

Julie Bach

Professor Daniel Youd

WRIT 215

December 12, 2023

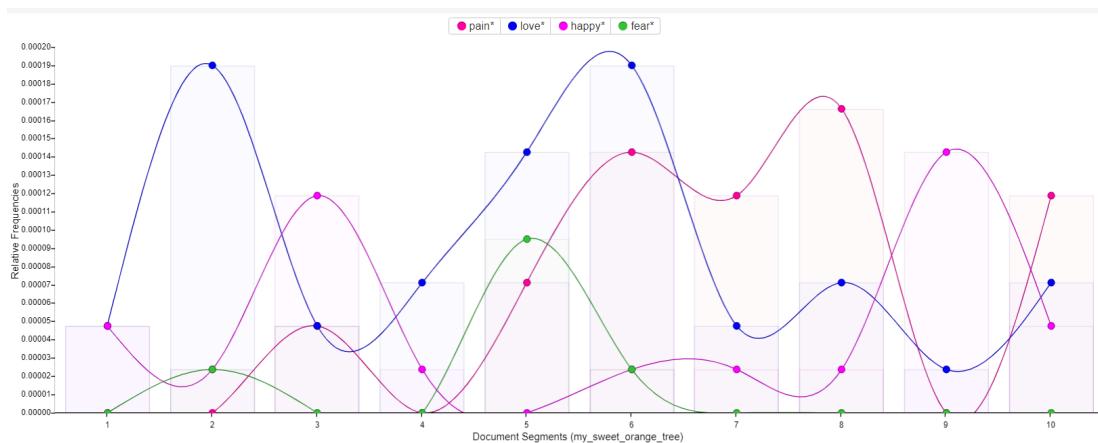
FINAL PAPER

With the emergence of digital sources, humans are processing data regardless of space and time to differentiate between bad and good information. It is undeniable that we naturally develop critical thinking from experience for better contextual understanding and data processing. When it comes to a data-is-everywhere world, an advanced computer operating system should be developed to cope with large databases and provide productive, less labor-intensive processes. With the aid of Natural Language Processing (NLP) techniques and programming languages, we can transform from qualitative methods—traditional document analysis—to quantitative text analysis. Hence, this paper aims to conduct both analysis methods on a novel called *My Sweet Orange Tree*, which was originally published in Brazil, by José Mauro de Vasconcelos. Going beyond a novel about child abuse, I believe that other themes including compassion, love, and friendship are more apparent. In other words, despite physical and psychological abuse, the development of Zezé's self-compassion development might increase his emotional kindness towards himself and others which could affect his language intensity. This hypothesis is proved through Voyant Tools, a web-based analysis environment for digital text, sentiment lexicons, and traditional text analysis.

Published in the mid-20th century, *My Sweet Orange Tree* is an autobiographical novel about a precocious, imaginative, and clever five-year-old boy named Zezé. He was born into a humble and poor family and usually experienced abuse and punishments from his family members. As miserable as he is, he later finds solace from nurturing friendships with a talking orange tree called Pinkie in his new house and a Portuguese man named Manuel. However, he becomes traumatized and suffers from heartbreak after losing his only source of adult tenderness and care. Apart from emphasizing the tragedy of poor childhood upbringing, this novel also highlights the healing power of self-compassion development, imagination, and nurturing friendship that could influence humans' emotional valence and mitigate negativity bias towards negative situations.

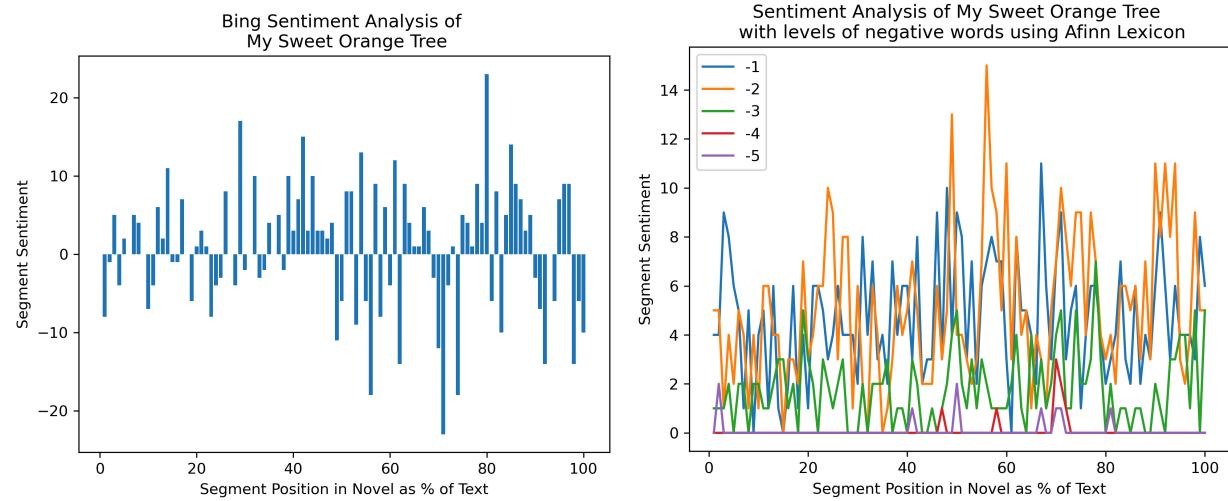
Self-compassion has been introduced as a philosophy of Buddhism for centuries that highlights the tenderness and understanding toward oneself when one suffers from inadequacy, failure, or misery. As stated by Dr. Kristin Nedd, it "is composed of three main elements: self-kindness, common humanity, and mindfulness". To derive the connection between self-compassion or self-competence and language density, I observed Zezé's relationships with family members and his best friends and then used lexicon-based sentiment analysis and valence with Bing, Afinn, and NRC sentiment lexicons. Based on the theory of negativity bias, our brain structure is designed to focus on the negative to survive, cope with threats, and take advantage of opportunities that encourage reproduction without seeing the beauty around us. Therefore, our propensity to "attend to, learn from, and use negative information far more than positive information" (Catherine Moore) seems to be more apparent. Despite the asymmetric information processing between negativity and positivity, I believe that other factors could outperform this

tendency like Zezé's high level of self-compassion helps him to view his family conditions better than other poor children, treat others nicely, and maintain emotional comfort. The contrary of this novel's general theme and the actual reading observations can be seen when looking at the trends of document terms on the Voyant Tools. The line graph of four words in which "fear" and "pain" represent the abuse theme, while the others are for the compassion and love genre has raised a question for me. The significant occurrences of positive words might go against our first impression of the novel summary, but the certainty is still low due to the small sample size.



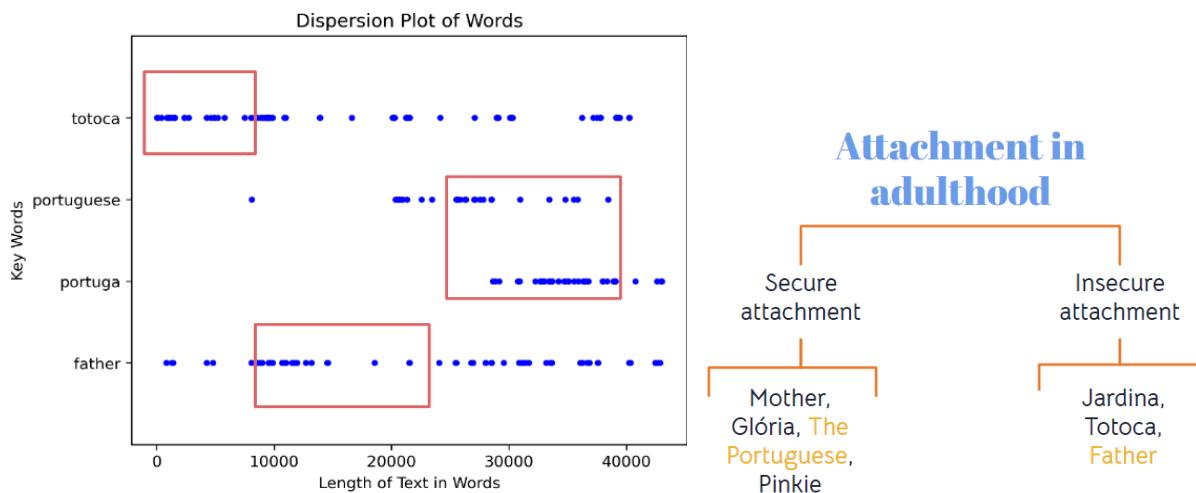
Therefore, I used the Bing sentiment lexicon, which categorizes words in a binary fashion, either positive or negative, to have a bigger picture of the emotional valences. I have raised some questions after seeing the same trend from the resulting chart below (the left-hand side graph): What truly happens in the story that contradicts the initial novel's theme? How did Zezé react to his tragic life? Did he change his behaviors? I also use the Afinn lexicon to derive a deeper understanding of the valence ratings of words that Zezé tends to use and get a surprising result (the right-hand side one). In particular, Zezé does not use many intensely negative words since the occurrences of negative words on the scale -2 are dominant compared to other ranges. Although given a broad level of the book's concept, the uncertainty of this trend when he

interacts with the abuser or his caregivers requires further focus on smaller segments to analyze potential aspects regarding more contextual information.



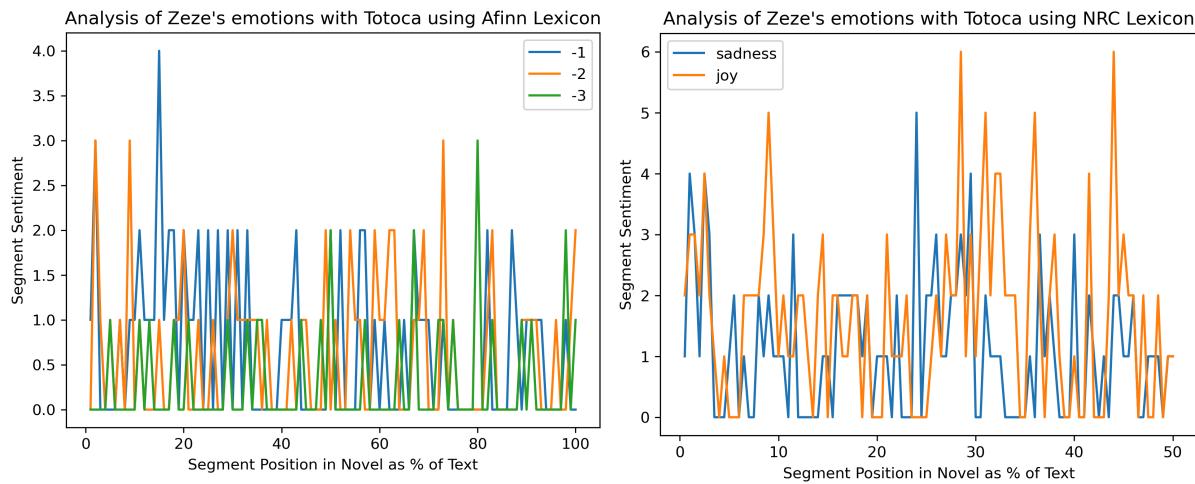
According to Bowlby, attachment behavior is completed at the age of two or three but the attachment that occurs and continues throughout life could contribute to determining emotions, thoughts, and behaviors. In a secure attachment, an individual feels secure, supported, and emotionally stable when being with the other so it is expected that individuals with secure attachments have high self-compassion. Conversely, individuals with insecure attachments tend to avoid close interactions, suppress negative emotions and thoughts, and develop emotional distance which suggests that their self-compassion levels are low. With the context of Zezé's life, he develops secure attachments with his mother, his sister, Glória, the Portuguese, Manuel, his teacher, and his talking orange tree, Pinkie, whom he receives love and support from, while his relationships with his father, his brother, Totoca, and another sister, Jardina are considered as insecure attachments because they always punish him arrogantly and hurt his feelings. It is obvious that people tend to have positive feelings towards their caregivers so to determine one's level of self-compassion, more quantitative measures should be done focusing on their negative

relationships. In this case, I will look at Zezé's reactions to the abusers, his father and brother, and the caregiver, his Portuguese friends, as the love and support he is searching for from the Portuga are what he deserves to receive from his father so their roles are quite similar in this novel. I tend to choose different segments for those three observations to avoid overlapping results as can be seen from this plot.

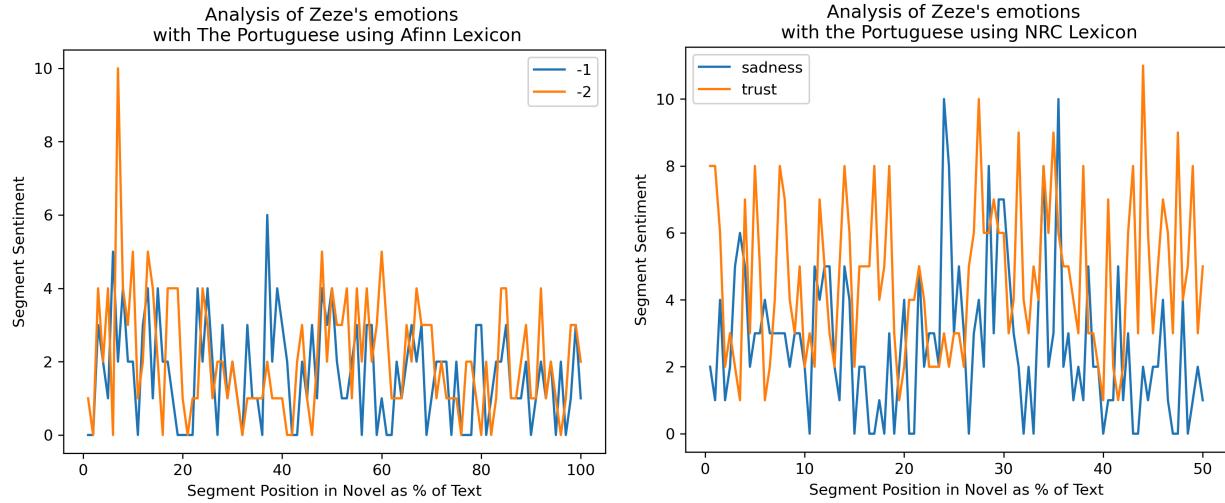


Regarding the negative words from the Afinn lexicon and the “sadness” and “trust” categories from the NRC lexicon, not only does he have a fairly low average level of negative valences, at -2, but also uses more “trust” words compared to sad ones. Those results go beyond my expectations as it does not seem reasonable for those experiencing physical and psychological abuse so there might be an underlying factor, which may result from his self-compassion development. Moreover, the range of his negative emotions with other insecure attachments (Totoca) experiences the same pattern. He might also have more positive reactions toward his brother because the sentiment analysis graph using the NRC lexicon shows that the occurrence of words with joyful characteristics is higher than the other. Apart from that, another interesting observation when using the Voyant Tool is that there is no “hate” word in this novel

because it seems to be the basic response to childhood trauma including intense and ongoing emotional upset, depressive symptoms, anxiety, and behavioral changes. However, we can only see those reactions when he faces the loss of his two best friends, Pinkie and Manuel, in the very last chapters. Therefore, Zezé's strong level of self-compassion might affect his emotions and perspectives towards negative situations and therefore mitigate the negativity bias.



On the other hand, observing the secure attachment between him and the Portuguese, the results surprisingly show the same trend with the average level of negative valence at -2 but have more peaks during some segments. Therefore, after calculating those exact segment positions and locating them in the novel, those strange trends result from some remarkable events including their first meet, when Zezé is caught by the Portuga because of his naughty pranks, and his traumatized period when his second father dies of a car accident.



It is undeniable that lexicon-based sentiment analysis has helped us to conduct quantitative data research on narrative writing, especially with long novels and substantial chunks of text, to discover common patterns, find averages, and make comparisons between factors like how I observed Zezé's lexical density for negative and positive valences. We could take advantage of those tools to categorize novels into sub-genres because this child-abuse novel does not bring negative and intense reading experience so readers might be offered various choices depending on their preferences. However, quantitative methodology might produce confounding results as we overlook the context or the setting of each scene. By way of illustration, without using traditional document analysis to have qualitative data, I could not know what happened and why the density of negative words significantly increased. Hence, it is important to combine these quantitative and qualitative analyses to avoid skewed distributions, and confounding variables, and deliver more insightful findings.

As interesting as the findings from sentiment analysis, I also encountered difficulties when choosing segments for each character. In particular, based on the dispersion plot, the

density of the words “father”, “portugal”, and “portuguese” appear at the same segment position so choosing the same for both will end up with overlapping and unclear results. This problem can be solved using the Text Encoding Initiative (TEI) to tag specific sentences and paragraphs belonging to each character which requires a substantial amount of time and effort. If I had enough time, I would continue working on this project to prove my hypothesis persuasively. Last but not least, I did not have a chance to compare the translation editions of this novel as I originally read this book in Vietnamese but analyzed the English version. I believe that different translation versions not only bring unique reading experiences but also affect how the author defines positive and negative valances and categorizes them into smaller scales.

In conclusion, My Sweet Orange Tree is a coming-of-age story that explores themes of life, love, loss, fear, and friendship. Through sentiment analysis of Zezé’s valences of words, I believed that self-compassion could contribute to mitigating the negativity bias in using intense words toward severe situations and negative relationships. I not only have a chance to discover different aspects of the novel by using qualitative and quantitative analysis but also understand some potential problems that require complex data preparation steps for better-structured databases and optimized analyzing processes.

References

KELEŞ, F. B., AK, M., & KESİÇİ, Ş. “Analysis of the Novel of My Sweet Orange Tree in the Context of Self-Compassion” *Research on Education and Psychology*, 30 Jun. 2023, <https://doi.org/10.54535/rep.128226>. Accessed 12 Dec. 2023.

Moore, Catherine. “What is Negativity Bias and How Can It Be Overcome?” *Positive Psychology*, <https://positivepsychology.com/3-steps-negativity-bias/#bias>. Accessed 12 Dec. 2023.

Neff, Kristin. “What is Self-Compassion?” *Self-compassion*, <https://self-compassion.org/the-three-elements-of-self-compassion-2/>. Accessed 12 Dec. 2023.

Vasconcelos, Jose. *My Sweet Orange Tree*, 1968.