

Project Oprah: Examining crowdsourced linguistic indicators of generality (#24512)

Created: 06/07/2019 01:37 PM (PT)

Public: 11/05/2020 12:11 PM (PT)

Author(s)

Ariana Orvell (University of Michigan) - aorvell@umich.edu

Ethan Kross (University of Michigan) - ekross@umich.edu

Susan Gelman (University of Michigan) - gelman@umich.edu

1) Have any data been collected for this study already?

No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?

We will be using crowdsourced data collected from Amazon Kindle's "Popular Highlights" feature to address the following questions:

Primary questions:

Q1. Are linguistic indicators that express generality (i.e., generic you, generic we, one, people) more likely to appear in literary passages that have been highlighted by readers compared to selected control passages that have not been highlighted?

Q2. Is generic-you more likely to appear in literary passages that have been highlighted by readers compared to selected control passages that have not been highlighted?

Secondary questions:

Q3. Are certain meaningful groups of linguistic indicators more likely to appear in highlighted versus non-highlighted control passages? Specifically:

Are indicators with parts of speech that could apply to the reader (i.e., "you" and "we") more likely to appear in highlighted passages vs. non-highlighted control passages?

Are indicators that are distanced from the speaker (you + one + people) more likely to appear in highlighted passages vs. non-highlighted control passages?

Is the magnitude of the effect between these two groups of indicators different?

3) Describe the key dependent variable(s) specifying how they will be measured.

Trained research assistants will code for instances of the following linguistic mechanisms:

Generic-you (i.e., "you" and any other contractions of "you" that are used to refer to people in general; e.g., "What doesn't kill you makes you stronger")

Generic-we (i.e., "we", "us", "our" and any other contractions that are used to refer to a group of people; e.g., "What doesn't kill us makes us stronger")

One (i.e., "one" to mean "anyone at all"; e.g., "What doesn't kill one makes one stronger")

"People" (i.e., "people" used to refer to people in general, also includes variants such as "all persons," "every person," "each person" "a person," e.g., "What doesn't kill people makes them stronger")

4) How many and which conditions will participants be assigned to?

Our design consists of two conditions: Highlighted passages and non-highlighted control passages (see response to question #6 for additional details on method).

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

Because our design involves passages nested within books, we will use multi-level modeling. Condition (Control passages vs. Highlighted passages) will be entered as a fixed effect. Book and passage number will be entered as random effects.

Our primary analyses will compare the usage of our selected linguistic mechanisms in the highlighted passages to the usage of the linguistic mechanisms in the non-highlighted control passages.

We will specify a model that is most appropriate for the data by examining the distribution of the dependent variable and through tests of model fit. All models will account for word length of a given passage.

Our primary questions will be tested with two models:

Q1. A model where the dependent variable is a composite index of the linguistic indicators that we coded for (i.e., generic you, generic we, one, people) will test whether linguistic indicators that express generality are more likely to appear in literary passages that have been highlighted by readers compared to selected control passages that have not been highlighted.

Q2. A model where the dependent variable is generic-you will test whether generic-you is more likely to appear in literary passages that have been

highlighted by readers compared to selected control passages that have not been highlighted.

5. Any secondary analyses?

We will conduct several exploratory analyses:

First, we will individually whether test each of the remaining indicators (i.e., "we", "one", and "person") appears more in highlighted vs. non-highlighted control passages.

Second, we will test whether meaningful groups of indicators are more likely to appear in highlighted vs. non-highlighted control passages. Specifically, we will examine:

- Indicators that use a linguistic shifter that could refer to the reader (i.e., you + we)
- Indicators that are distanced from the speaker (you + one + people)

As a secondary analyses, we will examine whether the year in which books were written are correlated with any of the linguistic indicators. If so, we will include year written as a covariate in secondary analysis.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Any passages that are not in the English language will be excluded.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

Our sample will consist of all highlighted passages (identified by Amazon's "Popular Highlights" feature) and control passages in books selected for Oprah's Book Club. To be included in our sample, books needed to meet the following criteria:

- Original text written in English
- Genre is adult fiction (neither non-fiction nor children's books)

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

For all books, we will standardize orientation to portrait and font size will be left as the default upon download (to ensure that the first sentence on the pages preceding the highlighted passage are consistent across devices and for replication purposes).

We will additionally collect the following information:

- Date downloaded to Amazon Kindle library & date highlights are collected (given that the number of Popular Highlights may change over time)
- Passage location provided by Amazon (for control and highlighted passages)
- Number of highlights (for highlighted passages only)
- Number of sentences (for control and highlighted passages)
- Number of words (for control and highlighted passages)

Control passages will be selected by navigating two pages back. The start of the third sentence that appears on the page will mark the beginning of the control passage. We are navigating three sentences in to allow for additional context. Control passages will be matched to the highlighted passage in number of sentences.

Special circumstances:

- If there are two highlighted passages on a given page, the control passages will have the same number of sentences separating them as did the two highlighted passages
- If the third sentence of the Control passage is also highlighted, the Research Assistant will navigate to the start of next full sentence.
- If the Highlighted passage contains an incomplete sentence, then it will be treated as one sentence (for purposes of selecting a matched control passage). We will record instances of this.
- If the highlighted passage is on the first two Kindle pages of the book, move forward two pages for the control passage
- If this happens and there is a highlighted section two pages AFTER the control page, copy from the page directly prior to highlighted passage (directly after the first control page).