**Name: Katherine Lee**

**Name of partner(s) if you worked with someone on this assignment: Elodia Lunn**

**Hands On 2 Reflection Questions**

1. **Public figures**
2. Which two public figures did you choose, and what kind of public figure are they (politician, singer, athlete, etc.)? In what key way(s) do these figures differ (e.g., politician vs. celebrity, Democrat vs. Republican, Gen Z vs. Boomer)?

**Greta Thunberg**

* **Climate activist, gen Z**

**Michelle Obama**

* **Former first lady, gen X,**

1. What social media platform are you using to pull these figures’ posts?

Instagram

1. Which of the three coding dimensions do you predict that the posts from these two public figures will differ on and why?

**I predict that the emotional tone will differ on between these public figure. Greta often has a more negative tone towards climate issues whereas Michelle focuses more on empowerment which has a more positive tone.**

1. **Using the coding scheme**
   1. How difficult was it to rate the posts on authenticity, analytic thinking, and emotional tone? Did you find certain dimensions easier to code than others, and why?

**Overall, it was hard difficult rating the posts in general because the scaling was so vast. I wouldn’t be able to differentiate between a 90 and 92. Or also it would be nice to have an anchor for what is 0 or what is 100.**

**I found the emotional tone easiest to code because the two figures were vary different in their tone and it was easier to see the difference between them. Analytic thinking was difficult I felt that some of the writing was formal but also personal (which made it ambiguous.**

* 1. Which parts of the coding scheme were ambiguous? How do you think this ambiguity would affect interrater reliability? (If you worked together, discuss how this ambiguity actually affected the interrater reliability between you and your partner.)

**Since I didn’t have much experience with coding, my interrater reliability is low, the more examples we went through, I felt my coding improved or was more accurate because I was able to compare each text-score to each other. When comparing me and my partner own ratings, I was more lower in scaling, so it also showed how our ratings differed from each other and also reduced our reliability.**

**In terms of ambiguity, I felt some cases (particularly analytical thinking) the text was both formal but personal, so it was difficult to rate ambiguity when there was also other similar text that had ambiguity. This also reduced interrater reliability because I didn’t have a concrete method when rating with more ambitious text.**

* 1. What is one way that you could change the coding scheme/process to increase reliability, according to the observing behavior lecture?

**Reduce the scale (make it a 1-7 scale) and also provide examples of what is 1 and what is 7.**

* 1. Is it possible that knowing who the posts were from could have affected your ratings of the posts, and why? How could this affect the validity of your ratings?

**Yes, we came in thinking that Geta would be more negative due to her presence online. I assumed Michelle would speak in a more formal tone (as she was a former lawyer) and more positive since I thought she would talk about more empowering topics. This reduced the validity of my ratings compared to if we had an unprimed (blind) rater. In my rating I subconsciously anchored Geta to have a lower score and for Michelle for a higher score.**

* 1. What is one way that you could change the coding scheme/process to increase validity, according to the observing behavior lecture?

**Using an unprimed rater (a rater who doesn’t know the identities of the figures) can increase validity. This approach allows for less influence from external factors.**

1. **LIWC text analysis**
   1. How do the LIWC results compare to your hand-coded ratings? In what ways are they similar or different?

**Our hand-coded ratings and the LIWC results are very different. For authenticity both results were around 50 points different from each other, but they were in the same direction**

* 1. Are you surprised by any of the LIWC scores for particular posts, and why?

**I was surprised by pretty much all of them. I thought Geta’s post would be more negative than what the LIWC program gauged. As for Analytic, I was particularly surprised that Michelle scored very high in analytic because she often used a personal tone. However, I can see that compared to other social media texts, hers can be more analytic than most. Additionally, I was shocked about authenticity being so low on LIWC. I am curious to see what would score 100 in authenticity.**

* 1. Look at the means for both your hand-coded ratings and LIWC’s calculations of authenticity, analytical thinking, and emotional tone for the two public figures you chose. How did their posts differ on these three dimensions? Did they differ in the way you originally expected?

**The authenticity and Analytic dimensions differed the most between the LIWC calculations and what we originally expected. I thought that Obama was less analytic because she was speaking in a personal voice, so our mean was at 22, but on LIWC, the program rated her at 87. Also, LIWC rated Obama as more analytic than Greta, whereas we rated them as the opposite.**

**Another thing was that we believed Geta was speaking in a more formal voice, and we rated her mean at 80, but on LIWC, she was rated at 70. They also differed on the authenticity dimension, While in terms of comparison against each other (Michelle vs. Greta), they had similar point distances to the LIWC mean comparisons, we were 50 points more than LIWCs mean. Lastly, the emotional tone was the dimension matched what I originally expected, but I thought Greta would have an even lower score though.**

1. **Correlations**
   1. How correlated were your ratings with the LIWC scores, both on the three individual dimensions and overall? Use the table in the instruction packet to interpret each of the four correlation coefficients.

Our overall correlation was a **weak negative correlation**, Our emotional tone had an **extremely strong correlation**, our analytical was a **negative correlation** and authenticity had a **positive weak correlation**.

* 1. Your ratings and the LIWC scores are two measures of the same three constructs (authenticity, analytic thinking, and emotional tone). Let’s say that your ratings are our best existing measure of these three constructs (the criterion). If your ratings are the criterion, what would the correlations between your ratings and the LIWC scores suggest about the criterion validity of the LIWC measures of each of these three constructs?

**There are varying degrees of alignment between our ratings and the LIWC scores between the constructs. The emotional tone showed an extremely strong correlation, this would mean there would be a high similarly and the LIWC scores would be close to our scales. But for the other constructs, they are differ: The analytical thinking had a negative corelation which is a disagreement between the LIWC scores and my scores. So that would result in the LIWC scores being very far off (opposite) from our measure. As for the authenticity, there was a weak positive correlation meaning that there was some agreement but not as similar as desired. The overall ratings had a weak negative correlation so overall, the LIWC scores has a lot of differences.**

<https://docs.google.com/spreadsheets/d/1PcHxWC-mvNHpiaQSw2-WDXcTiCv4Pqlt/edit#gid=1598074415>