Check for
updates

Linguistic Features and Psychological States: The Case of Virginia Woolf

Xiaowei Du*

Department of Foreign Language, Huazhong University of Science and Technology, Wuhan, China

This study investigated the relation between psychological states and linguistic features with the case of Virginia Woolf. We analyzed the data from *The Diary of Virginia Woolf* and *Virginia Woolf: Biography* by automatic text analysis and statistical analysis, including stepwise multiple regression and Deep Learning algorithm. The results suggested that the significant linguistic features can jointly predict the psychological states of Virginia Woolf, including the emotional value of *anger*, the absolutist word "*everything*," and the total of first-person plural pronouns. In addition, we found that the total use of first-person plural pronouns and the emotional value of *anger* were negatively related to mental health of Virginia Woolf. While the use of the absolutist word "*everything*" was positively related to mental health of Virginia Woolf. Meanwhile, we developed a model that can predict the psychological states of Virginia Woolf, with 86.9% accuracy. We discussed the findings and enumerated the limitations of this study at the end of the paper. The results not only complemented previous studies in the understanding of the relation between language and psychological health, but also facilitated timely identification, intervention, and prevention of mental disorders.

Keywords: psychological states, linguistic features, prediction, mental disorders, prevention

INTRODUCTION

The relation between psychological states and linguistic features recently has attracted the interest of researchers. The motivation of such a line of research is that the psychological states of a person can be examined by his/her linguistic features (Nguyen et al., 2017) since the language of a person seems a reliable device to interpret his/her internal thoughts, emotions, and feelings (Herbert et al., 2019).

Some studies have explored the relation between psychological states and linguistic features. These studies are roughly categorized into two types. First, psychological states are assessed by the way people write about their experiences (Barnes et al., 2007; Tausczik and Pennebaker, 2010; Al-Mosaiwi and Johnstone, 2018; Kim et al., 2019). Second, psychological states are predicted by factors such as the diachronic changes of linguistic features in written texts (Rodrigues et al., 2016; Ziemer and Korkmaz, 2017; Eichstaedt et al., 2018; Boukil et al., 2019).

However, mixed findings were obtained regarding the relation between linguistic features and psychological states. First, psychological states are found developing closely related with different linguistic feature. For example, Barnes et al. (2007) found that the suicide texts reflect a linguistic trend in negative emotion words and death words. In contrast, Kim et al. (2019) showed a linguistic

trend in modifier, numerals, first and second person pronouns, emotion words (positive, negative, sadness, and depression-related), and future tense verbs in the suicide texts. Second, the relation between linguistic features found in texts and psychological states are inconsistent. For example, Tausczik and Pennebaker (2010) found that people who consider suicide express more negative emotional words. Another example showed an opposite result that people who consider suicide express less negative emotional words (Barnes et al., 2007).

Therefore, the purpose of this study is to explore the relation between psychological states and linguistic features with the case of Virginia Woolf. The findings of this study may examine and complement previous studies and provide an approach to early identification, intervention, and prevention of mental disorders.

LITERATURE REVIEW

In this section, we introduce linguistic features that have been used to recognize psychological states in previous studies and a computational method that has been used to analyze sentiments and emotions.

Linguistic Features and Psychological States

Recent studies have shown that linguistic features such as personal pronouns, emotion words, absolutist words, color words, word count, and question marks may be used to identify psychological states.

The use of personal pronouns has served as a promising indicator of psychological states (Tausczik and Pennebaker, 2010). For example, Campbell and Pennebaker (2003) suggested that personal pronouns manifest psychological states from the perspective of a person's social connection or social isolation. Specifically, an excessive use of first-person singular pronouns may be related to a high degree of self-involvement, while an increased use of the other pronouns may indicate improvement of social engagement (Cohn et al., 2004; Simmons et al., 2008). In addition, people use more first-person singular pronouns when in grief or depression or attempting suicide (Rude et al., 2004; Boals and Klein, 2005; Eichstaedt et al., 2018). Last, it is also found that fewer first-person singular pronouns may involve deceptive communications (Newman et al., 2003), while more first-person plural pronouns may indicate that he or she was in a happy marriage (Simmons et al., 2008).

Similarly, the use of emotional words is also correlated with psychological states (Barnes et al., 2007), though mixed findings were found regarding the relation between emotions and psychological states. The findings can be roughly categorized into three types. First, both positive and negative emotional words are related with mental health, and more negative emotional words and less positive emotional words reflect a less healthy mental state (Pennebaker et al., 2003; Kahn et al., 2007). Second, only negative emotional words are related with mental health, and more negative emotional words reflect a less healthy mental health (Kahn et al., 2007; Herbert et al., 2019). Third, more emotional expressions improve psychological states, such as

positive emotional words, anger, or sadness (Rude et al., 2004; Graves et al., 2005; Barnes et al., 2007). It is worth noting that some studies stressed the impact of negative emotional expression on psychological states. Previous studies found that negative emotional words not only carry more information about mental health than positive emotional words (Garcia et al., 2012) but may also benefit psychological states (Lerner et al., 2003).

Absolutist words have also been employed to predict psychological states (Savekar et al., 2019) since absolutist words are believed to present an absolutist thinking (Al-Mosaiwi and Johnstone, 2018). Empirical studies have revealed that the absolutist thinking, a cognitive distortion related to extreme and rigid thoughts, may do harm to mental health (Savekar et al., 2019; Jones et al., 2020). More importantly, compared with negative emotional words, absolutist words may more accurately track the severity of the affective disorder (Al-Mosaiwi and Johnstone, 2018). To be specific, the language of suicidal ideation contains approximately 30% more absolutist words than that of anxiety and depression, and approximately 80% more than that of normal mental health (Al-Mosaiwi and Johnstone, 2018).

Other linguistic features such as color words, question marks, and word counts, may also be used to predict psychological states (Barnes et al., 2007; Tausczik and Pennebaker, 2010; Wadsworth et al., 2016). For example, Wadsworth et al. (2016) found that the use of *white* in Sylvia Plath's poems increased significantly while the use of *gray*, *yellow*, *purple*, and *brown* decreased rapidly before Sylvia Plath committed suicide. Furthermore, Barnes et al. (2007) analyzed letters and diaries of suicidal youth and found that the use of question marks increased rapidly before their suicide. Last, some studies indicated that word counts may be related to psychological states since it reflects how engaged people are in the expression or how much information people produced (Alvarez-Conrad et al., 2001; Tausczik and Pennebaker, 2010; Wadsworth et al., 2016).

Sentiment Analysis and Psychological States

Sentiment analysis has recently been used to study psychological states since it could take a step closer to understanding and anticipating an individual's physiological states and needs (Al-Thubaity et al., 2018; Chatterjee et al., 2019; Moreno-Blanco et al., 2020). To be specific, sentiment analysis quantifies or extracts the polarity of sentiments, attitudes, emotions, and opinions of a given text (Cambria, 2016; Rendalkar and Chandankhede, 2018). More importantly, sentiment analysis broadly covers two dimension, namely, sentiment analysis in a narrow sense and emotion analysis (Lei and Liu, 2021). On the one hand, sentiment analysis in a narrow sense mainly identifies the sentiment polarities such as positive, negative, or neutral (D'Andrea et al., 2015; Lennox et al., 2020; Zucco et al., 2020). Some studies have applied sentiment analysis in a narrow sense to identify psychological states (Tausczik and Pennebaker, 2010). For example, Herbert et al. (2019) found that people express more negative emotions in their notes before committing suicide. Similarly, Eichstaedt et al. (2018) found that depressed people express more negative emotions, which

TABLE 1 | The classification of variables.

The dependent variable	Psychological states	*1 normal, 2 recovery, 3 restlessness, anxiety, and irritability 4 physical discomfort, 5 mental illness*
The independent variables	Colors	*red, white, black, green, yellow, blue, purple, gray*
	Personal pronouns	*i, my, me, we, us, our, she, her, he, him, his, it, its, them, their, they, you, your, yours, i_my_me, we_our_us, you_your_yours, he_his_him, she_her, it_its, they_them_their*
	Absolutist word	*absolutely, all, always, entire, ever, every, everyone, everything, definitely, full, must, complete, completely, constant, constantly, whole never, nothing, totally, total absolutist words*
	Emotions	*anticipation, sadness, joy, trust, disgust, anger, surprise, fear*
	Question mark	
	Word counts	
	Sentiment value	

contributed negatively to their psychological states. In addition, Rodrigues et al. (2016) detected positive or negative emotions of cancer patients by analyzing their texts in online communities to support their treatment.

On the other hand, emotion analysis focuses on recognizing a set of basic emotions, which may extend a more accurate and reliable detection of psychological states and feelings (Calabrese and Cannataro, 2016; Molina Beltrán et al., 2019). Previous studies have developed several versions of basic emotions. Specifically, Ekman (1993) proposed six universal emotions including anger, disgust, fear, sadness, joy, and surprise. Besides, Plutchik (1980) developed eight basic emotions such as anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. Some studies have investigated the relation between basic emotions and psychological states (Desmet and Hoste, 2013; Ciullo et al., 2016). For instance, suicide individuals tend to express more emotional words such as anxiety, sadness, and depression in response to their feelings of anger or anxiety (Kim et al., 2019). Meanwhile, the use of joy or happiness words, revealing a sense of enjoyment, satisfaction, and pleasure, and these words are frequently used when an individual is in the situation of wellbeing, inner peace, love, safety, and contentment (Papapicco and Mininni, 2020). Additionally, the use of sadness words reflects the degree of social withdrawal or mood flattening, occurring with a higher frequency when an individual is most likely in grief, loss, frustration, depression, and suicide ideation (Barnes et al., 2007; Eichstaedt et al., 2018; Kim et al., 2019).

In fact, many studies have explored the relation between psychological states and linguistic features. However, a few limitations should be noted. First, most studies applied commercial tools such as Linguistic Inquiry and Word Count Program (LIWC) to analyze the psychological states of texts, which may limit the application of their findings in that such tools are not freely available. More importantly, these studies can only explore the linguistic features provided by LIWC, which may limit the comprehensiveness of the research. Second, few studies have examined the relation between psychological states and other linguistic features besides personal pronouns and positive and negative emotions (Wadsworth et al., 2016; Al-Mosaiwi and Johnstone, 2018). Third, previous studies yielded conflicting findings, such as different linguistic features that are related to psychological states (Eichstaedt et al., 2018; Kim et al., 2019), and a different relation between linguistic features and psychological states (Barnes et al., 2007; Tausczik and Pennebaker, 2010).

Therefore, this study aims to address the foregoing limitations and examine the relation between linguistic features and psychological states with the case of Virginia Woolf. More specifically, we use *Python* and *R* (version 3.6.0) to analyze the psychological states of texts, which can consider more linguistic features. In addition, this study examine linguistic features that are related to psychological states of Virginia Woolf. We also confirm the relation between linguistic features and the psychological states of Virginia Woolf. Finally, we aim to provide an equation to predict the psychological states of Virginia Woolf by linguistic features.

MATERIALS AND METHODS

In this section, we introduce the data, the methods for text analysis, and the statistical analysis in this study.

Data

The Diary of Virginia Woolf (Bell, 1984) was used for text analysis in the present study for three reasons. First, Virginia Woolf was a British novelist and essayist regarded as one of the major modernist literary figures of the twentieth century (Boeira et al., 2016). Second, she had mental disorders and attempted suicide several times, and killed herself at the age of 59 (Androutsopoulou et al., 2019). Third, *The Diary of Virginia Woolf* (Bell, 1984) contains 1,577 diaries in total and covers 26 years, from 1915 until 4 days before death in 1941, except in 1916, Virginia Woolf didn't write a diary due to a serious breakdown (Blodgett, 1989). It is a private text of self-presentation and provides more comprehensive and affluent information to reveal the relation between linguistic features and psychological states (Briggs, 2011; Androutsopoulou et al., 2019). We saved each diary piece as one text and obtained 1,577 texts in total.

Virginia Woolf: A Biography (Bell, 1972) was used in this study for three reasons. First, this book is more reliable than other biographies about Virginia as the writer is the nephew of Virginia Woolf. Second, the chronology record of this book is complete, from Virginia Woolf's birth (1882) to her death (1912). More importantly, it is a chronology that records the stages of composition of Virginia Woolf's books, her daily life, and especially her illness. We found the psychological states of Virginia Woolf are recorded including *normal*, *recovery*, *restlessness*, *anxiety*, *irritability*, *physical discomfort*, and *mental*

illness. Thus, we extracted the information on psychological states of Virginia Woolf and the corresponding time for analysis.

Data Search

In this study, we explored some linguistic features as shown in **Table 1**.

We calculated the sentiment and emotion values of each text with Jockers (2015) *syuzhet*, a lexicon-based sentiment analysis package in R (version 3.6.0). To help make the results more reliable, we first calculated the sentiment and emotion values at sentence levels and then computed the mean values of each text as the final value of the text. We chose the *Bing* lexicon (Liu, 2012) to calculate sentiment values and the *NRC* lexicon (Mohammad and Turney, 2013) to calculate the emotion values for the reason that the lexica contain comprehensive lists of sentiment and emotion words and have been widely used in sentiment and emotion research (Calabrese and Cannataro, 2016). With the *Bing* lexicon (Liu, 2012), each sentence was assigned a sentiment value, while each sentence was assigned emotion values from eight perspectives, i.e., anger, anticipation, disgust, fear, joy, sadness, surprise, and trust.

Besides the sentiment and emotion values, we also calculated other linguistic features of each text, such as the number of words, absolutist words, color words, personal pronouns, and question marks. First, following Al-Mosaiwi and Johnstone (2018), we counted the frequency of 19 absolutist words of a text. Second, we counted the frequency of all personal pronouns. It should be noted that we combined the frequency of subject personal pronoun, the object personal pronoun, and the adjective subject pronoun of each personal pronoun of a text (e.g., *i_my_me*) due to the very low occurrence of the pronouns. Third, following Wadsworth et al. (2016), we counted the frequency of each color word of a text, including *red*, *white*, *black*, *green*, *yellow*, *blue*, *purple*, and *gray*. The calculation of all the foregoing linguistic features was performed with a homemade Python script.

Last, due to different text lengths, we normalized the raw frequency of the linguistic features with the following formula.

$$Normalized\ frequency = \frac{Raw\ frequency}{Number\ of\ words\ in\ the\ text} \times 1,000$$

Data Analysis

The stepwise multiple linear regression analysis was conducted to identify the relation between linguistic features and psychological state of Virginia Woolf, which was implemented with Venables and Ripley's (2002) *MASS*, a powerful package for the statistical and graphical analysis in R (version 3.6.0). To be specific, the independent variables are the linguistics features of texts we extracted. Besides, the dependent variable is the psychological states of Virginia Woolf, which were expressed in numbers for identification and calculation (e.g., 1 represents normal). Our reasons for choosing the stepwise multiple linear regression are listed as follows. First, it identifies the independent variables that can serve as the influential predictors to identify the dependent variable, with the inclusion set at $P < 0.05$. More importantly, it also gains insight into the correlation

between the influential predictors and the dependent variable. A positive value of regression coefficient presents a positive relation, and vice versa.

Then, we used the significant linguistic features to build a model for identifying psychological state of Virginia Woolf based on Deep Learning algorithms with the *RapidMiner Studio* (the educational version 9.7). We did so for the following reasons.

First, Deep Learning, as an advanced machine learning algorithm, involves in deciphering the hidden and more complex but meaningful phenomena within the data that the traditional statistical methods struggled to show (Wongso et al., 2017). In simple words, *Deep Learning* may build an automatic but more effective model for identifying psychological state of Virginia Woolf. Second, we can yield an objective evaluated criteria, namely, *accuracy*, with higher values for a better performance (Savoy, 2020).

In conclusion, we determined the significant linguistic features in identifying psychological state of Virginia Woolf, examined their relations, respectively, and built and evaluated the model for identifying the psychological state of Virginia Woolf.

RESULTS

The results show some interesting findings. First, we found the significant linguistic features that jointly predicted the psychological states of Virginia Woolf after examining a total of 65 linguistic features, respectively, in the stepwise multiple regression analysis. The significant linguistic features were the total usage of first-person plural pronouns, i.e., "we_our_us," the emotional value of *anger*, and the absolutist word "*everything.*" The other linguistic features of this study had no significant effect on predicting the psychological states of Virginia Woolf (*p*-values were larger than 0.05).

Second, we found some correlations between the significant linguistic features and mental health of Virginia Woolf. Specifically, the total use of first-person plural pronouns (Coefficients = −0.007) and the use of emotional value of *anger* (Coefficients = −0.007) were negatively related to mental health of Virginia Woolf. In contrast, the use of the absolutist word "everything" (Coefficients = 0.074) was positively related to mental health of Virginia Woolf.

Last, we used the significant linguistic features to build a model for identifying psychological state of Virginia Woolf, which yielded 86.9% accuracy with Deep Learning algorithms.

DISCUSSION

Several points should be noted based on the results of the present study. First, the total use of first-person plural pronouns correlated negatively with the psychological states of Virginia Woolf. The reasons are listed as follows. First, the use of first-person plural pronouns serves as a mark of social connection, inclusiveness, and belongingness (Pennebaker et al., 2003). Second, fewer first-person plural pronouns serve as evidence

of greater self-involvement and selfishness (Pennebaker et al., 2005). Thus, it was the degree of connection and belongingness reflected the psychological states of Virginia Woolf with the total use of first-person plural pronouns. Inconsistent with some previous research, the total use of first-person plural pronouns outperformed than other personal pronouns in predicting psychological states of Virginia Woolf. The result is most likely due to the property of diaries, Virginia Woolf always concerned and recorded her personal activities, thoughts, and feelings. Thus, the use of first-person singular pronouns occurred with a higher frequency but little variation; the use of second and third personal pronouns occurred with a lower frequency; only the use of first-person plural pronouns occurred with a higher frequency and variation. In addition, the total use of first-person plural pronouns highlighted the fluctuations in variation and ensured a significant statistical analysis.

Second, the use of anger words correlated negatively with the psychological states of Virginia Woolf. Many researchers have studied the relation between emotions and psychological states since emotions are one of the key aspects that characterize many mental health conditions (Pennebaker et al., 2003, 2005). However, mixed findings were obtained regarding the relation between emotions and psychological states. For example, Pennebaker et al. (2003) found a linear relation between positive emotion and mental health, and a curvilinear relation between negative emotion and mental health. Some studies found a negative relation between negative emotion and mental health (Kahn et al., 2007; Herbert et al., 2019). They stressed the negative emotions have a detrimental impact on psychological states. Thus, this study complements previous studies, which proved that the anger expression could predict and benefit the psychological states of Virginia Woolf. The reasons are listed as follows. First, anger is an interpersonal feeling, which typically emerges in a close relation (Fehr et al., 1999). Virginia Woolf was involved in a close connection with others and expressed anger, which is beneficial to her psychological health. In addition, angry expression triggers more optimism and shows a higher expectation for the future (Lerner et al., 2003). Last, anger expression helps people realize mistakes and promote problem-solving (Aarts et al., 2010).

Third, the absolutist word "everything" correlated positively with the prediction of the psychological states of Virginia Woolf. Consistent with previous research, absolutist thinking performs an unhealthy and inflexible way of thinking with more absolutist words (Savekar et al., 2019). As expected, the thought patterns of unhealthy psychological individuals are concerned with the feeling of absolutism. It might also signify that absolutist thinking harmed the psychological health of Virginia Woolf. The reasons are listed as follows. First, absolutist thinking disrupts emotion regulation, promotes emotional distress, and make people vulnerable to poor psychological and physical health (Al-Mosaiwi and Johnstone, 2018). Moreover, absolutist

people are often perfectionists who see their values, goals, and outcomes as being right (Antoniou et al., 2017). They prefer anger, self-blame, and deprecatory thoughts (Savekar et al., 2019). It is of interest to note that the absolutist word "everything" outperformed than other absolutist words we extracted in predicting psychological states of Virginia Woolf. The possible reason is that the absolutist word "everything" represents an overly dichotomous thought. The dichotomous thought lacks tolerance and compromise, which may do harm to mental health (Jones et al., 2020). Moreover, "everything" served as an indefinite pronoun or premodifier in the diary of Virginia Woolf. Thus, the diachronic change of usage of "everything" was more significant than of other absolutist words.

Last, our study verifies the value of the proposed model for predicting the psychological states of Virginia Woolf. Consistent with previous studies, the results proved that a change of psychological states affects the words people used, and the words people used convey a great deal of information about their psychological health. More importantly, our study, similar to ample research, suggested that the model based on linguistic features for predicting psychological states may become increasingly feasible and more accurate. For example, Boukil et al. (2019) proposed an automatic system based on Deep learning algorithm to predict suicidal ideation through analyzing sentiment and feelings expressed on social media. Similarly, Nguyen et al. (2017) extract linguistic features and topics to discriminate depressed online communities from other groups based on machine learning algorithms. Particularly, it yielded 77.6% accuracy in the binary classification of depression vs. Bipolar by Lasso algorithms. Meanwhile, Eichstaedt et al. (2018) built a model with some linguistic predictors to identify depressed *Facebook* users, with a higher prediction accuracy (AUC = 0.72). In addition, de Ávila Berni et al. (2018) developed a model to identify texts proposed by suicidal individuals based on the Naïve-Bayes machine-learning algorithm. The model achieved a higher performance, with an accuracy of 80%. In all, we can develop our proposed model into a system or tool for assessing mental health, monitoring mental disorder, and preventing suicide.

CONCLUSION

In this study, we explored the relation between linguistic features and the psychological states of Virginia Woolf, which generated three findings. First, the result confirmed the total use of first-person plural pronouns, the emotional value of *anger*, and the absolutist word "*everything*" can jointly predict the psychological states of Virginia Woolf. Second, we provided an effective model to predict psychological

states of Virginia Woolf. We can further applied the model in various areas such as interventions, therapeutic protocols, and suicide preventions.

Two limitations of this study should be noted. First, we should consider more factors that affect psychological health to improve the explanatory power and usefulness of our equation, such as other linguistic features, social, environmental, economic, and political contexts. Second, we should examine whether our findings are generalizable in any way with more replicative studies on different people, language, and materials.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

XD: investigation, formal analysis, methodology, validation, and writing–review and editing.

REFERENCES

Aarts, H., Ruys, K. I., Veling, H., Renes, R. A., de Groot, J. H. B., van Nunen, A. M., et al. (2010). The art of anger: reward context turns avoidance responses to anger-related objects into approach. *Psychol. Sci.* 21, 1406–1410. doi: 10.1177/0956797610384152

Al-Mosaiwi, M., and Johnstone, T. (2018). In an absolute state: elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clin. Psychol. Sci.* 6, 529–542. doi: 10.1177/2167702617747074

Al-Thubaity, A., Alharbi, M., Alqahtani, S., and Aljandal, A. (2018). "A Saudi Dialect Twitter Corpus for Sentiment and Emotion Analysis," in *2018 21st Saudi Computer Society National Conference (NCC)*, (Piscataway: IEEE), 1–6. doi: 10.1109/NCG.2018.8592998

Alvarez-Conrad, J., Zoellner, L. A., and Foa, E. B. (2001). Linguistic predictors of trauma pathology and physical health. *Appl. Cogn. Psychol.* 15, S159–S170. doi: 10.20971/kcpmd.2007.50.2.1

Androutsopoulou, A., Rozou, E., and Vakondiou, M. (2019). Voices of hope and despair: a narrative-dialogical inquiry into the diaries, letters, and suicide notes of Virginia Woolf. *J. Constr. Psychol.* 33, 367–384. doi: 10.1080/10720537.2019.1615015

Antoniou, E. E., Bongers, P., and Jansen, A. (2017). The mediating role of dichotomous thinking and emotional eating in the relationship between depression and BMI. *Eating Behav.* 26, 55–60. doi: 10.1016/j.eatbeh.2017.01.007

Barnes, D. H., Lawal-Solarin, F. W., and Lester, D. (2007). Letters from a suicide. *Death Stud.* 31, 671–678. doi: 10.1080/07481180701405212

Bell, A. O. (1984). *The Diary of Virginia Woolf (1936-1941)*. San Diego: Harvest.

Bell, Q. (1972). *Virginia Woolf: A Biography*. New York: Harcourt Brace Jovanovich.

Blodgett, H. (1989). A woman writer's diary: Virginia woolf revisited. *Prose Stud.* 12, 57–71. doi: 10.1080/01440358908586360

Boals, A., and Klein, K. (2005). Word use in emotional narratives about failed romantic relationships and subsequent mental health. *J. Lang. Soc. Psychol.* 24, 252–268. doi: 10.1177/0261927X05278386

Boeira, M. V., Berni, G. Á, Passos, I. C., Kauer-Sant'Anna, M., and Kapczinski, F. (2016). Virginia Woolf, neuroprogression, and bipolar disorder. *Braz. J. Psychiatry* 39, 69–71. doi: 10.1590/1516-4446-2016-1962

Boukil, S., El Adnani, F., Cherrat, L., El Moutaouakkil, A. E., and Ezzyyani, M. (2019). "Deep learning algorithm for suicide sentiment prediction," in *Advanced Intelligent Systems for Sustainable Development (AI2SD'2018)*, ed. M. Ezzyyani (Berlin: Springer International Publishing), doi: 10.1007/978-3-030-11884-6_24

Briggs, K. (2011). The making of A Writer's Diary: being extracts from the diary of Virginia Woolf. *Textual Pract.* 25, 1033–1050. doi: 10.1080/0950236x.2011.618459

Calabrese, B., and Cannataro, M. (2016). "Sentiment analysis and affective computing: Methods and applications," in *Brain-Inspired Computing*, eds K. Amunts, L. Grandinetti, T. Lippert, and N. Petkov (Berlin: Springer International Publishing).

Cambria, E. (2016). Affective computing and sentiment analysis. *IEEE Intell. Syst.* 31, 102–107. doi: 10.1109/MIS.2016.31

Campbell, R. S., and Pennebaker, J. W. (2003). The secret life of pronouns flexibility in writing style and physical health. *Psychol. Sci.* 14, 60–65. doi: 10.1111/1467-9280.01419

Chatterjee, A., Gupta, U., Chinnakotla, M. K., Srikanth, R., Galley, M., and Agrawal, P. (2019). Understanding Emotions in Text Using Deep Learning and Big Data. *Comput. Hum. Behav.* 93, 309–317. doi: 10.1016/j.chb.2018.12.029

Ciullo, F., Zucco, C., Calabrese, B., Agapito, G., Guzzi, P. H., and Cannataro, M. (2016). "Computational challenges for sentiment analysis in life sciences," in *2016 International Conference on High Performance Computing & Simulation (HPCS)*, (Innsbruck: IEEE), 419–426. doi: 10.1109/HPCSim.2016.7568365

Cohn, M. A., Mehl, M. R., and Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychol. Sci.* 15, 687–693.

D'Andrea, A., Ferri, F., Grifoni, P., and Guzzo, T. (2015). Approaches, tools and applications for sentiment analysis implementation. *Int. J. Comput. Appl.* 125, 26–33. doi: 10.5120/ijca2015905866

de Ávila Berni, G., Rabelo-da-Ponte, F. D., Librenza-Garcia, D., Boeira, M. V., Kauer-Sant'Anna, M., Passos, I. C., et al. (2018). Potential use of text classification tools as signatures of suicidal behavior: a proof-of-concept study using Virginia Woolf's personal writings. *PLoS One* 13:e0204820. doi: 10.1371/journal.pone.0204820

Desmet, B., and Hoste, V. (2013). Emotion detection in suicide notes. *Expert Syst. Appl.* 40, 6351–6358. doi: 10.1016/j.eswa.2013.05.050

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preoțiuc-Pietro, D., et al. (2018). Facebook language predicts depression in medical records. *Proc. Natl. Acad. Sci.* 115, 11203–11208. doi: 10.1073/pnas.1802331115

Ekman, P. (1993). Facial expression and emotion. *Am. Psychol.* 48, 384–392.

Fehr, B., Baldwin, M., Collins, L., Patterson, S., and Benditt, R. (1999). Anger in close relationships: an interpersonal script analysis. *Pers. Soc. Psychol. Bull.* 25, 299–312. doi: 10.1177/0146167299025003003

Garcia, D., Garas, A., and Schweitzer, F. (2012). Positive words carry less information than negative words. *EPJ Data Sci.* 1:3. doi: 10.1140/epjds3

Graves, K. D., Schmidt, J. E., Bollmer, J., Fejfar, M., Langer, S., Blonder, L. X., et al. (2005). Emotional expression and emotional recognition in breast cancer survivors: a controlled comparison. *Psychol. Health* 20, 579–595. doi: 10.1080/0887044042000334742

Herbert, C., Bendig, E., and Rojas, R. (2019). My sadness – our happiness: writing about positive, negative, and neutral autobiographical life events reveals linguistic markers of self-positivity and individual well-being. *Front. Psychol.* 9:2522. doi: 10.3389/fpsyg.2018.02522

Jockers, M. (2015). *Syuzhet: Extract Sentiment and Plot Arcs from Text*. Available online at https://github.com/mjockers/syuzhet. (accessed on Apr 12, 2021).

Jones, L. S., Anderson, E., Loades, M., Barnes, R., and Crawley, E. (2020). Can linguistic analysis be used to identify whether adolescents with a chronic illness are depressed? *Clin. Psychol. Psychother.* 27, 179–192. doi: 10.1002/cpp.2417

Kahn, J. H., Tobin, R. M., Massey, A. E., and Anderson, J. A. (2007). Measuring emotional expression with the linguistic inquiry and word count. *Am. J. Psychol.* 120:263. doi: 10.2307/20445398

Kim, K., Choi, S., Lee, J., and Sea, J. (2019). Differences in linguistic and psychological characteristics between suicide notes and diaries. *J. Gen. Psychol.* 146, 391–416. doi: 10.1080/00221309.2019.1590304

Lei, L., and Liu, D. (2021). *Conducting Sentiment Analysis*, 1st Edn. Cambridge: Cambridge University Press.

Lennox, R. J., Veríssimo, D., Twardek, W. M., Davis, C. R., and Jarić, I. (2020). Sentiment analysis as a measure of conservation culture in scientific literature. *Conserv. Biol.* 34, 462–471. doi: 10.1111/cobi.13404

Lerner, J. S., Gonzalez, R. M., Small, D. A., and Fischhoff, B. (2003). Effects of fear and anger on perceived risks of terrorism: a national field experiment. *Psychol. Sci.* 14, 144–150. doi: 10.1111/1467-9280.01433

Liu, B. (2012). Sentiment analysis and opinion mining. *Synth. Lect. Hum. Lang. Technol.* 5, 1–167.

Mohammad, S. M., and Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Comput. Intell.* 29, 436–465. doi: 10.1111/j.1467-8640.2012.00460.x

Molina Beltrán, C., Segura Navarrete, A. A., Vidal-Castro, C., Rubio-Manzano, C., and Martínez-Araneda, C. (2019). Improving the affective analysis in texts: automatic method to detect affective intensity in lexicons based on Plutchik's wheel of emotions. *Electron. Libr.* 37, 984–1006. doi: 10.1108/EL-11-2018-0219

Moreno-Blanco, D., Ochoa-Ferreras, B., Gárate, F. J., Solana-Sánchez, J., Sánchez-González, P., and Gómez, E. J. (2020). "Evaluation and Comparison of Text Classifiers to Develop a Depression Detection Service," in *XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019*, eds J. Henriques, N. Neves, and P. de Carvalho (Berlin: Springer International Publishing), doi: 10.1007/978-3-030-31635-8_146

Newman, M. L., Pennebaker, J. W., Berry, D. S., and Richards, J. M. (2003). Lying words: predicting deception from linguistic styles. *Pers. Soc. Psychol. Bull.* 29, 665–675. doi: 10.1177/0146167203029005010

Nguyen, T., O'Dea, B., Larsen, M., Phung, D., Venkatesh, S., and Christensen, H. (2017). Using linguistic and topic analysis to classify sub-groups of online depression communities. *Multimed. Tools Appl.* 76, 10653–10676. doi: 10.1007/s11042-015-3128-x

Papapicco, C., and Mininni, G. (2020). Impact memes: PhDs HuMor(e). *Multimed. Tools Appl.* 79, 35973–35994. doi: 10.1007/s11042-020-09166-0

Pennebaker, J. W., Mehl, M. R., and Niederhoffer, K. G. (2003). Psychological aspects of natural language use: our words, our selves. *Ann. Rev. Psychol.* 54, 547–577. doi: 10.1146/annurev.psych.54.101601.145041

Pennebaker, J. W., Slatcher, R. B., and Chung, C. K. (2005). Linguistic markers of psychological state through media interviews: John Kerry and John Edwards in 2004, Al Gore in 2000. *Anal. Soc. Issues Public Policy* 5, 197–204. doi: 10.1111/j.1530-2415.2005.00065.x

Plutchik, R. (1980). "Theories of emotion," in *Emotion: Theory, Research and Experiences*, eds R. Plutchik and H. Kellerman (New York, NY: Academic Press).

Rendalkar, S., and Chandankhede, C. (2018). "Sarcasm detection of online comments using emotion detection," in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, (Piscataway: IEEE), 1244–1249. doi: 10.1109/ICIRCA.2018.8597368

Rodrigues, R. G., das Dores, R. M., Camilo-Junior, C. G., and Rosa, T. C. (2016). SentiHealth-Cancer: a sentiment analysis tool to help detecting mood of patients in online social networks. *Int. J. Med. Inform.* 85, 80–95. doi: 10.1016/j.ijmedinf.2015.09.007

Rude, S., Gortner, E.-M., and Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cogn. Emot.* 18, 1121–1133. doi: 10.1016/j.beth.2006.01.004

Savekar, A., Tarai, S., and Singh, M. (2019). "Linguistic markers in individuals with symptoms of depression in bi-multilingual context," in *Early Detection of Neurological Disorders Using Machine Learning Systems*, eds S. Paul, P. Bhattacharya, and A. Bit (Pennsylvania: IGI Global), 216–240. doi: 10.4018/978-1-5225-8567-1.ch012

Savoy, J. (2020). *Machine Learning Methods for Stylometry: Authorship Attribution and Author Profiling*. Berlin: Springer International Publishing.

Simmons, R. A., Chambless, D. L., and Gordon, P. C. (2008). How do hostile and emotionally overinvolved relatives view relationships?: what relatives' pronoun use tells us. *Fam. Process* 47, 405–419. doi: 10.1111/j.1545-5300.2008.00261

Tausczik, Y. R., and Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *J. Lang. Soc. Psychol.* 29, 24–54. doi: 10.1177/0261927X09351676

Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S.* New York: Springer.

Wadsworth, F. B., Vasseur, J., and Damby, D. E. (2016). Evolution of vocabulary in the poetry of sylvia plath. *Digit. Scholarsh. Humanit.* 32, 660–671.

Wongso, R., Luwinda, F. A., Trisnajaya, B. C., Rusli, O., and Rudy. (2017). News Article Text Classification in Indonesian Language. *Procedia Comput. Sci.* 116, 137–143. doi: 10.1016/j.procs.2017.10.039

Ziemer, K. S., and Korkmaz, G. (2017). Using text to predict psychological and physical health: a comparison of human raters and computerized text analysis. *Comput. Hum. Behav.* 76, 122–127. doi: 10.1016/j.chb.2017.06.038

Zucco, C., Calabrese, B., Agapito, G., Guzzi, P. H., and Cannataro, M. (2020). Sentiment analysis for mining texts and social networks data: methods and tools. *WIREs Data Min. Knowl. Discov.* 10:e1333. doi: 10.1002/widm.1333