

Analyzing Pronoun Usage in Yelp Reviews

The Power of **AND**

University of Wisconsin
Eau Claire

Nicole McMahon, Garrett Dekan, and Heather Amthauer | Computer Science

INTRODUCTION

The pronouns an individual chooses when speaking or writing can reveal a considerable amount about him or her. This research project examines whether the emotional aspect of an experience has any influence on the types of pronouns that a person uses. We accomplish this task by analyzing Yelp reviews and determining whether the pronouns a person chooses when writing a review correspond to the rating he or she gives that experience on a scale of one to five stars. This research is significant because any correlation we find would indicate that people use different types of pronouns when describing positive experiences as opposed to negative experiences. Our conclusions could be used to infer the emotional state of the author of any informal piece of writing, which would help readers determine if the author has any underlying bias toward the subject.

BACKGROUND

We used a dataset of 10,000 Yelp reviews written between 2005 and 2013 that we downloaded from Kaggle. Our pronoun categories are first person plural, first person singular, total first person, second person, third person plural, third person singular, and total third person pronouns. According to Catalina Toma and Jeffrey Hancock (2012), “first-person plural pronouns (‘we’ and ‘ours’) indicate inclusiveness, whereas second- and third-person pronouns (‘you’ and ‘they’) place the [speaker] in the outgroup.” In addition to indicating exclusion, evidence suggests that second person pronouns are also “correlated with negative interpersonal outcomes.”

According to Dan Jurafsky and Victor Chahuneau (2014), reviews given fewer stars had a greater concentration of first person plural and third person pronouns. They found evidence that these pronoun categories are associated with experiencing trauma or negative situations, which “suggests that one-star reviews are narratives of negative emotion.” In these experiences, people generally emphasize “their belonging to groups, using the words *we* or *us* with high frequency, as a sign of solidarity...and a way of achieving ‘collective closure.’”

Drew Margolin’s and David M. Markowitz’s (2018) analysis of pronouns in Yelp reviews found that “on average, more highly rated reviews contain approximately one fewer self-reference (e.g., I, me, my) for every seven reviews.” The conclusion that there are fewer first person pronouns in reviews with higher stars agrees with the results of the study by Jurafsky and Chahuneau.

METHODS

To determine whether emotional responses have an effect on the types of pronouns people use, we began by using Python to create a part-of-speech tagger using the Natural Language Toolkit (NLTK). We then split up our set of Yelp reviews by stars. For each Yelp review, we counted all of the pronouns and tagged each one based on the pronoun category it fit into. We normalized our pronoun counts to control for the wide variety of review lengths. We utilized one-sided t-tests to check for significant differences in our pronoun counts. We tested for significant differences in pronoun counts between reviews given low stars and reviews given high stars for each pronoun category. We also tested for significant differences in pronoun counts between different pronoun categories while holding the number of stars constant. We chose our significance level to be 0.05.

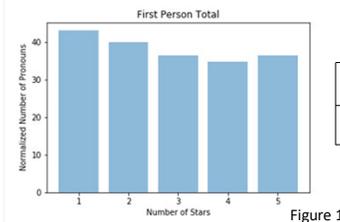


Figure 1

1 star vs. 5 star : $p = 1.34e-09$
2 star vs. 4 star : $p = 7.11e-09$

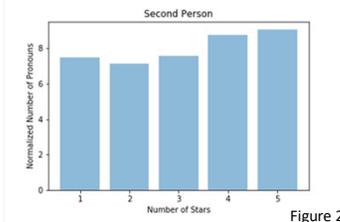


Figure 2

1 star vs. 5 star : $p = 0.0029$
2 star vs. 4 star : $p = 0.0006$

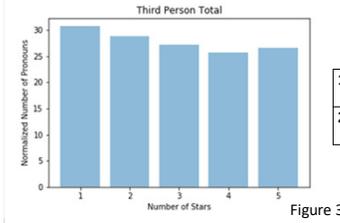


Figure 3

1 star vs. 5 star : $p = 9.74e-06$
2 star vs. 4 star : $p = 7.37e-05$

RESULTS

Based on our statistical analysis, we found that both first person singular and first person plural pronouns are generally more common in low star reviews. Naturally, total first person pronouns also show this trend (see figure 1). Similarly, third person singular pronouns also tend to appear more often in low star reviews. However, third person plural pronoun differences were not significant between star categories. Our data show that third person singular pronouns appear more frequently in reviews than third person plural pronouns. Because of this, third person singular pronouns had a more significant effect on the trend of total third person pronouns. Therefore, total third person pronouns are generally more common in low star reviews (see figure 3). Conversely, our data show that second person pronouns are generally more common in high star reviews (see figure 2).

When comparing first person plural pronouns with third person plural pronouns in each star group, we found that third person plural pronoun counts are always notably higher (see figure 4). When comparing first person singular pronouns with third person singular pronouns in each star group, we found that first person singular pronoun counts are always notably higher (see figure 5). Because of these large disparities in pronoun counts between categories on both graphs, every p-value we calculated when comparing first and third person pronoun categories is significant to the point of approaching zero.

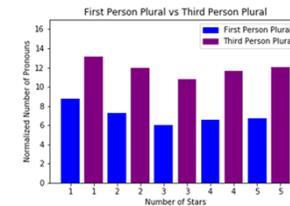


Figure 4

1: $p = 3.23e-08$
2: $p = 3.25e-13$
3: $p = 3.58e-25$
4: $p = 2.49e-50$
5: $p = 1.70e-43$

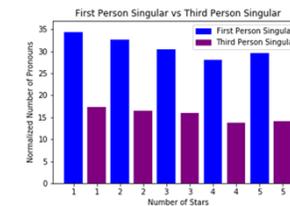


Figure 5

1: $p = 4.43e-43$
2: $p = 3.35e-59$
3: $p = 4.13e-77$
4: $p = 8.36e-184$
5: $p = 3.52e-162$

DISCUSSION AND CONCLUSION

The fact that first person pronouns are generally more common in low star reviews likely indicates that people have a tendency to be more focused on themselves when writing a review about a negative experience. We found that third person singular pronouns are also typically more common in low star reviews. This indicates that when people have had a negative experience, they are more likely to focus on a singular individual or aspect of the situation that contributed to the negative experience. Conversely, our data show that second person pronouns are generally more common in high star reviews. This likely indicates that people are typically more focused on the reader of the review when sharing their positive experiences.

When writing about groups of people, typically review writers are more likely to focus on the outgroup as opposed to the ingroup. This indicates that the group’s interaction with the people providing the service is the biggest contributing factor as to whether the reader would visit the establishment. On the other hand, when writing about individuals, a review writer typically focuses more on him or herself as opposed to a singular other individual. This shows that the writer is more likely to talk about his or her own personal experience in order to better convey the quality of experience the reader would have.

FUTURE

Because our dataset is a few years outdated and limited in size, we hope to replicate our process with a larger and more recent dataset in order to compare the new results with our current results. Furthermore, we acknowledge the fact that other variables, such as gender, age, geographic location, and price, could be skewing our data, and we hope to analyze the effects of these factors in the future. It would also be interesting to analyze how adjective use varies in relation to the number of stars given for reviews.

SOURCES

- Jurafsky, D., Chahuneau, V., Routledge, B. R., & Smith, N. A. (2014, April 7). Narrative framing of consumer sentiment in online restaurant reviews. Retrieved February 2, 2019, from <https://uncommonculture.org/ojs/index.php/fm/article/view/4944/3863>
- Margolin, D., & Markowitz, D. M. (2017). A Multitheoretical Approach to Big Text Data: Comparing Expressive and Rhetorical Logics in Yelp Reviews. *Communication Research*, 45(5), 688-718. doi:10.1177/0093650217719177
- Toma, C. L., & Hancock, J. T. (2012, February 02). What Lies Beneath: The Linguistic Traces of Deception in Online Dating Profiles. Retrieved February 2, 2019, from <https://online.library.wiley.com/doi/pdf/10.1111/j.1460-2466.2011.01619.x>