



RESEARCH ARTICLE

# The Relationship Between Linguistic Features and Psychological States: A Quantitative Approach

Du Xiaowei<sup>1</sup>; Zhou Xintong; He Xiaofei

<sup>1</sup>Dalian Maritime University, Dalian  
116026, China



OPEN ACCESS

## PUBLISHED

30 August 2024

## CITATION

Xiaowei, D., et al., 2024. The Relationship Between Linguistic Features and Psychological States: A Quantitative Approach. Medical Research Archives, [online] 12(8). <https://doi.org/10.18103/mra.v12i8.5685>

## COPYRIGHT

© 2024 European Society of Medicine. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## DOI

<https://doi.org/10.18103/mra.v12i8.5685>

## ISSN

2375-1924

## ABSTRACT

Linguistic features are crucial in identifying psychological states. Prior studies have attempted to investigate the influence of psychological states on individual language usage through quantitative analysis, as well as which linguistic features can effectively indicate an individual's psychological state. Conducting a systematic review of pertinent literature, this study initially outlines the linguistic features related to psychological states extracted quantitatively, encompassing three dimensions: vocabulary, syntax, and emotion. Subsequently, it summarizes the correlations between linguistic features across different levels and psychological states based on quantitative methods. Last, the limitations of existing research are discussed. This study contributes to a deeper understanding of the rationale, significance, and applicability of applying linguistic features in detecting psychological states and provides guidance for future research in this domain.

**Keywords:** psychological states; linguistic features; quantitative analysis

## 1. Introduction

Mental disorders have emerged as a leading cause of the global disease burden. Research, in recent, has focused on computationally treating language as a meaningful tracker of the evolution of psychological risks using quantitative methods.<sup>1</sup> The motivation for this line of research stems from the premise that individuals' psychological states can manifest in their writing and communication styles, potentially allowing for automatic identification.<sup>2</sup> This domain could contribute to recognizing changes in an individual's psychological state through digital channel, which serves as a critical aspect at the intersection between language and mental health studies. Meanwhile, these studies are expected to alleviate the scarcity of resources for mental health by providing timely, efficient, and reliable monitoring and diagnostic services.

Previous studies, based on quantitative analysis, have found that linguistic features at multiple levels are related to individuals' psychological states.<sup>3</sup> Linguistic features, especially at lexical, syntactic, and sentiment levels, exhibit significant differences in texts generated by individuals with different psychological states.<sup>4,5</sup> Specifically, word, as the basic unit of the human language system, is closely related to psychological symbols and concepts.<sup>2</sup> In addition, syntax, as the representation of cognitive mechanisms, reflects individuals' organizations of phrase positions, constructions of sentence structures, and expression of event relationships.<sup>6,7</sup> Furthermore, sentimental expression, involving the interaction between the individual's language system, cognition, and neural networks for emotional processing, conveys the ways and feelings of experiencing the world, feelings, needs, and anticipations.<sup>8</sup>

This study attempts to summarize the relationship between linguistic features from different dimensions and psychological states and have a systematical review of such line of research based on quantitative approach. The objective of our study is two-fold. First, the findings of previous studies in this area are summarized to provide a potential new approach

for future clinical transform and medical services. Second, the gaps or deficiencies in such line are identified to direct future studies.

## 2. Overview of linguistic features related to psychological states

The language system in the human brain is the carrier of human thoughts and the representation of cognition.<sup>9</sup> In other words, language serves as a window into the inner world of individuals with the reflection of their emotions, cognition, and psychological states. This section provides an overview of the linguistic features of individuals with mental disorders from the lexical, syntactic, and emotional levels based on quantitative methods (see Table 1).

### 2.1. LEXICAL LEVEL

Previous studies on the lexical features of texts produced by individuals with psychological disorders fall roughly into two groups based on their foci and methods used.<sup>1</sup> The first group of such research examines lexical features from a micro perspective. In other word, they focus on the use of specific words (personal pronoun, insight words, and absolutist words), which could reflect individuals' focus, social status, and attention habits.<sup>5,10</sup> For instance, high-frequency words could reflect individuals' focus and attention habits.<sup>5</sup> In addition, personal pronouns are found to reveal social status, like social engagement and social isolation.<sup>10</sup> Meanwhile, the frequent use of absolutist words (always, completely) may indicate extremist tendencies since these words are believed to represent absolutist thinking.<sup>11</sup>

The second group investigates lexical features from a macro perspective. Lexical features from a macro perspective refer to the overall lexical characteristics at the textual level, such as the ratio of unique words, cumulative frequency distribution, and repetition ratio.<sup>12</sup> For example, the more repeated words demonstrate the higher textual concentration and lower lexical richness. In contrast, the more rarely used words suggest the richer word types with more textual information.<sup>13</sup> Moreover, the average word length could reflect an individual's language ability,

discourse style, and textual complexity. Compared with specific words, lexical features from the macro perspective could capture objective, diverse, and a complete view of an individual's lexical characteristics.<sup>14</sup>

## 2.2. SYNTACTIC LEVEL

The investigations of syntactic features pertinent to psychological states could be divided into shallow and deep syntactic features. On the one hand, shallow syntactic features refer to simple indicators in a sentence, such as average sentence length, numbers of tokens in the past, present and future tenses, and the proportions of different part-of-speech tags.<sup>15</sup> For example, previous studies have shown that short sentences and frequent objects effectively convey the most relevant details and final indicative information with an individual's foci.<sup>16</sup> In addition, individuals produce more adverbs, adjectives, more digressive sentences, and descriptive content when experiencing cognitive impairments, such as thinking disorders, comprehension reduction, and distraction.<sup>17</sup>

On the other hand, deep syntactic features represent more complex syntactic structure measurement which focuses on grammatical subordination and embedding, such as types of clauses, phrases, and subordinate relationships.<sup>6</sup> In particular, the previous research revealed individuals' higher cognitive ability, language proficiency, and thinking patterns based on more embedding and subordinate sentences produced from given grammatical information.<sup>18</sup> It is worth noting that some software programs currently could calculate syntactic structure features of texts automatically and accurately, such as Biber Tagger,<sup>19</sup> L2SCA,<sup>6</sup> and Coh-Metrix.<sup>20</sup>

## 2.3. SENTIMENTAL LEVEL

Previous studies have primarily relied on sentiment analysis to extract sentimental features, which could reflect the polarity of emotions and the degree of emotionality.<sup>21</sup> In a narrow sense, sentiment analysis refers to the identification of emotional poles: positive, negative, and neutral. In a broad sense, sentiment analysis encompasses two dimensions:

emotion and sentiment, which facilitate in-depth understanding and identification of an individual's mental needs.<sup>22</sup> Emotion analysis, as a branch of sentiment analysis, focuses on identifying various basic emotions, such as anger, anticipation, disgust, and fear.<sup>23</sup> While sentiment is the effect of emotion. For instance, "happiness" is a category of emotion and "positive" is the corresponding sentiment.

The approach for sentiment analysis mainly includes two categories currently. The first is dictionary-based method.<sup>9</sup> To be specific, this method calculates the frequency of different emotional words to determine textual sentiment or emotion based on dictionaries.<sup>8</sup> The popular sentiment dictionaries include LIWC,<sup>24</sup> NRC,<sup>25</sup> and Bing.<sup>26</sup> The second method is based on machine learning.<sup>12</sup> The key of this method is to establish algorithms for recognizable models for textual sentiment and mood. It operates with two steps. First, algorithms are applied to part of the dataset (training set) to construct classifiers, such as Support Vector Machine (SVM) and Generalized Linear Model (GLM). Second, the classifiers are employed to test the rest of dataset (test set) to predict sentiment polarity.<sup>27</sup>

## 3. Materials and Methods

This section includes three parts. First, we summarize the materials related studies used. Meanwhile, we outline quantitative methods for extracting indices of linguistic features. In addition, we report the main quantitative methods for investigating the relationship between linguistic features and psychological states.

### 3.1 MATERIALS

Early studies in these areas have primarily utilized small datasets, such as spontaneous speech, questionnaires, or written texts. Over the past decade, there has been a significant shift towards using data collected from social media platforms, including Meta (formerly Facebook), Twitter, and Reddit, etc. This trend is attributed to the distinct advantages offered by social media. First, many social media platforms ensure a high level of user anonymity, making them a preferred medium for individuals with mental disorders to discuss or mention their issues more discreetly.

Second, social media provides vast amounts of multimodal data, presented in a chronological manner. More importantly, data from social media

can complement traditional clinical information in a cost-effective manner.

**Table 1.** Linguistic features pertinent to mental disorders extracted from quantitative methods

Level	Categories	Examples
Lexical features	a macro perspective	Lexical richness (TTR, Entropy, rare word ratio); Lexical density; Lexical complexity, etc.
	a micro perspective	Functional words (pronouns, articles, auxiliary); Cognitive process (insight, certainty, tentativeness); Perceptive process (feelings, vision, auditory); Biological process (body, health/diseases, digestion); Personal foci (work, entertainment, money, religion, death); Social words (family, friends); Punctuation marks (period, comma, colon, question mark); Personal pronoun, spatial, and temporal words ratio
Syntactic features	shallow syntactic features	The proportions of different parts of speech (nouns, adjectives, object, adverb); Discourse length ; Average sentence length
	deep syntactic features	Syntactic complexity (clause number, complex nouns number); Dependency distance; Dependency tree, etc.
Sentiment features	Sentiment Emotion	Positive, Negative, neural Anger, anticipation, disgust, fear, happiness, sadness, surprise, trust, etc.

Note: Most of the linguistic features in Table 1 are calculated by mean, medium, and standard deviation. See details of formulas in Calz'a et al.,<sup>8</sup> Du & Sun,<sup>12</sup> and Du.<sup>13</sup>

### 3.2 METHODS FOR EXTRACTING INDICES OF LINGUISTIC FEATURES

Methods for extracting indices of linguistic features generally has categorized into two types, namely, top-down and bottom-up.<sup>28</sup> Specifically, top-down methods aim to represent the psychological concepts in language, which are essential based on the method of word counting.<sup>9</sup> Word counting involves defining lists of words and phrases that are used to represent a concept of which contain markers of psychological construct, state, or trait, and then using algorithms to find and count their lists.<sup>11</sup> In other words, the more higher frequency words used, the more likely the user is thinking a certain way, experiencing a particular emotion, and focusing a specific topic.

In contrast, bottom-up methods, can also be referred to as the data-driven approach, view language as a

whole to computationally model the patterns that parse language into interpretable quantities for analysis.<sup>12</sup> The bottom-up methods apply advance techniques, such as Machine Learning, Artificial Intelligence, and Computational Linguistics, to contextual language. To be specific, techniques like *the Bag of Words*, *Term Frequency-Inverse Document Frequency (IF-IDF)*, *Word Embedding*, and *N-gram language modelling*, could conceptualize language through lens of words, the co-occurrence of words, and the statistics produced by more primitive variables.<sup>6,63,65</sup> By doing so, these methods could extract the varied meanings of words and sentences with different usage and functions across context.

### 3.3 METHODS FOR INVESTIGATING RELATION

Research methods, exploring the relationship between linguistic features of different dimensions

and psychological states, mainly includes two categories. In the first category of method, studies in this line employ statistical analysis methods. For example, some studies have used descriptive statistical analysis (means, variance) to quantify linguistic feature indicators from the overall structure and distribution of texts generated by individuals among different psychological states.<sup>29</sup> Moreover, inferential statistical analyses (ANOVA, T-test, regression analysis) were employed to test whether there are significant differences or linear relationships in various linguistic features and different psychological states.<sup>11</sup> Last, meta-analysis methods could systematically review previous research results to verify how psychological states affect individuals' linguistic features.<sup>29</sup>

In the second category, studies focused on classification methods to explore the predictive performance of linguistic features to psychological states. Specifically, this line of research was implemented with machine learning algorithms to achieve their classifications. Then, they reported the predictive performance of the selected language features, such as accuracy, precision, recall, and F1 scores.<sup>12</sup> The higher value indicates the better performance. Compared to traditional methods, machine learning has two merits. First, machine learning algorithms could decipher the hidden information since they are more sensitive to data structure.<sup>22</sup> Second, machine learning could manage and analyze massive datasets and variables automatically and fast.<sup>29</sup>

## 4. Results and Discussion

This section systematically reviews the relationship between linguistic features at multiple levels and psychological states based on different quantitative methods. The previous research has revealed the effects of psychological states on individuals' linguistic features, as well as the predictive performances of linguistic features on individuals' psychological states.

### 4.1 LEXICAL FEATURES AND PSYCHOLOGICAL STATES

Many prior studies have explored the relationship between lexical features and psychological states

based on statistical analysis methods. However, the results of some studies were lack of generalizability.<sup>21</sup> The following thus primarily reports on repeatable research findings in four categories.

First, first-person singular pronouns show significant differences among different psychological states. For instance, first-person singular pronouns occur frequently in texts generated by individuals with mental disorders, such as emotional vulnerability, narcissism, bipolar disorder, self-doubt, depression, suicidal tendencies, and suicide.<sup>1,31,32</sup> In addition, Tølbøll<sup>30</sup> verified that a small but positive correlation exists between depression and the use of first-person singular pronouns through meta-analysis. The reason is that high-frequency first-person singular pronouns reflects individuals' self-focus, which further embodies lower dominance and adaptability, higher self-importance, and more severe emotional distance.<sup>33</sup>

Second, absolutist words with absolutism thinking differs significantly in different psychological states. Specifically, previous studies have indicated that individuals with mental disorders use more absolutist words, such as borderline personality disorder (BPD), emotional eating disorder, depression, and suicide.<sup>34</sup> Meanwhile, the higher frequency of absolutist words suggests the more severe mental disorders.<sup>11</sup> More importantly, Al-Mosaiwi & Johnstone<sup>11</sup> revealed that absolutist words may be more accurate than pronouns and negative emotion words in recognizing mental disorders. The reason may be that the calculation of absolutist words is not involved in context.<sup>35</sup> Last, frequent use of absolutist words could reflect individuals' irrational thinking patterns to indicate their mental disorders.<sup>11</sup>

Third, lexical features at textual level could effectively identify psychological states. The possible reason has two perspectives. One issue is that the lexical richness and the severity of mental disorders have a negative correlation.<sup>13</sup> To be specific, Litvinova et al.<sup>36</sup> found that individuals produce texts with lower lexical richness when scoring higher in introversion, sadness, aggressiveness, emotional instability, depression, PTSD, and suicide. Another issue, the

more severe mental disorders an individual has, the lower uniformity of word frequency distribution and the more fragmented writing patterns occur.<sup>13</sup> For instance, Kim et al.<sup>37</sup> found that when individuals in PTSD recovered, the organizational structure of their expressions regarding trauma memories enhanced, and fragmentation decreased. The reason is that lexical features at textual level could present individuals' comprehensive ability of language organization, which reflects the degree of their focus on the world.

Last, the research based on classification methods have revealed that lexical features are effective for identifying psychological states. Specifically, Du & Sun<sup>8</sup> and Du<sup>13</sup> applied binary classification based on machine learning to acquire first-person pronouns accuracies of 76.2%, 74.7%, and 82.6% in the prediction of anxiety, depression, and suicidal ideation respectively. Similarly, absolutist words and lexical richness indices obtained accuracies of 61.5%, 65.8%, 74.4% and 72.0%, 75.3%, 82.5% respectively.

#### 4.2 SYNTACTIC FEATURES AND PSYCHOLOGICAL STATES

Prior studies could fall into three categories in exploring the relationship between syntactic features and psychological states based on statistical analysis.

First, different psychological states affect the proportion of parts of speech in an individual's text.<sup>38</sup> For instance, individuals with suicidal ideation use more nouns as objects to express their decisions.<sup>39</sup> In addition, texts generated by schizophrenia individuals contain more nouns and fewer verbs as predicates.<sup>40</sup> Moreover, the depressed use more adverbs and adjective to organize more descriptive sentences.<sup>7</sup> Last, compared to non-suicide notes, Kim et al.<sup>37</sup> found that suicide notes contain fewer modifiers and future tense verbs.

Second, individuals with different psychological states produce significant differences in the average sentence length.<sup>41</sup> On the one hand, some studies found that individuals construct simplified sentences

when in stress, suicidal ideation, and schizophrenia.<sup>9</sup> The possible reason is that they focus on conveying the most relevant details to express final information and intentions. Meanwhile, Bhatia et al.<sup>42</sup> presented that the average length of suicide notes was 150 words in their study. On the other hand, other studies show opposite findings, that is the higher total word number indicates the more physical discomforts. For example, suicide survivors and individuals impending suicide use longer sentences and increase textual length rapidly.<sup>43,44</sup>

Last, the previous research indicates that syntactic complexity is negatively correlated with the severity of psychological states.<sup>45</sup> In other words, individuals with mental disorders yield texts with lower syntactic complexity, such as borderline personality disorder, schizophrenia, PTSD, suicidal ideation, and suicide.<sup>18,40</sup> To be specific, they have difficulties in discourse expression with the lack of referential cohesion since their impaired cognitive abilities and thinking disorders.<sup>45</sup> Furthermore, these disorders affect individuals' language skills resulting in ambiguous, sparse, or irrelevant sentences.

#### 4.3 SENTIMENT FEATURES AND PSYCHOLOGICAL STATES

Previous research pertinent to linguistic features at sentiment level could be roughly divided into three categories to explore their relationship to psychological states based on traditional statistical analysis findings.

The first category indicated that the positive correlation exists between the degree of mental disorders and the negative words as well as the degree of happiness and the positive words.<sup>21</sup> For example, individuals with mental disorders (depression, suicide) use more negative words and fewer positive words.<sup>37,46,47</sup> In addition, Tølbøll<sup>30</sup> employed meta-analysis to test that depression has a small but positive and a small negative correlation to the use of negative and positive words, respectively.

Second, some studies suggested that only negative words are related to mental disorders. Specifically,

the more negative words reflect the less mental health.<sup>48</sup> For instance, individuals with mental disorders (depression, pre-suicide, suicide) use more negative words, while the use of positive words present no significant difference compared to the control group.<sup>49</sup> Furthermore, individuals who commit suicide, attempt suicide and suffer from depression produce more negative words (such as anxiety, sadness, anger, self-blame, and fear).<sup>50,51</sup>

The third category indicated that only positive words have positive correlation to mental disorders. For example, Holmes et al.<sup>52</sup> showed that high-frequency positive words have slightly positive relation to aggravated pain symptoms and the degree of depression, respectively. Meanwhile, Handelman & Lester<sup>53</sup> found that the notes of individuals who had committed suicide contained more positive sentiment than those of individuals who had attempted suicide. In addition, Leenaars<sup>54</sup> revealed that real suicides use a more positive tone since they believe that the suicide note is the last opportunity to express gratitude, love, and care for survivors. Last, Pennebaker & Stone<sup>55</sup> found that a young woman named Katie used more positive sentimental words near death since the suicide decision may improve her emotion temporarily.

It is worth noting that the linguistic features at sentiment level show a good performance in the recognition of psychological states based on the research findings of classification methods. For example, De Choudhury et al.<sup>56</sup> applied emotional words to detect the risk of postpartum depression with the accuracy of 71.21%. Moreover, Tsugawa et al.<sup>57</sup> combined positive and negative words with machine learning algorithm to obtain a predictive accuracy of 79% on depression. Last, Du & Sun<sup>8</sup> presented that the emotional intensity has an accuracy of 78.5%, 76.7% and 82.1% in predicting anxiety, depression, and suicidal ideation, respectively.

## 5. Research limitations and future directions

To date, new liberal arts and modern computational technologies have developed rapidly. The research

pertinent to the analysis of an individual' linguistic features is on the verge of a significant revolution in the prediction of psychological states.<sup>58</sup> However, these previous studies may still have several limitations.

First, some studies relied on simplified indices like word frequency for the extraction of linguistic features, rather than fine-grained quantitative measures. The previous research indicates that linguistic features extracted from simplified indices have lower predictive power in identifying psychological states.<sup>8</sup> In addition, the lack of precise measurements may lead to negative outcomes such as misdiagnosis, improper treatment, and undue panic. Therefore, future research should employ advanced techniques to extract comprehensive linguistic features for in-depth understanding of language use across different psychological states.

Second, there is few studies focusing on non-English languages when the research on the relationship between psychological states and linguistic features are increasing. More importantly, cross-linguistic feature extraction may affect research results in their reliability, comparability, and transferability due to typological peculiarities (e.g., morphological structure).<sup>12</sup> Consequently, future research should establish a multimodal corpus specifically for non-English speaking individuals with psychological disorders. Moreover, it is essential to develop detailed statistical analyses and performance assessments to use linguistic features in non-English languages for the identification of psychological states.

Last, previous studies have mainly focused on the description of the relationship between language and psychological states. However, fewer studies have explored the principle rational of this relationship from cognitive mechanisms and neurology.<sup>17</sup> The traditional statistical analysis methods may result in overfitting, which needs further validation.<sup>59</sup> Meanwhile, research results based on sole classification methods might struggle to balance predictive accuracy with interpretability.<sup>22</sup> Therefore, future research should leverage interdisciplinary advantages for theoretical development based on

multi-theoretical approach involving cognitive, pragmatic, and clinical perspectives.

## 6. Conclusion

This study first reviews the current research on the relationship between linguistic features and psychological states. Subsequently, it summarizes various linguistic features automatically extracted from lexical, syntactic, and emotional levels. In addition, our study outlines the main methods and findings of previous research. Moreover, we give advice for future studies and researchers. First, future research should enrich the repertoire of linguistic features, validate the generalizability of statistical analysis results, and improve the accuracy and robustness of machine learning algorithms in identifying an individual's psychological states. Second, researchers should expand the scope of research, enhance theoretical innovation, and address practical issues. In particular, they should improve the research on the relationship between linguistic features and mental health by using artificial intelligence, big data technologies, and natural language processing. Last, there is potential to develop sensitive, low-cost, and non-invasive screening tools for mental disorders and monitoring psychological states in clinical practices.

## Acknowledgments:

We thank the referees and the editors for their insightful comments. Their suggestions have significantly enhanced the quality of the initial manuscripts.

## Disclosure Statement:

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Funding:

This study was Supported by "Natural Science Foundation of Liaoning Province" (Grant No.2024-BS-011), "Dalian Federation of Social Sciences" (Grant No. 2024dlskzd046), and "The Fundamental Research Funds for the Central Universities" (Grant No. 3132024346).



## References:

- [1] ADAM-TROIAN J, BONETTO E & ARCISZEWSKI T. Using absolutist word frequency from online searches to measure population mental health dynamics [J]. *Scientific Reports*, 2022, 12 (1):2619. <https://doi.org/10.1038/s41598-022-06392-4>.
- [2] ALLGOOD S M, SEEDALL R B & WILLIAMS R B. Expressive writing and marital satisfaction: A Writing sample analysis [J]. *Family Relations*, 2020, 69(2): 380-391. <https://doi.org/10.1111/fare.12416>.
- [3] AL-MOSAIWI M & JOHNSTONE T. In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation [J]. *Clinical Psychological Science*, 2018, 6(4):529-542. <https://doi.org/10.1177/2167702617747074>.
- [4] ANTONIOU E, EBONGERS P & JANSEN A. The mediating role of dichotomous thinking and emotional eating in the relationship between depression and BMI [J]. *Eating Behaviors*, 2017, 26:55-60. <https://doi.org/10.1016/j.eatbeh.2017.01.007>.
- [5] BARNES D H, LAWAL-SOLARIN F W & LESTER D. Letters from a suicide [J]. *Death Studies*, 2007, 31(7): 671–678. <https://doi.org/10.1080/07481180701405212>.
- [6] BHATIA M S, VERMA S K & MURTY O P. Suicide notes: Psychological and clinical profile [J]. *The International Journal of Psychiatry in Medicine*, 2006, 36(2): 163–170. <https://doi.org/10.2190/5690-CMGX-6A1C-Q28H>.
- [7] BIBER D. *Variation across speech and writing* [M]. Cambridge: Cambridge University Press, 1988.
- [8] BLACKBURN K G, WANG W, PEDLER R, THOMPSON R & GONZALES D. Linguistic markers in women's discussions on miscarriage and abortion illustrate psychological responses to their experiences [J]. *Journal of Language and Social Psychology*, 2021, 40(3): 398–411. <https://doi.org/10.1177/0261927X20965643>.
- [9] BOUKIL S, EL ADNANI F, CHERRAT L, et al. Deep learning algorithm for suicide sentiment prediction [A]. In M Ezziyani (Ed.), *Advanced Intelligent Systems for Sustainable Development (AI2SD'2018)* (Vol. 914, pp. 261–272) [C], 2019, Springer International Publishing. [https://doi.org/10.1007/978-3-030-11884-6\\_24](https://doi.org/10.1007/978-3-030-11884-6_24).
- [10] BOYD R L & SCHWARTZ H A. Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field [J]. *Journal of Language and Social Psychology*, 2021, 40(1):21–41. <https://doi.org/10.1177/0261927X20967028>.
- [11] CALZÀ L, GAGLIARDI G, ROSSINI FAVRETTI R, et al. Linguistic features and automatic classifiers for identifying mild cognitive impairment and dementia [J]. *Computer Speech & Language*, 2021, 65, 101113. <https://doi.org/10.1016/j.csl.2020.101113>.
- [12] CAMBRIA E. Affective computing and sentiment analysis [J]. *IEEE Intelligent Systems*, 2016, 31(2):102-107. <https://doi.org/10.1109/MIS.2016.31>.
- [13] CHATTERJEE A, GUPTA U, CHINNAKOTLA M K, et al. Understanding emotions in text using deep learning and big data [J]. *Computers in Human Behavior*, 2019, 93:309–317. <https://doi.org/10.1016/j.chb.2018.12.029>.
- [14] CHEN X, SYKORA M D, JACKSON T W, et al. What about mood swings: identifying depression on twitter with temporal measures of emotions [A]. *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18* [C], 2018, 1653–1660. <https://doi.org/10.1145/3184558.31916244>.
- [15] CHENG Q, LI T M, KWOK, C-L, et al. Assessing suicide risk and emotional distress in chinese social media: A text mining and machine learning study [J]. *Journal of Medical Internet Research*, 2017, 19(7):e243. <https://doi.org/10.2196/jmir.7276>.
- [16] COHN M A, MEHL M R & PENNEBAKER J W. Linguistic markers of psychological change surrounding September 11, 2001 [J]. *Psychological Science*, 2004, 15(10):687-693.

<https://doi.org/10.1111/j.0956-7976.2004.00741.x>.

[17] COPPERSMITH G, DREDZE M, HARMAN C, et al. From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses [A]. *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* [C]. Denver, Colorado. Association for Computational Linguistics, 2015, 1–10. <https://doi.org/10.3115/v1/W15-1201>.

[18] CROSSLEY S A & MCNAMARA D S. Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing [J]. *International Journal of Continuing Engineering Education and Life-Long Learning*, 2011, 21(2/3):170.

<https://doi.org/10.1504/IJCEELL.2011.040197>.

[19] CUMMINGS L. Pragmatic disorders in the twenty-first century [A]. In L. Cummings (Ed.), *Handbook of Pragmatic Language Disorders* (pp. 1–22) [C]. Springer International Publishing, 2021. [https://doi.org/10.1007/978-3-030-74985-9\\_1](https://doi.org/10.1007/978-3-030-74985-9_1).

[20] D'ANDREA A, FERRI F, GRIFONI P, et al. Approaches, tools and applications for sentiment analysis implementation [J]. *International Journal of Computer Applications*, 2015, 125(3):26-33. <https://doi.org/10.5120/ijca2015905866>.

[21] DE BEER C, WARTENBURGER I, HUTTENLAUCH C & HANNE S. A systematic review on production and comprehension of linguistic prosody in people with acquired language and communication disorders resulting from unilateral brain lesions [J]. *Journal of Communication Disorders*, 2023, 101:106298. <https://doi.org/10.1016/j.jcomdis.2022.106298>.

[22] DE CHOUDHURY M, COUNTS S & HORVITZ E. Predicting postpartum changes in emotion and behavior via social media [A]. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 3267 [C]. 2013. <https://doi.org/10.1145/2470654.2466447>.

[23] DEMIRAY Ç K & GENÇÖZ T. Linguistic reflections on psychotherapy: Change in usage of the first person pronoun in information structure

positions [J]. *Journal of Psycholinguistic Research*, 2018, 47(4):959–973.

<https://doi.org/10.1007/s10936-018-9569-4>.

[24] DESMET B & HOSTE V. Emotion detection in suicide notes [J]. *Expert Systems with Applications*, 2013, 40(16):6351–6358.

<https://doi.org/10.1016/j.eswa.2013.05.050>.

[25] DU, X. Lexical Features and psychological states: A quantitative linguistic approach [J]. *Journal of Quantitative Linguistics*, 2023, 1-23.

<https://doi.org/10.1080/09296174.2023.2256211>.

[26] DU X, SUN Y. Linguistic features and psychological states: A machine-learning based approach [J]. *Frontiers in Psychology*, 2022, 12.

[27] EICHSTAEDT J C, SMITH R J, MERCHANT R M, et al. Facebook language predicts depression in medical records [A]. *Proceedings of the National Academy of Sciences* [C], 2018, 115(44):11203-11208. <https://doi.org/10.1073/pnas.1802331115>.

[28] GRAESSER A C, MCNAMARA D S, LOUWERSE M M, et al. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 2004, 36(2): 193–202. <https://doi.org/10.3758/BF03195564>.

[29] GREGORY A. The decision to die: The psychology of the suicide note [A]. In D Canter & L Alison (Eds.), *Interviewing and deception* (pp. 127-156) [C]. Aldershot, UK: Ashgate, 1999.

[30] HANDELMAN L D & LESTER D. The content of suicide notes from attempters and completers [J]. *Crisis*, 2007, 28(2):102-104.

<https://doi.org/10.1027/0227-5910.28.2.102>.

[31] HOLMES D, ALPERS G W, ISMAILJI T, et al. Cognitive and emotional processing in narratives of women abused by intimate partners [J]. *Violence Against Women*, 2007, 13(11): 1192-1205.

<https://doi.org/10.1177/1077801207307801>.

[32] HOMAN C, JOHAR R, LIU T, et al. Toward macro-insights for suicide prevention: Analyzing fine-grained distress at scale [A]. In P. Resnik, R. Resnik & M. Mitchell (Eds.), *Proceedings of the Workshop on Computational Linguistics and Clinical*

*Psychology: From Linguistic Signal to Clinical Reality* (pp. 107-117) [C]. *Association for Computational Linguistics*. 2014. <https://doi.org/10.3115/v1/W14-3213>.

[33] HOU R, YANG J & JIANG M. A study on Chinese quantitative stylistic features and relation among different styles based on text clustering [J]. *Journal of Quantitative Linguistics*, 2014, 21(3): 246-280. <https://doi.org/10.1080/09296174.2014.911508>.

[34] HU M and LIU B. Mining and summarizing customer reviews [A]. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '04)* [C]. Association for Computing Machinery, New York, NY, USA, 2004, 168-177. <https://doi.org/10.1145/1014052.1014073>.

[35] JAKOVLJEV I & MILIN P. The relationship between thematic, lexical, and syntactic features of written texts and personality traits [J]. *Psihologija*, 2017, 50(1):67-84. <https://doi.org/10.2298/PSI161012006J>.

[36] JI S, YU C P, FUNG S, et al. Supervised learning for suicidal ideation detection in online user content [J]. *Complexity*, 2018, 1-10. <https://doi.org/10.1155/2018/6157249>.

[37] JONES L S, ANDERSON E, LOADES M, et al. Can linguistic analysis be used to identify whether adolescents with a chronic illness are depressed? [J]. *Clinical Psychology & Psychotherapy*, 2020, cpp.2417. <https://doi.org/10.1002/cpp.2417>.

[38] JUOLA P, MIKROS G K & VINSICK S. Correlations and potential cross-linguistic indicators of writing style [J]. *Journal of Quantitative Linguistics*, 2019, 26(2):146-171. <https://doi.org/10.1080/09296174.2018.1458395>.

[39] KIM K, CHOI S, LEE J, et al. Differences in linguistic and psychological characteristics between suicide notes and diaries [J]. *The Journal of General Psychology*, 2019, 146(4):391-416. <https://doi.org/10.1080/00221309.2019.1590304>.

[40] KIM M & CROSSLEY S A. Modeling second language writing quality: A structural equation

investigation of lexical, syntactic, and cohesive features in source-based and independent writing [J]. *Assessing Writing*, 2018,37:39-56. <https://doi.org/10.1016/j.asw.2018.03.002>.

[41] KOTU V & DESHPANDE B. *Predictive analytics and data mining: Concepts and practice with RapidMiner* [M]. Elsevier/Morgan Kaufmann, Morgan Kaufmann is an imprint of Elsevier, 2015.

[42] LE X, LANCASHIRE I, HIRST G, et al. Longitudinal detection of dementia through lexical and syntactic changes in writing: A case study of three British novelists [J]. *Literary and Linguistic Computing*, 2011, 26(4):435-461. <https://doi.org/10.1093/lc/fqr013>.

[43] LEENAARS, A. A. *Suicide notes: Predictive clues and patterns* [M]. New York: Human Sciences Press, 1988.

[44] LESTER, D. Bereavement after suicide: A study of memorials on the internet [J]. *OMEGA - Journal of Death and Dying*, 2012, 65(3):189-194. <https://doi.org/10.2190/OM.65.3.b>.

[45] LITVINOVA T, ZAGOROVSKAYA O, LITVINOVA O, et al. Profiling a set of personality traits of a text's author: A corpus-based approach [A]. In A Ronzhin, R Potapova & G Nemeth (Eds.), *Speech and Computer: Proceedings of the 18th International Conference, SPECOM 2016* (pp. 555-562) [C]. Springer International Publishing.

[46] LYONS M, AKSAYLI N D & BREWER G. Mental distress and language use: Linguistic analysis of discussion forum posts [J]. *Computers in Human Behavior*, 2018, 87:207-211. <https://doi.org/10.1016/j.chb.2018.05.035>.

[47] LU X. Automatic analysis of syntactic complexity in second language writing [J]. *International Journal of Corpus Linguistics*, 2010, 15(4):474-496. <https://doi.org/10.1075/ijcl.15.4.02lu>.

[48] MOHAMMAD S M, AND TURNEY P D. Crowdsourcing a word-emotion association lexicon [J]. *Comput. Intell*, 2013, 29:436-465. doi: 10.1111/j.1467-8640.2012.00460.x.

- [49] MORALES M R & LEVITAN R. Speech vs. text: A comparative analysis of features for depression detection systems [A]. *2016 IEEE Spoken Language Technology Workshop (SLT)* [C], 2016, 136-143. <https://doi.org/10.1109/SLT.2016.7846256>.
- [50] ÖZCAN VURAL A & KURUOĞLU G. Nominal and verbal predicate use in schizophrenia [J]. *PSYCHOLINGUISTICS*, 2020, 27(2):213-228. <https://doi.org/10.31470/2309-1797-2020-27-2-213-228>.
- [51] PENNEBAKER J W, CHUNG C K, FRAZEE J, et al. When small words foretell academic success: The case of college admissions essays [J]. *PLoS ONE*, 2014, 9(12):e115844. <https://doi.org/10.1371/journal.pone.0115844>.
- [52] PENNEBAKER J W & STONE L D. Words of wisdom: Language use over the life span [J]. *Journal of Personality and Social Psychology*, 2003, 85(2), 291–301. DOI: 10.1037/0022-3514.85.2.291.
- [53] PULVERMAN C S, LORENZ T A & MESTON C M. Linguistic changes in expressive writing predict psychological outcomes in women with history of childhood sexual abuse and adult sexual dysfunction [J]. *Psychological Trauma: Theory, Research, Practice, and Policy*, 2015, 7(1):50-57. <https://doi.org/10.1037/a0036462>.
- [54] RAMÍREZ-ESPARZA N & CHUNG C, KACEWICZ E, et al. The psychology of word use in depression forums in English and in Spanish: testing two text analytic approaches [A]. *ICWSM 2008 - Proceedings of the 2nd International Conference on Weblogs and Social Media* [C]. ICWSM 2008, Seattle, WA.
- [55] RUDE S, GORTNER E-M & PENNEBAKER J. Language use of depressed and depression-vulnerable college students [J]. *Cognition & Emotion*, 2004, 18(8):1121-1133. <https://doi.org/10.1080/02699930441000030>.
- [56] SAILUNAZ K, DHALIWAL M, ROKNE J, et al. Emotion detection from text and speech: A survey [J]. *Social Network Analysis and Mining*, 2018, 8(1):28. <https://doi.org/10.1007/s13278-018-0505-2>.
- [57] SCHOENE A M, TURNER A, DE MEL G R, et al. Hierarchical multiscale recurrent neural networks for detecting suicide Notes [J]. *IEEE Transactions on Affective Computing*, 2021, 1-1. <https://doi.org/10.1109/TAFFC.2021.3057105>.
- [58] SHEFFLER J L, JOINER T E & SACHS-ERICSSON N J. The interpersonal and psychological impacts of COVID-19 on risk for late-life suicide [J]. *The Gerontologist*, 2021, 61(1):23-29. <https://doi.org/10.1093/geront/gnaa103>.
- [59] TAUSCZIK Y R & PENNEBAKER J W. The psychological meaning of words : LIWC and computerized text analysis methods [J]. *Journal of Language and Social Psychology*, 2010, 29(1):24-54. <https://doi.org/10.1177/0261927X09351676>.
- [60] TEN THIJ M, BATHINA K, RUTTER L A, et al. Depression alters the circadian pattern of online activity [J]. *Scientific Reports*, 2020, 10(1):17272. <https://doi.org/10.1038/s41598-020-74314-3>.
- [61] TØLBØLL K B. Linguistic features in depression:A meta-analysis [J]. *Journal of Language Works*, 2019, 4:22.
- [62] TSUGAWA S, KIKUCHI Y, KISHINO F, et al. Recognizing depression from Twitter activity [A]. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15* [C], 2015,3187–3196. <https://doi.org/10.1145/2702123.2702280>.
- [63] XIAO W & SUN S. Dynamic lexical features of PhD theses across disciplines: A text mining approach [J]. *Journal of Quantitative Linguistics*, 2020, 27(2):114–133. <https://doi.org/10.1080/09296174.2018.1531618>.
- [64] ZINKEN J, ZINKEN K, WILSON J C, et al. Analysis of syntax and word use to predict successful participation in guided self-help for anxiety and depression [J]. *Psychiatry Research*, 2010, 179(2): 181-186. <https://doi.org/10.1016/j.psychres.2010.04.011>.
- [65] ZÖRNIG P & ALTMANN G. A sequential activity measure for texts and speeches [J]. *Glottology*, 2016, 7(2). <https://doi.org/10.1515/glott-2016-0015>.