Embedded Personalities: Word Embeddings and the "Big Five" Personality Model

Oliver Müller
Saarland University
s8olmuel@uni-saarland.de

Stefania Degaetano-Ortlieb
Saarland University
s.degaetano@mx.uni-saarland.de

Abstract

The Big Five personality model (OCEAN: Openness to Experience, Conscientiousness, Extraversion, Agreeableness & Neuroticism) has been a cornerstone in psychology (McCrae and John, 1992), offering robust cross-cultural validity for understanding personality traits. Traditionally, these dimensions are derived from factor analyses of self-assessment questionnaires, where participants were asked to rank themselves on adjective scales. The present study explores a novel approach by using word embeddings to represent adjectives associated with the Big Five as vectors in a multi-dimensional space. Using a pre-trained Word2Vec model, we mapped 100 adjectives onto a high-dimensional vector space. After dimensionality reduction and clustering with PCA and K-means, results successfully recreated the Big Five dimensions. Our method demonstrates potential for expanding personality analysis to other fields of study such as literary studies or on historical data where self-assessment approaches are not applicable and possibly uncovering new insights into personality research.

1 Introduction

The Big Five personality model, encompassing Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN), is widely regarded as one of the most robust and cross-culturally valid frameworks to investigate personality traits in experimental psychology research (McCrae and John, 1992; Goldberg, 1993). Traditionally, these dimensions are derived from factor analyses of self-assessment questionnaires, where participants are asked to rank themselves on adjective scales (Goldberg, 1992) or phrase-based statements (McCrae and John, 1992) (see also John et al. (1999) for a historical overview). Adjective-based studies stem from the *lexical hypothesis*, which posits that the most significant personality traits are encoded in natural language (Allport and Odbert, 1936). Moreover, there also exist covariation patterns, i.e. people that tend to rate themselves as high on adjectives like *happy* would also rate themselves as high on *social*. Using these patterns of covariation, the results of these adjective questionnaires were then correlated using an exploratory factor analysis leading to the clustering of the given five factors (cf. Goldberg (1992)). While these methods have yielded consistent results, they rely heavily on subjective reporting and assume linear relationships between traits (John et al., 1999).

In recent years, word embeddings have been shown to capture semantic and relational properties of language (Mikolov et al., 2013). We apply word embeddings to explore their potential for modeling psychological constructs like personality traits.

In this paper, our aim is to model the Big Five dimensions using adjective word embeddings, which code adjectives as vectors in a multi-dimensional space. This allows for clustering and visualization of relationships between traits without reliance on self-reported data, which may sometimes be skewed by errors of the participants' subjective perception, also referred to as the introspection illusion by Pronin and Kugler (2007). By applying Principal Component Analysis (PCA) and clustering (with K-means), we aim to recreate the Big Five dimensions and evaluate their representation within the embedding space. The motivation behind this study is the idea that the investigation of personality traits could be expanded to other scenarios where self-assessment questionnaires cannot be applied as in the case of characters in novels or historical correspondences between individuals. Also, given that word embeddings, by their nature, encode relationships between words as vectors in a multi-dimensional space, we get to see how adjectives cluster into traits and how closely related they are to one another. This is particularly valuable

Proceedings of LaTeCH-CLfL 2025, pages 205–215
May 4, 2025 ©2025 Association for Computational Linguistics

since the vast majority of studies usually exclusively rely on exploratory or confirmatory factor analyses (EFA, CFA) as their primary evaluation methods for the Big Five clusterings, which do not naturally lend themselves to such visualizations. Here, our research question is whether there is a significant difference between the number and the clustering of the Big Five personality dimensions when applying a word embeddings approach instead of a factor analysis.

Results show that we can replicate findings, which gives value to the traditional approach and validity to applying a word embedding approach on scenarios beyond self-assessment based ones. Thus, the word embedding approach provides a scalable alternative for analyzing language in the view of personality traits.

2 The Big Five Personality Model

2.1 Background

Tupes and Christal (1961) achieved a breakthrough in personality research by creating a robust and generalizable model of personality traits. Through eight experiments analyzing intercorrelation matrices, they identified the Big Five dimensions: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience. Because their research was intended to improve personnel management and performance within the military, they had access to a relatively controlled set of participants. Although controlled, their study included diverse participants (from assessment programs, military training, airmen, undergraduate and graduate students) with varying levels of familiarity with the Air Force (days to several years) and a wide range of raters (from novices in psychology to trained and seasoned clinical psychologists and psychiatrists), ensuring broad applicability of the findings. Traits were rated in bipolar pairs (e.g., extroverted vs. introverted), aiming to capture the full spectrum of personality dimensions.

Building on this, Goldberg (1992) refined the Big Five framework by formalizing a concise set of unambiguous English-language adjectives to represent the five dimensions. The goal was to find exactly such a set of adjectives that was both rather small but at the same time produced the Big Five factor clustering as uniformly as possible. Through a series of four studies, Goldberg (1992) demonstrated that unipolar adjective scales (e.g., *friendly* rather than *friendly* vs. *unfriendly*) produced clearer and more robust factor structures than bipolar scales, which were used previously. His efforts culminated in a list of 100 unipolar adjectives that consistently reproduced the Big Five dimensions across diverse datasets.

DeYoung et al. (2007) further expanded the understanding of the Big Five by identifying two correlated subdimensions (or aspects) within each domain, supported by biological and genetic evidence. Their studies validated the Big Five Aspect Scales (BFAS) and found significant genetic correlations for these subdimensions using genetic factors from a previous study by Jang et al. (2002) and correlating them with each of the 10 aspects, highlighting the complexity and nuanced structure of the Big Five traits. These findings supplied further evidence for the hypothesis that the Big Five dimensions of personality and their 10 aspects developed from both environmental and genetic factors.

2.2 The Big Five dimensions

Human personality might very well be far too complex for a five-factor model to sufficiently and exhaustively encompass its entire scope and complexity. Despite that, the Big Five Model is the closest approximation that personality scientists were ever able to come up with in order to objectively measure and categorize significant trait dimensions. A relevant fact to point out beforehand is that these dimensions are not mutually exclusive and that they are individually measured on scales of 1 to 100. This allows for many interesting combinations of traits such as people who are very high in positive emotion and negative emotion, simultaneously.

Openness to Experience is subdivided into *Openness* and *Intellect* relating to two important aspects of this dimension which are aesthetics (interest in beauty) and ideas (interest in truth), respectively (DeYoung et al., 2007; Johnson, 1994). In general, people in this dimension were described by high degree of intellectual capacity, enjoying aesthetic impressions, having wide interests, and having unusual, unconventional thought (McCrae and John, 1992, 198), i.e. they experience the need for variety, novelty, and change and can be described with adjectives such as artistic, curious, imaginative, insightful, and original (McCrae and John, 1992, 179).

Conscientiousness is characterized by a high sense of diligence and dutifulness and governed by conscience, with people being thorough, neat, well-organized, diligent, and achievement-oriented

(McCrae and John, 1992, 197) as well as efficient, planful, reliable, and responsible (McCrae and John, 1992, 178). It encompasses both proactive aspects, such as the need for achievement and commitment to work, and inhibitive aspects, such as moral scrupulousness and cautiousness (DeYoung et al., 2007, 881). It splits into the two aspects of *Industriousness* and *Orderliness*, i.e. industrious people being keen to carry out their plans, finish what they start, get things done quickly and knowing what they are doing, and orderly people who besides liking order also keep things tidy, and like to follow a schedule (DeYoung et al., 2007, 888). Adjectives used to describe this dimension of personality are for example systematic, thorough, meticulous, analytical, efficient and orderly.

Extraversion is characterized by agency or dominance and sociability. DeYoung et al. (2007) suggest two aspects of Extraversion: *Assertiveness* and *Enthusiasm*. While *Assertiveness* relates to taking charge of things, having a strong personality, knowing how to captivate others, and seeing oneself as a good leader, *Enthusiasm* relates to easily making friends, showing feelings when happy, and having fun (DeYoung et al., 2007, 888). People with the Extraversion trait can be described by adjectives such as active, assertive, energetic, enthusiastic, outgoing and talkative (McCrae and John, 1992, 178).

Agreeableness is the dimension that captures how likely people are to quite literally agree or disagree with other people. People at the higher end of this dimension have characteristics such as altruism, nurturance, caring, and emotional support (Digman, 1990, 422). It is subdivided into the aspects *Compassion* and *Politeness*. While for *Compassion* people indicate to feel others emotions and inquire about others' well-being as well as sympathize with others' feelings, i.e. generally taking an interest in other people's lives, *Politeness* is related to respecting authority and avoiding to seem pushy, imposing one's will on others or taking advantage of others (DeYoung et al., 2007, 887). Adjectives used within this dimension are appreciative, forgiving, generous, kind, sympathetic and trusting (McCrae and John, 1992, 178).

Neuroticism is related to experiencing distress with recurrent nervous tension, depression, frustration, guilt, and self-consiousness often associated with irrational thinking, low self-esteem, and poor control of impulses and cravings (McCrae and John, 1992, 195). This dimension is subcategorized into the aspects *Volatility* and *Withdrawal*. While *Volatility* relates to getting upset or angry easily and change moods a lot, *Withdrawal* denotes being filled with doubts about things, feeling easily threatened, worrying about things and being easily discouraged (DeYoung et al., 2007, 887). Adjectives used for this personality type are anxious, self-pitying, tense, touchy, unstable, and worrying (McCrae and John, 1992, 179).

2.3 Previous work and Contribution

Research on personality traits using textual data spans a range of approaches and several recent studies have demonstrated the potential of computational methods in this domain.

Pizzolli and Strapparava (2019) applied personality trait recognition to theater scripts, focusing on specific utterances within dialogues. Using supervised learning models, such as Support Vector Machines and Random Forests, based on bag-of-words and linguistic features they classify characters based on the Big Five personality traits. Recently, Tiuleneva et al. (2024) have published a novel textual dataset of fiction characters' utterances based on the characters' gender and Big Five personality traits. They were able to show that imagined personae mirror language categories of real people, but did so in a more expressive manner. While effective for analyzing fictional characters, this method is tailored to a specific genre and has to rely heavily on manually annotated datasets, with limits in the generalizability across diverse textual domains.

Similarly, Carducci et al. (2018) used supervised learning to predict Big Five traits from Twitter data, emphasizing real-world social media language. This approach successfully demonstrated the applicability of personality trait analysis in short, informal texts but required labeled data and focused primarily on individual-level predictions.

Several recent studies have applied word embeddings to personality analysis, though their objectives and methods differ from our work.

Kazameini et al. (2020) developed a model combining BERT-derived contextualized embeddings with psycholinguistic features, utilizing a Bagged Support Vector Machine (SVM) classifier to predict Big Five personality traits from text. Other studies have examined the biases embedded in word representations. For example, Agarwal et al. (2019) explored implicit biases in word embeddings related to personality stereotypes. While this research highlighted the biases embedded in pre-trained models, it did not use these embeddings to explore or map personality traits in textual corpora.

Multi-modal approaches, such as Ouarka et al. (2024), combine text, audio, and visual data using advanced deep learning architectures to predict personality traits. These methods achieve impressive results in multi-modal settings but require extensive computational resources, which limits their accessibility for humanities researchers working with text-only corpora.

Lastly, Siddique et al. (2019) developed GlobalTrait, a multilingual embedding-based model for aligning personality traits across languages. While this approach addressed multilingual settings, it did not explore the semantic relationships within monolingual corpora or their application to cultural and historical analyses.

A systematic review by Ahmad et al. (2020) provides a broad overview of both supervised and unsupervised methods for personality classification from text, emphasizing their application to structured and labeled datasets. Although comprehensive, the review highlighted the need for flexible, exploratory methods suitable for domains where labeled data may not exist.

Our study differs from the above approaches and presents a first step in meeting these needs in that it applies an unsupervised methodology to explore the semantic relationships among adjectives associated with the Big Five personality traits. By employing clustering techniques and Principal Component Analysis (PCA) on pre-trained word embeddings, we uncover latent structures without relying on labeled datasets. Unlike supervised models, which primarily aim to predict personality traits, our approach focuses on mapping their semantic organization. This allows for exploratory analyses that are particularly beneficial in digital humanities, historical linguistics, and cross-cultural studies, where labeled data is often unavailable. Another key distinction is that supervised approaches necessitate extensive labeled datasets, which are resource-intensive to compile and may not exist for all languages or contexts. Our method circumvents this limitation, aiming for scalability and applicability across diverse textual sources without the need for manual annotation. This makes it especially useful for studying personality traits in corpora where traditional survey-based approaches cannot be applied.

Furthermore, our methodology focuses on semantic relationships among adjectives and emphasizes visualization, making the relationships between traits and adjectives intuitively accessible for interdisciplinary collaboration within and beyond the humanities. While we present a first step towards an exploratory framework of personality traits for texts, a long-term aim would be to provide humanities researchers with a scalable and interpretable tool to uncover semantic patterns in text, bridging computational linguistics and cultural analysis.

3 Methods

To analyze whether the original personality dimensions would emerge using the word embeddings model, a list of 100 adjectives is compiled. This list includes both the original adjectives from Tupes and Christal (1961) and newly selected adjectives, with 20 adjectives allocated to each of the five personality dimensions (10 for each of the two aspects; see Section 2.2). The original studies often used bipolar adjective scales (i.e., *unconventional* vs. *conventional*, *silent* vs. *talkative*), which may work well for methods relying on the number of participants in questionnaire-based experiments rather than the frequency of the items. However, since our approach relies on word-embedding modeling, where the frequency of adjectives matters, we need a more diverse and balanced selection of adjectives. To ensure comprehensive coverage of the dimensions, half of the adjectives are drawn from the original study, while the other half is generated using the Large Language Model ChatGPT-4o, after briefing it to compile 50 additional Big Five adjectives, evenly distributed across the 10 aspects of the Big Five dimensions. This design choice was aimed to enhance diversity of the adjective list and ensure a broad representation of the personality dimensions in the model since the authors in the original often used the previously mentioned bipolar adjectives (e.g., supervised vs. unsupervised), which differed only in their polarity but not their semantic content (see Appendix A for the list of adjectives and Appendix B for the prompt used to generate adjectives).

The word embeddings are calculated using a pretrained Word2Vec model (Google-News-300) accessed via the gensim Python library. The Google-News-300 model was chosen for its extensive training on a large and diverse corpus, ensuring broad coverage of personality-descriptive terms. Additionally, pre-trained embeddings offer a scalable and computationally efficient alternative to training embeddings from scratch. Each of the 100 adjectives corresponding to the Big Five dimensions is encoded into 300-dimensional vector representations.[1] These embeddings are converted into a data frame for easier manipulation.

A principal component analysis (PCA), implemented with scikit-learn, reduced the data to a visualizable three-dimension space, capturing the most significant variance in the 300-dimensional word embedding space. The first three components were selected as they represented the most meaningful structure in the data while balancing interpretability and dimensionality reduction. While alternative dimensionality reduction techniques like t-SNE or UMAP could have been used, PCA was selected for its ability to maintain global structure and provide interpretable linear projections, which are critical for analyzing relationships between personality traits. PCA was complemented with a K-means clustering. Similarly, K-means clustering was chosen for its efficiency and simplicity in identifying distinct groups in high-dimensional spaces. Unlike supervised methods, which require annotated datasets and focus on prediction, our unsupervised approach is better suited to uncover latent semantic patterns in unlabeled data.

To determine the optimal number of natural clusters and assess clustering quality, rather than relying on the assumption of having five clusters as in the Big Five, we conducted a silhouette analysis using the scikit-learn library again before creating a K-means clustering of the embeddings. The results in Figure 1 reveal that five clusters provided the highest average silhouette score, which supports the hypothesis that the Big Five dimensions are reflected in the embeddings. We then used a 3D-visualization of the five clusters to represent results. The selection of three components follows common practice in high-dimensional semantic space analysis, where the goal is to retain as much meaningful structure as possible while avoiding overfitting to noise. Although additional components could capture residual variance, the first three already provide a robust and interpretable organization of personality-related words.[2]

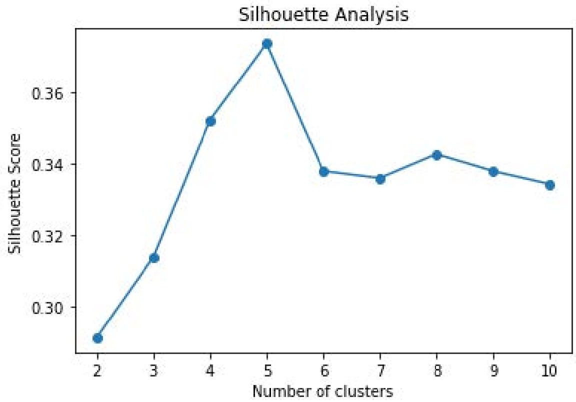

Figure 1: Silhouette analysis indicating five clusters

To test for significant differences between the dimensions, a one-way ANOVA is performed in Python using spyder for each of the three PCA dimensions.

4 Results on Big Five from Word embeddings

Results of the adjective clustering based on the Word2Vec embeddings are shown in Figure 2 in a three dimensional space. Considering the adjectives in each cluster, they in fact group themselves into the established Big Five personality dimensions with only a few rare outliers at the edges, which can be expected given that the Big Five are rather heterogeneous dimensions.

In the following, we are going to give a detailed description of the insights that can be deduced from this type of visualizations. We then move on with a statistical analysis of the findings applying ANOVA. Finally, we consider the contributions of personality aspects to the principal component analysis.

4.1 Trends of personality clusters in the 3D PCA space

Starting from the top left-hand side, the blue dots represent the *Conscientiousness* dimension, which subsumes the aspects of *Orderliness* and *Industriousness* with adjectives such as *thorough*, *orderly*, *efficient*. Continuing on the same plane to the right-hand side, we can see the yellow dots representing

[1]This specific model was chosen due to its generalizability and popularity. In future work we want to apply different models with even higher numbers of dimensions to compare the clusterings.

[2]In a first attempt at a visualization, the PCA was used to reduce the data to a 2D-model. While this visualization already showed a clear clustering of the adjectives into the original Big Five traits, it was far too cluttered to make out many of the individual adjectives and it lacked the depth of a third dimension to better distinguish between a lot of the positions of the traits and the adjectives within them. For these reasons, we opted to visualize a 3D version of the clustering.

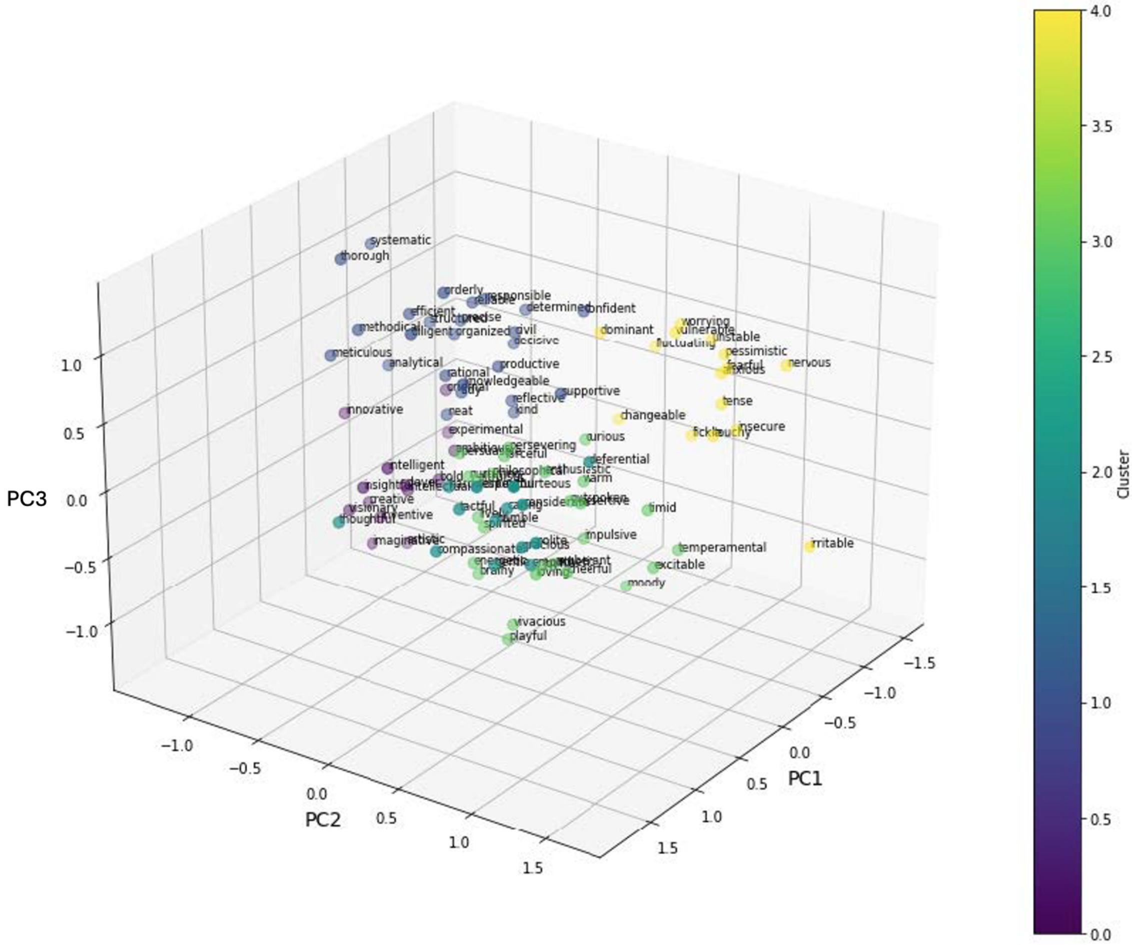

Figure 2: Adjective clustering based on Word2Vec embeddings (3D PCA) indicating Big Five dimensions (blue Conscientiousness, yellow Neuroticism, light green Extraversion, dark green Agreeableness, purple Openness to Experience)

the *Neuroticism* dimension, subsuming *Volatility* and *Withdrawal* with adjectives such as *dominant*, *nervous*, *vulnerable*. Venturing downward and to the left again, the *Extraversion* dimension, which subsumes the aspects of *Enthusiasm* and *Assertiveness* is represented by the light green dots with, e.g., *curious*, *impulsive*, *vivacious*. Continuing further to the left and slightly to the front, we can observe the dimension of *Agreeableness* through the dark green dots, which represents *Compassion* and *Politeness* with adjectives such as *tactful*, *compassionate*, *caring*. Lastly, moving even further to the left and slightly upward, the *Openness to Experience* dimension, representing *Openness* and *Intellect*, manifests itself through the purple dots with adjectives such as *experimental*, *intelligent*, *creative*.

Figure 2 clearly shows a separation between the negative emotion dimension Neuroticism (yellow) on the right and all of the others. Furthermore, there is a major overlap between the dimensions of Extraversion and Agreeableness in the middle of the plot (green colors). Another visible disconnect can be seen between Conscientiousness (blue) and Openness to Experience (purple) on PC3. Upon closer inspection, we can also see a few outliers on the edges of some of the dimensions. For Conscientiousness, *systematic* and *thorough* are positioned far higher on the y-axis than most items in the cluster. As for Neuroticism, *irritable* is way off on PC2 and PC3, arguably being positioned proximally closer to the edge of the Extraversion dimension than the Neuroticism cluster. Concerning Extraversion itself, the distribution of the adjectives is rather spread out, with *vivacious* and *playful* being the only arguable outliers further outside on PC1. The only proper outlier in the Agreeableness dimension is *thoughtful* on the very left-hand side of PC2. In the Openness dimension, we can see *innovative* as an outlier higher up on PC3.

4.2 Statistical assessment of the validity of the clusters

Table 1 shows the means and standard deviations of the three PCAs for each of the 10 aspects of the Big Five dimensions (e.g., Agreeableness_Compassion, Agreeableness_Politeness, etc.). To quantitatively assess the validity of the clustering results, we conducted a one-way ANOVA for each of the three principal components (PC1, PC2, PC3). The analysis tested whether the mean values of each principal component significantly differed across the clusters derived from the K-means algorithm. The results showed strong statistical significance for all three dimensions (PC1: F=17.2629, $p < 0.0001$; PC2: F=26.2739 $p < 0.0001$; PC3: F=11.8351, $p < 0.0001$), indicating that the clusters are well-separated in PCA space and that the trait-associated adjectives form distinct groups.

4.3 Contribution of personality aspects to PCA

Figures 3 to 5 show the contribution of aspects to the principal components, which allow us to further inspect how clusters separate from each other. The contributions are visualized as positive and negative values, indicating potential alignment or opposition of each aspect with the corresponding PC. PC1 (see Figure 3) shows to have positive contributions from both aspects of Agreeableness and negative contributions from both Neuroticism aspects. This component captures opposition between positive and negative emotional traits, consistent with previous work (Costa and McCrae, 1992). PC2 (see Figure 4) has positive contributions from Neuroticism_Withdrawal and Agreeableness-Politeness and negative contributions from Openness to Experience and Conscientiousness_Orderliness.PC3 (see Figure 5) has positive contributions from Conscientiousness and negative ones from Extraversion (Enthusiasm) and Openness to Experience, distinguishing structured and orderly traits against spontaneity and enthusiasm.

These findings suggest that the Word2Vec embeddings successfully capture the semantic relationships between personality aspects, with the three principal components providing a structured and interpretable representation of the main variance in personality-related word meanings. The principal components appear to reflect interpretable dimensions that align with the psychological constructs of the Big Five dimensions. The visualization of contributions offers insights into the clustering structure and validates the embeddings' capacity to model personality traits on the basis of adjectives.

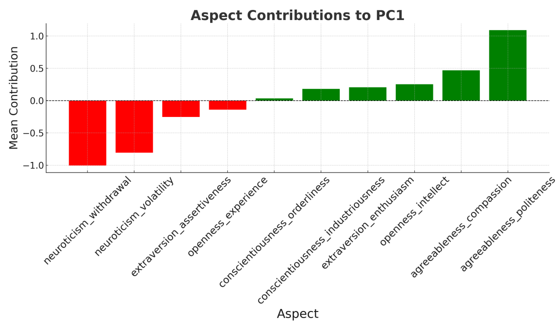

Figure 3: Aspects contribution to PC1

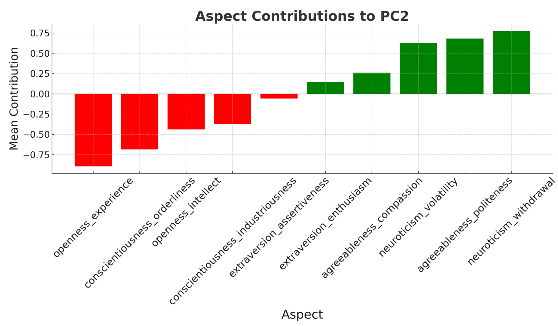

Figure 4: Aspects contribution to PC2

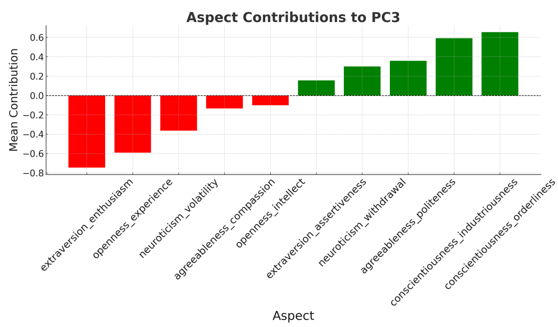

Figure 5: Aspects contribution to PC3

5 Summary and Conclusion

The focus of this study was to explore the relationships among adjectives associated with the Big Five personality traits in textual corpora. Since traditional supervised methods require labeled datasets, which are often unavailable for historical or literary texts, we opted for unsupervised methods. Using a pre-trained word embeddings model (GoogleNews-300), the Big Five dimensions and their 10

Table 1: Means and standard deviations of the top three principal components (PC1, PC2, PC3) for word2vec embeddings of Big Five personality aspects. Higher or lower mean values indicate stronger alignment of words in each aspect with the respective principal component, while the standard deviation reflects the variability in this alignment. This provides insight into how different personality traits are structured in semantic space and how consistently their associated words cluster together.

Big Five Aspect	PC1 (Mean $\pm$ Std)	PC2 (Mean $\pm$ Std)	PC3 (Mean $\pm$ Std)
Agreeableness_Compassion	0.4679 $\pm$ 0.5114	0.2617 $\pm$ 0.3207	-0.1338 $\pm$ 0.3117
Agreeableness_Politeness	1.0900 $\pm$ 0.6036	0.6824 $\pm$ 0.4126	0.3596 $\pm$ 0.3142
Conscientiousness_Industriousness	0.1805 $\pm$ 0.4555	-0.3677 $\pm$ 0.3391	0.5918 $\pm$ 0.5711
Conscientiousness_Orderliness	0.0346 $\pm$ 0.3865	-0.6832 $\pm$ 0.2815	0.6544 $\pm$ 0.3697
Extraversion_Assertiveness	-0.2526 $\pm$ 0.4429	-0.0583 $\pm$ 0.4122	0.1563 $\pm$ 0.4776
Extraversion_Enthusiasm	0.2065 $\pm$ 0.1998	0.1453 $\pm$ 0.2456	-0.7446 $\pm$ 0.3690
Neuroticism_Volatility	-0.8064 $\pm$ 0.4568	0.6265 $\pm$ 0.4731	-0.3629 $\pm$ 0.4888
Neuroticism_Withdrawal	-1.0067 $\pm$ 0.3205	0.7759 $\pm$ 0.2196	0.3006 $\pm$ 0.3019
Openness_Experience	-0.1412 $\pm$ 0.4316	-0.8940 $\pm$ 0.4150	-0.5901 $\pm$ 0.4175
Openness_Intellect	0.2526 $\pm$ 0.4899	-0.4391 $\pm$ 0.3241	-0.0996 $\pm$ 0.5237

aspects were successfully recreated and visualized in a 3D vector space. A principal component analysis (PCA) and K-means clustering were employed to analyze and visualize the relationships among personality-descriptive adjectives. Clustering and PCA enable exploratory analysis, allowing us to uncover latent patterns and relationships in the data without pre-existing labels. Quantitative evaluation through a one-way ANOVA demonstrated statistically significant results for all of three PCA dimensions. These findings suggest that the principal components reflect interpretable psychological dimensions, mostly consistent with traditional personality research (Goldberg, 1992; McCrae and John, 1992).

Word embeddings enable the identification of semantic patterns that are not easily captured by static mappings of adjectives to personality traits. For instance, adjectives such as *spirited* and *warm* – related to the aspects Enthusiasm (Extraversion) and Compassion (Agreeableness) – cluster closely, reflecting their shared semantic connotations. Similarly, *moody* (low Extraversion) and *irritable* (Volatility aspect of Neuroticism) are proximate, highlighting overlapping associations with mood variability.

This study demonstrated that word embeddings can effectively capture the semantic structure of the Big Five personality traits, with clustering and PCA revealing meaningful relationships between adjectives. While it is expected that these adjectives would group according to their original psychometric categories, our findings provide an unsupervised validation of personality trait associations as they emerge from naturally occurring language use rather than self-assessment data. This approach highlights the potential for exploring personality traits in corpora where traditional survey methods are not applicable, such as historical texts, literary works, or social discourse.

The approach offers potential for exploring personality traits in a range of humanities contexts, such as character analysis in literature, trait evolution, and comparative analyses across texts.

6 Future Directions and Applications

In future work we want to validate these findings by comparing results with randomly sampled adjectives to ensure that clustering is not an artifact of the preselected lists.

We also aim to expand the list of personality-descriptive adjectives to include a broader and more comprehensive set of terms. This would allow to inspect if one should move towards enhancing the granularity of trait analysis and whether this might provide richer insights into personality dimensions. A larger, more inclusive list could also mitigate biases in current adjective sets, which may not fully capture the diversity of language use across different contexts. For example, applying an expanded adjective list to literary texts could reveal nuanced personality profiles of characters.

Calculating embeddings directly from domain-specific textual corpora, rather than relying solely on pre-trained models like Google-News-300, would allow for a more accurate and contextsensitive analysis. This might allow to explore the portrayal of personality traits across genres (e.g., literature, political rhetoric, historical correspondence) or time periods. For example, word embeddings can be used to trace diachronic semantic shifts, revealing how adjectives like *noble* or *ambitious* have changed in meaning over centuries. Such analyses align with prior work on semantic change (Hamilton et al., 2016; Dubossarsky et al., 2017) and provide insights into broader cultural and societal transformations. For instance, a diachronic study comparing political speeches from different eras could highlight shifts in the use of adjectives associated with traits like Confidence (Extraversion) or Conscientiousness, reflecting changing norms in political communication. Or embeddings derived from historical correspondence, such as letters exchanged between suffragettes, might reveal how rhetorical styles evolved during moments of activism. Traits like Politeness (Agreeableness) and Assertiveness (Extraversion) could be mapped to demonstrate how individuals adapted their language to align with social norms or achieve persuasive goals. Also, it would be interesting investigating how linguistic means besides adjectives might correlate with personality traits (see, e.g., Degaetano-Ortlieb et al. (2021) on registerial adaptation vs innovation across linguistic levels for women of the 18th century during periods of cultural transformation).

Considering visualization, future work could also focus on visualizing how personality traits evolve over time within narratives or rhetorical contexts. Segmenting texts temporally allows for the tracking of shifts in personality descriptors, providing dynamic insights into the development of traits. For example, analyzing political speeches segmented by key events could uncover shifts in rhetorical strategies. Traits like Compassion (Agreeableness) may dominate during times of national crisis, while Assertiveness (Extraversion) might be more prominent in competitive electoral campaigns. Such visualizations would offer a compelling view of how traits fluctuate in response to external pressures.

7 Limitations

Despite its advantages, the application of word embeddings has certain limitations. Adjectives with context-dependent meanings may pose challenges, as static embeddings lack the ability to account for sentence-level nuances. For instance, the word *reserved* might align with Introversion in one context but with Conscientiousness in another. While word embeddings capture general semantic relationships effectively, they may fail to handle such variability with precision. Contextualized embeddings, such as those produced by BERT, could address this limitation by incorporating sentence-level context, but their computational demands are significantly higher. While limitations such as the inability to capture contextual nuance remain, this first attempt can offer a substantial improvement over static adjective-to-trait mapping, bringing quantitative rigor to the study of personality traits in text. Future work integrating contextual embeddings may further enhance the capacity to analyze complex and nuanced textual data.

References

Oshin Agarwal, Funda Durupınar, Norman I. Badler, and Ani Nenkova. 2019. Word embeddings (also) encode human personality stereotypes. In *Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (* SEM 2019)*, pages 205–211.

Hussain Ahmad, Muhammad Zubair Asghar, Alam Sher Khan, and Anam Habib. 2020. A systematic literature review of personality trait classification from textual content. *Open Computer Science*, 10(1):175–193.

Gordon W. Allport and Henry S. Odbert. 1936. Trait-names: A psycho-lexical study. *Psychological Monographs*, 47(1):171.

Giulio Carducci, Giuseppe Rizzo, Diego Monti, Enrico Palumbo, and Maurizio Morisio. 2018. Twitpersonality: Computing personality traits from tweets using word embeddings and supervised learning. *Information*, 9(5):127.

Paul T. Costa and Robert R. McCrae. 1992. *NEO PI-R Professional Manual*. Psychological Assessment Resources, Odessa, FL.

Stefania Degaetano-Ortlieb, Tanja Säily, and Yuri Bizzoni. 2021. Registerial adaptation vs. innovation across situational contexts: 18th century women in transition. *Frontiers in Artificial Intelligence*, 4.

Colin G. DeYoung, Lena C. Quilty, and Jordan B. Peterson. 2007. Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, 93(5):880–896.

John M. Digman. 1990. Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41:417–440.

Haim Dubossarsky, Daphna Weinshall, and Eitan Grossman. 2017. Outta control: Laws of semantic change and inherent biases in word representation models. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1136–1145, Copenhagen, Denmark. Association for Computational Linguistics.

Lewis R. Goldberg. 1992. The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4:26–42.

Lewis R. Goldberg. 1993. The structure of phenotypic personality traits. *American Psychologist*, 48(1):26–34.

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.

K. L. Jang, W. J. Livesley, A. Angleitner, R. Reimann, and P. A. Vernon. 2002. Genetic and environmental influences on the covariance of facets defining the domains of the five-factor model of personality. *Personality and Individual Differences*, 33:83–101.

Oliver P. John, Laura P. Naumann, and Christopher J. Soto. 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In *Handbook of Personality: Theory and Research*, pages 114–158. Guilford Press.

John A. Johnson. 1994. Clarification of factor five with the help of the AB5C model. *European Journal of Personality*, 8:311–334.

Amirmohammad Kazameini, Samin Fatehi, Yash Mehta, Sauleh Eetemadi, and Erik Cambria. 2020. Personality Trait Detection Using Bagged SVM over BERT Word Embedding Ensembles. *arXiv preprint arXiv:2010.01309*.

Robert R. McCrae and Oliver P. John. 1992. An introduction to the five-factor model and its applications. *Journal of Personality*, 60:175–215.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *Preprint*, arXiv:1301.3781.

Ayoub Ouarka, Tarek Ait Baha, Youssef Es-Saady, and Mohamed El Hajji. 2024. A deep multimodal fusion method for personality traits prediction. *Multimedia Tools and Applications*, pages 1–23.

Daniele Pizzolli and Carlo Strapparava. 2019. Personality traits recognition in literary texts. In *Proceedings of the Second Workshop on Storytelling*, pages 107–111, Florence, Italy. Association for Computational Linguistics.

Emily Pronin and Matthew B. Kugler. 2007. Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology*, 43(4):565–578.

Farhad Bin Siddique, Dario Bertero, and Pascale Fung. 2019. Globaltrait: Personality alignment of multilingual word embeddings. In *Proceedings of the aaai conference on artificial intelligence*, volume 33, pages 7015–7022.

Marina Tiuleneva, Vadim A. Porvatov, and Carlo Strapparava. 2024. Big-five backstage: A dramatic dataset for characters personality traits & gender analysis. In *Proceedings of the Workshop on Cognitive Aspects of the Lexicon @ LREC-COLING 2024*, pages 114–119, Torino, Italia. ELRA and ICCL.

Ernest C. Tupes and Raymond E. Christal. 1961. Recurrent personality factors based on trait ratings. Technical Report ASD-TR-61-97, Lackland Air Force Base, TX.

A List of Adjectives used for Big Five Dimensions and Aspects

Openness *Openness to Experience*: imaginative, creative, original, artistic, inventive, innovative, curious, insightful, visionary, experimental
Intellect: intelligent, intellectual, clever, analytical, philosophical, reflective, rational, knowledgeable, thoughtful, brainy

Conscientiousness *Orderliness*: organized, neat, tidy, systematic, meticulous, precise, methodical, orderly, well-organized, structured
Industriousness: efficient, hardworking, diligent, responsible, reliable, productive, persevering, ambitious, thorough, goal-oriented

Extraversion *Enthusiasm*: energetic, enthusiastic, lively, cheerful, spirited, vivacious, fun-loving, joyful, playful, exuberant
Assertiveness: assertive, bold, confident, dominant, forceful, outspoken, persuasive, self-assured, determined, decisive

Agreeableness *Compassion*: compassionate, kind, caring, warm, gentle, empathetic, altruistic, supportive, nurturing, loving
Politeness: polite, courteous, respectful, considerate, tactful, gracious, well-mannered, civil, deferential, humble

Neuroticism *Volatility*: temperamental, moody, irritable, touchy, unstable, impulsive, excitable, fickle, changeable, fluctuating
Withdrawal: anxious, fearful, nervous, insecure, self-conscious, worrying, pessimistic, vulnerable, tense, timid

B Prompt to Generate Adjectives

I am conducting a study about the Big Five personality model, where I want to use word embeddings instead of the traditional factor analyses to display the clustering of the personality dimensions. I extracted 50 adjectives from Tupes and Christal (1961), out of which 5 adjectives were extracted for each of the 10 aspects of the Big Five dimensions. You will compile 50 additional Big Five adjectives that are also evenly distributed across the 10 aspects of the Big Five so that we will end up with a list of 100 adjectives in total.