



# Who Killed Sanmao and Virginia Woolf? A Comparative Study of Writers with Suicidal Attempt Based on a Quantitative Linguistic Method

Shucheng Zhu<sup>1</sup>, Xi Wang<sup>2</sup>, and Pengyuan Liu<sup>1,3</sup>(✉)

<sup>1</sup> School of Information Science, Beijing Language and Culture University, 15th Xueyuan Road, Haidian District, Beijing 100083, China

201721198327@stu.blcu.edu.cn, liupengyuan@pku.edu.cn

<sup>2</sup> Faculty of Foreign Studies, Beijing Language and Culture University, 15th Xueyuan Road, Haidian District, Beijing 100083, China

wangxi@xiao100.com

<sup>3</sup> Center of National Language Recourse and Monitoring Research Print Media Branch, Beijing Language and Culture University, 15th Xueyuan Road, Haidian District, Beijing 100083, China

**Abstract.** Writing style can reflect a writer's change of mood. It is thus important to discover quantitative linguistic markers of writing styles to be used to analyze the works of writers with suicidal attempt. Using five quantitative linguistic markers, this paper investigates the works of Sanmao and Virginia Woolf -- two writers with suicidal attempt -- diachronically. Although they wrote in different languages, the two writers shared some similarities in terms of these quantitative linguistic markers. Via analyzing quantitative linguistic markers of these two writers' works, one can clearly see their change of mood and how their writing styles were influenced. This paper also attempts to find out the reasons why they chose to end their lives with this quantitative linguistic method. These quantitative linguistic markers can be utilized in future studies, such as suicidal ideation detection.

**Keywords:** Quantitative linguistics · Suicidal attempt · Sanmao · Virginia Woolf

## 1 Introduction

On 14<sup>th</sup> October 2019, the famous Korean superstar Sulli committed suicide in her own house. Her premature death at the age of 25 drew public attention back to the topic of suicide. Suicide has become the third chief cause of death in today's world. It is estimated that there are 1.6 million suicidal attempts each year, and approximately 0.8 million people die of suicide [1]. Suicide has always been a hot topic for discussion. In the 1900s, the suicides of two famous female writers -- Sanmao and Virginia Woolf, from China and the U.K., shocked the literary world. Therefore, it is crucial to find methods to detect suicidal ideation or tendency in order to prevent suicide.

Many significant efforts have been made recently. As a first step, several psychological dictionaries have been compiled [2–6]. With the help of these resources and computing technologies, remarkable progress has been made in detecting suicidal ideation

and attempts. Most of such research has been conducted on social media materials or has analyzed suicide notes, and has helped to prevent people with suicidal tendencies. Most of the psychological dictionaries have compiled specific words related to different themes, such as death, love, anger, etc. However, language is full of duplicity, especially when it is employed to express negative emotions. Therefore, it is problematic to merely use specific words to detect suicidal ideation. With the help of quantitative linguistics techniques, we can find markers reflecting language phenomena that cannot be consciously hidden by writers. These markers may be used to analyze works in different languages, simultaneously manifesting the general nature of language and the psychology of human beings.

In this paper, we use five quantitative linguistic markers –Richness, Pronoun, Sensory, Lambda and Zipf’s Exponent -- to analyze works of two female writers, Virginia Woolf and Sanmao. Virginia Woolf’s works are in English whereas Sanmao wrote in Chinese. We can clearly see how big events have influenced their writing styles and their courses of life, respectively. After analysing the values of these markers, we also try to find out different reasons why they committed suicide.

## 2 Related Works

Literary scholars tend to consider a writer’s life as auxiliary information for textual interpretation and analysis. Few of them have attempted to probe into the mental conditions of creators of literary works by analyzing their texts. Most of them have applied Freudian psychoanalysis in their studies. Here we take the works of Virginia Woolf as an example. Virginia Woolf’s life and novels are both fascinating and perplexing to literary scholars. In tracing Woolf’s protagonists’ remarks about death, notably in her later works, Brombert has proclaimed that Woolf considered writing “as a form of salvation”, “a struggle against death, an act of defiance, a way of protecting her sanity” [7]. Jean Thomson has focused particularly on *Mrs. Dalloway* and believed that Woolf weaved her own experiences of excitement and depression into this book in order to give an authentic portrayal of the mind of an ex-soldier, Septimus Smith. Thomson has also seen *Mrs. Dalloway* as a kind of “Dance of Death” [8]. Other critical biographers, underlining Woolf’s manic-depressive side, have denied her intention of control over her life [9]. C. Caramagno has also expressed his complaints about Woolf’s biographies in his contentious essay in the rather subjective nature [10]. Although literary scholars have tended to concentrate on biographical details, especially suicidal attempts, most of their focuses have been put on plot and characterization.

With the help of combining linguistic features and computing technologies, linguists and computing experts are making efforts to study the relation between features of writing style and writers with suicidal ideation. Earlier studies have focused on which professions are more prone to mental disorder [11] or used simple linguistic features to distinguish between ordinary letters and suicidal notes [12]. After psychological dictionaries have been compiled, especially LIWC, the analysis of suicide and language has become a problem of classification by using natural language processing methods. These works include detecting whether the users have suicidal ideation [13–16], or mental disorder [17], and analyzing the emotions and language features of people with

suicidal ideation and attempts [18–20]. By using Suicide Probability Scale (SPS) and a classification model, scholars can distinguish people with a high suicidal probability and those with a low probability [21–23]. In distinguishing suicide notes and ordinary notes, computers do a better job than mental health professionals [24]. Technologies like sentiment analysis, opinion mining and depression detection are not perfect for use with suicide [25]. Besides, most studies on detecting suicide in real life have focused on the language used in social-media blogs. The method by using psychological dictionaries is just sentiment analysis. No real analysis of language is performed. In addition, no studies has focused on the chronological changes in the writing styles of the writers with suicidal ideation, as if they somehow all had suicidal ideation throughout their lifetimes.

In suicide detection studies, some scholars have used dictionaries to study the change of words in the blogs of a Chinese adolescent one year before s/he committed suicide [26]. Distinctively, this study concentrates on such change in chronological order, but the time is too short to reflect the subject's lifetime. Some others have mainly used four quantitative linguistic markers: Richness, Pronoun, Activity Power and Sensory (RPAS) [27, 28], as well as Critical Slowing Down [29], to compare two female English writers, Iris Murdoch and P. D. James, to see how big events and the mental disease Alzheimer influence writers' writing style. The quantitative markers RPAS were used in identifying personality in previous studies [30–32] and were proved to have good results. It is also believed that these markers can be used to identify mental diseases like Alzheimer, depression and extreme behaviour like terrorism. But these studies have only focused on the English language. Therefore, it is necessary to prove that these measures are also effective with works in different languages.

### 3 Methodology

#### 3.1 Materials

In this paper, we have chosen the works of Sanmao and Virginia Woolf to study their change of writing style by using quantitative linguistic markers. Sanmao and Virginia Woolf were both famous female writers who died of suicide. Tables 1 and 2 show the works of Virginia Woolf and Sanmao that we have chosen.

We have only chosen full-length novels of Virginia Woolf to control the variable of genre. All the works are from Project Gutenberg<sup>1</sup> and their contents are in English. The Python package NLTK is used to tag parts-of-speech. We have only selected the novels and proses of Sanmao to control the variable of genre. It should be noticed that the first published novel of Sanmao is *The Story of the Sahara* and the second one is *Rainy Season Won't Come Again*. But *Rainy Season Won't Come Again* is a collection of essays and short stories Sanmao wrote between the age of seventeen and twenty-two. So, we have placed *Rainy Season Won't Come Again* as her first book to analyze. In order to keep the purity of the selected works, everything written by others (including forewords, epilogues, etc.) is removed. The Python package Jieba is used to segment and to tag the parts-of-speech.

<sup>1</sup> <http://gutenberg.net.au>.

### 3.2 Markers

Quantitative Index Text Analyzer (QUITA) [33] is used to compute the markers R4, Hapax Percentage and Lambda. Altmann Fitter is used to compute Zipf's Exponent [34].

**Richness.** Richness is a measure of a person's ability to use vocabulary (with focus on its size). It reflects not only the person's age and education background, but also his/her mood and psychological activities [27, 28, 31, 32]. People in a normal or joyful mood tend to have higher Richness than people in a bad mood, such as people with suicidal ideation [12]. The traditional Richness measure Type-Token Ratio (TTR) is subject to the influence of text size. Therefore, R4 and Hapax Percentage are chosen as the Richness markers as these two measures can eliminate the influence of text size.

**Table 1.** Works of Virginia Woolf we have chosen to analyze in this paper.

Number	Name	Publication time
V1	<i>The Voyage Out</i>	March, 1915
V2	<i>Night and Day</i>	October, 1919
V3	<i>Jacob's Room</i>	October, 1922
V4	<i>Mrs. Dalloway</i>	May, 1925
V5	<i>To the Lighthouse</i>	May, 1927
V6	<i>Orlando: A Biography</i>	June, 1928
V7	<i>The Waves</i>	October, 1931
V8	<i>The Years</i>	March, 1937
V9	<i>Between the Acts</i>	April, 1941

**R4.** R4 is one of the many ways of calculating the vocabulary richness of a text. The bigger it is, the higher Richness the text has. Formula (1) shows how R4 is calculated. Here V is the total number of types in a text, N is the total number of tokens in the text, r is the rank and f(r) is the frequency in the rank-frequency distribution of the text.

$$R4 = 1 - \frac{1}{V} \left( V + 1 - \frac{2}{N} \sum_{r=1}^V rf(r) \right) \quad (1)$$

**Hapax.** Hapax Percentage is the simple ratio of the number of hapax legomena in a text to the number of tokens. Hapax legomena are words that occur only once in a text. Formula (2) shows how Hapax Percentage is calculated. Here, NH is the number of hapax legomena in a text.

$$Hapax = \frac{NH}{N} \quad (2)$$

**Table 2.** Works of Sanmao we have chosen to analyze in this paper.

Number	Name in English	Publication time
S1	<i>Rainy Season Won't Come Again</i>	July, 1976
S2	<i>The Story of the Sahara</i>	May, 1976
S3	<i>The Scarecrow Notes</i>	June, 1977
S4	<i>Crying Camel</i>	August, 1977
S5	<i>Tender Night</i>	February, 1979
S6	<i>Back</i>	August, 1981
S7	<i>Dream Whispering Color</i>	August, 1981
S8	<i>Travelling Far to Travel Throughout</i>	May, 1982
S9	<i>Give You A Horse</i>	July, 1983
S10	<i>Fallen City</i>	March, 1983
S11	<i>My Baby</i>	July, 1987
S12	<i>Trouble Learning to Remember</i>	July, 1988

**Pronoun.** Pronoun is generally used to identify the author's gender [27, 28, 31, 32]. It has been shown that females prefer to use more personal pronouns and that the more pronouns a writer uses, the more emotional s/he is [35, 36]. In addition, people with suicidal ideation use more personal pronouns in their social media [18]. Formula (3) shows how Pronoun Percentage is calculated. Here P is the total number of personal pronouns in a text.

$$\text{Pronoun} = \frac{P}{N} \quad (3)$$

**Sensory.** Sensory words refer to visual, auditory, haptic, olfactory and gustatory adjectives [37]. Sensory words have been proved to have the capacity to stimulate our cerebral cortex [38]. Less sensory adjectives and smaller values of exclusivity can also be a signal of mental disease [28, 32]. In this paper, we use formula (4) to calculate Sensory [32]. Here  $\varphi_i$  is the number of the i-th adjective in the text and  $\vartheta_i$  is the value of exclusivity of the adjective, NK is the total number of the sensory adjective list. For the English texts, the sensory adjective list with exclusivity is from [39] and for the Chinese texts, the sensory adjective list with exclusivity is from [40]. To eliminate the influence of text size, all of the terms in the summation are normalized.

$$\text{Sensory} = \sum_{i=1}^{i=NK} \frac{\varphi_i \vartheta_i}{N} \quad (4)$$

**Lambda.** Lambda is an indicator that deals with the frequency structure of a text. It reflects both vocabulary richness and the relationship between the neighboring frequencies in a text. This measure also eliminates the influence of text size. Formula (5) shows its calculation.

$$\text{Lambda} = \frac{\sum_{r=1}^{V-1} [(f(r) - f(r+1))^2 + 1]^{\frac{1}{2}} (\log_{10} N)}{N} \quad (5)$$

**Zipf's Exponent.** Formula (6) shows Zipf's formula. Here  $P_r$  is the frequency of the word with rank  $r$ .  $C$  and  $b$  are parameters with  $b$  being an exponent. Studies have proved that the exponent  $b$  can distinguish languages. The exponent of English is around 1.11 which is the highest among natural languages [41, 42]. Other studies also found that Zipf's exponent might be related to language cognition and other aspects [43–46]. Zipf's exponent is related to language complexity, meaning that children's Zipf's exponents are higher than adults' exponents [47]. We believe that Zipf's exponent is also related to one's mood. If the writer is in the good condition, the Zipf's exponent of his/her language should be low.

$$P_r = Cr^{-b} \quad (6)$$

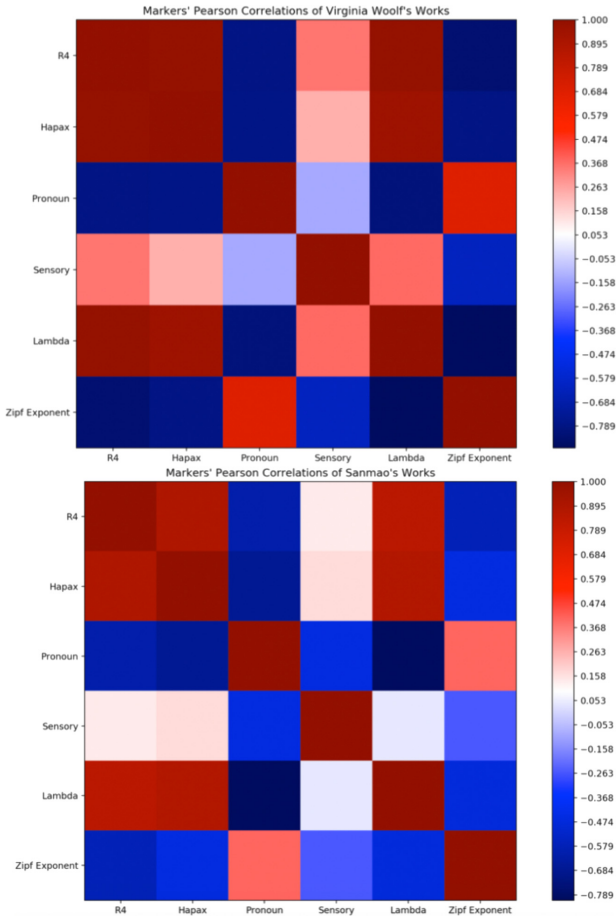
## 4 Analysis

### 4.1 Correlation Analysis

The two matrices in Fig. 1 show the Pearson correlations between the quantitative linguistic markers of Virginia Woolf's works and Sanmao's works respectively. It is clearly seen that the markers Richness, Sensory and Lambda show homogeneity as the higher the markers are, the better mood and mental condition the writers are in. The markers Pronoun and Zipf's Exponent also exhibit homogeneity as the higher the markers are, the worse mental conditions the writers are in. However, it is improper to suggest that the higher the Pearson Correlations are, the better efforts the quantitative linguistic measures give as the different quantitative linguistic markers represent different aspects.

### 4.2 Markers Visualization

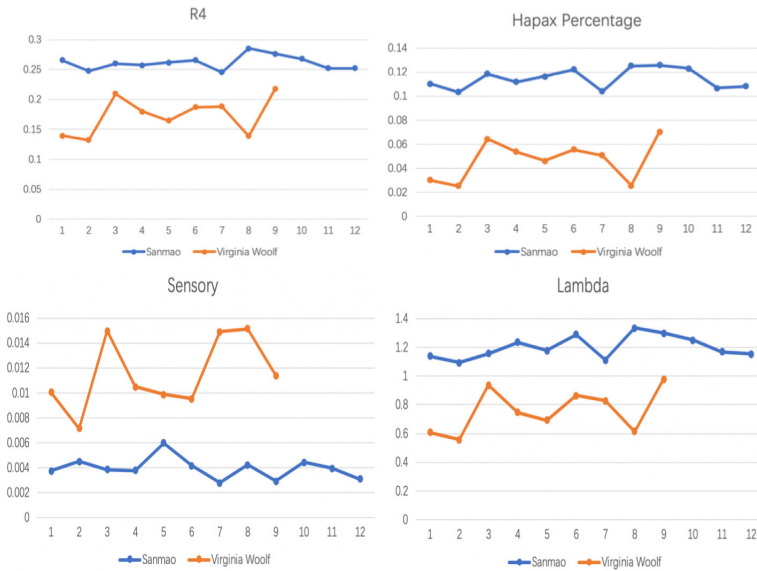
**Richness, Sensory and Lambda Visualization.** Figure 2 shows how the markers' values change diachronically in the works of Virginia Woolf and Sanmao. R4, Hapax percentage, Sensory and Lambda share a similar tendency in Woolf's works. In the beginning, when she wrote her first two novels (V1 and V2), the values of the markers were relatively low. It was painstaking and difficult for her both physically and mentally. Leonard Woolf, her husband, believed *The Voyage Out* (V1) "to be the root cause of Virginia's distress" [48]. While waiting for the publication of *The Voyage Out*, Woolf was extremely anxious about its critical reception. In September 1913, she attempted to commit suicide [49]. On 23<sup>rd</sup> February 1915, Woolf became hallucinated, seeing her mother in the room talking wildly to her. It was only until September that she began to lead a normal life. Woolf remained relatively stable in 1916 and 1917, but afterwards, Vanessa's desperate wish to get pregnant and the need to finish *Night and Day* (V2) made her depression deteriorate [48]. The next decade witnessed the prime time of Woolf's literary career. The values of V3 to V7 were at a high level, and Woolf's condition was relatively stable. Though the Sensory values of V4 to V5 and the values of three other markers of V5 were low, they were still relatively high compared to the values of V1, V2 and V8. Then, when she wrote and revised her eighth novel (V8), an anti-war novel, she was horrified and threatened by the imminent war. It made her extremely angst-ridden.



**Fig. 1.** Pearson Correlations between quantitative linguistic markers: Virginia Woolf’s Works (first image) and Sanmao’s Works (second image)

The pressure of outside cruelty and the stress of finishing the novel forced Virginia Woolf into insomnia, headaches and a breakdown that followed [48]. For V9, the values of the three markers other than Sensory were higher than the values of the corresponding markers of V8. This is because the mood of Virginia Woolf got better in her last years. It was a sudden event -- the horrific war (the World War II) destroyed her beloved city London -- that accelerated Woolf’s suicidal tendency.

With Sanmao, we can see a very low point in *Dream Whispering Color* (S7). In this book, Sanmao wrote down her sad feelings about the loss of her husband Jose (Hexi). She suffered a lot during that time. Before S7, Sanmao wrote about her happy marriage life in the Sahara. After S7, she travelled to Latin America and Europe. This, in a sense, alleviated her pain. So, the Richness values during this period were higher. At the end of her writing career (S10, S11 and S12) the marker values fall as they predict her suicide. This also means that there was no sudden event that led to her death like Woolf.

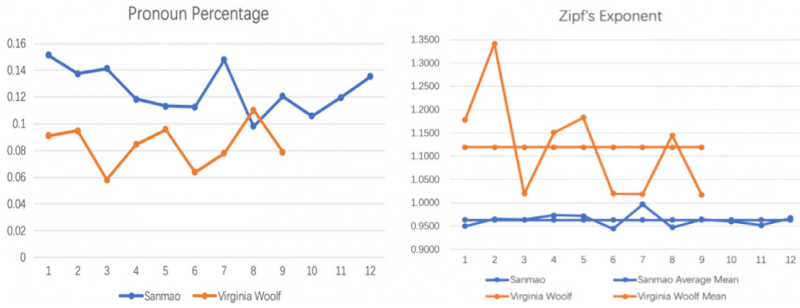


**Fig. 2.** R4, Hapax, Sensory and Lambda by work number for Virginia Woolf and Sanmao. Sanmao's lowest marker values are at S7, and the marker values of her last several works the go down. Woolf's lowest marker values are at V1, V2 and V8, and the marker values of her last work go up. Sensory is an exception. It reflects her worse mood while the other markers show that her sudden suicide was because of the shock of the war.

**Pronoun and Zipf's Exponent Visualization.** Figure 3 shows the Pronoun and Zipf's Exponent values changing diachronically in the works of Virginia Woolf and Sanmao. Pronoun and Zipf's Exponent values share a similar tendency in Woolf's works. The value of markers indicates high emotional level at the beginning (V1 & V2), in the middle (V5) and almost at the end (V8). Emotional elation can be caused by personal temperament and high involvement in the process of writing. As we have shown above, when Woolf wrote *The Voyage Out* (V1), *Night and Day* (V2), and *The Years* (V8), she suffered from cyclothymic depression and her mood alternated between extreme depression and euphoria. In these periods, Woolf was highly depressive. As for the case of *To the Lighthouse* (V5), it is Woolf's most "self-revealing" novel [7]. This novel is personal, for it is an elegy for Woolf's parents. As for V4, even though she was in a mentally stable condition, the content of her writing was disturbing: Mrs. Dalloway dealt with an ex-soldier, who suffered from PTSD and committed suicide at the end of the novel. The emotional depth Woolf revealed through these works is profound.

Sanmao's Pronoun Percentage and Zipf's Exponent values reach their highest points in S7. She wrote S7 when she was suffering badly from the loss of her husband. The marker values of her last two books have rising slopes. This shows that Sanmao's mental condition was deteriorating. The Zipf's Exponent graphs also shows that the exponent of English work is higher, confirming the findings of previous study [42].

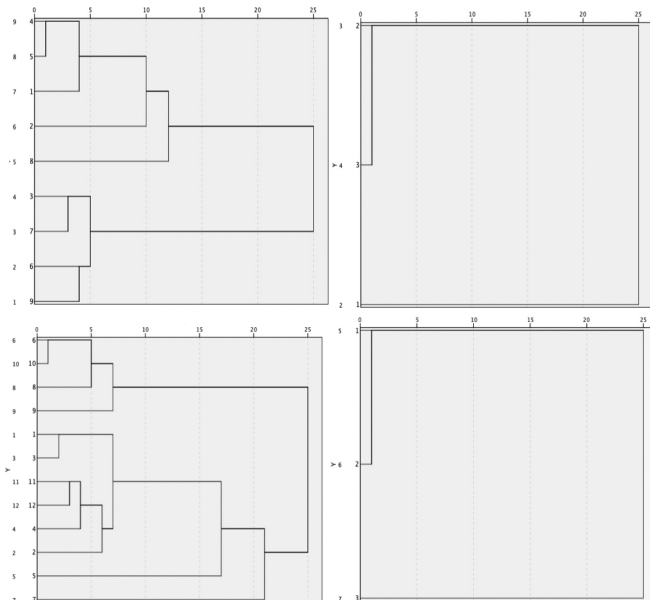




**Fig. 3.** Pronoun and Zipf’s Exponent by work number for Virginia Woolf and Sanmao. For Sanmao’s works, her highest values are at S7 and the values of her last 2 works have rising slopes. For Virginia Woolf, her highest values are at V1, V2, V5 and V8 and the values of her last two works have downward slopes. The means of the Zipf’s Exponent values show that Zipf’s Exponent can distinguish different languages.

### 4.3 Sequence Clustering

After normalization of all the markers, system clustering is used for sequence clustering [50]. Figure 4 shows the sequence clustering of the works of Sanmao and Woolf.



**Fig. 4.** System clustering of Virginia Woolf’s (top 2) and Sanmao’s (bottom 2) works

For Virginia Woolf, we understand:

V1, V2: Virginia Woolf's earlier writings. As a novice in literature, Woolf suffered greatly from the pressure to finish the novels. So, her mood alternated between depression and euphoria.

V3, V4, V5: These three books can be seen as Woolf's best literary works. Also, this period was the primetime of her career and personal life.

V6, V7: These two books are generally considered as the best of her later career.

V8: As the outcry of antiwar by Woolf, *The Years* was consuming for Woolf's sensitive nerves, unstable mentality and vulnerable physical condition.

V9: We can see a recovery tendency in the period between Woolf's last two novels. However, the horrific and inexorable war accelerated Woolf's suicidal tendency.

For Sanmao, we understand:

S1, S2, S3, S4, S5, S6: These are her early works and the reflection of her life in the Sahara.

S7: It reveals the darkness of her life because of her beloved husband's death.

S8, S9, S10: She created these works during or after her travels in Latin America and Europe.

S11, S12: Her last 2 works indicate her suicidal tendency because of her negative emotion.

## 5 Conclusion

We use five quantitative linguistic markers to analyze the English and Chinese works of two female writers. Our results show that all the markers we have chosen can reflect the psychological conditions of the two writers. With the help of these markers, we can find out the psychological changes in the minds of writers who never told the public why they chose to commit suicide to end their lives. We also find that big events, such as mental disorder and the loss of spouse and war, can influence one's life. These influences may be hidden by using emotional words, but cannot be hidden when the language of a writer's works is analyzed using quantitative linguistic markers. However, the markers we use are far from enough. More and more materials should be gathered to analyze the people's psychological activities. We hope that more and more useful features will be found for detecting suicidal ideation. We hope that anyone suffering from mental problems will indeed get the help they need -- indeed and in need.

**Acknowledgments.** This work is supported by the Fundamental Research Funds for the Central Universities, Research Funds of Beijing Language and Culture University 20YCX152, and Ministry of Education of Humanities and Social Science Project 18YJA740030. Thanks to Prof. Su Qi, Dr. Wan Mingyu and Zhou Jie from Peking University, Zhang Ying, Pan Yue, Du Bingjie and Zhang Sanle from Beijing Language and Culture University for their discussions.

## References

1. World Health Organization: Mental Health Action Plan 2013–2020. World Health Organization, Geneva, Switzerland (2013)

2. Coltheart, M.: The MRC psycholinguistic database. *Q. J. Exp. Psychol. Sect. A* **33**(4), 497–505 (1981)
3. Yao, Z., Wu, J., Zhang, Y., et al.: Norms of valence, arousal, concreteness, familiarity, imageability, and context availability for 1,100 Chinese words. *Behav. Res. Methods* **49**(4), 1374–1385 (2017)
4. Bucci, W., Kabasakalian, R.: The RA Research Group: Instructions for scoring Referential Activity (RA) in transcripts of spoken narrative texts. Ulm, Germany: Ulmer Textbank (1992)
5. Bucci, W., Maskit, B.: Building a Weighted Dictionary for Referential Activity. *Computing Attitude and Affect in Text*, pp. 49–60 (2005)
6. Pennebaker, J.W., Boyd, R.L., Jordan, K., et al.: The Development and Psychometric Properties of LIWC2015 (2015)
7. Brombert, V.: Virginia woolf – “death is the enemy.” *Hudson Rev.* **63**(3), 429–444 (2010)
8. Thomson, J.: Virginia Woolf and the case of Septimus Smith. *San Francisco Jung Inst. Libr. J.* **23**(3), 55–71 (2004)
9. Kenney M.S., Kenney, Jr. J.E.: Virginia woolf and the art of madness. *Massachusetts Rev.* **23**(1), 161–185 (1982)
10. Caramagno, T.C.: Manic-depressive psychosis and critical approaches to Virginia Woolf’s life and work. *PMLA* **103**(1), 10–23 (1988)
11. Post, F.: Verbal creativity, depression and alcoholism: an investigation of one hundred American and British writers. *Br. J. Psychiatry* **168**(5), 545–555 (1996)
12. Osgood, C.E., Walker, E.G.: Motivation and language behavior: a content analysis of suicide notes. *J. Ab. Soci. Psychol.* **59**(1), 58 (1959)
13. Xu, L., Song, W.: Suicidal idea detection based on chinese microblog language features. *J. North Univ. China (Nat. Sci. Ed.)* **40**(4), 350–357 (2019)
14. Coppersmith, G., Leary, R., Crutchley, P., et al.: Natural language processing of social media as screening for suicide risk. *Biomed. Inf. Insights* **10**, 1178222618792860 (2018)
15. Guan, L., Hao, B., Liu, T., Cheng, Q., et al.: A pilot study of differences in behavioral and linguistic characteristics between Sina suicide microblog users and Sina microblog users without suicide idea. *Chin. J. Epidemiol.* **36**(5), 421–425 (2015)
16. Huang, X., Zhang, L., Chiu, D., et al.: Detecting suicidal ideation in chinese microblogs with psychological lexicons. In: 2014 IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops, pp. 844–849. IEEE (2014)
17. Wang, T., Zhou, Z., Zhu, T., et al.: Mood disorder patients’ language features on their microblogs. *Int. J. Embedded Syst.* **7**(1), 34–42 (2014)
18. Cheng, Q., Li, T.M.H., Kwok, C.L., et al.: Assessing suicide risk and emotional distress in chinese social media: a text mining and machine learning study. *J. Medical Internet Res.* **19**(7), e243 (2017)
19. Egnoto, M.J., Griffin, D.J.: Analyzing Language in Suicide Notes and Legacy Tokens. *Crisis* (2016)
20. Pestian, J., et al.: Sentiment analysis of suicide notes: a shared task. *Biomed. Inf. Insights* **5s1**, BII.S9042 (2012). <https://doi.org/10.4137/BII.S9042>
21. Guan, L., Hao, B., Cheng, Q., et al.: Behavioral and linguistic characteristics of microblog users with various suicide ideation level: an explanatory study. *Chin. J. Public Health* **31**(3), 349–352 (2015)
22. Guan, L., Hao, B., Cheng, Q., et al.: Identifying Chinese microblog users with high suicide probability using internet-based profile and linguistic features: classification model. *JMIR Mental Health* **2**(2), e17 (2015)

23. Zhang, L., Huang, X., Liu, T., Li, A., Chen, Z., Zhu, T.: Using linguistic features to estimate suicide probability of chinese microblog users. In: Qiaohong, Zu., Bo, Hu., Ning, Gu., Seng, Sopheap (eds.) HCC 2014. LNCS, vol. 8944, pp. 549–559. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-15554-8\\_45](https://doi.org/10.1007/978-3-319-15554-8_45)
24. Pestian, J., Nasrallah, H., Matykiewicz, P., Bennett, A., Leenaars, A.: Suicide note classification using natural language processing: a content analysis. *Biomed. Inf. Insights* **3**, BII.S4706 (2010). <https://doi.org/10.4137/BII.S4706>
25. Gómez, J.M.: Language technologies for suicide prevention in social media. In: Proceedings of the Workshop on Natural Language Processing in the 5th Information Systems Research Working Days (JISIC), pp. 21–29 (2014)
26. Li, T., Chau, M., Yip, P., Wong, P.: Temporal and computerized psycholinguistic analysis of the blog of a Chinese adolescent suicide. *Crisis* **35**(3), 168–175 (2014). <https://doi.org/10.1027/0227-5910/a000248>
27. Kernot, D., Bossomaier, T., Bradbury, R.: The impact of depression and apathy on sensory language. *Open J. Modern Linguist.* **7**(1), 8–32 (2016)
28. Kernot, D., Bossomaier, T., Bradbury, R.: The stylometric impacts of ageing and life events on identity. *J. Quant. Linguist.* **26**(1), 1–21 (2019)
29. Dakos, V., Bascompte, J.: Critical slowing down as early warning for the onset of collapse in mutualistic communities. *Proc. Natl. Acad. Sci.* **111**(49), 17546–17551 (2014)
30. Kernot, D.: The Identification of Authors using Cross Document Co-Referencing. The University of New South Wales Master Degree Thesis (2013)
31. Kernot, D., Bossomaier, T., Bradbury, R.: Novel text analysis for investigating personality: identifying the dark lady in Shakespeare’s sonnets. *J. Quant. Linguist.* **24**(4), 255–272 (2017)
32. Kernot, D., Bossomaier, T., Bradbury, R.: Using Shakespeare’s sotto voce to determine true identity from text. *Front. Psychol.* **9**, 289 (2018)
33. Kubát, M., Matlach, V., Čech, R.: Quantitative Index Text Analyzer. RAM-Verlag, Germany (2014)
34. Altmann-Fitter. Altmann-Fitter User Guide (2013)
35. Rayson, P., Leech, G.N., Hodges, M.: Social differentiation in the use of English vocabulary: some analyses of the conversational component of the British national corpus. *Int. J. Corpus Linguist.* **2**(1), 133–152 (1997)
36. Schmid, H.J., Fauth, J.: Women’s and Men’s style: fact or fiction? new grammatical evidence. In: *Corpus Linguistics Conference*, Lancaster (2003)
37. Zhou, J., Su, Q., Liu, P.: A metaphorical analysis of five senses and emotions in mandarin Chinese. In: Hong, Jia-Fei., Zhang, Yangsen, Liu, Pengyuan (eds.) CLSW 2019. LNCS (LNAI), vol. 11831, pp. 607–617. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-38189-9\\_62](https://doi.org/10.1007/978-3-030-38189-9_62)
38. Fernandino, L., Binder, J.R., Desai, R.H., et al.: Concept representation reflects multimodal abstraction: a framework for embodied semantics. *Cereb. Cortex* **26**(5), 2018–2034 (2015)
39. Lynott, D., Connell, L.: Modality exclusivity norms for 423 object properties. *Behav. Res. Methods* **41**(2), 558–564 (2009)
40. Chen, I.H., Zhao, Q., Long, Y., et al.: Mandarin Chinese modality exclusivity norms. *PLoS one* **14**(2), (2019)
41. Liu, H., Huang, W.: Quantitative linguistics: state of the art, theories and methods. *J. Zhejiang Univ. (Human. Soc. Sci.)* **42**(2), 178–192 (2012)
42. Mehri, A., Jamaati, M.: Variation of Zipf’s exponent in one hundred live languages: a study of the holy bible translations. *Phys. Lett. A* **381**(31), 2470–2477 (2017)
43. McCowan, B., Doyle, L.R., Hanser, S.F.: Using information theory to assess the diversity, complexity, and development of communicative repertoires. *J. Comp. Psychol.* **116**(2), 166 (2002)

44. Ferrer-i-Cancho, R., Elvevåg, B.: Random texts do not exhibit the real Zipf's law-like rank distribution. *PLoS One* **5**(3) (2010)
45. Kalankesh, L.R., Stevens, R., Brass, A.: The language of gene ontology: a Zipf's law analysis. *BMC Bioinf.* **13**(1), 127 (2012)
46. Perline, R.: The Random Division of the Unit Interval and the Approximate -1 Exponent in the Monkey-at-the-Typewriter Model of Zipf's Law. *arXiv preprint 1512.02177* (2015)
47. Baixeries, J., Elvevåg, B., Ferrer-i-Cancho, R.: The evolution of the exponent of Zipf's law in language ontogeny. *PloS one* **8**(3) (2013)
48. Dally, P.: *The Marriage of Heaven and Hell: Manic Depression and the Life of Virginia Woolf*. St Martin's Press, New York (1999)
49. Bishop, E.: *A Virginia Woolf Chronology*. Macmillan, Basingstoke (1989)
50. Yan, G., Lu, Y.: Study on clustering method of multidimensional sequence samples. *Stat. Decis.* **256**(4), 29–30 (2008)