

How well do we know ourselves? Identifying suicide markers in online communication: A case study of a graduate student's writing

Tetiana Tytko

University of Maryland, College Park
tтыtko@umd.edu

Mary Catherine Augstkalns

Ohio University
ma256208@ohio.edu

Linguistic features of suicide notes have been analyzed and compared to forged suicide notes and ordinary letters (Osgood & Walker 1959), diaries, and reports (Baddeley, Daniel, & Pennebaker 2011). However, the real focus should be on determining suicide warning signs in one's writing prior to the suicide note. Contemporarily, text messages can illustrate informal communication and showcase the emotional state a person experiences at the time of texting (Xie & Kang 2015). This research investigates linguistic features of genuine suicide notes present in the online communication of a suicidal person before attempting suicide. Also, this paper determines common themes discussed in these texts. Personal correspondence of a suicidal person was analyzed over five months. Two corpora (Corpus 1 and Corpus 2) were studied using the following measures: Osgood and Walker's (1959) allness terms, Hu and Liu's (2004) negative sentiment lexicon, and English pronouns. AntConc (Anthony 2019) along with the Linguistic Inquiry and Word Count program (Pennebaker, Booth & Francis 2001) were utilized to find frequencies and categorize the person's psychological states. Findings indicate that the number of allness terms was significantly higher in Corpus 2, immediately prior to the suicide attempt, whereas first-person singular pronouns were equally used in both corpora. Further, Corpus 2 contained more words of positive evaluation which contradicts previous research findings. The discourse analysis of the participant's messages demonstrated an equal quotient of themes connected to blame discussed in both corpora. These results show that personal writing reflects people's emotional states better than they think, as the analysis of the texts revealed that themes discussed in Corpus 2 before the suicide attempt were present even five months before, in Corpus 1, which contradicts the participant's self-report. Therefore, it is the interlocutor's duty to utilize the tools at hand to assist such people in crisis.

1. Introduction

According to the American Psychiatric Association (as cited in Weir 2019), suicide is the second leading cause of death among youth and the fourth leading cause of death among people who reached the age of 35 in the U.S. Because of social media and the means of telecommunication, modern society is currently witnessing an increased number of suicide cases. This depressing news becomes viral and is exhaustively discussed by the public. Reporting such news can escalate this issue by amplifying the number of suicide cases (Stack 2020). Instead of passively observing this phenomenon, the population needs to receive real methods to take social actions and identify proactive factors in preventing suicide (Suicide Prevention 2019).

In response, scientists from different fields have been investigating this equally social and individual issue for more than half a century, and yet they have only concluded that the causes of suicide are complex, and there are no identical suicides (e.g., Baldessarini & Hennen 2004; Harwitz & Ravizza 2000; Furr et al. 2001; Bennell et al. 2011; Mandrusiak et al. 2006; Sanderson & Sanderson 1997; Stirman & Pennebaker 2001). Although “predicting when someone will commit suicide has been nearly impossible” (American Psychiatric Association 2003:3), numerous researchers around the globe continue to work independently on reducing such a high suicide rate (e.g., Huang, Goh & Liew 2007; Hinduja & Patchin 2010; Pestian et al. 2017). By addressing this problem in their research, they can contribute to the comprehension of suicide cases and inform the public about when suicide may occur as well as how it can be prevented.

While clinicians, psychiatrists, and counselors have extensively analyzed suicide-related factors such as the time when suicide occurred and motives and methods, linguists’ infrequent studies have generally been limited to investigation of suicide notes (e.g., Baldessarini & Hennen 2004; Osgood & Walker 1959) or literary works (e.g., Berman & Schiff 2000; Lester 1993; Stirman & Pennebaker 2001). This is unfortunate since written communication can be a valuable source of information reflecting mental states. Moreover, it is accessible and understandable for most people as they interact daily through informal written messages. From the linguistic perspective, the text analysis of written discourse produced before a suicide and examination of its content can identify the linguistic markers and typical themes indicating someone may commit suicide.

For this reason, suicide markers must first be identified by linguists in the written discourse of a suicidal person and presented for the public so any person has competence to spot these linguistic patterns signaling the shift in emotional states that can lead to suicide. Family members and close friends whose loved ones suffer from depression need to be provided tangible mechanisms to be able to detect factors for suicidal behavior in order to seek specialists' treatment while help is still available. Therefore, this paper aims to investigate this issue by identifying and presenting common suicide markers in online communication along with the themes discussed by a person who shows signs of depression and is inclined to attempt suicide.

2. An overview of recent suicide discourse

Normally, written works can provide linguists with invaluable information about the reasons for people's suicidal acts and their psychological state before death. Research in psychology and clinical linguistics has found that the way people communicate can reflect their emotional and mental states (Osgood 1960; Schoene & Dethlefs 2016). To uncover the change in linguistic expressions associated with shifting psychological states, some researchers have examined samples written by suicidal authors (Sanderson & Sanderson 1997; Stirman & Pennebaker 2001). Most of them were well-known public figures producing literary works such as novels or poems who kept entries in a diary. Alternatively, other studies conducted a thorough investigation using a database of suicide notes in order to reveal typical themes discussed by suicidal people experiencing heightened emotions prior to committing suicide (Al-Mosaiwi & Johnstone 2018; Osgood & Walker 1959; Pestian et al. 2010).

For over half a century, the language of suicide notes has been of great interest to linguists. One of the first interdisciplinary studies in psychology employing linguistic text analysis compared over a hundred genuine suicide notes to a collection of ordinary personal letters and simulated suicide notes (Osgood & Walker 1959). The authors found that suicide notes, in general, were not considerably different from letters to family and friends in the degree of stereotypy. For example, many people expressed their gratitude and love to spouses and parents, and often the overall tone of the note was positive. Still, genuine suicide notes significantly differed from forged suicide notes as they reflected suicide motives as well as revealed personal conflict. In genuine suicide notes, Osgood & Walker emphasized the increase of "evaluative assertions about Ego" and absolute terms such as *always*, *ever*, *never*, *totally*, etc. (1959:62). Although the linguistic examination of suicide notes provided

researchers with the comprehensible classification of linguistic features which are reliable predictors of suicidal intent (Jones & Benell 2007; Osgood & Walker 1959; Schoene & Dethlefs 2016), the real focus should be on determining suicide warning signs in one's writing prior to the suicide note.

The other approach to identifying suicide markers lies in analyzing written artistic works produced by songwriters (e.g., Lightman et al. 2007), writers (e.g., Sanderson & Sanderson 1997), and poets (e.g., Stirman & Pennebaker 2001). By means of text and discourse analysis, written samples of suicidal and nonsuicidal artists were analyzed and compared for suicidal linguistic markers. The overall findings suggested that suicidal artists use significantly more words pertaining to self-reference than their nonsuicidal counterparts. Additionally, the frequency of first-person singular pronouns increases as their suicidal thoughts become more prominent with the progression of their self-absorption, while references to other people considerably decrease. These studies showcase suicidal warning signs that are frequently overlooked due to the stereotypes of popular culture. Going back to the days of Aristotle, who wrote of the state of "melancholia" and linked it to enhanced creativity, many songwriters, authors, and poets have been known to experience depressive states (Bailey, 2003). These results are not representative of the population as the research was focused upon a specific group of people susceptible to depressive states. The linguistic features of suicide notes therefore need to be further investigated in written discourse of the average person.

Responding to this call, Baddeley et al. (2011) conducted a thorough case study on language use and suicide prediction. The researchers studied the suicide of Henry Hellyer, an Australian traveler who died at the age of 42. Hellyer's written works (e.g., letters, diaries, and reports) were analyzed in order to detect salient linguistic features that changed over the course of seven years of his life. Similarly, Baddeley et al. noticed a significant spike in first-person singular pronoun use with a decrease in the use of the first-person plural pronoun, which supports the findings of the artists' written work analysis. Another marker of suicidality detected by the scholars was the extensive use of words with negative emotional coloring or negative sentiment words (Al-Mosaiwi & Johnstone 2018). Because these words are frequent signs of depression, the negative sentiment words were identified as another marker signaling emotional shift. Still, contradictions exist.

Some prior studies have shown that the use of positive emotional tone can increase before the act of suicide, whereas negative word use may decrease. In other words, some individuals suffering from an increased

depression may face moments of hopefulness right before committing suicide (Barnes, Lawal-Solarin & Lester 2007; Lester 2014). This explains why numerous suicidal acts were not foretold. Despite the fact that verbal signs can be misleading in predicting suicide, Baddeley et al. (2011) believe that text analysis is one of the most effective approaches in predicting suicide and preventing people from taking their own lives. Although their document sample size was limited, their findings can inform clinicians regarding the content of the words that depressed clients use, which mental specialists can take into consideration while providing treatment (Pestian et al. 2017). Still, despite some discrepancy in the number of positive or negative emotion words that would indicate suicidal thoughts, the researchers collectively argued that informal writing samples such as diary entries, letters, essays, and poems can reflect indicators of amplified self-focus, breakdown in social interaction, and depressive states reflected in their writing.

As suggested by Baddeley et al. (2011), a specific genre affects to what extent people can express their emotions. The scientists observed the quotient of suicide markers was higher in personal letters and diaries. This means that these markers tend to appear more frequently in the informal register. For instance, there was a shift in Hellyer's writings from professional to personal works that, in turn, stipulated the change in his language patterns (Baddeley et al. 2011). Based on these findings, in the modern multimodal world, signs of suicidal thoughts cannot generally be found in professional discourse (e.g., reports, formal letters, papers). However, more informal channels of communication (e.g., social media, text messages, blogs) can better serve as modes for expressing people's distress and disappointment via language use.

Overall, the results obtained through the analysis of suicide notes and written samples produced by suicidal authors comply with each other in the augmented use of "I" in the suicidal state, self-reference, extensive use of words with a negative connotation, and absolute terms. The researchers further agreed that these linguistic suicide markers are prevalent in informal register conversations. On the other hand, the results were inconclusive as for the prevalence of negative or positive emotions. Therefore, more research is needed to shed light on the correlation between suicide markers found in literary works and suicide notes. Likewise, future research can bridge the gap between scientific findings and simple tools people can use to identify warning signs in a loved one's writing. While some forms of informal writing are inaccessible or unethical to access, other channels of written communication can be selected for in-depth examination.

Presently, computer-mediated communication is most commonly used for personal correspondence. Thus, most informal communication is done online via text messages, emails, and social media posts. Despite the fact that in recent years there has been an increased number of suicides among young people desperately seeking help by uploading videos on YouTube or trying to draw public attention by posting “this hurts too much,” on their blog (White & Kral 2014:122), there has been little research on the online communication of suicidal people. However, this is starting to change. A few recent studies have compared hand-written suicide notes with suicide posts (Leis et al. 2019) and attempted to study signs of depression in tweets that could indicate suicidal behavior (O’Dea et al. 2017). In fact, the in-depth investigation of text messages with family members and friends is of consequence since it may show how one’s thoughts, feelings, and language change over time. Thus, this study aims to fill the gap in the previous body of research by narrowing down the data to be analyzed to online communication, i.e., text messages to family and friends.

To further justify the analysis of text messages, this type of communication is an excellent illustration of the informal register through which a person can express their true emotions. The informal register is that which is used when talking with loved ones, typically family and friends. It is conversational, casual, and intimate when compared to the professionalism of the formal register or the factuality of the neutral register, and thus creates an environment of genuineness that does not exist in other registers (Formal and informal language, n.d.). Because most text messages, as well as social media posts, are usually shared instantly, they reflect authentic feelings, especially when they are sent under heightened emotions. But unlike posts that can be deleted, text messages cannot be unsent. Thus, they can serve as evidence of the actual emotional state one experienced at the time of sending a text (Xie & Kang 2015).

Taking into consideration the extensive amount of research on the language of suicide notes and the insufficient number of empirical studies on suicide markers in online communication, this research aims to investigate the linguistic features of genuine suicide notes that are present in the online communication of a suicidal person before the suicide as well as discover common themes discussed in these text messages. Based on the findings of the abovementioned studies, this research proposes the following hypotheses: 1) linguistic features of genuine suicide notes will be present in the online communication of the suicidal person immediately prior to the suicide attempt; 2) linguistic features of genuine suicide notes will not be present to a significant degree in the online communication when no suicide attempt was made. Additionally, this paper aims to

answer the research question: What common themes and topics are discussed by the suicidal person in their online communication prior to the suicide attempt?

3. Methods

The participant of this case study is a female graduate student in a public university who attempted suicide and in the past year has voluntarily committed herself for psychiatric treatment. She gave consent to analyze her electronic messages that she sent over the course of five months.

The raw data for this research consists of the participant's personal correspondence, text messages to family (one of her parents) and friends (her best friend), in two months: September (15,995 words) and January (32,317 words). In Corpus 1, September, the participant reported having no suicidal thoughts, whereas in Corpus 2, January, she attempted suicide after being depressed and experiencing self-injurious behaviors. Given that the two corpora vary significantly in length, they were equalized by truncating Corpus 1 and Corpus 2 to 10,000 words. Thus, the first 10,000 words of the September corpus, which were most distant from the suicide attempt, were chosen for Corpus 1, and the last 10,000 words of January, leading up to the suicide attempt immediately following the last text message, were chosen for Corpus 2. This was seen as justified in order to produce the widest contrast. In general, abbreviation of corpora is a common practice for reasons such as technical limitations, time, and copyright (Biemann et al 2013; Wynne, 2005).

In order to analyze the data, three measures were selected from the analyzed literature: Osgood and Walker's (1959) allness/absolute terms, Hu and Liu's (2004) negative sentiment/emotion lexicon, and three types of English pronouns (subject personal pronouns, object personal pronouns, and reflexive pronouns). Additionally, to find the frequencies of the abovementioned measures in each corpus, AntConc (Anthony 2019) was used as the concordancer. Later, using Rayson's (2016) log-likelihood calculator, frequencies for all categories in the two corpora were compared in order to determine whether differences were statistically significant. Also, the Linguistic Inquiry and Word Count (LIWC) program (Pennebaker, Booth & Francis 2001) was utilized to categorize the person's emotional and psychological states as well as to track the change of the parts of speech use. While AntConc was used for individual analysis and comparison of the frequencies of suicide notes' measures such as pronouns, negative sentiment words, and allness terms, the text analysis program LIWC was utilized for a more extensive analysis of the

written text. On the basis of numerous language dimensions, the LIWC counted and contrasted first person singular pronouns such as *I* and *me* with first person plural pronouns such as *us* and *our*, compared negative emotion words and positive emotion words, as well as identified the saliency of terms specific to suicide discourse such as *death*, *die*, etc. Thus, the LIWC program complemented the results obtained through individual frequency search considering their contextualized usage. Finally, discourse analysis was used in determining and categorizing the common themes discussed in online communication in both corpora.

4. Results and discussion

The data analysis started with identifying frequencies of Osgood and Walker's (1959) allness/absolute terms in both corpora and then comparing them. Osgood and Walker used this measure to identify elevated emotions and strong emotional drive. They claimed that such words like *always*, *never*, *all*, *everything*, *nothing* were common in suicide messages and relatively less so in letters to relatives and friends. To recognize the frequencies results trending towards significance, Rayson's (2016) log-likelihood calculator was used. The higher the log-likelihood (*LL*) value, the more significant the difference between the two scores is: $LL^3 3.84$ is significant at $p < 0.05$; $LL^3 6.63$ is significant at $p < 0.01$; $LL^3 10.83$ is significant at $p < 0.001$; and $LL^3 15.13$ is significant at $p < 0.0001$. The *Effect Size for Log Likelihood (ELL)* measure (Johnston et al. 2006) was used, included with Rayson's calculator, to determine the effect sizes of the *LL* values. The results obtained are presented in Table 1.

Allness terms	Corpus 1 (ptw)	Corpus 2 (ptw)	LL	ELL
Absolutely	1	2	0.01	0.00004
All	46	38	0.76	0.00001
Always	8	16	2.27	0.00005
Completely	2	4	0.68	0.00003
Constantly	1	0	1.39	0.00010
Definitely	7	0	9.70**	0.00039
Entire	1	0	1.39	0.00010
Ever	6	54	44.17****	0.00065
Every	15	24	2.10	0.00004
Everyone	5	5	0.00	0.00000
Everything	3	4	0.14	0.00001
Full	4	6	0.40	0.00001
Must	1	4	1.93	0.00011
Never	8	21	6.04*	0.00011
Nothing	6	4	0.40	0.00001
Totally	1	4	1.93	0.00011
Whole	6	11	1.49	0.00003
Total	122	199	18.65***	0.00018

Note: ptw = per 10000 words; LL = log-likelihood value; ELL = effect size for log likelihood; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; **** = $p < 0.0001$.

Table 1. The frequencies of the allness terms in Corpus 1 and Corpus 2

Log-likelihood tests show that the overall frequency of allness terms was significantly greater in Corpus 2. Corpus 1 contains only one allness term, *definitely*, that occurs significantly more frequently. Corpus 2, however, includes two more terms, *ever* and *never*, significantly more frequently than in Corpus 1. These results indicate that the participant used considerably more absolute words in Corpus 2 than in Corpus 1. In other words, in the month when she started thinking about suicide, she also started using more extreme and rigid terms. Thus, these findings support Hypothesis 1, as the linguistic features of genuine suicide notes, allness terms, were present in the online communication of the suicidal person prior to the suicide attempt. However, Hypothesis 2 was not supported as a considerable number of allness terms was found in Corpus 1, when the participant self-reported no suicidal feelings to a professional.

Another strong marker of suicidal thinking is elevated use of negative sentiment words (Osgood & Walker 1959; Baddeley et al. 2011). However, some studies found that words with negative evaluation such as *sad*, *bad*, *depressed*, *miserable* were not prevalent in a person's writing

right before the suicide (Barnes, Lawal-Solarin & Lester 2007; Lester 2014). The frequencies of Hu and Liu's (2004) negative sentiment words were found in both corpora and they are shown in Table 2.

Negative sentiment words	Corpus 1 (ptw)	Corpus 2 (ptw)	LL	ELL
Afraid	2	0	2.77	-
Alone	0	2	2.77	-
Bad	11	13	0.17	0.00000
Depressed	2	4	0.68	0.00003
Down	2	0	2.77	-
Hurting	0	1	0.00	-8.52
Lost	1	3	1.05	0.00008
Sad	6	4	0.40	0.00001
Stuck	1	2	0.34	0.00004
Stupid	2	0	2.77	-
Terrible	2	5	1.33	0.00005
Unsure	0	2	2.77	-
Useless	0	1	0.00	-8.52
Total	29	37	0.97	0.00001

Note: ptw = per 10000 words; LL = log-likelihood value; ELL = effect size for log likelihood; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; **** = $p < 0.0001$.

Table 2. The frequencies of negative sentiment words in Corpus 1 and Corpus 2

Table 2 shows that there is no significant difference between the frequencies of the negative sentiment words. The individual frequencies of the negative sentiment words, such as *terrible*, *depressed*, and *bad* are higher in Corpus 2. Also, Corpus 2 does contain more negative emotion words overall, but this difference is not significant. This is aligned with the claim that suicidal people are inclined to give a more negative evaluation. Likewise, the obtained findings support Hypothesis 1, as there was a high occurrence of negative sentiment words before the suicide attempt. Surprisingly, Hypothesis 2 was not borne out, as there was an equally high amount of negative emotion words in both corpora.

Regarding the types of pronouns, numerous studies reported a significant increase in the use of personal pronouns (e.g., *I*, *me*, *my*), as people in a suicidal state often focus on themselves and their own problems and struggles and become disconnected with other people (Baddeley et al. 2011; Osgood & Walker 1959). Table 3 shows the frequency of pronouns in each corpus and their comparison.

Pronouns	Corpus 1 (ptw)	Corpus 2 (ptw)	LL	ELL
I	619	575	1.62	0.00001
you	83	142	15.65***	0.00017
he	50	107	21.17****	0.00024
she	20	124	83.58****	0.00098
it	230	242	0.31	0.00000
we	26	29		
they	64	48	0.16	0.00000
me	70	70	0.00	0.00000
him	22	26	0.33	0.00001
her	25	61	15.54***	0.00021
us	9	3	3.14*	0.00009
them	23	23	0.00	0.00000
myself	7	13	1.83	0.00004
yourself	0	2	2.77	-7.13
herself	0	4	5.55*	-4.36
himself	0	1	1.39	-8.52
itself	1	1	0.00	0.00000
themselves	2	0	2.77	-7.13
Total 1st person singular	693	658	0.91	0.00001
Total other pronouns	555	813	48.95	0.00037

Note: ptw = per 10000 words; LL = log-likelihood value; ELL = effect size for log likelihood; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; **** = $p < 0.0001$.

Table 3. The frequencies of pronouns in Corpus 1 and Corpus 2

Table 3 reveals that, in fact, the total number of first-person singular pronouns is higher in Corpus 1 and lower in comparison to other pronouns in Corpus 2, but this difference is not significant. At the same time, Log-likelihood tests show that the overall frequency of pronouns is significantly greater in Corpus 2, as well as such pronouns as *you*, *he*, *she*, *her*, and *herself*. Indeed, the total number of first-person singular pronouns is higher in Corpus 1 and considerably lower in comparison to other pronouns in Corpus 2. These findings are contrary to expectations and do not support either of the Hypotheses.

Finally, after both corpora were analyzed through the LIWC (Pennebaker, Booth & Francis 2001), the percentages of total words within the text were provided and compared to the average for online, text-based

communication using a 100-point scale (0 equals a low dimension, 100 equals a high dimension). Table 4 shows the obtained results.

Traditional LIWC Dimension	Participant Data		Average for Online Communication
	Corpus 1	Corpus 2	
I-words (I, me, my)	7.6	5.2	5.51
Social words	11.8	8.9	9.71
Positive emotions	2.8	3.4	4.57
Negative emotions	2.9	1.0	2.10
Cognitive processes	13.8	10.5	10.77
Summary Variables			
Analytic	14.8	47.8	44.88
Clout	43.9	54.0	43.16
Authenticity	42.2	33.1	55.66
Emotional tone	24.1	70.9	63.35

Table 4. The LIWC analysis of the text of the first and second corpora

These numbers support the previous findings. The quotient of I-words is higher in Corpus 2 as well as the negative emotion words. Surprisingly, this analysis also reveals that if positive emotion words are considered, their percentages are higher in Corpus 2. This supports Lester (2014) who detected the increase of words reflecting positive emotions as well as the decrease of personal pronoun use in a female college student's diary who died by suicide. Two more dimensions in the summary variables stand out from the rest due to their noticeable deviation from the average and each other. Analytic dimension reflects how a person's thinking adheres to formal, logical, and analytical thinking. The participant's analytical thinking was significantly lower in Corpus 1 than in Corpus 2 and when compared to the average. This appears to indicate that in Corpus 1, the participant wrote more narratively without making any connections to the future or the past. According to Osgood and Walker (1959), the present tense is one of the salient features of the discourse of suicide notes. Finally, the emotional tone includes the overall contexts of positive and negative dimensions, where numbers that are higher than 50 suggest more positive tone, and numbers below 50 signify a more negative tone. Therefore, because of a significant difference between the two corpora, a conclusion may be drawn that the person used more positive words in her online communication before attempting suicide than five months before.

These results are consistent with some research and contradict others. Contrary to expectations, this study did not find any salient linguistic features of suicide notes, except for the allness terms, that were present in the online communication of the suicidal person prior to the suicide attempt that were not found to a significant degree in her online communication five months before. In turn, the percentages of negative sentiment words and first-person pronouns were higher in Corpus 1 than Corpus 2. Thus, Hypothesis 1 was supported partially by the finding of this research, whereas Hypothesis 2 was not borne out.

In order to answer the research question about the common patterns in online communication of the suicidal person, discourse analysis was used. As mentioned by Osgood and Walker (1959), genuine suicide notes include specific criticism of a person or a situation, and they rarely are vague and ambiguous. Table 5 shows the determined themes the participant discussed either with her friend or her parent in communication. One unanticipated finding was that these themes were discussed in both corpora; hence, even five months before her suicide attempt the participant was complaining about her life and blaming people and circumstances to an unexpected degree outside the norm.

Themes	Corpus 1	Corpus 2
Me, myself, and I	<i>I have to do all my own work myself</i>	<i>I still exist</i>
Boyfriend	<i>He wouldn't say a word to me</i>	<i>He probably lacks the mental capacity for shame, you're right</i>
Boyfriends' parents	<i>We were waiting forever for them to come...</i>	<i>They could have done better job judging this.</i>
Dad	<i>It's so frustrating because he accuses me of being hysterical when there is a crisis. He wouldn't listen to me...</i>	<i>My boyfriend and dad both just fart whenever but I can't do that.</i>

School	<i>Will school ever not consume me whole?</i>	<i>I'm not a college person so the environment is shitty for me. I'm surprised I'm not depressed yet.</i>
Friends	<i>I don't want people to think like I'm trying to act like I'm worse off.</i>	<i>I can't dump everything on my friends because then they would resent me and stop being my friends I think?</i>
Loneliness/Hopelessness	<i>How sad and empty must your life be?</i>	<i>But sometimes I don't have anyone else to talk to</i>

Table 5. Common themes discussed in Corpus 1 and Corpus 2

These results indicate that the participant's complaints were more often specific than not. Based on the analysis of the participant's personal text messages in both months, more often she, without realizing, was equally blaming specific people, rather than complaining about the loneliness and hopelessness of her life. Most importantly, this means that she was in a suicidal state even five months prior to her actual suicide attempt, even though she was not aware of that.

5. Conclusion

This research attempts to contribute to the growing understanding of suicidal language by analyzing the online communication of a suicidal person. Specifically, this research aimed to identify if some of the most prominent features of suicide notes (allness terms, negative sentiment words, and first-person pronouns) are present in significant enough amounts to be noticed in text messages of suicidal people. The present results are significant in at least two major respects.

First, because the findings were contrary to expectations, further examination will be required to identify the time when the participant experienced the shift in her emotional state and started having suicidal thoughts. So far, only the absolutist dichotomous thinking and the increased use of allness words were noticeable in Corpus 2; however, I-

pronouns and negative words were more ubiquitous in Corpus 1. In fact, there were significantly more positive emotion words before the suicide attempt. These findings do not permit the researchers to draw conclusions on linguistic markers that can signal a suicidal state. But these limitations can be addressed in the future.

The second claim that can be made is that writing does reflect people's emotional and mental states better than they think. The discourse analysis revealed that the participant experienced negative sentiments regarding her family and personal life five months before the suicide attempt. Evidently, depression and possibly suicidal thoughts appeared even before Corpus 1, but she was not fully aware of them. Thus, additional samples of the participant's communication (including post-therapy) will be needed to establish the trend in the development of suicidal behavior, and the full extent of the existing corpora, especially the larger Corpus 2, can be utilized to identify trends in the participant's suicidal ideation. Plus, a semi-structured interview may be conducted in the future for achieving triangulation on the present research. Additionally, the construction of control corpora from someone who has self-reported no suicidal thoughts or behavior may be necessary in each of these time frames, ideally from a person of the same gender, age, and occupation (grad student) as the current participant.

Further limitations of this study include a lack of counterfactual corpora, coupled with a lack of concrete positive emotion analysis. Additionally, the two corpora were truncated in order to equalize the data. While it is often not desirable to shorten already seemingly small corpora, since a long artefact in any small corpus can exert undue influence (Wynne, 2005), the individual text messages in each corpus were judged to generally be of short enough length to justify the truncation process for ease of comparability, although there still exists the possibility that this modified the results. It is also important to note that this study is currently ongoing and many of these limitations can still be addressed.

Still, this research has important implications for developing a framework with specific measures for the analysis of suicidal language in people's writing. Eventually, these results could also inform the designers of a machine learning approach that can identify the thought markers of suicidal subjects (Pestian et al. 2017). Therefore, suicide prevention can and should start with the investigation of suicide from different fields, especially because research from different disciplines can be complementary. The collaboration of linguists, clinicians, counselors, and programmers can contribute to developing innovative tools determining the written clues that will help stop deaths by suicide.

Despite the limitations, the findings could inform scientists and practitioners across many disciplines regarding words and expressions that suicidal people use. The appearance of some of the linguistic features in one's writing can reflect the psychological change that a person undergoes. These can be feasible warning signs that should be noticed not only by specialists but also by family and friends for predicting and preventing the suicide of the people they love.

REFERENCES

- American Psychiatric Association. 2003. Practice guideline for the assessment and treatment of patients with suicidal behaviors. *American Journal of Psychiatry*, 160(Suppl). 1-60.
- Al-Mosaiwi, Mohammed & Johnstone, Tom. 2018. In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science*, 6(4). 529-542.
- Anthony, Laurence. 2019. AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/software.html>. (18 September, 2020).
- Baddeley, Jenna L. & Daniel, Gwyneth R. & Pennebaker, James W. 2011. How Henry Hellyer's use of language foretold his suicide. *Crisis*.
- Bailey, Deborah Smith. 2003. Considering creativity – The 'Sylvia Path' effect. *Monitor on Psychology* 34(10). <https://www.apa.org/monitor/nov03/plath>. (10 October, 2020).
- Baldessarini, Ross & John Hennen. 2004. Genetics of Suicide: An Overview. *Harvard Review of Psychiatry* 12(1). 1-13. doi:10.1080/714858479.
- Barnes, Donna H. & Lawal-Solarin, Foluso W. & Lester, David. 2007. Letters from a suicide. *Death Studies* 31(7). 671-678.
- Berman, Jeffrey & Schiff, Jonathan. 2000. Writing about suicide. In Anderson, Charles & MacCurdy, Marian (eds.), *Writing and healing: Toward an informed practice*. Refiguring. 301-322.
- Bennell, Craig & Jones, Natalie J. & Taylor, Alyssa. 2011. Determining the authenticity of suicide notes: Can training improve human judgment? *Criminal Justice and Behavior*, 38(7). 669-689.
- Biemann, Chris & Bildhauer, Felix & Evert, Stefan & Goldhahn, Dirk & Quasthof, Uwe & Schäfer, Roland & Simon, Johannes & Swiezinski, Leonard & Zesch, Torsten. 2013. Scalable construction of high-quality web corpora. *J. Lang. Technol. Comput. Linguistics* 28(2). 23-59.

- Formal and informal language. n.d. In *Cambridge Online Dictionary*. Retrieved from <https://dictionary.cambridge.org/us/grammar/british-grammar/formal-and-informal-language> (10 October, 2020).
- Harwitz, David & Ravizza, Luigi. 2000. Suicide and depression. *Emergency medicine clinics of North America*, 18(2). 263-271.
- Hinduja, Sameer & Patchin, Justin W. 2010. Bullying, cyberbullying, and suicide. *Archives of Suicide Research* 14(3). 206-221. doi:10.1080/13811118.2010.494133.
- Hu, Mingqing & Liu, Bing. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. 168-177.
- Johnston, Janis E. & Berry, Kenneth J. & Mielke, Paul W. 2006. Measures of effect size for chi-squared and likelihood-ratio goodness-of-fit tests. *Perceptual and Motor Skills*, 103. 412-414.
- Leis, Angela & Ronzano, Francesco & Mayer, Miguel A. & Furlong, Laura I. & Sanz, Ferran. 2019. Detecting signs of depression in tweets in Spanish: Behavioral and linguistic analysis. *Journal of medical Internet research*, 21(6). e14199.
- Lester, David. 1993. Suicide in creative women. *Suicide and the Creative Arts*. 189-201.
- Lester, David. 2014. *The "I" of the storm: Understanding the suicidal mind*. Walter de Gruyter GmbH & Co KG.
- Lightman, Erin J. & McCarthy, Philip M. & Dufty, David F. & McNamara, Danielle S. 2007. Using computational text analysis tools to compare the lyrics of suicidal and non-suicidal songwriters. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 29(29). 1217-1222.
- Mandrusiak, Michael & Rudd, David M. & Joiner, Thomas E. & Berman, Alan L. & van Orden, Kimberly A. & Witte, Tracy. 2006. Warning signs for suicide on the Internet: A descriptive study. *Suicide and Life-Threatening Behavior*, 36(3). 263-271.
- O'Dea, Bridianne & Larsen, Mark E. & Batterham, Philip J. & Cleave, Alison L. & Christensen, Helen. 2017. A linguistic analysis of suicide-related Twitter posts. *Crisis* 38(5). 319-329. doi:10.1027/0227-5910/a000443.
- Osgood, Charles E. 1960. The cross-cultural generality of visual-verbal synesthetic tendencies. *Behavioral Science* 5(2). 146-169. doi:10.1002/bs.3830050204.
- Osgood, Charles E. & Walker, Evelyn G. 1959. Motivation and language behavior: A content analysis of suicide notes. *The Journal of Abnormal and Social Psychology*, 59(1). 58.
- Pestian, John & Nasrallah, Henry & Matykiewicz, Pawel & Bennett, Aurora & Leenaars, Antoon. 2010. Suicide note classification using

- natural language processing: A content analysis. *Biomedical Informatics Insights* 3. BII.S4706. doi:10.4137/bii.s4706.
- Pestian, John P. & Sorter, Michael & Connolly, Brian & Cohen, Kevin B. & McCullumsmith, Cheryl & Gee, Jeffrey T. ... & STM Research Group. 2017. A machine learning approach to identifying the thought markers of suicidal subjects: A prospective multicenter trial. *Suicide and Life-Threatening Behavior*, 47(1). 112-121.
- Rayson, Paul. 2016. Log-likelihood and effect size calculator. Retrieved from <http://ucrel.lancs.ac.uk/llwizard.html>. (1 May, 2020).
- Sanderson, Richard K. & Sanderson, Rena. 1997. Suicide and literary biography: The case of Hemingway. *Biography*. 405-436.
- Schoene, Annika Marie & Dethlefs, Nina. 2016. Automatic identification of suicide notes from linguistic and sentiment features. In *10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*. 128-133.
- Stack, Steven. 2020. Media guidelines and suicide: A critical review. *Social Science & Medicine*, 112690.
- Stirman, Shannon W. & Pennebaker, James W. 2001. Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic medicine*, 63(4). 517-522.
- Suicide Prevention. 2019, September. *Center for disease control and prevention*. Retrieved from www.cdc.gov. (11 November, 2019).
- Weir, Kirsten. 2019, March. Worrying trends in U.S. suicide rates. *Monitor on Psychology*, 50(3). Retrieved from <http://www.apa.org/monitor/2019/03/trends-suicide>. (11 November, 2019).
- White, Jennifer & Kral, Michael J. 2014. Re-thinking youth suicide: Language, culture, and power. *Journal for Social Action in Counseling & Psychology*, 6(1). 122-142.
- Wynne, Martin. 2005. *Developing linguistic corpora*. Oxford: Oxbow Books.
- Xie, Wenjing & Kang, Cheeyoun. 2015. See you, see me: Teenagers' self-disclosure and regret of posting on social network site. *Computers in Human Behavior*, 52. 398-407.