# Analyzing Student Think-Aloud Responses Using Natural Language Processing

Amanda Jensen

Department of Educational Psychology, University of Minnesota

PSY 8012: Data Science in Psychology

Dr. Richard Landers

4 May 2024

# Analyzing Student Think-Aloud Responses Using Natural Language Processing

This project was meant to analyze the student responses from a think-aloud study where students said their thoughts out loud after each sentence of a text. In the original study where this data was collected, the researchers were piloting a new online system in order to collect student think-aloud data digitally. In the initial study, the research questions revolved around the way the text was presented on the screen. Students were randomly assigned to one of two conditions – iterative and consecutive. For iterative, students only saw the current sentence displayed on the screen. For consecutive, they saw the current sentence plus all previous sentences. This was meant to determine if one display method was superior to the other. In the present study there is still a possibility to examine the data with the conditions in mind; however, the primary research questions in this current study is more about the student responses themselves and using Natural Language Processing (NLP) to examine the amount and types of words students were using.

## Hypotheses and Research Questions

**RQ1**: Do some sentences generate responses that demonstrate more complex lexical and syntactic properties than others?

**H1**: I predict that there would be some variation between sentences regarding the complexity of the sentences and words that students use. For example, sentences from the text with higher syntactic complexity and abstract vocabulary may encourage more complex responses to students when compared to simpler sentences.

**RQ2**: Which words are students using the most to describe their thinking while reading the texts? Does this vary when examining subgroups within the sample?

**H2**: In general, I predict that some of the most frequently used words by students may be those making reference to thought such as “think”, “know”, and “understand.” However, I think there will be difference in the words used between subgroups, particularly when comparing students between the grades and multilingual learners compared to non-multilingual learners.

**RQ3**: Which emotions are students demonstrating in their responses while reading the texts? Does this vary when examining subgroups within the sample?

**H3**: I hypothesize that students will demonstrate mostly positive emotions in response to the texts they are reading. I do not predict there being any major differences between different subgroups related to their emotions.

# Method

## Open Science Materials

There are several methods intentionally used in this study to practice Open Science. The first is a web-based binder. A binder is an interactive and reproducible research environment from repositories like GitHub ensuring that computational environments and the code within can be easily accessed and used by others. The link to the binder can be found [here](https://hub.binder.curvenote.dev/user/rnlanders-datascience-q32p8rfh/rstudio/).

As previously mentioned, a GitHub repository was utilized for this project. GitHub is a widely-used platform for version control and collaboration that allows users to manage and store revisions of projects. You can go [here](https://github.com/about) to learn more about GitHub. The GitHub repository for this project can be accessed [here](https://github.com/jens2482/psy8712-final). Within the repository, you can navigate to the README.md file. This file provides an overview of the project, including an introduction to the research, instructions on how to navigate the repository, and how to reproduce the results. It is recommended that you view this file before doing anything else.

## Participants

The study involved 78 fourth and fifth grade students from a public charter school. The mean age was 10.6 (SD  = 0.65). 15% of the students were multilingual learners (MLs).

## Measures

The first measure is the mean length of sentence. This serves as a proxy for the complexity of the sentences that students are using. The second related measure, length of words, is again a proxy, this time for the linguistic complexity of the student’s response. The word frequency measure represents the count of each word used by students while saying their thoughts after each sentence during the session. Lastly the sentiment measure assesses the emotional tone of student responses, categorizing responses into emotional states such as positive, negative, anticipation, trust, joy, surprise, anger, fear, disgust, and sadness.

## Procedure

Students were prompted to say their thoughts out loud after each sentence in a text. The student responses were transcribed using automated speech recognition (ASR) technology.

# Analyses

## Descriptive Statistics and Static Visualizations

| Sentence | Average Words Per Sentence | SD Words Per Sentence | Average Length of Word | SD Length of Word |
| --- | --- | --- | --- | --- |
| 0 | 28.11 | 27.81 | 4.03 | 2.09 |
| 1 | 37.26 | 45.74 | 3.96 | 1.82 |
| 2 | 56.52 | 81.09 | 3.97 | 1.91 |
| 3 | 25.87 | 21.80 | 4.02 | 1.89 |
| 4 | 22.06 | 11.15 | 3.98 | 1.72 |
| 5 | 21.64 | 15.85 | 3.76 | 1.64 |
| 6 | 22.20 | 10.23 | 3.89 | 1.88 |
| 7 | 26.34 | 16.75 | 3.91 | 1.82 |
| 8 | 22.49 | 12.07 | 3.81 | 1.63 |
| 9 | 38.76 | 51.26 | 3.83 | 1.68 |
| 10 | 23.50 | 12.42 | 3.76 | 1.81 |
| 11 | 19.24 | 7.70 | 3.97 | 1.87 |
| 12 | 18.53 | 8.33 | 3.87 | 1.79 |
| 13 | 18.23 | 7.48 | 3.94 | 1.86 |
| 14 | 10.54 | 5.96 | 3.77 | 1.74 |
| 15 | 13.65 | 8.17 | 3.83 | 1.75 |
| 16 | 9.62 | 5.13 | 3.89 | 1.64 |
| 17 | 8.10 | 4.43 | 4.30 | 2.18 |

|  |  |
| --- | --- |
|  |  |

## Interactive Visualization

For this project, I created an interactive Shiny web app to explore more about the research questions, specifically research questions 2 and 3. Examining words and sentences that are used the most frequently overall and then comparing it to different subgroups within the study would require a lot of different graphs to be inserted into the report. Instead, I can link to an interactive app where reviewers and those viewing the report can explore all the possible combinations among the groups to see how things vary. Users can filter data to see how the results change by manipulating one or any combination of the following variables: condition, gender, grade, and multilingual learner (ML) status of the data being represented. Users can also switch between views that show representations of the descriptive, frequency, and sentiment analyses to explore the three research questions. The shiny app can be accessed at [this link](https://jens2482.shinyapps.io/app_final/).

## Data Cleaning

In the data cleaning process, the raw data is first transformed by tokenizing responses into individual words, changing all words to lowercase, and removing punctuation to facilitate textual analysis. Each word’s length is calculated to assess lexical complexity. Words are also matched against a stop word list, tagging each as a stop or non-stop word, allowing for optional filtering in later analyses. Additionally, words are assigned sentiments from a sentiment lexicon. A summary metric, ‘words per sentence’, is also generated for each sentence by counting words, grouped by user and sentence identifiers.

## Analysis

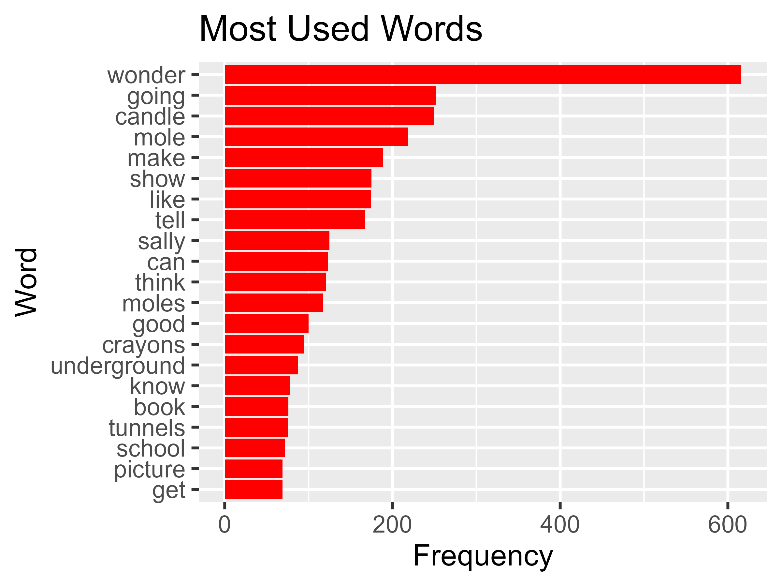
#### RQ1/H1: Do some sentences generate responses that demonstrate more complex lexical and syntactic properties than others?

The analysis of variance showed that the number of words per sentence did vary significantly by sentence (F = 1104.60, p = .00). The analysis of variance showed that the number of letters per word did not vary significantly by sentence (F = 3.38, p = .07).

|  |  |
| --- | --- |
|  |  |

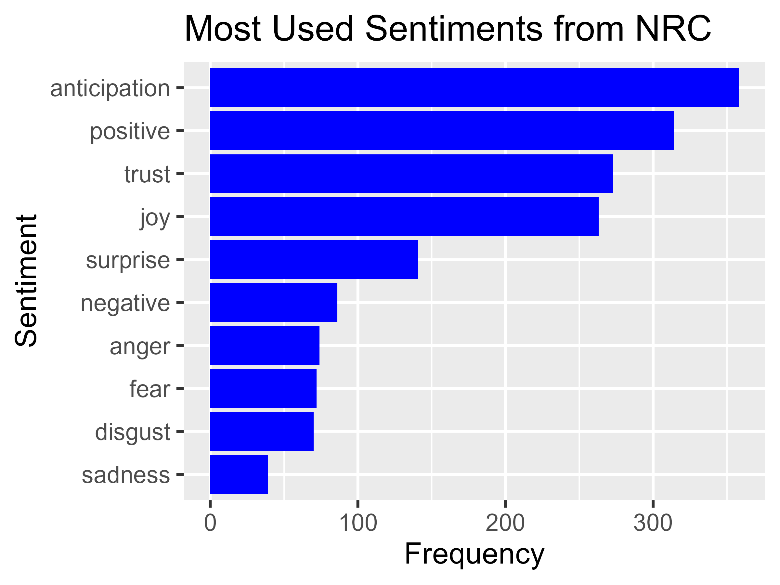
#### RQ2/H2: Which words are students using the most to describe their thinking while reading the texts? Does this vary when examining subgroups within the sample?

It was hypothesized that students would use more words related to thought, and that does appear to be true to some extent. The most used word by far was ‘wonder’. The words ‘think’ and ‘know’ also make the top 20. Other words like ‘candle’ and ‘mole’ make sense because those were the topics of the text. Otherwise, most of the other top words are verbs like ‘going’, ‘make’, ‘show’, ‘like’, and ‘tell. We also hypothesized that there would be a difference between fourth and fifth grade students. There is a difference the most used word between the two grades: for fourth grade it is ‘wonder’ and for fifth grade it is ‘candle’, but otherwise the lists stayed mostly the same. The top frequency lists for ML and non-MLs were also similar.



#### RQ3/H3: Which emotions are students demonstrating in their responses while reading the texts? Does this vary when examining subgroups within the sample?

It was hypothesized that students would use more positive words than negative and that there would not be much variation between different groups of students. There were actually some interesting differences, but overall the top sentiments were positive ones and the negative sentiments remained toward the bottom of the list.



# Reflection

Despite the strange way I go about coding things from time to time, this assignment may be included in that, I learned so much from this class! I really feel like I know R so well now since we started from the group-up. I think learning skills in Data Camp definitely helped me to become a more proficient coder and I am planning to continue to use that as long as possible. I also liked learning how to learn the idiomatic and industry standard practices for using R so I plan to keep that up as well. I am also fully on team tidyverse so I would like to continue honing my skills with the packages in that set. My next step is trying to learn a little bit more about Python because that has many different tools for text analysis, which is where I want to use a lot of my coding skills for the future. I will definitely try to start using loops more because of this class, as previously I was often copying and pasting a lot of my code. I think the setup of this class covered so many topics that I feel like I have a base knowledge in so many data science-related areas now that many of my peers in my program lack, so I’d like to also help build up their skills with what I know. Thank you so much for everything this semester!