

## The Language of Suicide Notes: A Corpus-based Stylistic Analysis

**Haya Abdul-Salam Jasim**

Department of English, College of Education for Women  
University of Baghdad  
[haya.s.aljumaily@gmail.com](mailto:haya.s.aljumaily@gmail.com)

**Eman Adil Jaafar**

Department of English, College of Education for Women  
University of Baghdad  
[eman\\_jafer@coeduw.uobaghdad.edu.iq](mailto:eman_jafer@coeduw.uobaghdad.edu.iq)

DOI: <http://doi.org/10.36892/ijlls.v4i1.715>

**Received:**  
21/09/2021

**Accepted:**  
1/12/2021

**Keywords:**  
Corpus-based,  
Stylistic Analysis,  
Suicide notes,  
WordSmith

### *Abstract*

*This study aimed to contribute to the literature of suicide notes studies, especially those who are mainly interested in analyzing the content of suicide notes and the stylistics studies intended to extract the linguistics features. It was designed to investigate and compare a set of linguistics stylistic features of a collection of suicide notes collected by Shneidman and Farberow in 1957 and a collection of suicide notes posted on an online website (Reddit) in 2018-2020. This study aimed to examine whether there is any change in the language of suicide notes between these two collections of data due to the different periods of time and medium (the first collection includes written suicide notes while the second includes posted online suicide notes). So, to achieve the aim of the study, two types of corpora were compiled, one dedicated to Shneidman and Farberow's collection of suicide notes (1957) and the second for the online suicide notes (2018-2020). The two corpora were examined for their most frequent POS, distribution of content words, TTR, and words and sentences' length. The analysis was conducted using CLAWS (UCREL, 2020), "the Constituent Likelihood Automatic Word-tagging System" for part-of-speech tagging, and WordSmith 8.0 software (Scott, 2020). Such analysis indicated that: both types employed an informal style of writing, but the words and sentences used in OSN tend to be longer and more complicated than those in SSN. Both corpora had lower TTR and STTR, which indicated that they were not lexically diverse. Their words were highly repetitive because they concentrated on one or two themes only. Finally, concerning content word distributions, the verbs were the most frequent type, followed by nouns, pronouns, adjectives, and adverbs.*

## INTRODUCTION

According to (World Health Organization, 2019), suicide is the third leading cause of death among people whose ages range between 15 to 19 years old. Annually, 800 000 people die from suicide. Primitively, a suitable clarification of the three concepts about which the suicide studies revolve around should be presented; “Suicidal people” are people who want/wish to kill themselves; according to the Collins Concise English Online Dictionary, “a person who kills himself or herself intentionally.” Meanwhile, “Suicide” is “the act or an instance of killing oneself intentionally.” While a suicide note is “a note that a person writes and leaves before killing himself or herself.”

Generally speaking, suicidal people tend to leave suicide notes before they commit suicide. So, as a definition of suicide notes that the researchers deal with in this study; Suicide notes are defined as a piece of writing written/posted online by people who dare to end their lives and admit that they did it intentionally and willingly for different reasons, which vary according to their circumstances. Suicidal people usually explain this behavior (committing suicide) by leaving suicide notes in which they will either express their feelings to their lovers or blame someone or just have a chance to say their last words.

Different studies deal with suicide notes from different perspectives other than a corpus stylistic study. Jasim and Jaafar (2022) show an overview of the past studies that dealt with suicide language. This study is mainly concerned with identifying the linguistic stylistic features of the suicide notes language by means of a comparison between two collections of suicide notes. The first one belongs to Shneidman and Farberow (1957) that have been retrieved from their book entitled *Clues of Suicide* in 1957 (Shneidman & Farberow, 1957). While the second is a collection of suicide notes that have been collected by the researchers from the online website Reddit. Thus, this study aims to examine the change in the linguistics stylistic features of suicide notes language using corpus tools.

## DATA, METHODOLOGY, AND PROCEDURES

### Data

This study has employed two collections of suicide notes, the first one presented by Shneidman and Farberow in their book entitled *Clues of Suicide* in 1957 (the genuine notes only), while the second has been retrieved from the online website Reddit. The online suicide notes have been collected from particular sub-Reddit like (r/SuicideWatch) and (r/Suicidal\_Thoughts). To achieve the aims of the study that data is organized in the form of two corpora; the process of compiling each corpus has two main steps;

#### 1. Sampling

Concerning the first corpus, which is called (SSN), 33 suicide notes have been transcribed from Shneidman and Farberow’s (1957) book and stored electronically in “txt.” file format as a self-build corpus so that it can be easily dealt with by using the corpus software and tools. The second, which is called (OSN), 33 suicide notes have been collected randomly from the online website Reddit. This process is done by typing the keywords, which are frequently used by the suicide notes written as titles to their notes, in the searching box, and then copying and pasting the data into a “txt.” file. Phrases like “My suicide note,”

“My suicide letter,” “My last words” have been used by researchers in this study for data collection.

What unified these letters are, all of them have been posted by their authors in subreddit dedicated to suicide, and the title of these posts include such terms (My Suicide notes, My suicide letters, last Thing to Say, etc.); which means that the authors intentionally have posted them in this website as their suicide notes. The researchers have checked all these users’ accounts for four months regarding their active statuses after posting their suicide notes. Finally, they ensure that no more comments or posts have been added by them in their personal accounts or the other accounts on this website. It is worth mentioning here that Reddit provides an announcement environment, so the author’s real name does not appear only the fake ones that have been chosen by the person when he/she has signed up in this website, but for the privacy concerns and to make sure that these data do not violate the ethical considerations the researchers have replaced all the names and addresses or phone numbers with a fake one. After collecting them and making the pre-mentioned changes, the researchers have stored them in the “txt” file format as a self-build corpus.

As for the legal consideration, Reddit is an open-access public source for data; any user in Reddit has access to the data and can use it, considering Reddit terms, but using sensitive data like the ones retrieved from particular Subreddits like (r/SuicideWatch) (r/depreession) is subject to the rules of these subreddits (Proferes et al., 2021). The r/SuicideWatch subreddit, for example, prevents any use of its data without taking permission and revising the research by the moderators. So, the researchers have contacted the moderators, and permission has been taken to perform this study. Table 1 below illustrates the two corpora used in this study.

<i>Data</i>			
Abbreviation	Corpus	Size	Year
SSN	Shneidman and Farberow’s suicide notes	33 notes	1957
OSN	Online suicide notes from Reddit	33 notes	2018-2020

Table 1 Corpora types and Abbreviation

The two corpora used in this study have specific features in common: specialized (established for a particular purpose in mind), written, monologue, and diachronic (from successive periods) corpora.

## 2. Annotation

The second step in compiling the corpora is the annotation, which means marking each word in these corpora with its part of speech; this process is called POS tagging. The researchers have used an online Free CLAWS web tagger that can be reached from this link <http://ucrel-api.lancaster.ac.uk/claws/free.html>. CLAWS (the Constituent Likelihood Automatic Word-tagging System) is a software developed in 1980 by University Centre for Computer Corpus Research on Language (UCREL) with 96-97% accuracy. This step is of great help; it simplified the process of dealing with the texts using corpus tools and software.

## Methodology

This study is influenced by the steps presented by McIntyre and Walker (2019) in their book *Corpus Stylistics theory and practice* to conduct a detailed corpus-based analysis of its data. This book presents a guide to all researchers and stylisticians, especially those who have limited or primitive experiences of using the corpus to conduct a corpus stylistic analysis dealing with a large amount of data. It offers the researchers the chance to first: familiarize themselves with the available and suitable corpus linguistic software and tools

### **The Language of Suicide Notes: A Corpus-based Stylistic Analysis**

that can be used effectively to serve the aims of corpus stylistic studies, which is, concerning this study, WordSmith 8.0.

Second, learn about the principles of data collection, sampling, and compiling the corpora by setting non-linguistic boundaries to decide in advance what to include in each corpus. The non-linguistic boundaries used in this study are the geographical boundaries (Shneidman and Farberow's book (1957) and specific online platform, Reddit), medium (written and posted online), domain (written suicide notes and internet post from specific Sub-Reddits), genre (suicide language), date(s) (1957 and from 2012-2020) author (genuine suicide notes writers and randomly selected member from this particular platform Reddit), user status (in the case of the second corpus, all the user's account should not be in an active status since, at less, four months ago).

Third, learn about the final steps in compiling the corpus and the available ways to annotate it. The current study uses CLAWS for POS tagging.

The analysis in this study is carried out following the checklist presented by Leech and Short (2007) in their book entitled *style in Fiction*. This heuristic checklist of linguistic and stylistic categories can be used as a model or theory to analyze the linguistic and stylistic features of a particular text/corpus; it can be applied to any type of text (whether literary or non-literary). It includes four major categories: lexical categories, grammatical categories, figures of speech, and context and cohesion. Each one of these four major categories has its own subcategories, as the table below shows;

<i>Leech and Short's checklist of style markers</i>	
Lexical categories	General, nouns, adjectives, verbs, and adverbs
Grammatical categories	Sentence types, sentence complexity, clause types, clause structure, noun phrases, verb phrases, other phrase types, word classes, and general
Figures of speech	Grammatical and lexical schemes, phonological schemes, and tropes
Context and cohesion	Context and cohesion

Table 2. Leech and Short's checklist of style markers

However, this study did not adopt the entire categories presented in this list, but rather it uses selected ones. These selected categories have been chosen based on the findings of the prior studies that have dealt with suicide notes (illustrated in section 2.1). As for the quantitative analysis, the analytical software WordSmith 8.0 is used.

## **PROCEDURES**

### **WordSmith**

WordSmith 8.0 is paid analytical software developed by Mike Scott, the first version of this software release in 1996, which is WordSmith 1.0 version (Scott, 2012, p.2), and the last version, which is WordSmith 8.0, is released in 2020. The latter is used in this current study. This software can be used to perform a different analytical process; however, what makes it special is that it presents statistics that are very helpful in dealing with corpora texts, like the words and sentences length, the number of words in the sentences, and the Standard token type ratio.

In this study, the quantitative analysis will be conducted using WordSmith software to check particular style markers selected from Leech and Short's checklist of style markers. The focus in this study is on the lexical level only, and due to the nature of the texts under

investigation (the suicide notes) that conventionally tend to be simple, short, and far from been complicated, and based on what the previous study indicates about the content of suicide note, a set of style markers will be presented in this study as a checklist to analyze the corpora based on it. The checklist includes words and sentences length, lexical distribution in general, and content words distribution in particular (nouns, verbs, adjectives, and adverbs). Then, a stylistic interpretation of the results will be presented to each corpus separately, and after that, a comparison will be made based on the results of such analysis.

**FINDINGS AND DISCUSSION**

**Words and Sentences’ length:** The two corpora are examined using WordSmith 8.0 Analytical software to check the length of the most frequent words. The words are classified based on the number of letters they contain by using the Wordlist tool. As shown in figure (1) below, the SSN corpus includes a higher frequency of words that contain 2-3-4- letters and a lower frequency of words that include 10-11-12 letters;

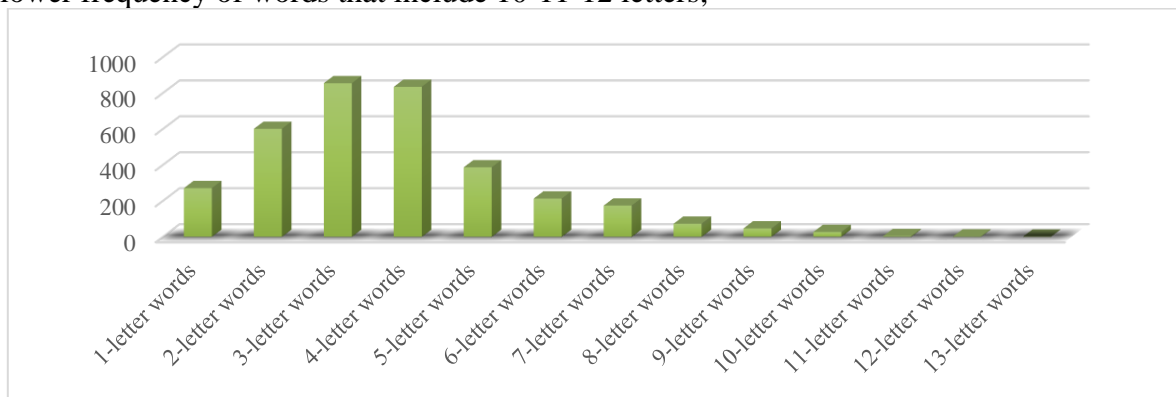


Figure (1) SSN words’ length

Meanwhile, as shown in figure (2), the OSN corpus includes a higher frequency of words that contain 2-3-4 and a relatively lower frequency to words that include 5-6-7, while the lowest frequency is for words that contain 8-9-10-11-12-13-14 letters.

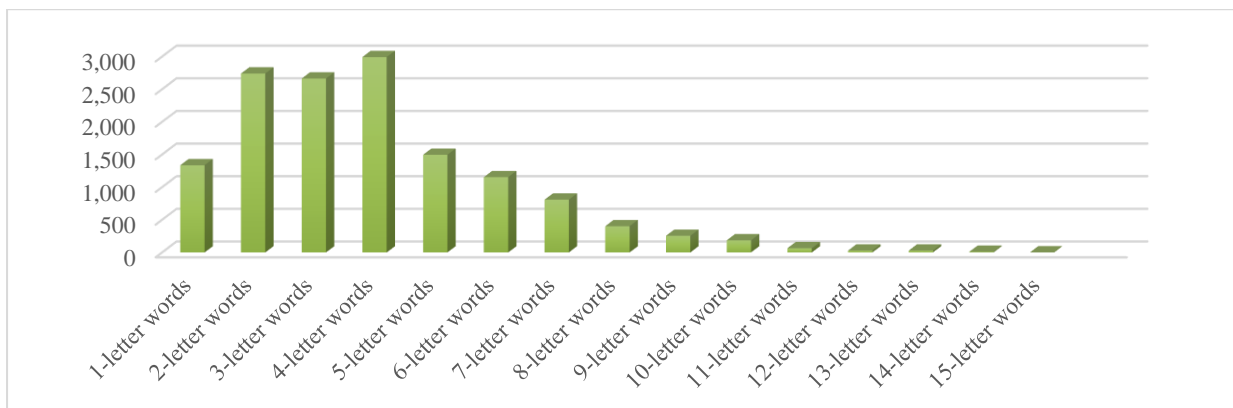


Figure (2) OSN words’ length

Generally speaking, words are the basic components of any type of communication (whether written or spoken). So, the simpler and shorter the words are, the more obvious and uncomplex the writing is. On the other hand, the longer and more complex words are, the more complicated and formal the writing is. According to the results infer from Baddeley et al.’s (1975) experimental study, the words that include one to five letters are regarded as short words, while the words that include six or more letters are regarded as long words.

### The Language of Suicide Notes: A Corpus-based Stylistic Analysis

As illustrated in figures 1 and 2 above, the words with 2,3,4 letters are the more frequent words in both corpora, and the length of the words in the SSN corpus ranges between 1-13 letters, while the length of the words in the OSN corpus ranges between 1-15 letters.

In other words, the online suicide notes tend to include more complex and longer words than the old suicide notes written in 1957. The latter tend to include simple and shorter words; such results can be interpreted as follow; first, the notes in the OSN corpus are written for strangers people, so the online suicide notes include more practical and detailed information, by which the writers to express themselves thoroughly. Meanwhile, the notes in the SSN corpus are written to family members, relatives, or friends, who are in close relation to the suicide note writers. So, they use simple and easy language to express themselves thoroughly. Second, this difference in the length of words between the two corpora can also be explained regarding the language change during the last two centuries; as Bochkarev et al. (2012) argue, the average word length has increased systematically during the last two centuries as a consequence to the processes of words creation, and verbalized certain words by presenting multi-syllables and complex new words with the help of affixation (p.8).

Thus, the length of the average word in the SSN corpus is 1.86, while in the OSN corpus is 2.17. These results confirm the information presented in figure 1 and 2.

**Sentence length:** Sentence length depends on the number of words that each sentence contains, starting from the first word with the capital letter and ending with the full stop, exclamation, or question marks. Using WordSmith 8.0, the data can be organized by numbers to show how many sentences each corpus has and the average sentences' length.

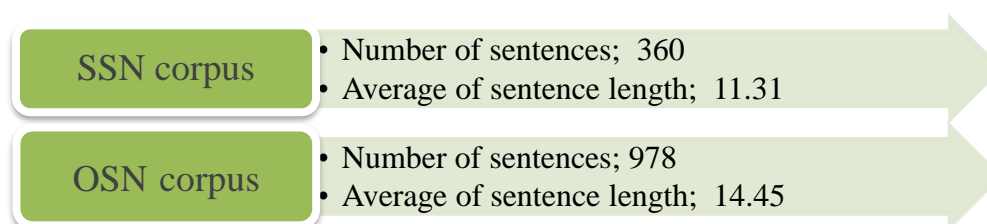


Figure (3) Average of sentence length

In general, any piece of writing that consists of short sentences is easy to be understood. While in contrast, any piece of writing that consists of long sentences is more complicated and hard to be understood by the readers. Short sentences include simple structure, and they are used to present the information directly and obviously. Meanwhile, longer sentences have complicated structures that are hard to understand, especially if the readers have inadequate knowledge; they usually misinterpret them. Longer sentences are used to convey rich details and diverse contents.

The information presented in figure 3 above can be interpreted as follows: the sentences in SSN corpus, which include eleven words or more, are regarded as “easy” sentences. Meanwhile, the sentences in the OSN corpus, which include fourteen words or more, are regarded as “fairly easy” sentences. In other words, the suicide notes language in the two corpora is not completely easy, nor is it completely difficult. It includes simple and easy words, but the writers’ desire to convey all that they want to say as their last words made them use relatively long sentences.

The results in figure 3 above show immense inequality in the numbers of sentences in the two corpora, which can be explained easily since online suicide notes tend to be longer

than the old written suicide notes. Even though each corpus contains the same number of suicide notes (33 for each), the number of tokens each corpus include are immensely different (the SSN corpus has 3,499 tokens while the ONS corpus has 14,183 tokens). Below are a few examples to show the disparity in the length of the sentences between the two corpora:

**A. SSN corpus;**

- (SN-10) " I'm tired. There must be something fine for you. Love."
- (SN-19) "I am talking over this Cyanide deal to myself. God knows what I'll do. I have it here. Possibly 20 grains-5 more than is necessary. I still love you. Be good to X Please."
- (SN-3) "I hope this is what you wanted."

**B. OSN corpus;**

1. (SN-9) "I have been pondering over this question again and again why do we live. We live to make a difference, to make our lives better and to prosper. I have been thinking what I have achieved so far. I have achieved regrets, worthlessness, self-pity and remorse. I had not done a single right thing in my life but I am changing that right now. By ending my life, I am rectifying my 24 years of mistake. I am very happy to do this. This is the best decision ever I could take in my entire life (well, my life was very short!). Ending my life gives me the greatest pleasure ever; reliving the pain daily hurt so much....."
2. (SN-13) I'm right here , in my room , waiting to die , it's funny , because I'm turning 17 in less than 24hours , it just makes me sad that I would die on my birthday, but It is what it is for now . I know how insignificant i am in this whole existance , but I just wanted to mark somewhere, that I felt betrayed by everyone , I have no one , literally.
3. (SN-15) " Hello all, I hope you are all doing better than me. I've made the long decision to finally kill myself and end my suffering as it has reached far beyond my limit to the point I hear voices and see hallucinations of people in my house and its too late now to get help....."

Obviously, the average of sentences' length in the OSN corpus is higher than the average sentences length in the SSN corpus, and it conventionally means that the sentences in OSN are more complicated and detailed than in the SSN corpus. Furthermore, as illustrated above, Online suicide notes tend to be longer and convey more details because they address public readers. The writers in the OSN corpus try to express themselves entirely by mentioning more details. In contrast to the notes in the SSN corpus that have been written during the past periods, which tend to be shorter and do not include many details. The writers tend to use short and straightforward sentences. McIntyre and Walker's study (2019) in their book have emphasized this step and considered it crucial to distinguish between the style of different authors/texts.

**Lexical distribution**

**Token Type Ratio**

Scott (2014) argues, identifying the different parts of speech and their frequencies of occurring in each corpus do not accurately visualize its lexical diversity while calculating the Type Token Ratio, which usually works by calculating the ratio between the numbers of tokens in the corpus and the different types used, can achieve that (p.398). McIntyre and Walker (2019) have emphasized this step and added that knowing the words and sentences length in a particular corpus/text alone is not very helpful in showing a full understanding of

its content. But rather, identifying the different types of the words, the texts include and showing the most frequently repeated ones help to identify its stylistic features (p.125).

The token type ratio (TTR), according to Scott (2014), is “The ratio between types and tokens in the text” (Scott, 2020, p.398). In this study, the researchers have used WordSmith software to generate the TTR by counting the number of the tokens and the number of different types in the text. Such analysis can be helpful for the equal-sized corpus only since it is highly dependent on the text size. So it is better to employ the Standardized TTR (STTR), which presents an average of TTR by counting the average of TTR for every 1,000 words.

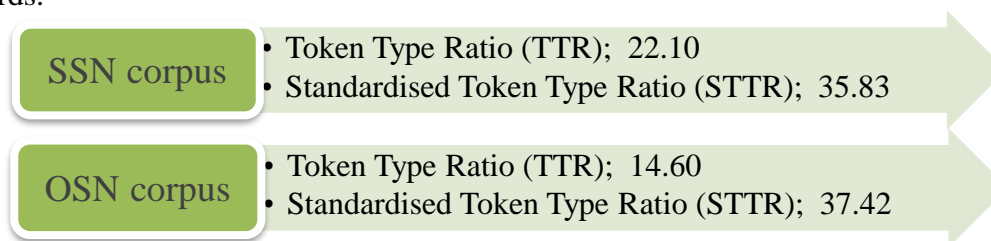


Figure (4) TTR and STTR for the two corpora

As shown in figure 4 above, The TTR of the SSN corpus is 22.10, while in the OSN corpus, the TTR is 14.60. Such results can be explained since the OSN corpus consists of higher numbers of tokens, its TTR is lower than the SSN. Meanwhile, STTR is relatively similar in both corpora. Generally speaking, if the STTR of a particular corpus/text is high, this usually means that this corpus/text is lexically diverse. In contrast, if the STTR of a particular corpus/text is low, this usually means that this corpus/text is highly repetitive.

Both corpora show a relatively lower TTR and STTR, which means that they are not lexically diverse. Such results can be explained because the content of suicide notes generally includes only one or two themes. The suicide notes writers in both corpora employ highly repetitive words, focusing on one theme in their writings, either blaming, thanking, advising, requesting, instructing others, just saying goodbye, or expressing their feelings. In contrast to the writers in any other type of writing, they mix more than one theme.

After presenting a clear image of the lexical diversity in suicide notes writing, Now, it is essential to shedding light on the distribution of content words, in particular, because it will help in characterizing the text into a nominalize or verbalize text style to indicate whether the suicide notes are nominal preferences or verbal preferences.

### **Distribution of content words**

With the aim of analyzing a text based on its lexical usage, the first step is to check its lexical diversity and then to check the distribution of content words, which are nouns, verbs, adjectives, adverbs, and pronouns. This step helps to identify whether this text has a Nominal or verbal style (formal or informal). This step is one of the crucial features that help identify the style of the texts that stylisticians can check.

In order to identify the distribution of content words in the two corpora, it is useful to use the annotation txt. format file that has been marked for POS using CLAWS. Using Microsoft Excel, The two corpora are categorized into different parts of speech, as figure (5) shows. The light gray is used to refer to the OSN corpus, while the dark gray is used to refer to the SSN corpus. It is worth mentioning that the lower frequencies of the SSN corpus are related to the fact that it has fewer tokens than the OSN corpus.



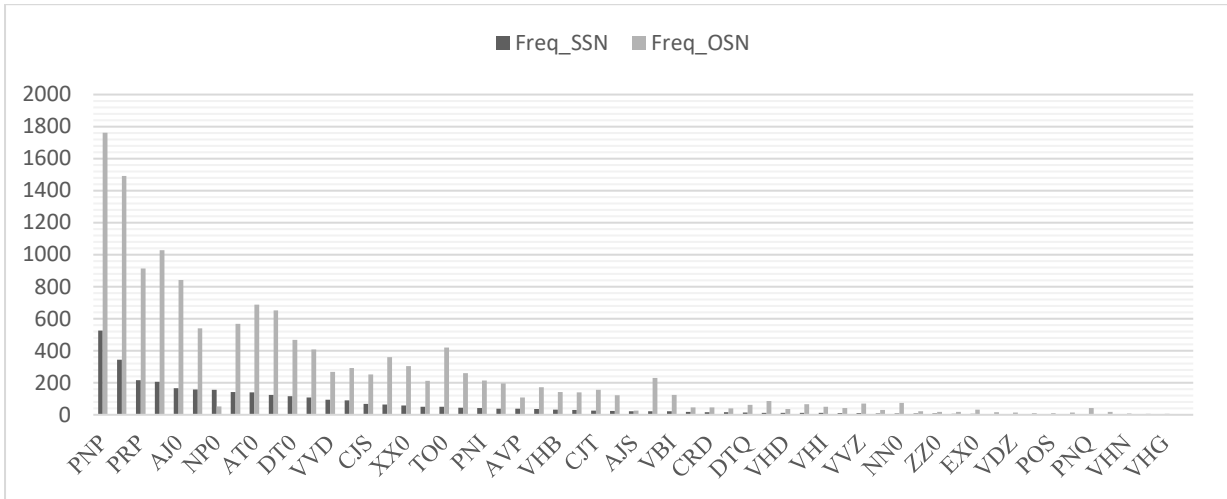


Figure (5) POS tagging

Using WordSmith software, the frequency of each type of content words has been identified in each corpus. As shown in figure (6) below, the verbs are the most frequently used type of content words (750 freq.). Sorting them based on the highest frequency, verbs like “have,” “do,” “love,” “know,” and “tell” get the highest numbers of frequency in the verbs list. Meanwhile, the nouns with (665 freq.) are the second most frequently used type of content words. Words like “way,” “thing,” “wife,” “darling,” and “love” are at the top of the most frequently used nouns in this corpus. Furthermore, adverbs with (274 freq.) are the third most frequently used type of content words in this corpus; words like “not,” “so,” “ever,” “now,” and “well” are at the top of the most frequently used adverbs in this corpus. Finally, the adjectives with (202 freq.) are at the end of the most frequently used content type list. Adjectives like “good,” “dear,” “only,” “sorry,” and “wonderful” are the most frequently used adjectives in the corpus. It is worth mentioning that the pronouns show a very high usage in this corpus; they represent (665 freq.) of the total number of its tokens, pronouns like “I,” “you,” “my,” “me,” and “it” are at the top of the list of most frequently used pronouns in this corpus.

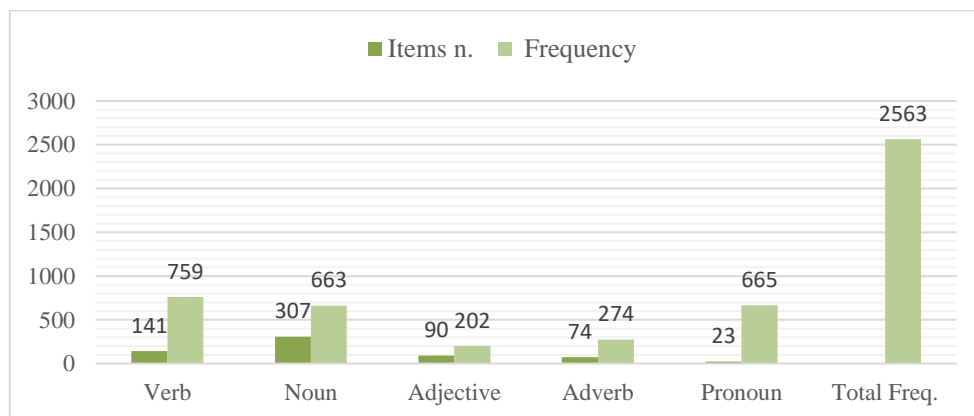


Figure (6) total frequency of content words in SSN

As shown in figure (7) below, the same procedures have been applied to the OSN corpus, and the results also indicate that the most frequently used type of content word is the verbs, representing (3,383 freq.) of the total number of ONS corpus. Verbs like “have,” “do,” “want,” “know,” and “feel” are at the top of the most frequently used verbs in this corpus. Meanwhile, nouns are the second frequently used type of the content words. Words like “life,” “people,” “time,” “thing,” and “friend” occur at the top of the most frequently used

### The Language of Suicide Notes: A Corpus-based Stylistic Analysis

nouns in this corpus. Furthermore, adverbs with (1,348 freq.) are the third most frequently used type of content words in the ONS corpus. Words like “not,” “just,” “so,” “always,” and “never” are the most frequently used adverbs in this corpus. Finally, At the end of the most frequently used content type list, the adjectives occur with (1,002 freq.). Adjectives like “good,” “sorry,” “bad,” “much,” and “short” are the most frequently used adjectives in this corpus. The same can be said concerning the use of pronouns in the ONS corpus; they show a high frequency of occurring (2,332 freq.). Pronouns like “I,” “my,” “you,” “me,” and “it” are at the top of the most frequently used pronouns in this corpus.

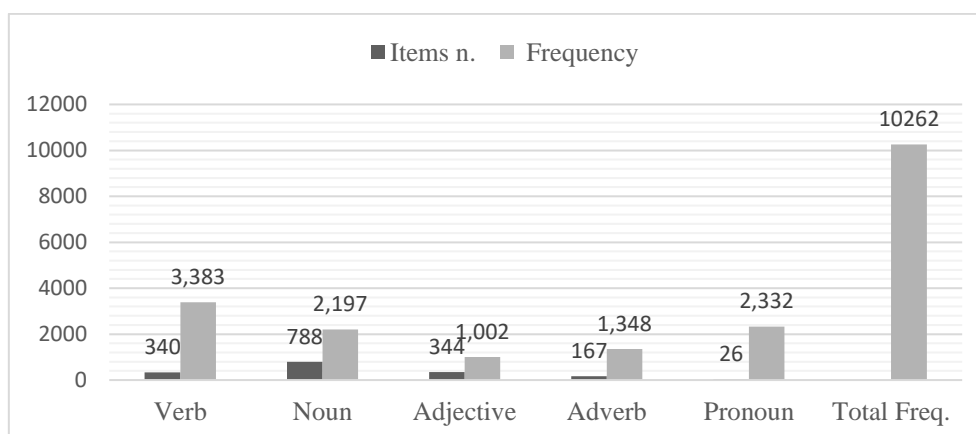


Figure (7) total frequency of content words in OSN

From the above figures, it appears that the SSN and the OSN are similar to some extent in the distribution of content words, the verbs are the most frequent type to use in the two corpora, then the pronouns and nouns cover a higher frequency in both, and regarding the adjective and adverbs both cover lower frequency in the two corpora. Moreover, the total frequency of the content word distribution is relatively similar regardless of the differences in tokens numbers that cause some variations in the figure. So, roughly speaking, the language of suicide notes in both corpora are of verbal preferences. Such results can be explained because the verbs are used more frequently than the nouns and pronouns in both corpora.

With regard to the results obtained from Heylighen and Dewaele’s (1970) study, the more frequent verbs, adverbs, and pronouns are in a text, the more informal it will be. In contrast, the more frequent nouns, adjectives, prepositions, and articles are in a text, the more formal it will be. Informal style tends to be more flexible, more direct, and less informative and accurate; meanwhile, the formal style tends to be precise, direct, and cognitive loaded (Heylighen & Dewaele, 1999). So, based on the above-stated information and results, one can confidently state that suicide notes language in both corpora usually employs an informal style. It includes more verbs (because they usually want several actions to be done because this is their last words), more pronouns, more nouns (due to the fact they want to be direct and easy to be understood), and fewer numbers of adjectives (almost all the adjectives are negative once) and adverbs.

## CONCLUSION

In this paper, a corpus-based stylistic analysis has been done with the aid of WordSmith software and annotation statistic software to two types of corpora that included

two collections of suicide notes. The results of the analysis have emphasized the role of the corpus tools to aid the stylistic analysis; they facilitate the analysis process and help to present accurate information concerning the quantitative analysis. A comparison has been made based on a pre-prepared checklist about specific categories that have been selected from Leech and Short's (2007) checklist. This study aimed to identify whether there is a difference in the language of suicide notes from the stylistic features due to the change in time and the medium (to be posted online). The results of such comparison indicate that:

- The average of the word length in the OSN corpus tend to be longer than the average of words length in the SSN corpus; therefore, the words in SSN seem to be more simple, short, and easy to be understood in contrast with the words in OSN which tend to be complicated and lengthy.
- The average of sentences' length in the OSN corpus is higher than in the SSN corpus, and it conventionally means that the sentences in OSN are more complicated and detailed than in the SSN corpus.
- Concerning STTR, both corpora present a relatively lower TTR and STTR, which means that they are not lexically diverse, and their words are highly repetitive.
- Concerning the distribution of content words, the verbs are the most frequent type to use in the two corpora, then the pronouns and nouns cover a higher frequency in both, and regarding the adjectives and adverbs, both cover lower frequency in the two corpora.

## REFERENCES

- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14, 575-589.
- Bochkarev, V. V., Shevlyakova, A. V., & Solovyev, V. D. (2012). Average word length dynamics as indicator of cultural changes in society. *Social Evolution & History*, 14(2), 153-175. [https://www.sociostudies.org/journal/files/seh/2015\\_2/153-175.pdf](https://www.sociostudies.org/journal/files/seh/2015_2/153-175.pdf)
- Heylighen, F., & Dewaele, J. (1999). Formality of language : definition, measurement, and behavioral determinants. *Interneter Bericht, Center "Leo Apostel", Vrije Universiteit Brüssel*, 4.
- Ho, Y. (2011). *Corpus Stylistics in Principles and Practice*. London Continuum.
- Jasim, H.A. and Jaafar, E.A. (2022) Studies on Linguistic Stylistic Analysis of Suicide Notes and Suicidal Thoughts Posts. *International Journal of Research in Social Sciences & Humanities*, January- March 2022 Vol. 12, Issue 1; 100-124  
<https://www.ijrssh.com/admin/upload/06%20Haya%20Abdul%2001289.pdf>
- Leech, G., & Short, M. (2007). *Style in Fiction*. Pearson Education Limited.  
<https://doi.org/10.4324/9781315835525>
- McIntyre, D., & Walker, B. (2019). *Corpus stylistics: Theory and practices*. Edinburgh University Press.
- McMenamin, G. R. (2002). *Forensic linguistics: Advances in forensic stylistics*. CRC press.
- Osgood, C. E., & Walker, E. G. (1959). Motivation and language behavior: A content analysis of suicide notes. *Journal of Abnormal and Social Psychology*, 59(1), 58–67.
- Proferes, N., Jones, N., Gilbert, S., Fiesler, C., & Zimmer, M. (2021). Studying Reddit: a systematic overview of disciplines, approaches, methods, and ethics. *Social Media+ Society*, 7(2).
- Scott, M. (2012). *WordSmith Tools Help Manual*, Lexical Analysis Software
- Scott, M. (2014). *WordSmith Tools Help Manual*. Lexical Analysis Software.
- Scott, M. (2021). *WordSmith Tools [Computer software]*. Stroud: Lexical Analysis Software.  
<https://www.lexically.net/wordsmith/downloads/>

Shneidman, E. S., & Farberow, N. L. (1957). *Clues to Suicide*. Public health reports.

UCREL. (2020). *CLAWS tagger* (No. C5). <http://ucrel-api.lancaster.ac.uk/claws/free.html>

World Health Organization. (n.d.). *Suicide*. (2019 Sep 2). Retrieved December 28, 2020, from <https://www.who.int/news-room/fact-sheets/detail/suicide>

**AUTHORS' BIO**

---

**Haya Abdul-Salam Jasim** is MA student at the Department of English/ College of Education for Women / University of Baghdad. Her research interests include Forensic linguistics, Stylistics, and Corpus stylistics.

**Eman Adil Jaafar** is Assistant professor of linguistics at the Department of English/ College of Education for Women / University of Baghdad. Her research interests include but not restricted to stylistics, corpus and cognitive stylistics and applied linguistics.

---