Understanding the Language Use Differences between Reviews Associated with Star Rated Restaurants on Yelp

1st Oluwadara Kalejaiye
Howard University

2nd Edward Williams
Howard University

3rd Farhan Rashid
Howard University

4th Danae Dunlap
Howard University

5th Praise-EL Michaels
Howard University

6th Legand Burge
Howard University

7th Anietie Andy
Howard University

***Abstract*—On Yelp, users do the following: (a) rate businesses/organizations e.g. restaurants, on a scale of 1 to 5 – where 1 indicates that the business is not good and 5 indicates it is really good and (b) write reviews about their views, experiences and/or interactions with these businesses/organizations. Using Yelp reviews, prior works applied natural language processing methods to compare the reviews associated with 1–star rated businesses to those associated with 5–star businesses and determined that there were significant differences in language use in reviews associated with these star rated businesses. In the analyses conducted in these prior works, reviews associated with 2–star and 4–star rated businesses, respectively, were not analyzed. This was in part due to the assumption that the language used in reviews associated with 1–star rated businesses are different from those used in reviews associated with 2–star rated businesses; the same applies to reviews associated with 4–star and 5–star rated businesses. However, no work has validated this assumption by quantifying these language use differences. This work aims to address this. Specifically, using a psycholinguistic dictionary, Linguistic Inquiry and Word Count (LIWC) and word-score based methods, we quantify the differences between the Yelp reviews associated with 1 and 2 star rated businesses and 4 and 5 star rated businesses, respectively. For example, we find that the LIWC categories on *Negative Emotion* (Cohen's D = 0.492) and *Anger* (Cohen's D = 0.369) are more associated with 1–star reviews when compared to 2–star reviews and the LIWC categories on *Positive Emotion* (Cohen's D = 0.628) and *Sadness* (Cohen's D = 0.133) are more associated with 2–star reviews when compared to 1–star reviews. Also, we find that the LIWC category on *Focus Present* is more associated with reviews associated with 1–star rated businesses and 5–star rated businesses, thereby suggesting that reviewers tend to express their dislike (1–star reviews) and approval (5–star reviews) while they are present at the business. Sometimes, the same users write reviews associated with different star rated businesses. We conduct further analyses to compare the reviews (by the same users) written for different star rated businesses. We discuss the findings of this work and their implications.**

***Index Terms*—Yelp, ratings, LIWC**

I. Introduction

Yelp provides an online platform for users to rate businesses on a scale of 1 to 5, where 1 indicates that the business is bad and 5 indicates that the business is very good. Also, on Yelp, users write reviews about their views, experiences, and/or interactions with these businesses. Prior works [1]–[3] used natural language processing (NLP) methods to compare the reviews associated with 1–star rated businesses to those associated with 5–star ratings; it was determined that there are significant language use differences between reviews associated with 1 and 5 star rated businesses. In these prior works [1]–[3], reviews associated with 2-star, 3-star, and 4-star rated businesses, respectively, were not analyzed. This was partly because of the following assumptions: (a) the language used in reviews associated with 1 star and 2 star rated businesses are significantly different (b) the language used in reviews associated with 4 star and 5 star rated businesses are significantly different (c) the reviews associated with 3 star rated businesses are either neutral, not as negative as 1 star reviews, or not as positive as 5 star reviews. No prior work has validated these assumptions. Hence, this work aims to address this. In this work, we focus on comparing the language used in reviews associated with a category of business, specifically, restaurants, that have a 1–star rating compared to those with 2–star ratings; we do the same for reviews associated with restaurants with 4–star ratings compared to 5–star ratings. The rationale for this is that the reviews associated with 1–star rated businesses are more similar to reviews associated with 2–star rated businesses compared to 3–star businesses. Similarly, we assume that the reviews associated with 5–star rated businesses are more similar to those associated with 4–star rated businesses compared to 3–star businesses. Therefore, in the analyses conducted in this work, we do not use the reviews associated with 3–star rated businesses.

There are several benefits to gaining insights as to whether there are language use differences (and if these differences are significant) between reviews associated with 1–star and 2–star and 4–star and 5–star rated businesses, respectively. For example, a given business category e.g. restaurants may have few reviews associated with 1–star rated businesses, however, they may have more reviews associated with 2–star rated businesses. If the language use differences between the reviews associated with 1 and 2 star rated businesses in the given business category are not significant or there are no language use differences, then perhaps the reviews associated with the 2–star businesses could be used as complimentary data with the reviews of the 1–star rated businesses for analyses to gain insights about the given category of businesses. Below are examples of reviews associated with 1, 2, 4, and 5 star rated businesses, respectively.

1.

1) *It's baffling how a fast-food joint can decline over time yet remain in business. Enticed by ads for their "new and improved" chicken nuggets at a bargain price, I should've anticipated disappointment. The service at the drive-through was painfully slow, and to my dismay, they were out of BBQ sauce. I settled for mustard and, upon finding out they were out of Diet Coke or encountering a problem, requested a bottled water instead. Upon arrival at the window, I was ignored for about 5 minutes, and then, without any acknowledgment, a hand appeared to take my card. As I waited, I noticed a bag on the counter, suspecting it was mine, and indeed, my order was left sitting there for nearly 8 minutes, resulting in a lukewarm meal. To add insult to injury, they gave me ketchup instead of mustard and forgot the water altogether. Frustrated and unwilling to confront them, I took my order and left, regretting giving them another chance and resolving to never return to this location—or any of their other locations, for that matter.*
2) *I've visited this place on several occasions. The food is average, not particularly noteworthy. While the waitresses, generally appealing in their uniforms, are noticeably sluggish in service. On my most recent visit, my waitress was nowhere to be seen until it was time to pay. It's unfortunate; the place has a great location and concept but seems to suffer from either insufficient staffing or poor management, perhaps even both. Additionally, paying $15 for lunch feels a bit excessive.*
4) *I found my second visit more satisfying than my initial one right after their opening. Back then, the BBQ sandwich felt too meat-heavy for my taste. Service was decent on a Friday night. My lunch experience today, featuring a delightful eggplant poboy, was exactly what I needed. My meat-loving boyfriend even acknowledged the club sandwich as a dish he'd order again. The mac cheese was average, missing some seasoning. However, the service was notably slow. Quick lunch spots should recognize that diners typically have only an hour. It's surprising, especially given the number of raw dishes they offer, that they're on the slower side. I'm definitely planning a return visit, eager to try the nachos and spaghetti, which looked appetizing today as many patrons ordered them.*
5) *This was my first visit, following a suggestion from a workmate. I enjoyed one of the finest Cuban sandwiches I've ever tasted. I'm keen on exploring more of their menu, particularly the yellow rice black beans, and the picadillo. The personal touch added by the family that runs this establishment elevates the dining experience. The owner and matriarch, especially, ensures a warm, engaging atmosphere for every patron. For anyone in search of genuine Cuban cuisine, this place is a must-visit.*

In this work, we use the following NLP methods: (a) the Linguistic Inquiry Word Count (LIWC) [4], which is a psycholinguistic dictionary made up of several categories such as health, and a collection of curated words associated with each of these categories and (b) the following word score based methods [5] *Valence*, *Arousal*, and *Dominance* to identify and quantify the language use differences between reviews associated with 1 and 2 star businesses, respectively and reviews associated with 4 and 5 star rated businesses. *Valence* refers to the extent to which a word conveys a positive or negative meaning; for instance, the word *"agreeable"* has a high valence, while the word *"disaster"* has a low valence. *Arousal* assesses the level of emotional intensity a word conveys; for example, *"elate"* is a word with high arousal, while *"nap"* is a word with low arousal. *Dominance* measures the degree of control or influence that a word suggests; for instance, the word *"influential"* has high dominance, while *"dull"* has low dominance.

Sometimes, individuals write reviews associated with different star rated businesses (e.g. one individual may write reviews for 4–star rated businesses and also for 5–star rated businesses). Are there language use differences between these reviews? Understanding these language use differences can help determine if for example, 1–star reviews and 2–star reviews by the same users could be used for analyses to gain insights about a given business category. We conduct additional analysis using LIWC and the word score based methods *Valence*, *Arousal*, and *Dominance* to determine if there are language use differences in the reviews written by the same users for different star rated businesses.

This paper addresses the following research questions:

font=

(R1) **What are (if any) the language use differences between reviews associated with 1–star and 2–star rated businesses and 4 –star and 5–star rated businesses, respectively?**

(R2) **Given Yelp reviews written by the same users for different star rated businesses, are there/what are the language use differences associated with these reviews?**

In the analyses conducted in this work, we measure effect sizes using Cohen's D – which quantifies the standardized difference between two means; we only include findings where Cohen's D meets or exceeds a predefined threshold i.e. 0.10.

II. Related Work

The influence of online review platforms, such as Yelp, on consumer behavior has led to extensive research into using NLP and machine learning (ML) techniques to gain insights about the user-generated content on these platforms. For example, NLP and ML techniques have been used to (a) understand customer satisfaction levels and identify areas needing improvement and (b) for seeking to understand cultural values or influences on the way people interact on these platforms.

Sometimes, before making a purchase or visiting a business establishment/organization, individuals analyze information – associated with the business, from online reviews platforms such as Yelp, before making a purchase/visiting the business. Following a purchase, some of these individuals write reviews about their experience interacting with the business on these online review platforms [6]. [7] investigates how trust in the review system on these online review platforms impacts consumer decision-making in the initial stages of purchasing. Consumers tend to be more skeptical of highly rated reviews than those with lower ratings, requiring substantial numerical proof in the form of many reviews to support their decisions. In contrast, they need less numerical proof for lower-rated reviews. Thereby suggesting that the number of reviews has more influence on establishing trustworthiness of the ratings when consumers first encounter them. [8] addresses the task of sifting through large numbers of reviews on these online review platforms, to identify those that are reliable and useful; it suggests assessing the written content of the reviews and gauging the agreement among the reviewers about a product/business.

The analysis of sentiments expressed in Yelp reviews has become increasingly sophisticated with the advent of deep learning techniques. Studies employing models such as the hybrid bidirectional recurrent convolutional neural network with an attention-based mechanism (BRCAN) have demonstrated remarkable accuracy in classifying the sentiment of textual content, catering to the nuances embedded within user-generated reviews [9]. These advancements underscore the pivotal role of sentiment analysis in extracting valuable data from the vast expanse of Yelp's review dataset, providing a quantitative basis for understanding consumer sentiment at scale [9], [10].

Additionally, the utilization of Python libraries such as NLTK [1], has facilitated the application of sentiment analysis in NLP research, enabling the extraction and categorization of emotional tones from textual reviews with heightened precision [11]. This technical progression supports a granular analysis of user sentiments, contributing to a more refined understanding of consumer satisfaction and areas for business improvement.

Predictive modeling in the context of Yelp reviews has seen substantial developments, particularly in rating prediction. The application of Tensor Factorization (TF) and the Personalized Neural Usefulness Network (PUN) model illustrates the industry's shift towards contextual and nuanced analysis of review content for predicting ratings [12], [13]. This created an inherent usefulness of sentences within reviews to predict user ratings more accurately and these models represent a significant leap forward in accuracy and reliability, leveraging contextual factors.

Hybrid models that combine sentiment analysis with recommender systems offer another avenue for enhancing rating prediction. Incorporating sentiment as a factor in prediction models, researchers have been able to improve the accuracy of rating predictions, showing an intricate relationship between sentiment and user rating behaviors [14]. This multifaceted approach not only advances our understanding of how users rate experiences but also aids businesses in identifying the impact of consumer sentiment on their ratings [15], [16].

Aspect-Based Sentiment Analysis (ABSA) has provided a more detailed lens through which to view Yelp reviews, identifying specific aspects of businesses that customers discuss and evaluating the sentiment associated with each aspect [16], [17]. This method allows for a detailed breakdown of consumer feedback, highlighting specific areas of success or concern. AYLIEN Text Analysis API tools have been instrumental in deploying ABSA at scale, enabling businesses to derive actionable insights from complex datasets [18].

Recognizing the influence of visual content in reviews, multimodal sentiment analysis merges textual and image analysis to offer a comprehensive sentiment evaluation. This approach acknowledges that images accompanying reviews can significantly sway the overall sentiment expressed [19]. Techniques for image feature extraction and classification, when combined with textual sentiment analysis, underscore the potential for a holistic assessment of user reviews [20].

The global nature of consumer sentiment, cross-cultural analysis has become increasingly important and studies utilizing transfer-learning frameworks have shown that sentiment analysis models can be effectively adapted to diverse cultural contexts, enhancing the applicability and accuracy of these models across different linguistic and cultural datasets [21].

In addition to textual analysis, the emergence of multimodal sentiment analysis acknowledges the growing significance of visual content in reviews. When text and image analysis is integrated, researchers have begun to explore the compound effect of multimodal content on overall sentiment, paving the way for a more comprehensive understanding of user reviews [20]. This approach is particularly relevant in the context of Yelp, where visual content can significantly influence a review's sentiment.

Recent advancements in sentiment analysis have capitalized on the power of deep learning to interpret the complexities of human language with unprecedented depth and accuracy. The integration of models such as Convolutional Neural Networks (CNNs) with Word2Vec embeddings has shown promising results in classifying text sentiment, demonstrating the capacity of these models to capture semantic nuances in large datasets

[1] https://www.nltk.org/

[22]. This approach has been particularly effective in analyzing Yelp reviews, where the variety and volume of data necessitate sophisticated models to accurately gauge consumer sentiment.

The determination of review helpfulness represents a novel area of exploration within the domain of Yelp review analysis. Research efforts have focused on developing models that predict the helpfulness of reviews based on textual and meta-textual features, aiding users in navigating the extensive repository of reviews [23], [24]. This line of research extends the utility of predictive modeling beyond rating prediction, offering insights into the factors that contribute to the perceived value of reviews by the Yelp community [15], [25].

In prior work [1], [2], NLP methods were used to analyze 1–star and 5–star online reviews associated with healthcare facilities.

Our work is different from prior work. In this work, using NLP techniques, we aim to determine the language use differences between reviews associated with (a) 1–star reviews compared to 2–star reviews and (b) 4–star reviews compared to 5–star reviews.

III. Dataset

In this work, we use a Yelp dataset [2]. Specifically, we only use reviews associated with restaurants; Table I shows the dataset used in this work.

TABLE I
YELP RESTAURANT DATASET

Rating	# of Restaurants	# of Users	# of Reviews
1-star	247	3105	3255
2-star	2492	10843	12163
4-star	13128	232677	495596
5-star	1554	40040	46375

This paper is formatted as follows: using LIWC and the word score based methods [5] *Valence*, *Arousal*, and *Dominance*: (a) in section **R1 – Analysis: All Reviews**, we compare 1 and 2 star Yelp reviews and 4 and 5 star Yelp reviews, respectively (b) in section **R2 – Analysis: Reviews by the Same Users**, we compare reviews (written by the same users) associated with different star rated businesses (c) in the **Discussion** section, we discuss the findings of this work (d) in the **Limitations** section, the limitations of this work are discussed (e) in the **Future Work** section, we discuss the future work and in the **Conclusion** section, we conclude the paper.

IV. R1 – Analysis: All Reviews

In this section, LIWC, *Valence*, *Arousal*, and *Dominance* are used to identify the differences in language use and communication strategies used by users when writing Yelp reviews associated with (a) 1 and 2 star rated restaurants and (b) 4 and 5 star rated restaurants. The dataset from Table I is used for all the analyses in this section.

[2]https://www.yelp.com/dataset

A. LIWC Analysis

Similar to prior work [26], [27] which used LIWC to identify the LIWC categories most associated with posts in one group compared to those in a different group, we determine the differences in the LIWC categories associated with 1-star Yelp reviews when compared to 2–star Yelp reviews and vice versa. Also we determine the LIWC categories associated with 4–star Yelp reviews when compared to 5–star Yelp reviews and vice versa. Specifically, we determine the proportion of words associated with the LIWC categories in reviews associated with restaurants with a particular star rating (e.g. 1–star) compared to the words associated with the LIWC categories in reviews associated with restaurants with a different star rating (e.g. 2–stars). Table II shows the LIWC categories most associated with 1–star reviews when compared to 2–star reviews and Table III shows the LIWC categories that are most associated with 2–star reviews when compared to 1–star reviews. Similarly, Table IV shows the LIWC categories most associated with 4–star reviews when compared to 5–star reviews and Table V shows the LIWC categories that are most associated with 5–star reviews when compared to 4–star reviews.

TABLE II
LIWC CATEGORIES MORE ASSOCIATED WITH 1-STAR REVIEWS COMPARED TO 2-STAR REVIEWS

#	LIWC	Cohen's D
1	Negative Emotions	0.492
2	Anger	0.369
3	Power	0.364
4	Certain	0.326
5	Risk	0.322
6	Social Processes	0.250
7	Time	0.244
8	Personal Pronouns	0.241
9	First Person Singular Pronoun	0.219
10	Hear	0.218
11	Third Person Plural Pronoun	0.208
12	Negations	0.199
13	Common Verbs	0.199
14	Focus Present	0.180
15	Anxiety	0.173
16	Adverb	0.125

B. Word Score based Methods

Prior works have used the word score based methods *Valence*, *Arousal*, and *Dominance* to understand the various communication strategies users utilize when communicating on social media and online forums [28], [29]. Using a lexicon of 20,000 human rated words [5] for *Valence*, *Arousal*, and *Dominance*, for each of the reviews associated with 1 and 2 star Yelp reviews (and for 4 and 5 star reviews), the average ratings of the content words associated with each of the *Valence*, *Arousal*, and *Dominance* word category was computed. We do the following: (a) for 1 and 2 star restaurant reviews, we create features that represents the the average *Valence*, *Arousal*, and *Dominance* scores, respectively for

TABLE III
LIWC CATEGORIES MORE ASSOCIATED WITH 2-STAR REVIEWS COMPARED TO 1-STAR REVIEWS

#	LIWC	Cohen's D
1	Positive Emotion	0.628
2	Ingest	0.401
3	Differentiation	0.356
4	Leisure	0.286
5	Reward	0.273
6	Affiliations	0.249
7	Quantities	0.215
8	Conjunctions	0.182
9	Assent	0.173
10	First Person Plural Pronoun	0.173
11	Cognitive Processes	0.139
12	Sadness	0.133
13	Friend	0.125

TABLE IV
LIWC CATEGORIES MORE ASSOCIATED WITH 4-STAR REVIEWS COMPARED TO 5-STAR REVIEWS

#	LIWC	Cohen's D
1	Differentiation	0.481
2	Articles	0.239
3	Focus Past	0.229
4	Cognitive Processes	0.197
5	Negation	0.162
6	Leisure	0.131
7	Negative Emotion	0.128
8	Hear	0.127
9	Function	0.112

reviews and another feature that represents if a review is associated with a 1–star restaurant or 2–star restaurant (b) for 4 and 5 star restaurant reviews, we create features that represents the average *Valence*, *Arousal*, and *Dominance* scores, respectively for reviews and another feature that represents if a review is associated with a 4–star restaurant or 5–star restaurant.

C. Results

1) 1 and 2 star Reviews Analyses:: As it relates to the analyses on reviews associated with 1–star and 2–star restaurants, respectively, we find that the LIWC categories on *Negative Emotion*, *Anger*, *Personal pronoun*, *Negation*, *Focus Present*, and *Anxiety* (i.e. Table II) are more associated with reviews for 1–star rated restaurants when compared to 2–star rated restaurants. Also, we find that the LIWC categories on *Positive Emotion*, *Differentiation*, *First Person Plural Pronoun*, and *Sadness* are more associated with reviews for 2–star rated restaurans compared to 1–star rated restaurants.

2) 4 and 5 star Reviews Analyses:: As it relates to the analyses on reviews for 4–star restaurants and 5–star restaurants, we find that the LIWC categories on *Focus Past*, *Negation*, and *Negative Emotion* are more associated with reviews for 4–star restaurants compared to 5–star restaurants. Also, the LIWC categories on *Positive Emotion*, *Focus Present*, and *Focus Future*, are more associated with 5–star reviews when compared to 4–star reviews.

TABLE V
LIWC CATEGORIES MORE ASSOCIATED WITH 5-STAR REVIEWS COMPARED TO 4-STAR REVIEWS

#	LIWC	Cohen's D
1	Achieve	0.337
2	Certain	0.337
3	Social	0.321
4	Positive Emotion	0.278
5	Focus Present	0.245
6	Power	0.231
7	Adjective	0.223
8	Focus Future	0.183
9	Friends	0.172
10	Drives	0.160
11	Family	0.142
12	Affiliation	0.127

TABLE VI
VALENCE, AROUSAL, AND DOMINANCE BETWEEN 1-STAR REVIEWS COMPARED TO 2-STAR REVIEWS

Attribute	Cohen's D
Valence	20.717
Arousal	37.624
Dominance	1.910

3) Word Score based Methods Analyses:: Table VI shows the effect size (Cohen'D) between the features that represent the average *Valence*, *Arousal*, and *Dominance* scores, respectively, for reviews and the feature that represents if a review is associated with a 1–star restaurant or 2–star restaurant. Table VII shows the effect size (Cohen'D) between the features that represent the average *Valence*, *Arousal*, and *Dominance* scores, respectively, for reviews and the feature that represents if a review is associated with a 4–star restaurant or 5–star restaurant.

We discuss these results in the discussions section.

V. R2 – ANALYSIS: REVIEWS BY THE SAME USERS

In the previous section, it was shown that there are differences in language use and communication strategies used in Yelp reviews associated with 1 and 2 star rated restaurants and 4 and 5 star rated restaurants, respectively. However, sometimes, the same user writes reviews associated with different star rated businesses. For example, below are reviews by the same user for a 4–star rated restaurant and a 5–star rated restaurant, respectively.

1.

4) *"Cute interior and owner gave us tour of upcoming patio/rooftop area which will be great on beautiful days like today. Cheese curds were very good and very filling. Really like that sandwiches come w salad, esp after eating too many curds! Had the onion, gruyere, tomato sandwich. Wasn't too much cheese which I liked. Needed something else...pepper jelly maybe. Would like to see more menu options added such as salads w fun cheeses.*

TABLE VII
VALENCE, AROUSAL, AND DOMINANCE BETWEEN 4-STAR REVIEWS COMPARED TO 5-STAR REVIEWS

Attribute	Cohen's D
Valence	2.488
Arousal	1.682
Dominance	1.89

Lots of beer and wine as well as limited cocktails. Next time I will try one of the draft wines"

5) *"Love this place. Tried a few of the lunch meats and crab cakes but the maccheese and lasagna were extra amazing. They did not get shared bc wanted all to my self! Excited for my next order."*

In this section, we aim to answer the following research question: given Yelp reviews (associated with restaurants) written by the same users for different star rated restaurants, are there/what are the differences in communication strategies utilized and language use associated with these reviews? There are several benefits to conducting this analyses; below is an example.

- Do the same users who write Yelp reviews for different star rated businesses/organizations e.g. 4 and 5 star rated businesses/organizations, respectively, use the same language in these reviews? If this is the case, then in cases where there are few reviews associated with a business category e.g. restaurants, with a given star rating e.g. (5–stars), then the reviews associated with 4–star rated businesses in the same category of business e.g. restaurant, can be used as complimentary data in addition to the reviews for the 5–star rated businesses, to gain insight about the given business category.

In this section, similar to the previous section **R1 – Analysis: All Reviews**, LIWC, *Valence*, *Arousal*, and *Dominance* are used to identify the differences in communication strategies utilized and language use in reviews (written by the same users) associated with different star rated restaurants on Yelp.

A. Dataset

From Table I, we identified users who write Yelp reviews associated with 1 and 2 star rated restaurants, respectively and we collect all their reviews associated with these 1 and 2 star rated restaurants. Also, we identify users who write Yelp reviews associated with 4 and 5 star rated restaurants and we collect their reviews associated with these 4 and 5 star rated restaurants. Table VIII shows the dataset of users who write Yelp reviews associated with 1 and 2 star rated restaurants and Table IX shows the dataset of users who write reviews for 4 and 5 star rated restaurants, respectively. Given that there are few reviews (by the same users) associated with 1 and 2 star rated restaurants (i.e.Table VIII), in this section, we only conduct analyses on the reviews associated with the 4 and 5 star rated restaurants (Table IX).

a) :

TABLE VIII
YELP RESTAURANT DATASET FOR 1 AND 2 STAR RESTAURANTS – SAME USERS

Rating	# of Restaurants	# of Users	# of Reviews
1 star	111	183	210
2 star	240	183	271

TABLE IX
YELP RESTAURANT DATASET FOR 4 AND 5 STAR RESTAURANTS – SAME USERS

Rating	# of Restaurants	# of Users	# of Reviews
4 star	10915	12235	82640
5 star	1503	12235	16927

B. LIWC Analysis

In this section, using the exact same approach for the LIWC analysis from section **R1 – Analysis: All Reviews**, we determine the differences in the LIWC categories associated with 4-star Yelp reviews when compared to 5–star Yelp reviews, by the same users and vice versa. Table X shows the LIWC categories most associated with 4–star reviews when compared to 5–star reviews and Table XI shows the LIWC categories that are most associated with 5–star reviews when compared to 4–star reviews.

TABLE X
LIWC CATEGORIES MORE ASSOCIATED WITH 4-STAR REVIEWS COMPARED TO 5-STAR REVIEWS BY THE SAME USERS

#	LIWC	Cohen's D
1	Differentiation	0.420
2	Article	0.246
3	Focus Past	0.216
4	Tentative	0.202
5	Cognitive processes	0.192
6	Negations	0.148
7	Total function words	0.121
8	Leisure	0.118
9	Prepositions	0.114
10	Negative emotion	0.111

C. Word Score based Methods

Here, using the same approach for the word score based methods analyses in section **R1 – Analysis: All Reviews**, for each review associated with 4 and 5 star Yelp reviews, the average ratings of the content words associated with each of the *Valence*, *Arousal*, and *Dominance* word category is computed. We create features that represents the average *Valence*, *Arousal*, and *Dominance* scores, respectively for reviews and another feature that represents if a review is associated with a 4–star restaurant or 5–star restaurant. Table XII shows the results.

D. Result

1) 4 and 5 star Reviews Analyses:: As it relates to the analyses on reviews – by the same users, associated with

TABLE XI
LIWC CATEGORIES MORE ASSOCIATED WITH 5-STAR REVIEWS COMPARED TO 4-STAR REVIEWS BY THE SAME USERS

#	LIWC	Cohen's D
1	Social processes	0.308
2	Achievement	0.307
3	Certain	0.297
4	Positive emotion	0.256
5	Focus Present	0.256
6	Affect	0.249
7	Common adjectives	0.210
8	Power	0.198
9	Second person pronoun	0.168
10	Drives	0.160
11	Friends	0.147
12	Focus Future	0.146
13	Money	0.145
14	Male references	0.142
15	Family	0.134
16	Work	0.129
17	3rd person singular	0.124
18	Perception	0.122
19	Home	0.107
20	Causation	0.101

4 – star restaurants and 5 – star restaurants, we find that the LIWC categories on *Focus Past*, *Negation*, and *Negative Emotion*, for example are more associated with 4–star reviews compared to 5–star reviews (i.e. Table X). Also, we find that the LIWC categories on *Social Processes*, *Positive Emotion*, *Focus Present*, and *Second Person Pronoun*, for example are more associated with 5–star reviews compared to 4–star reviews (i.e. Table XI).

We discuss these results in the discussions section.

TABLE XII
VALENCE, AROUSAL, AND DOMINANCE BETWEEN 4-STAR REVIEWS COMPARED TO 5-STAR REVIEWS BY SAME USERS

Attribute	Cohen's D
Valence	0.941
Arousal	1.543
Dominance	1.383

VI. DISCUSSION

The aim of this work is to gain insights about the language use differences between reviews associated with 1–star rated restaurants and 2–star rated restaurants and between 4–star and 5–star rated restaurants. Specifically, this work uses the following language features, LIWC and the word score based methods *Valence*, *Arousal*, and *Dominance* to determine the language use differences between Yelp reviews associated with (a) 1–star rated restaurants compared to 2–star rated restaurants and (b) 4–star rated restaurants compared to 5–star rated restaurants. This work also uses LIWC, *Valence*, *Arousal*, and *Dominance* to analyze reviews, written by the same users, for 4 and 5 star rated restaurants, to determine the differences in language use in these reviews. In this section, we discuss the findings from the analyses conducted in this work.

A. Analysis: All Reviews

LIWC Analysis: Regarding reviews associated with 1 and 2 star rated restaurants, we find that (a) the LIWC categories on *Negative Emotion*, *Anger*, *Negation*, *Focus Present*, and *Anxiety* are associated with reviews for 1–star restaurants when compared to 2–star restaurants (Table II) (b) the LIWC categories on *Positive Emotion*, *Differentiation*, *First Person Plural Pronoun*, and *Sadness* (Table III) where more associated with reviews for 2–star restaurants when compared to reviews for 1–star restaurants. As stated in the introduction section, a rating of 1–star for a business indicates that the business is bad; also, a rating of 2–stars for a business indicates that the business is bad – but not as bad as the 1–star rated business. Hence, the LIWC categories on *Negative Emotion*, *Negation*, and *Anger* are more associated with 1–star reviews compared to 2–star reviews. Also, the LIWC category on *Focus Present* is more associated with reviews of 1–star rated restaurants; potentially, this indicates that these reviews tend to be written while these individuals (i.e. reviewers) are still in the restaurants. Compared to the reviews associated with 1–star rated restaurants, the reviews associated with 2–star restaurants are more associated with the LIWC categories on *Positive Emotion*, *Differentiation*, *First Person Plural Pronoun*, and *Sadness*. A potential reason for why the LIWC category on *Positive Emotion* is more associated with reviews for 2–star rated restaurants compared to 1–star rated restaurants is that while 2–star rated restaurants are not that good, they are better than the 1–star rated restaurants, hence the LIWC *Positive Emotion*.

Regarding reviews associated with 4 and 5 star rated restaurants, we find that (a) the LIWC categories on *Focus Past*, *Negation*, and *Negative Emotion*, for example are associated with reviews for 4–star restaurants when compared to reviews for 5–star restaurants (Table IV) (b) the LIWC categories on *Positive Emotion*, *Focus Present*, and *Focus Future* where more associated with reviews for 5–star restaurants when compared to 4–star restaurants (Table V). Potentially, the reason why thhe LIWC categories on *Negation* and *Negative Emotion* are more associated with 4–star rated restaurants compared to 5–star rated restaurants is that while the 4–star restaurants are good, they are not as good as the 5–star rated restaurants. Similar to the reviews associated with 1–star rated restaurants, the LIWC category on *Focus Present* is more associated with reviews for 5–star rated restaurants, thereby indicating that reviewers tend to write reviews for 5–star rated restaurants while they are still in the restaurant. Also, the LIWC category on *Focus Future* are more associated with reviews for 5–star restaurants; this potentially indicates that these reviewers state in their reviews that they will patronize these 5–star rated restaurants in the future.

Word Score based Methods: From Table VI we find that the Word score based methods, *Valence*, *Arousal*, and *Dominance* were more associated with the reviews for the 1–star rated restaurants when compared to the 2–star rated restaurants. A potential reason for *Valence* being more associated with reviews for 1–star rated restaurants is that the words used in reviews associated with 1–star rated restaurants convey negative meanings. *Arousal* is more associated with reviews for 1–star rated restaurants; this is potentially because, the words used in reviews for 1–star restaurants express low arousal. Also, a potential reason for why *Dominance* is more associated with reviews for 1–star rated restaurants, is that low dominance words are used more in reviews for 1–star rated restaurants. The Cohen's D for *Valence* i.e. 20.717 and *Arousal* i.e. 37.624, are high, thereby indicating that a lot of low *Valence* and *Arousal* words, respectively, tend to be used in reviews associated with 1–star rated restaurants compared to those associated with 2–star restaurants.

From Table VII, similar to reviews associated with 1–star and 2–star rated restaurants, we find that *Valence*, *Arousal*, and *Dominance* were more associated with the reviews for the 5–star rated restaurants when compared to the 4–star rated restaurants. A potential reason for this is that the words used in reviews associated with 5–star rated restaurants, tend to express high *Valence*, *Arousal*, and *Dominance*.

B. Analysis: Reviews by the Same Users Analysis

LIWC Analysis: Regarding reviews – written by the same users, for 4–star and 5–star rated restaurants, we find that the LIWC categories on *Focus Past*, *Negation*, and *Negative Emotion*, for example are more associated with reviews for 4–star rated restaurants compared to 5–star rated restaurants. Also, the LIWC categories on *Positive Emotion*, *Focus Present*, and *Focus Future* were more associated with reviews for 5 – star rated restaurants. These results indicate that the language used in reviews (written by the same users) for 4–star and 5–star restaurants, respectively are different.

Word Score based Methods: From Table XII, it is observed that *Valence*, *Arousal*, and *Dominance* were more associated with the reviews for 5–star rated restaurants compared to those for 4–star rated restaurants. Thereby indicating that there are differences in language use in reviews (written by the same users) for 4–star and 5–star rated restaurants, respectively.

The findings from this work quantifies the language use differences between reviews associated with 1–star and 2–star rated restaurants and 4–star and 5–star restaurants, respectively. This work also shows that, there are language use differences in reviews, written by the same users, for 4–star and 5–star rated restaurants. Hence, based on the findings from this work, when analyzing Yelp reviews associated with businesses, especially restaurants, reviews for 1–star rated restaurants and 2–star rated restaurants should be analyzed separately and the same applies to reviews for 4–star arated restaurants and 5–star rated restaurants.

VII. Limitations

This work has several limitations; here we outline some of these limitations.

- The analyses conducted in this work focused on Yelp reviews associated with restaurants. It is possible that the Yelp reviews associated with other categories of businesses/organizations (other than restaurants) are different; therefore, the findings from this work may not apply to Yelp reviews associated with other business categories that are not restaurants.
- Furthermore, we acknowledge that the motivation for this study stems from previous work in the healthcare domain [1]–[3], which may also not apply to the restaurant industry.
- There are other online review platforms other than Yelp e.g. Tripadvisor. The language use and communication strategies used by users when writing reviews on these other online review platforms may vary from the language use and communication strategies used in Yelp reviews. Hence, the findings from this work may not apply to restaurant reviews (or other business categories reviews) from other online review platforms.
- Restaurants in different locations attract different clientele; for example, restaurants in New York City attract tourists in addition to people who leave in New York City compared to restaurants in a small town which mostly attracts local residents. Also, different restaurants serve different cuisines and meals. The reviews associated with restaurants that attract tourists and local residents may be different from restaurants that attract mostly local residents. The analyses in this work does not take these factors such as location, type of cuisine, and clientele into consideration.
- The dataset used for this work is a small subset of restaurants in the following metropolitan areas [3]: *Montreal*, *Calgary*, *Toronto*, *Pittsburgh*, *Charlotte*, *Urbana-Champaign*, *Phoenix*, *Las Vegas*, *Madison*, and *Cleveland*, and so the findings from this work cannot be generalized to all the restaurants in North America or other countries.

VIII. Future Work

In the future, we will extend this work by conducting several analyses; below we describe our future work:

- The analyses in this work focused on Yelp reviews associated with one business category i.e. restaurants. In the future, we will analyze Yelp reviews associated with several business categories such as hotels and hospitals; with these analyses, we will aim to answer research questions such as (a) are there language use differences in reviews (from the same business categories) associated with 1–star and 2–star (and 4–star and 5–star) rated businesses? (b) This work analyzed reviews from one online review platform i.e. Yelp. In the future, we will

[3]https://www.yelp.com/dataset/documentation/faq

analyze reviews from other online review platforms such as Tripadvisor, to determine if the findings from the analyses in this work, translates to the reviews on other online review platforms.

- Reviews associated with business categories e.g. restaurants, may (or may not) vary across geographical regions and cultural backgrounds. Previous work [30] has demonstrated the importance of cross-cultural differences in analyzing user disclosures on social networks. In the future, we will analyze reviews associated with different business categories from various online review platforms to determine if online reviews for businesses vary across geographical regions and cultural backgrounds.

IX. Conclusion

In this work, we show that there are language use differences between Yelp reviews associated with 1-star rated businesses compared to those associated with 2-star rated businesses; we show that the same applies to reviews associated with 4-star rated businesses compared to 5-star rated businesses. We also show that for Yelp reviews, written by the same users, there are language use differences, as well. Hence, when conducting analyses on Yelp reviews, specifically, restaurant reviews, reviews for 1–star rated restaurants and those for 2–star rated restaurants should not be analyzed together; the same applies to reviews associated with 4–star rated restaurants and 5–star rated restaurants.

References

[1] Abrams, Matthew P., Raina M. Merchant, Zachary F. Meisel, Arthur P. Pelullo, Sharath Chandra Guntuku, and Anish K. Agarwal. "Association Between Online Reviews of Substance Use Disorder Treatment Facilities and Drug-Induced Mortality Rates: Cross-Sectional Analysis." JMIR AI 2 (2023): e46317.

[2] Seltzer, Emily K., Sharath Chandra Guntuku, Amy L. Lanza, Christopher Tufts, Sindhu K. Srinivas, Elissa V. Klinger, David A. Asch, Nick Fausti, Lyle H. Ungar, and Raina M. Merchant. "Patient experience and satisfaction in online reviews of obstetric care: observational study." JMIR Formative Research 6, no. 3 (2022): e28379.

[3] Stokes, Daniel C., Rachel Kishton, Haley J. McCalpin, Arthur P. Pelullo, Zachary F. Meisel, Rinad S. Beidas, and Raina M. Merchant. "Online reviews of mental health treatment facilities: narrative themes associated with positive and negative ratings." Psychiatric services 72, no. 7 (2021): 776-783.

[4] Pennebaker, James W., Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. "The development and psychometric properties of LIWC2015." (2015).

[5] Mohammad, Saif. "Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words." In Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers), pp. 174-184. 2018.

[6] Mo, Zan, Yan-Fei Li, and Peng Fan. "Effect of online reviews on consumer purchase behavior." Journal of Service Science and Management 8, no. 03 (2015): 419.

[7] Gavilan, Diana, Maria Avello, and Gema Martinez-Navarro. "The influence of online ratings and reviews on hotel booking consideration." Tourism Management 66 (2018): 53-61.

[8] Jiménez, Fernando R., and Norma A. Mendoza. "Too popular to ignore: The influence of online reviews on purchase intentions of search and experience products." Journal of interactive marketing 27, no. 3 (2013): 226-235.

[9] Alamoudi, Eman Saeed, and Norah Saleh Alghamdi. "Sentiment classification and aspect-based sentiment analysis on yelp reviews using deep learning and word embeddings." Journal of Decision Systems 30, no. 2-3 (2021): 259-281.

[10] Faisol, Hussein, Kevin Djajadinata, and Muljono Muljono. "Sentiment analysis of yelp review." In 2020 International seminar on application for technology of information and communication (iSemantic), pp. 179-184. IEEE, 2020.

[11] Sadikin, Mujiono, and Abi Fauzan. "Evaluation of Machine Learning Approach for Sentiment Analysis using Yelp Dataset." European Journal of Electrical Engineering and Computer Science 7, no. 6 (2023): 58-64.

[12] Chen, Yifan, and Fanzeng Xia. "Restaurants' rating prediction using Yelp dataset." In 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), pp. 113-117. IEEE, 2020.

[13] Anthal, Jyotsna, Bhavna Sharma, and Jatinder Manhas. "Optimization-enabled deep learning for sentiment rating prediction using review data." Service Oriented Computing and Applications 17, no. 1 (2023): 39-58.

[14] Asghar, Nabiha. "Yelp dataset challenge: Review rating prediction." arXiv preprint arXiv:1605.05362 (2016).

[15] Liang, Ying. "Yelp Rating Prediction with Sentiment and Topic Models." (2017).

[16] Liu, Zefang. "Yelp review rating prediction: Machine learning and deep learning models." arXiv preprint arXiv:2012.06690 (2020).

[17] Luo, Manman, and Xiangming Mu. "A New Ontology for Restaurant Review Sentiment Analysis." Proceedings of the Association for Information Science and Technology 60, no. 1 (2023): 1068-1070.

[18] D'Aniello, Giuseppe, Matteo Gaeta, and Ilaria La Rocca. "KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis." Artificial Intelligence Review 55, no. 7 (2022): 5543-5574.

[19] Hung, Lai Po, and Suraya Alias. "Beyond sentiment analysis: A review of recent trends in text based sentiment analysis and emotion detection." Journal of Advanced Computational Intelligence and Intelligent Informatics 27, no. 1 (2023): 84-95.

[20] Bensoltane, Rajae, and Taher Zaki. "Towards Arabic aspect-based sentiment analysis: A transfer learning-based approach." Social Network Analysis and Mining 12 (2022): 1-16.

[21] Verma, Sharad, Ashish Kumar, and Aditi Sharan. "IAN-BERT: combining post-trained BERT with interactive attention network for aspect-based sentiment analysis." SN Computer Science 4, no. 6 (2023): 756.

[22] Rafay, Abdul, M. Suleman, and Affan Alim. "Robust review rating prediction model based on machine and deep learning: Yelp dataset." In 2020 International Conference on Emerging Trends in Smart Technologies (ICETST), pp. 8138-8143. IEEE, 2020.

[23] Qiao, Rui. Yelp review rating prediction: Sentiment analysis and the neighborhood-based recommender. University of California, Los Angeles, 2019.

[24] Zhang, Yinshi. "Semantic Feature Analysis and Mining for Yelp Rating Prediction."

[25] Zhang, Sonya, Linda Ly, Norman Mach, and Christian Amaya. "Topic modeling and sentiment analysis of yelp restaurant reviews." International Journal of Information Systems in the Service Sector (IJISSS) 14, no. 1 (2022): 1-16.

[26] Andy, Anietie, and Uduak Andy. "Understanding communication in an online cancer forum: content analysis study." JMIR cancer 7, no. 3 (2021): e29555.

[27] Andy, Anietie, Garrick Sherman, and Sharath Chandra Guntuku. "Understanding the expression of loneliness on Twitter across age groups and genders." Plos one 17, no. 9 (2022): e0273636.

[28] Tan, Chenhao, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. "Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions." In Proceedings of the 25th international conference on world wide web, pp. 613-624. 2016.

[29] Andy, Anietie. "Studying how individuals who express the feeling of loneliness in an online loneliness forum communicate in a nonloneliness forum: observational study." JMIR Formative Research 5, no. 7 (2021): e28738.

[30] Munmun De Choudhury, Sanket S. Sharma, Tomaz Logar, Wouter Eekhout, and René Clausen Nielsen. 2017. Gender and Cross-Cultural Differences in Social Media Disclosures of Mental Illness. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17). Association for Computing Machinery, New York, NY, USA, 353–369. https://doi.org/10.1145/2998181.2998220