounts of textual stimuli such as text messages, social media posts, work emails, reminder notes, online news articles, and calendar planners. Despite the prevalence of text in everyday human behavior, relatively little research has focused on text as a primary verbal operant. This is likely because researchers did not have efficient methods to aggregate, clean, organize, or analyze data on published textual stimuli. But, recent advances in technology allow researchers to aggregate, clean, organize, and analyze published textual stimuli at scales previously impossible. In this study, we show how researchers can use advances in technology to analyze emitted textual stimuli and some of the resulting empirical possibilities. For this study, we downloaded all publicly available articles (*N* = 10,405) since the inception of five behavior analytic journals and from one control journal. We then used computational techniques referred to as Natural Language Processing (NLP) to turn the collection of raw text into a dataset that could be used for descriptive and exploratory analyses. We show how NLP techniques allowed us to describe trends over time in behavior analytic publications such as the length of articles, topics researched, differences between journals, and differences between emitted textual stimuli in applied and basic research literatures. Behavior analytic researchers interested in efficiently collecting and analyzing textual stimuli published in journals or online outlets may want to consider adding NLP techniques to their toolkit.

Keywords: textual stimuli; verbal behavior; natural language processing; quantitative analysis; computational analysis

**Introduction**

Text is all around us. Text can come in many forms such as a manuscript, a book, text message, social media post, or work email. Every day, people respond to and produce a lot of text. For example, every day, approximately 500 million tweets are tweeted (Twitter Usage Statistics, 2017), 300 billion emails are emailed (Johnson, 2021), and 18 billion text messages are texted (U.S. Texting Statistics, 2021). For many humans worldwide, emitting and responding to text is how they get paid, how they connect socially with friends and family, and how they contact the events that unfold in our world.

Despite the ubiquitous presence of text, the behaviors involved in generating text have received comparatively less attention from researchers. For example, past researchers have studied the social acceptability of behavior-analytic terms (Becirevic et al., 2016; Critchfield et al., 2018) as well as public perceptions of job titles (Boydston & Hirst, 2020). Alternatively, researchers have used textual prompts when teaching various verbal behavior skills such as appropriate social interactions (e.g., Yamamoto & Isawa, 2019), instruction following (e.g., Phillips et al., 2019) and reading (e.g., Staats et al., 1962). But, this research is typically focused on using text as a stimulus for something else rather than being focused on the behaviors involved in generating text.

One reason that analyzing the behaviors involved in generating text might be understudied is because analyzing the resulting permanent product manually is a cumbersome task. One classic operant approach to understanding response strength is to count the number of times a response occurs within an interval (i.e., response rate as strength; e.g., Skinner, 1932, 1938). For text, this would be counting the number of times each word appears in a paragraph, chapter, or larger body of text (Skinner, 1957). Once counted, the researcher has a sense of the relative strength of different behaviors at the time of writing or editing. But counting the number of times that all unique words occur in a paragraph or chapter takes a lot of time and effort – let alone if we’re interested in an entire book or scientific journal. Consider the sentence:

*“I have also read books, not for what they said about verbal behavior, but as records of verbal behavior. I have done my share of comma-counting. I have listened to people speaking and jotted down slips, curious phrases, or interesting intraverbal sequences.”* (Skinner, 1957; p. 454)

How many times does the word ‘behavior’ appear? How many unique words are in the passage? How many times did you read the sentence to verify the accuracy of your answers? Scaling such a manual approach to analyze a paragraph, paper, or journal is impractical.

A second reason that analyzing the behaviors associated with generating text might be understudied is because analyzing this behavior is complex. Taking a functional approach requires that text be examined as a chain of behaviors in context. Fortunately, generating text creates a permanent product that forces a consistent method of recording chains of behavior. This consistent method of data collection makes the data for analyzing such chains of behavior easy to access and interpret. But, from a functional perspective, how exactly does one determine context within a set of textual stimuli? How do we identify the proper unit of analysis based on the function of the author’s behavior (e.g., the word, the sentence, the phrase)? Further, readers have some amount of learned history with different words. How do we account for that history to understand the effect of a text on the reader or listener? How might the response of a reader or listener shape the textual stimuli emitted by an author?

Rather than answering questions about behavioral function *a priori*, behavior analysts have a long history of inductively examining behavior-environment and behavior-behavior relations. That is, a science of human behavior involves observing behavior in different ways and at different scales to discover order in our data (Skinner, 1953; p. 6). Once found, orderly relations can be used to make predictions about the occurrence of behavior and, if the predictions are consistently accurate, the orderly relations may lead to the control of behavior. Thus, one initial approach to examine the generation of textual stimuli by authors might be through exploratory analysis of published text to identify orderly relations worth pursuing. To search for orderly relations in published text, however, requires that researchers have efficient and robust methods for analyzing large amounts of textual stimuli.

Recent advances in technology and computation have reduced the time required for researchers to analyze published textual stimuli. Natural Language Processing (NLP) is the epitome of those advancements. NLP can be defined as the use of computers to systematically examine the relations between published textual stimuli and the surrounding verbal environment (e.g., Bird et al., 2009). On the simple end, NLP can be used to efficiently count and rank the number of words in large bodies of text (e.g., word clouds; DePaolo & Wilkinson, 2014; Heimerl et al., 2014), determine the similarity between two articles or books (e.g., clustering techniques; Blei et al., 2003), or estimate the likelihood that a text has a positive or negative tone about the topic (e.g., sentiment analysis; Pang et al., 2002). On the more complex end, NLP can be used to automate conversations with customers (e.g., chatbots; Shevat, 2017), change music or order groceries using our voice (e.g., Alexa, Siri), write fictional stories for entertainment (e.g., GPT-2; Zeigler et al., 2019), predict patient response to treatment (Cox et al., 2021), and detect hate speech (Davidson et al., 2017).

Another area where NLP has been successfully used is to analyze publication trends in scientific fields. Here, researchers have used NLP to understand citation patterns and relationships between published articles (e.g., Mariani et al., 2019b), identify trends and changes in published topics over time (e.g., Mariani et al., 2019a), to predict emerging research trends (Krenn & Zeilinger, 2020), and to highlight important areas for future research (Rzhetsky et al., 2015). For example, Mariani and colleagues (2019) analyzed 65,000 documents produced by 50,000 authors to explore open access publications about NLP. Through this, the researchers were able to display the evolution of research topics over time and identified authors who contributed innovative topics within their field.

Analyzing published research in behavior analysis using NLP could be beneficial for several reasons. First, past researchers have manually analyzed trends in the published behavior analytic literature (e.g., author gender – Kranak et al., 2021; Li et al., 2018; scholarly productivity – Shabani et al., 2004; Dixon et al., 2015). Such manual analyses are time-intensive, can lack analytic-generalizability, and can be influenced by expectancy (e.g., Atkinson & Delamont, 2006; Rice & Ezzy, 2000). NLP provides efficient, generalizable, and more objective methods for analyzing published textual stimuli. Second, practically analyzing published research literature manually requires that researchers reduce the total number of articles examined via well-defined inclusion criteria for topics, articles, or years. NLP allows researchers to practically analyze millions of articles with the same effort as analyzing five articles. Such efficiency of scale increases the possible research questions that can be asked about the published literature. In turn, researchers can take more of an inductive, exploratory approach to analyzing published behavior analytic literature.

The purpose of this proof-of-concept experiment was twofold. First, we sought to demonstrate how researchers can use NLP techniques to efficiently analyze published textual stimuli at scale in an objective manner allowing for easy replication. Second, we sought to demonstrate how a preliminary and descriptive analysis of peer-reviewed behavior analytic research articles allows researchers to identify: trends in publication topics over time; similarities and differences in the published textual stimuli published in different journals; and how the scientific literature of behavior analysis compares to one journal publishing health service research outside of behavior analysis. In total, this study demonstrates what NLP is and how researchers might use NLP to expand the questions they ask about human published textual stimuli and the field of behavior analysis.

**Method**

*Journals.* We gathered 10,405 published articles from behavior analytic journals and one control journal. For behavior analytic journals, we gathered articles from The Analysis of Verbal Behavior, Behavior Analysis in Practice, the Journal of Applied Behavior Analysis, the Journal of the Experimental Analysis of Behavior, and The Behavior Analyst. For the control journal, we randomly selected *Health Service Research* because much published behavior analytic writing centers around the delivery of health-related services. We gathered all published articles that were freely available online through the National Center for Biotechnology Information (NCBI). This allowed us to include all articles from the inception of the journal to the onset of any embargo period required by copyright law.

Table 1 shows the pseudocode used to download and process the Portable Document Format (PDF) of each article collected for this study. At a high-level, we used computer programming techniques (Python packages Selenium - Salunke, 2014; Beautiful Soup – Richardson, 2007) to automate the PDF download process and store each PDF in a folder corresponding to journal name and year. Next, we extracted the raw text from each PDF using tesseract-ocr for Python (Lee, 2021). Next, we preprocessed the text for analysis using similar methods as Mariani et al. (2019a, 2019b) via the Python packages NLTK (Bird et al., 2009) and Pandas (McKinney, 2010). Once extracted, the raw text from each article was entered and preprocessed as a unique row in a dataset along with publication year and the journal name (e.g., figure 2). Once the data were in a tidy format (Wickham, 2014), used existing labels in the dataset to subset data into various groups (e.g., by year published, by journal) to plot the data for visual analysis. A full implementation of these processes and analyses using Python can be found at <https://github.com/Behavioral-Data-Science-Research-Lab/research-rover>. All statistical analyses were conducted using the scipy.stats package for Python 3.8.8 (Virtanen et al., 2020).

**Results**

Figure 1 shows box-and-strip plots of article word counts (*x*-axis) for each journal in the data set (*y*-axis) and by decade of publication (color of markers)[[1]](#footnote-1). Box-and-strip plots are comprised of a box-and-whisker plot (i.e., the square box for each journal with the lines extending from it) with a strip plot overlain to show the individual datum that comprise the boxplot. For the box-and-whisker plots, the box shows the 25th percentile of article word counts (left-side of the box), the 50th percentile of article word counts (i.e., the median; middle line within the box), and the 75th percentile of article word counts (right-side of the box). The lines extending out from the box extend to the minimum (left extension) and maximum (right extension) word counts in the dataset excluding outliers. Outliers are defined as word counts that fall 1.5 times beyond the interquartile range (i.e., 75th percentile value – 25th percentile value).

Figure 1 highlights the differences in amount of published textual stimuli as a function of the journal in which the manuscript was published (*F*(5, 10,401) = 72.11; *p* < .001). For example, the journal *Health Service Research* had the smallest distribution of word counts with the 25th-to-75th percentiles of word counts ranging from 5,979 to 7,802 words. In contrast, *The Behavior Analyst* had the largest distribution of word counts with the 25th-to-75th percentiles of word counts ranging from 2,228 to 8,034. Visually, a final theme that seems to be present in Figure 1 is a shift from darker colors (earlier decades) to lighter colors (recent decades) moving from left to right along the *x*-axis for the journals *JABA*, *JEAB*, and *The Behavior Analyst*. This suggests the length of articles has gotten longer over time. Of note, these analyses and the generation of the graph took approximately 30 seconds to conduct compared to the years (decades?) this would have taken without NLP techniques.

Figure 2 was designed to look at trends in article length over time more directly. Figure 2 shows box-and-strip plots of article word counts (*y*-axis) by year (*x*-axis)and for each journal (marker color). For example, the word count of the 75th percentile trends upward from 1968 to 2005 (range 3,121 words to 9,767 words) and has since leveled off. The trend in the 25th percentile takes an inverted U shape from 1958 to 1996 where it remained level through 2015 before increasing to the present day. Additionally, the median word count per article was lowest in 1964 with 2,141 words per article, and the median word count reached its highest point in 2019 with 7,222 words per article. Like Figure 1, this description of the published literature took approximately 30 seconds using NLP techniques.

Though word count may give a measure of the general strength of different textual responses, it might be challenging to interpret the emission of words in isolation. That is, published textual stimuli might influence reader behavior based on their chained relation to other textual stimuli (e.g., location of a word within a phrase or sentences). Another way to analyze published textual stimuli might be to examine how often specific chained responses appear in text (e.g., two-word phrase a.k.a. bigrams).

Figure 3is a bar chart illustrating the count (*x*-axis) of the 15 most prevalent two-word pairings (i.e., bigrams; *y*-axis) in the entire dataset. For example, the bigram “response rate” appears most frequently in behavior analytic journals (*n*=36,291), “problem behavior” appears second most (*n*=21,308), and “behavior analysis” appears third most (*n*=19,185). Relative to the three-term contingency, one bigram in the top 15 refers to stimuli antecedent to responding (“stimulus control”), six refer to behavior of some kind (e.g., “response rate”, “problem behavior”), two refer to consequence schedules of some kind (“multiple schedule”, “terminal link”), and the remaining appear to refer to orienting the reader to other stimuli emitted by the author (e.g., “figure show”) or providing context for other textual stimuli (e.g., “per minute”, “present study”, “behavior analysis”). In sum, Figure 3 suggests behavior analysts emit the most published textual stimuli about behavior (e.g., response rates, problem behavior), and perhaps discussing our own science.

Figure 4shows the samedata as Figure 3 but separated by journal. Separating the data in this way allows us to analyze differences in relative chained-response strength across different journals. For example, in *Health Service Research* “health care”, and “primary care” and “statistically significant” occur within the top 15 of most prevalent bigrams. In *JEAB*, “response rate” and “multiple schedule” were the top two most prevalent bigrams (*n*=33,820 and 8,289, respectively). In *JABA*, the bigrams, “problem behavior” and “functional analysis” were the top two most prevalent bigrams (*n*=16,953 and 9,742, respectively). And, in *BAP*, “behavior analysis” and “behavior analyst” were the top two most prevalent bigrams (*n*=3,630 and 3,115, respectively). As a whole, this bigram analysis highlights how the central focus of these journals was captured using NLP techniques and basic word counts of two-word behavior chains.

Researchers can also use scatterplots as an alternative method to analyze how the verbal behavior published in different journals compares to one another (e.g., Figure 5). In these “scatter-text” plots, the frequency that words are emitted are graphed wherein each axis corresponds to the frequency that word appears in a specific document. For two similar data sets of textual stimuli, scatterplots of the word frequency would have data located primarily in the the lower-left quadrant (used infrequently in both data sets) or the upper-right quadrant (used frequently in both data sets). For two dissimilar data sets of textual stimuli, scatterplots of the word frequency would have data located primarily in the upper-left or lower-right quadrants (used infrequently in one and frequently in another).

Figure 5 shows a scatterplot of the 500 most prevalent words in behavior analytic journals as a whole and the 500 most prevalent words in *Health Service Research*. Given, *Health Service Research* is a control journal, we expected data to fall along the diagonal from the upper-left portion of the plot to the lower right portion of the plot. This pattern is what we observed and suggests the verbal behavior emitted by authors publishing in *Health Service Research* include words that infrequently appear in behavior analytic publications. Similarly, verbal behavior published by authors publishing in behavior analytic journals appear very infrequently in *Health Service Research*  publications (e.g., reinforcer appears 1,107 times per 25,000 words in behavior analytic publications compared to 0 within *Health Service Research*).

Figure 5 can also be used to identify the overlapping variability in word use between *Health Service Research* and behavior analytic journals. Here, readers should orient to the overall spread of the data (i.e., the area covered on the plot). As the frequency of words increased in behavior analytic journals (lower-right quadrant of Figure 5), the frequency those same words were used in *Health Service Research* declined linearly. In contrast, as the frequency of words increased in *Health Service Research* (upper-left quadrant of Figure 5), no common trend or pattern is observed in the frequency the words are used in behavior analytic journals. For example, “policy” and “range” are used with similar frequency in behavior analytic journals but used almost most frequently (“policy”) and below average frequency (“range”) in *Health Services Research*. The primary exception here is the upper-left-most section of Figure 5 that takes on closer to a linear trend, albeit not as well defined as the linear trend for behavior analytic journals.

Figure 6 shows the same type of scatterplot comparing the *Journal of Applied Behavior Analysis* and the *Journal of the Experimental Analysis of Behavior*. The first notable feature is that the words primarily fall in the upper-left and lower-right quadrants. This suggests that the words being used most frequently in *JABA* are being used infrequently by authors in *JEAB* and vice versa. However, unlike Figure 5, the relative spread of the data is much greater in each quadrant of Figure 6. This suggests that—though clear differences in verbal communities exist—the degree of dissimilarity is less than compared with behavior analytic journals and *Health Services Research*.

**Discussion**

We used NLP techniques to conduct simple descriptive analyses of 10,405 published journal articles from five behavior analytic journals and one control journal. We found that the overall amount of verbal behavior, and the variability of word count, published as textual stimuli in behavior analytic journals differed with journal articles getting longer and of greater variability in length over time. We also demonstrated how NLP could be used to analyze the content of data sets containing large amounts of textual stimuli through counts of verbal behavior chains as well as through scatterplots comparing verbal behavior published across different journals. In applying NLP techniques to this large data set, we demonstrated that NLP might be a tool to gain novel insights into underexplored areas of verbal behavior such as writing or authoring manuscripts for publication and for comparing the verbal behavior of different communities within and external to behavior analysis. We discuss each of these findings in turn.

Figures 1 and 2 provide data on simple word counts for different journals. We were not present to observe, measure, or analyze the specific contingencies that led to the word counts published in every one of the manuscripts shown as markers across these plots. Nevertheless, the potentially idiosyncratic and unique contingencies surrounding each author’s writing behavior led to a permanent product. One potential function of these authors’ writing behavior might be to inform a specific audience about the results of an experiment focused on a specific topic. The data presented here suggest that the contingencies surrounding publishing in different journals (e.g., journal policies, instructions for editors and reviewers) has led to differing amounts of verbal behavior across journals and with all journal contingencies increasing the amount of emitted verbal behavior over time. Of note, the increased variability in article word length over time (Figure 2) was likely influenced, but not solely determined, by the increase in number of journals being published as the original publication years for each journal were as follows: *JEAB*: 1958; *JABA*: 1968; *The Behavior Analyst*: 1976; *The Analysis of Verbal Behavior*: 1982; *Health Services Research*: 2005; and, Behavior *Analysis in Practice*: 2008. The variability present in Figure 2 was relatively similar the year before and after each of these calendar years suggesting something else led to increased variability in article length.

It is unclear what the ‘ideal’ length of a journal article should be. We currently live in a climate where many readers consistently contact 250-character Tweets, millions of webpages of text spanning millions of varied topics and of varied length, and with competition for reader attention coming from many outlets. If we assume writers write to be read, a unique question that follows from our analysis is whether the contingencies that have consistently led to longer journal articles will increase or decrease the probability that any one published article is read by any given reader; and, how article length changes the probability of the total number of readers likely to read an article. We hedge our bets on the assumption that decreasing the probability that articles are read will likely decrease the relevance and survival of a journal or the topic of focus in minimally read journal articles. Every article contains a unique set of verbal stimuli. Future research might seek to test our assumption and identify the relevant set of textual stimuli that increase the likelihood ana article is being read and influences practitioner or researcher behavior. Importantly, NLP techniques makes answering questions such as that just posed much more practical than manual coding and analysis.

Figures 3 and 4 highlight one answer to an interesting challenge facing researchers interested in understanding scientific writing behavior: response chains. The frequency with which a behavior occurs has long be considered an indication of response strength (Skinner, 1932; 1938; 1957). However, with textual stimuli, single word utterances are challenging to analyze in isolation. For example, consider the word “hungry”. This response preceded by “am” likely influences listener/reader behavior differently than if preceded by “not”. By analyzing two-word response chains (i.e., bigrams), we were able to capture some of the surrounding context for the emission of any one word in any one publication.

Figure 3 highlighted how focusing on two-word response chains for describing the permanent products from writing carried some face validity when analyzing all behavior analytic publications. Bigrams related to antecedents, behaviors, and consequences all appeared in the 15 most frequent response chain emissions—as well as the bigram “functional analysis”. Functional analysis via antecedent-behavior-consequence relations are what operant analyses are all about. Thus, the results we obtained suggest that the approach of analyzing two-word response chains might be useful for finding orderly relations in published verbal behavior that allow us to accurately describe sets of published verbal stimuli.

Figure 4 highlighted how focusing on two-word response chains for describing the permanent products from writing carried some face validity when analyzing individual journals. For example, “behavior analyst” and “behavior analysis” were the top two most common bigrams in the journal *Behavior Analysis in Practice*, “verbal behavior” was the top bigram for the journal *The Analysis of Verbal Behavior*, “problem behavior” and “functional analysis” were the top bigrams in *JABA*, and “response rate” was the top bigram in *JEAB*. Each set of top bigram(s) for each journal—at least to these authors—seems to capture the primary function of the content that journal publishes. Thus, similar to the overall dataset, the results of the two-word behavior chain analysis suggest this method accurately captures important differences between sets of published verbal stimuli.

Despite such face validity, it is unclear that functionally describing and analyzing written behavior is best accomplished by using a single length of chained written responses. If we assume writing is controlled by contingencies in a similar manner as all other types of verbal behavior, then we would expect the functional unit for a written response chain to involve a single word in some cases, two-word written response chains in some cases, and perhaps other varied lengths of written response chains in other cases. Two-word written response chains seemed to perform well for the purpose of our experimental question. But, it may not be the best approach for different research questions and other approaches to topic modeling exist (e.g., Kherwa & Bansal, 2020; Mulunda et al., 2018). Manually analyzing all varied combinations of verbal behavior chains to identify where patterns emerge would be a cumbersome task. Leveraging NLP techniques make such exploratory research questions more tractable than previously.

Figures 5 and 6 show another method by which researchers can use NLP to compare data sets comprised of written responses. Scatterplots of the most frequently used words and phrases allow you to identify whether the two data sets contain similar emissions of written behavior (observed as running along the diagonal from lower-left to upper-right quadrant), the datasets contain dissimilar emissions of written behavior (observed as running along the diagonal from upper-left to lower-right quadrant), or somewhere between these extremes (observed as one big cloud of data with no clear trends or patterns).

We used scatterplots of the 500 most frequently used words or phrases to compare similarities between the control journal (*Health Service Research*)and all behavior analytic writing. As expected, the data fell along the diagonal suggesting the authors publishing in *Health Service Research* are emitting written textual responses that are different than those emitted by authors publishing in behavior analytic writing. “Of course, you found this”, one might respond, “The journals have different scope.” However, plotting the data in this way allows us to see more directly the extent of overlap between the different sets of published text, and the linear relation where each unit increase in word or phrase frequency in behavior analytic journals corresponded to a unit decrease in the frequency that same word or phrase was used in *Health Service Research*. Stated differently, the scatterplots of verbal behavior open the door for more sophisticated and precise methods of quantifying the degree of overlap rather than simply stating the journals are likely different.

Perhaps most surprising was a similar pattern of clear differences when *JABA* and *JEAB* were plotted against each other (Figure 6). All data from the top 500 most frequently used words and phrases fell in the upper-left and lower-right quadrants of the scatterplot. This, again, suggests that the words being used frequently in *JABA* are being used infrequently in *JEAB* and the words being used frequently in *JEAB* are being used infrequently in *JABA*. Authors have historically lamented the growing differences between applied and basic communities in behavior analysis (Baer, 1981; Hayes et al., 1980; Johnston, 1996; Michael, 1980). The data presented in Figure 6 provide a visual methodology to support those claims relative to the *JABA* and *JEAB* verbal communities. It is unclear whether this is the best approach to observing, measuring, and analyzing the trends and differences in different behavior analytic verbal communities over time. However, the NLP techniques used in the present analysis provide researchers with a more robust approach for observing and measuring these trends than historically were available.

There were several limitations to the present analysis. One limitation of the present study is that our data set consisted of all articles published from only five behavior analytic journals. We chose these publications mainly out of personal contact with these journals being the primary journals that publish research squarely under the label “behavior analytic”. It is possible that the results of the present study would differ if we included other journals that publish behavior analytic content alongside research from related fields (e.g., *Behavioural Processes*, *Psychological Record*, *Behavior Modification*). Nevertheless, the tools used in this study allow future researchers to easily incorporate those journals and run those analyses if they so choose.

A second limitation to the present analysis is that we included all articles from the journals regardless of their type. For example, we included articles like “technical notes” (e.g., Sprott et al., 1970), “on terms” (e.g., Merbitz et al., 2016), “memoirs” (e.g., Lattal, 1983), “erratums”, and special edition articles (e.g., Horne & Lowe, 1996). We chose to include these articles because this is the first known approach at using NLP methods to analyze all publications and in the ways discussed above. Thus, we felt it better to be overly inclusive rather than exclusive. Depending on the question future researchers ask, changing the types of articles included might be an important variable that could impact the results that follow from the techniques used in the present study.

**Conclusion**

Natural language processing (NLP) is an exciting and unique technology that has not yet been used in published behavior analytic research. NLP techniques allow researchers the ability to efficiently collect, process, and analyze written behavior at scales that were previously unapproachable. In this proof-of-concept study, we used NLP techniques to show how researchers can ask and answer questions about the published behavior analytic literature. We found the amount of written verbal behavior published per article has increased and become more variable over time. Further, analyzing the frequency two-word written response chains (bigrams) allowed us to accurately capture the functional approach that is the hallmark of behavior analytic publications as a field. Analyzing bigrams also allowed us to accurately capture the differences between journals based on their published scope of content. Lastly, we demonstrated how scatterplots comparing text-based data sets confirmed the expected differences between the control journal and behavior analytic journals; and also gave empirical support for historically noted differences between applied and basic verbal communities in behavior analysis.

Developing robust analytic techniques to analyze published written behavior from an operant and respondent framework might be advantageous for behavior analytic researchers. First, textual stimuli can be disseminated globally with the click of a button to influence reader behavior worldwide. Understanding the variables that control and predict the emission of and response to textual stimuli may carry significant global impact. Second, speech-to-text software allows researchers to easily transform vocal verbal behavior into text (Johnson et al., 2014). Developing techniques to analyze published textual stimuli may allow for easier data collection and analysis of more complex vocal verbal repertoires. Behavior analysts who build their skills in this domain can likely ask novel and unique questions about verbal behavior than were previously possible.

**References**

Atkinson, P., & Delamont, S. (2006). Rescuing narrative from qualitative research. *Narrative Inquiry*, *16*, 164–72. <https://doi.org/10.1075/ni.16.1.21atk>.

Baer, D.M. (1981). A flight of behavior analysis. *The Behavior Analyst, 4*(2), 85-91. <https://dx.doi.org/10.1007/FBF03391857>

Becirevic, A., Critchfield, T. S., & Reed, D. D. (2016). On the social acceptability of behavior-analytic terms: Crowdsourced comparisons of lay and technical language. *The Behavior Analyst, 39,* 305-317. <https://doi.org/10.1007/s40614-016-0067-4>

Bird, S., Loper, E., & Klein, W. (2009). Natural language processing with python. O’Reilly Media Inc.

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

Boydston, P.S., Hirst, E. S. (2020). Public perception and understanding of job titles related to behavior analysis. *Behavior Analysis in Practice.* 13, 394-401. <https://doi.org/10.1007/s40617-19-00384-z>

Critchfield TS. An emotional appeal for the development of empirical research on narrative. *Perspectives on Behavior Science.*2018;41(2):575–590. <https://doi.org/10.1007/s40614-018-0170-9>

Cox, D. J., Garcia-Romeu, A., Johnson M. W. (2021). Predicting changes in substance use following psychedelic experiences: natural language processing of psychedelic session narratives. *The American Journal of Drug and Alcohol Abuse*. 47(4), 444-454. <https://doi.org/10.1080/00952990.2021.1910830>

Davidson, T., Warmsley, D., Macy, M., Weber, I. (2017). Automated hate speech detection and the problem of offensive language. *Computation and Language.*

DePaolo, C.A., & Wilkinson, K. (2014). Get your head into the clouds: Using word clouds for analyzing qualitative assessment data. *TechTrends, 58,* 38-44. <https://doi.org/10.1007/s11528-014-0750-9>

Dixon, D.R., Vogel T., & Tarbox, J. (2012). A brief history of functional analysis and applied behavior analysis. https://doi.org/[10.1007/978-1-4614-3037-7\_2](http://dx.doi.org/10.1007/978-1-4614-3037-7_2)

Dixon, M. R., Reed, D. D., Smith, T., Belisle, J., & Jackson, R. E. (2015). Research Rankings of Behavior Analytic Graduate Training Programs and Their Faculty. Behavior analysis in practice, 8(1), 7–15. <https://doi.org/10.1007/s40617-015-0057-0>

Hayes, S.C., Rincover, A., & Solnick, J.V. (1980). The technical drift of applied behavior analysis. *Journal of Applied Behavior Analysis, 13*(2), 275-285.

Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014). Word cloud explorer: Text analytics based on word clouds. *IEEE* *47th Hawaii International Conference on System Science*. <https://doi.org/10.1109/HICSS.2014.231>

Horne, P. J., Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior.* 65(1). 185-241. <https://doi.org/10.1901/jeab.1996.65-185>

Johnson, J. (2021). Number of sent and received e-mails per day worldwide from 2017 to 2025. Retrieved from <https://www.statista.com/statistics/456500/daily-number-of-e-mails-worldwide/>

Johnston, J.M. (1996). Distinguishing between applied research and practice. *The Behavior Analyst, 19*(1), 35-47.

Kessler, J.S. (2017). Scattertext: a browser-based tool for visualizing how corpora differ. Association for Computational Linguistics. GitHub repository, <https://github.com/JasonKessler/scattertext#citation>

Kherwa, P., & Bansal, P. (2020). Topic modeling: A comprehensive review. *EAI Endorsed Transactions on Scalable Information Systems, 7,* e2. <https://doi.org/10.4108/eai.13-7-2018.159623>

Kranak, M.P., Rooker, G.W., Carr, C.J., Bradtke, P., Falligant, J.M., & Hausman, N.L. (2021). Evaluation of accepted and rejected submissions in the *Journal of Applied Behavior Analysis*: Gender and experience. *Journal of Applied Behavior Analysis, 54*(3), 1175-1187. <https://doi.org/10.1002/jaba.828>

Krenn, M., Zeilinger, A. (2020). Predicting research trends with semantic and neural networks with an application in quantum physics. *Proceedings of the National Academy of Sciences of the United States of America*. 117(4) 1910-1916. <https://doi.org/10.1073/pnas.1914370116>

Lattal, K. A., (1983). Don Hake, Behavior analyst: a memorial. *Journal of the Experimental Analysis of Behavior* . 39(1). 4-5. <https://doi.org/10.1901/jeab.1983.39-4>

Lee, M. (2021). Pytesseract 0.3.8. Documentation can be retrieved from: <https://pypi.org/project/pytesseract/>

Li, A., Curiel, H., Pritchard, J., & Poling, A. (2018). Participation of Women in Behavior Analysis Research: Some Recent and Relevant Data. Behavior analysis in practice, 11(2), 160–164. <https://doi.org/10.1007/s40617-018-0211-6>

Mariani, J., Francopoulo, G., & Paroubek, P. (2019a). The NLP4NLP Corpus (I): 50 years of publication, collaboration, and citation in speech and language processing. *Frontiers in Research Metrics and Analytics, 3*, 36. <https://doi.org/10.3389/frma.2018.00036>

Mariani, J., Francopoulo, G., Paroubek, P. & Vernier, F. (2019b). The NLP4NLP Corpus (II): 50 Years of Research in Speech and Language Processing. *Frontiers in Research Metrics and Analytics, 3,* 37. <https://doi/org/10.3389/frma.2018.00037>

McKinney, W. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference.* 51-56. <https://doi.org/10.25090/Majora-92bf1922-00a>

Merbitz, C. T., Merbitz, N. H., Pennypacker, H. S. (2016). On terms: frequency and rate in applied behavior analysis. *The Behavior Analyst.*39. 333-338. <https://doi.org/10.1007/s40614-15-0048-z>

Michael, J. (1980). Flight from behavior analysis presidential address ABA 1980. *The Behavior Analyst, 3*(2), 1-22. <https://dx.doi.org/10.1007/FBF03391838>

Mulunda, C. K., Wagacha, P. W., & Muchemi, L. (2018). Review of Trends in Topic Modeling Techniques, Tools, Inference Algorithms and Applications. *5th International Conference on Soft Computing & Machine Intelligence (ISCMI),* 28-37. <https://doi.org//10.1109/ISCMI.2018.8703231>

Pang, B., Lee, L., Vaithyanathan, S. (2002). Thumbs up?: sentiment classifaction using machine learning techniques.Association for Computational Linquistics, 79-82. https://doi.org/10.3115/1118693.1118704

Phillips, C.L., Vollmer, T.R., & Porter, A. (2019). An evaluation of textual prompts and generalized textual instruction-following. *Journal of Applied Behavior Analysis, 52*(4), 1140-1160. <https://doi.org/10.1002/jaba.649>

Rice, P.L., & Ezzy, D. (2000). Qualitative research methods: A health focus. South Melbourne, VIC, Autstralia: OUP Australia.

Richardson, L. (2007). Beautiful soup documentation. April.

Rzhetsky, A., Foster, J. G., Evans, J. A. (2015). Choosing experiments to accelerate collective discovery. *Proceedings of the National Academy of Sciences of the United States of America.* 112(47), 14569-14574.

Salunke, S. S. (2014). Selenium webdriver in Python: learn with examples. CreateSpace Independent Publishing Platform. ISBN: 978-1-4973-3736-9.

Shabani, D. B., Carr, J. E., Petursdottir, A. I., Esch, B. E., & Gillett, J. N. (2004). Scholarly productivity in behavior analysis: The most prolific authors and institutions from 1992 to 2001. The Behavior Analyst Today, 5(3), 235-243. <http://dx.doi.org/10.1037/h0100035>

Shevat, A. (2017). Designing bots: Creating conversational experiences. O’Reilly Media.

Skinner, B. F. (1932). Drive and reflex strength: II. *Journal of General Psychology, 6*, 38–48.

Skinner, B. F. (1938). The behavior of organisms: An experimental analysis. New York: Appleton-Century-Crofts.

Skinner, B.F. (1953). Science and Human Behavior. New York: Appleton-Century-Crofts.

Staats, A.W., Staats, C.K., Schutz, R.E., & Wolf, M. (1962). The conditioning of textual responses using “extrinsic” reinforcers. *Journal of the Experimental Analysis of Behavior, 5*(1), 33-40. <https://doi.org/10.1901/jeab.1962.5-33>

Sprott, R.L., Clark, F.W., & Wimer, R.E. (1970). A chamber for the instrumental control of licking behavior in the laboratory mouse. *Journal of the Experimental Analysis of Behavior. 14(3),* 341-343. <https://doi.org/10.1901/jeab.1970.14-341>

Twitter Usage Statistics (2021). Retrieved from <https://www.internetlivestats.com/twitter-statistics/#sources>

U.S. Texting Statistics (2017). Retrieved from <https://www>.localproject.net/docs/texting-stats/

Wickham, H. (2014). Tidy data. Journal of Statistical Software, 59(10), 1-23.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Yamamoto, S., & Isawa, S. (2019). Effects of textual prompts and feedback on social niceties of adolescents with autism spectrum disorder in a simulated workspace. *Journal of Applied Behavior Analysis, 53*(3), 1404-1418. <https://doi.org/10.1002/jaba.667>

Ziegler, D. M., Stiennon, N., Wu, J., Brown, T. B., Radford, A., Amodei, D., Christiano, P. F., & Irving, G. (2019). Fine-tuning language models from human preferences. CoRR, abs/1909.09593. 1-26. http://arxiv.org/abs/1909.08593

1. NB: We set the *x*-axis maximum value to 25,000 words for this figure to allow the reader to more easily see the distributions across journals. Trimming the *x*-axis to 25,000 words, however, removed 10 data points from *JEAB* with word counts of 39,116; 37,519; 36,476; 32,592; 31,170; 28,415; 28,341; 28,112; 27,850; and 27,090. [↑](#footnote-ref-1)