



# INTERNATIONAL JOURNAL OF RESEARCH IN SOCIAL SCIENCES & HUMANITIES

An International Open-Access Peer Reviewed Referred Journal

Impact Factor: 6.064

E-ISSN : 2249 – 4642

P-ISSN: 2454 - 4671

## STUDIES ON LINGUISTIC STYLISTIC ANALYSIS OF SUICIDE NOTES AND SUICIDAL THOUGHT POSTS

**Haya Abdul-Salam Jasim, Eman Adil Jaafar**

University of Baghdad, College of Education for Women  
Department of English, Iraq

DOI: <http://doi.org/10.37648/ijrssh.v12i01.006>

**Paper Received:**

04<sup>th</sup> December, 2021

**Paper Accepted:**

1<sup>st</sup> January, 2022

**Paper Received After Correction:**

11<sup>th</sup> January, 2022

**Paper Published:**

12<sup>th</sup> January, 2022



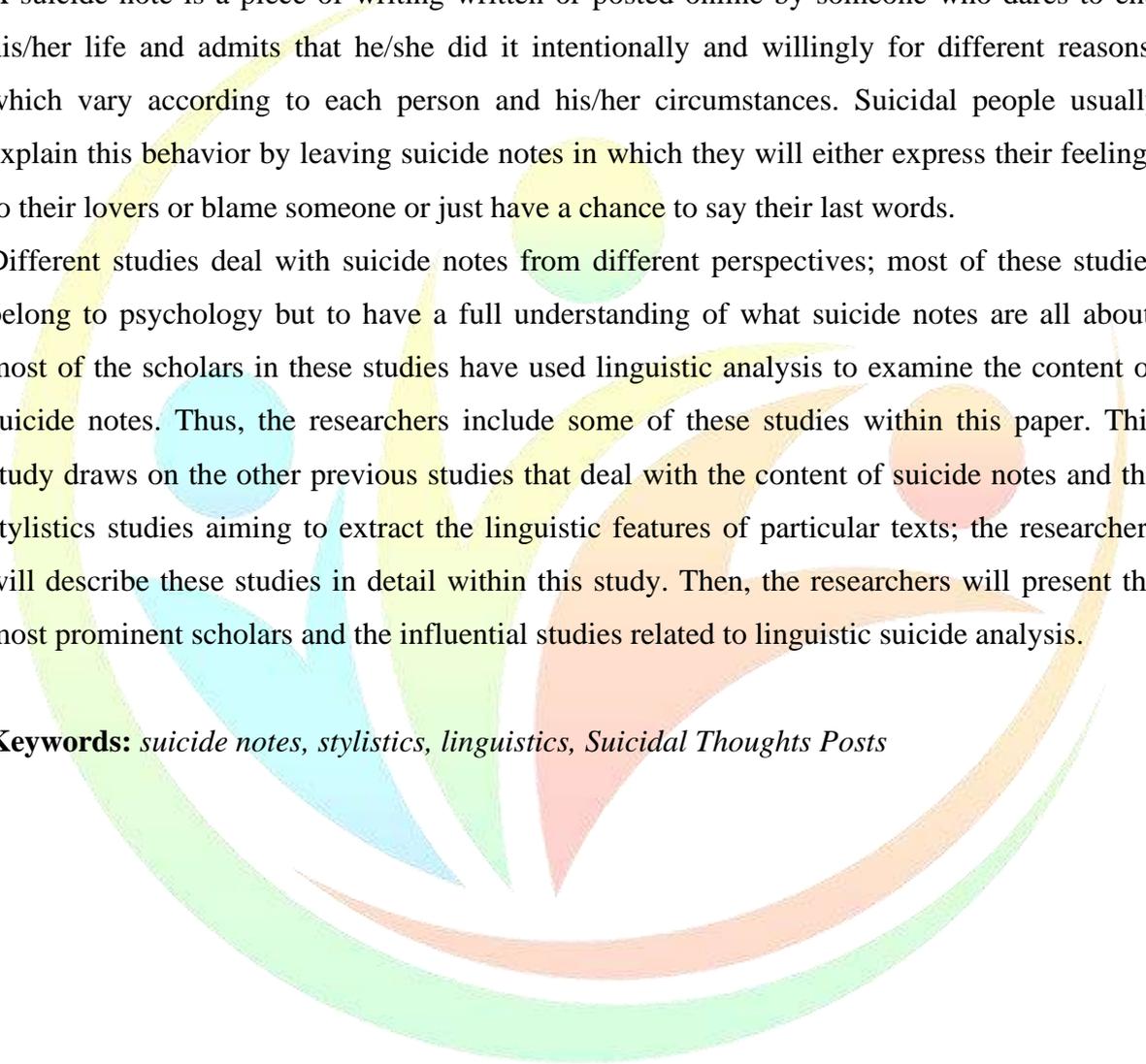
**How to cite the article:** Jasim H.A. and Jaafar E.A. (2022) Studies on Linguistic Stylistic Analysis of Suicide Notes and Suicidal Thoughts Posts. *International Journal of Research in Social Sciences & Humanities*, January- March 2022 Vol. 12, Issue 1; 100-124 DOI: <http://doi.org/10.37648/ijrssh.v12i01.006>

## ABSTRACT

A suicide note is a piece of writing written or posted online by someone who dares to end his/her life and admits that he/she did it intentionally and willingly for different reasons, which vary according to each person and his/her circumstances. Suicidal people usually explain this behavior 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; most of these studies belong to psychology but to have a full understanding of what suicide notes are all about, most of the scholars in these studies have used linguistic analysis to examine the content of suicide notes. Thus, the researchers include some of these studies within this paper. This study draws on the other previous studies that deal with the content of suicide notes and the stylistics studies aiming to extract the linguistic features of particular texts; the researchers will describe these studies in detail within this study. Then, the researchers will present the most prominent scholars and the influential studies related to linguistic suicide analysis.

**Keywords:** *suicide notes, stylistics, linguistics, Suicidal Thoughts Posts*

The logo for the International Journal of Research in Social Sciences and Humanities (IJRSSH) is a large, stylized graphic. It features a central figure that resembles a person with arms raised in a 'V' shape, composed of several overlapping, curved, leaf-like or flame-like shapes in shades of blue, green, yellow, and orange. This central figure is set against a background of a large, light-colored circle. Below the circle, the acronym 'IJRSSH' is written in a bold, orange, sans-serif font.

IJRSSH

## INTRODUCTION

This paper aims at presenting the most prominent studies that dealt with analysing the language of suicide notes. It also aims at examining the different methodologies in conducting the analyses. Some of these studies include:

### **Shneidman and Farberow (1957)**

Many linguistic and psychological scholars consider Shneidman and Farberow's (1957) study as the basis of the other suicide notes studies. They regard it as the seed from which other studies have blossomed. This study is psychological in nature, but it includes a linguistic analysis of the content of suicide notes. Shneidman and Farberow have presented their book entitled *Clues of Suicide* in 1957. This book includes a comparison between genuine and simulated suicide notes. Shneidman and Farberow's (1957) study deals with suicide in general and attempts to study different factors like the sociologic, ecologic, cultural, psychological, and many other factors, which consider motivational factors to suicide. Within this study, the researchers linguistically analyze suicide notes by using Mowrer's method, which is "The Discomfort-Relief Quotient". In general, a researcher uses this method to compare two subjects or two different texts by dividing each text into thoughts units. Then, these thoughts units are classified

into three categories: discomfort units, relief units, and neutral units. Their study investigates the usefulness of such a technique in distinguishing two different types of texts. Thus, it helps them to identify the differences between genuine and stimulating suicide notes (p.251).

The importance of Shneidman and Farberow's (1957) study is that it uses several suicide notes from (721) people who have committed suicide in Los Angeles, and after that, several researchers (as this study will reveal) use the same data to conduct their studies, but they deal with suicide notes from different perspectives. This study uses two types of suicide notes, which are the genuine and simulated notes; genuine suicide notes (which are pieces of writing left behind by people who commit suicide) while simulated suicide notes (which are pieces of writing written, intentionally, by either non-suicidal or non-depression people to achieve the aim of this study). In this study, Shneidman and Farberow divide the data into two groups; each one contains 33 suicide notes. The first group contains only genuine suicide notes, and the second one contains the simulated ones. Then, they examine the two groups using Mowrer's method (p.252).

After accounting for the total number of the thoughts units in all the 66 suicide notes under investigation, the researchers

have found that there are 553 thoughts units in both pairs, 369 of these units in the 33 genuine suicide notes, and 184 in the simulated suicide notes. This result indicates that genuine suicide notes are more verbose than simulated suicide notes. Shneidman and Farberow use Dr. Mower's personal communication as a model to discuss these numbers resulting from comparing genuine and simulated suicide notes. For example, this study indicates that genuine suicide notes have 226 discomfort statements, 65 relief statements, and 78 neutral statements. In contrast, the simulated suicide notes have 137 discomfort statements, 34 relief statements, and 13 neutral statements (p.254).

According to Shneidman and Farberow (1957), the Ratio of DRQ in the comparison between the two types of suicide notes is not significant. Moreover, the discussion of the result shows that the writers of genuine suicide notes need to say more words than the writers of the simulated suicide notes because they are sure that these words will be their last words, and this can be confirmed more by looking at the number of neutral statements in both types. The neutral statement usually includes instructions and guides to their lovers for the next levels of their lives because the writers of genuine suicide notes are sure that they will not be

there with them. The discussion also indicates slight differences in the quality of discomfort thoughts units between the two types of notes, while in the relief thoughts units number, there are no significant differences between the two types. Lastly, Shneidman and Farberow's (1957) study includes many other aspects of suicide to deal with; one of them is the linguistic analysis of suicide notes. The limitation of this study is that it uses a small sample of suicide notes (only 33), and it uses only one technique to analyze them: the DRQ technique.

#### **Osgood and Walker (1959)**

Osgood and Walker (1959) use the same suicide Notes that Shneidman and Farberow (1957) have collected from Los Angeles and have used in their study. The researchers choose only 100 suicide notes to conduct their study, which compares the suicide notes with other types of notes/messages that any normal person sends.

Osgood and Walker (1959) analyse 40 suicide notes written by males and 29 suicide notes written by females for the actual suicide notes data. While for the neutral notes, they use 13 notes written by males and 59 notes written by females. Concerning the comparison between the simulated and genuine suicide notes, they use 33 pairs of notes from Shneidman and Farberow (1957) study. To achieve the

aims of the study, Osgood and Walker present four hypotheses for comparison purposes. As a summary for these four hypotheses are presented in the following list:

1. Suicide notes have a greater stereotype than ordinary notes: the suicide notes should have more repetitive words, few adjectives, few adverbs, and less lexical diversity.

2. Suicide notes have greater disorganization in language than ordinary notes: the suicide notes should have more language errors, shorter sentences, and small word units.

3. Suicide notes have a higher frequency of words and grammatical choices that indicate self-destruction: the suicide notes should have more words that indicate a command, request, or any other type of words required from the listener to perform a particular behavior.

4. Suicide notes have more words and grammatical choices that indicate conflicts than ordinary notes: suicide notes should have a high frequency of constructions like (but, if, however, etc). (Osgood & Walker, 1959, pp. 58-59).

Osgood and walker (1959) employ 16 measures to compare the genuine suicide notes and ordinary notes, and they use 11 more measures with the 16 previous ones to compare between the genuine and

simulated ones. Then, they apply the analysis of words and count the frequency of a particular grammatical construction manually. The result of the comparison between suicide notes and ordinary notes shows that: three of the previous hypotheses prove to be right, while the fourth one, which states that the suicide notes should be disorganized, prove to be wrong. Meanwhile, the results of the comparison between the genuine and simulated notes indicate that: the genuine suicide notes are shorter than the simulated notes in the sentence length, and they usually have more simple action verbs and positive concepts, unlike the simulated suicide notes, which tend to have more mental state verbs and more abstract concepts. Furthermore, Osgood and Walker's (1959) work indicates that genuine notes writers usually use a higher percentage of depressing words, precise nouns, and verbs, and the male writers, in particular, tend to use shorter sentences in the genuine suicide notes than in the simulated one.

#### **Gottschalk and Gleser (1960)**

Gottschalk and Gleser (1960) conduct a study that aims to measure the personality variables by conducting a systematic analysis of verbal behavior in a particular text. The researchers choose the same suicide notes that Shneidman and Farberow (1957) have employed as data to

conduct their study. They use the same 33 pairs of genuine and simulated suicide notes and apply two different types of analysis, a grammatical and psychological analysis, to quote their words: "Our analyses of speech have been at two levels: an atomistic level, i.e. types of words used, and a 'molecular' level, i.e., types of themes and combinations of themes used." (Gottschalk & Gleser, 1960, p.195).

The method of analysis in Gottschalk and Gleser's (1960) work is to categorize each word in the texts into a particular grammatical and psychological category. This study is the first application of this method in a written text. Regarding the suicide notes, the number of words in suicide notes is counted, and each word classifies within a particular category. The 33 pairs of genuine and simulated notes appear to have a number of words that range from 7 to 400, with a median of 78 in the genuine notes and 48 in the simulated notes. The scholars use two statistical procedures to compare the two pairs: the sign and median test. The sign test is a statistical test that includes putting a plus or minus sign in the individual pair that consists of a higher number of the words in a particular category being investigated, and after that, the number of the plus and minus signs are also being counted. Meanwhile, the median test is also a statistical test that numerically

presents the results by depending on the middle value, if the results show more than two numbers or present the mean of the two results. The researchers, then, take each category median number in the simulated and the genuine notes within their analyses and make their Comparisons (Gottschalk & Gleser, 1960, pp.196-197).

Each word in the notes classifies into a particular grammatical and psychological category. The grammatical categories include verbs, adverbs, adjectives, prepositions, conjunctions, substantives, and interjections. Meanwhile, the psychological categories, which aim to classify the words according to the meaning that each word denotes from its place within the sentence, are words denoting feeling, perceptual process, motivation, cognitive process, words indicating time or space, words indicating some measure of quantity, and negation. After applying this analysis to the simulated and genuine notes, the number of words in each category counts against the total number of words in the notes to get the percentage of occurrences of each category in the notes (Gottschalk & Gleser, 1960, p. 196).

The result of this statistical analysis concerning the grammatical analysis of suicide notes indicates that the median test results do not present significant differences between the two types of

suicide notes under investigation. In contrast, the sign test results indicate that genuine suicide notes have a greater average of words than the simulated ones; the genuine suicide notes writers tend to use more words that indicate a direct reference to people, places, things, and lower percentages of words represent cognitive processes. Furthermore, they identify five of these thirty-four categories that they use to be significantly crucial in distinguishing between the genuine and simulated suicide notes. To quote their words:

“In all thirty-four categories were compared. For example, five were found to yield a significant difference between the two sets of notes, with or without regard to matching (Table 1). These were: the percentage references to substantives (nouns and pronouns); to others (third person); to others including the person addressed in the note; inanimate objects; and the percentage use of auxiliary words” (Gottschalk & Gleser, 1960, p.197).

Gottschalk and Gleser’s (1960) study has significantly contributed to the history of suicide notes study, especially the distinction between genuine and simulate

notes. Using categories is similar to what is, nowadays, known as semantic tagging, so this study can be considered the first attempt to apply this type of analysis to suicide notes.

#### **Ogilvie et al. (1966)**

Ogilvie et al.’s (1966) work, a chapter within a book entitled “The Content Analysis Reader,” is one of the studies that also analyze the 66 pairs of Shneidman’s notes but using a different procedure, which is the computer analysis. Utilizing the General Inquirer Software and connecting it with the “Harvard II Psycho-Sociological Dictionary” help them to specify singular words inside different classifications in the dictionary. These categories associate with sociology and psychology; they provide a new way to analyze the context of a particular word by putting it into some sort of degree according to sophisticated “disambiguation” standards. Recognizing the genuine suicide notes from the simulated one by employing computer analysis can be regarded as their significant contribution. According to Ogilvie et al. (1966), the dictionary is classified into two categories: first-order tags which are independent categories, and second-order tags, which are the non-independent categories. To quote their words:

“the first-order tags were subdivided into roles, objects, emotional states, and actions. all things were either roles or objects and all processes were either emotional states or actions. the second-order tags were classified as referring to either institutions, statuses, qualities, or symbolic referents.” (p.404).

Furthermore, Ogilvie et al. (1966) mention the other types of categories, and then they employ the sub-categories to distinguish between actual and fake notes and recognize the differences of word-count inside sub-corpora. General Inquirer Software functions in counting the word in each type of sub-categories and in identifying text for a grammatical function like a verb, adjective, etc. and part of speech like a pronoun; for example, the word “think” involves many words like consider, reason, recall, decide, remember, etc.

The results indicate that the simulated suicide notes show lower diversity in the categories of the word, while genuine suicide notes have higher diversity than the former. Moreover, they find that genuine suicide notes record higher numbers of words in almost all categories except for (the Emotional States and institutions). In comparison, the simulated suicide notes record a higher number of words that are classified under specific categories like (Objects, States, and qualities) and fewer

words in the category of Roles (Ogilvie et al., 1966, p. 405).

The results also indicate that the simulated suicide notes contain more words that refer to the cognitive process than the genuine suicide notes, which, in contrast, tend to use simple words. Another difference that Ogilvie et al.’s (1966) study have come up with is that in genuine suicide notes, the writer uses the pronoun (I) to refer to himself and (he/she) to refer to the others, meanwhile in simulated suicide notes, the writer uses (we) when referring to himself and a particular person. Their results also indicate that within genuine suicide notes, the writer refers to women more often than within the simulated suicide notes. The genuine suicide notes show higher percentages of occurring of particular words like “love,” “know,” “woman,” and more sentences that imply specific instructions to their lovers. To quote their words: “154 occurrences of female as a subject for the genuine notes compare to 64 occurrences for the simulated notes.” (Ogilvie et al., 1966, p. 406).

To summarize their finding, Ogilvie et al. (1966) argue that simulated suicide notes usually include higher numbers of cognitive and process words and lower numbers of words referring to a particular name, place, and time. In contrast, the genuine suicide notes include higher numbers of words that refer to a particular

name, place, and time and lower numbers of words that belong to cognitive and process categories.

Ogilvie et al.'s (1966) study confirms the results obtained by Gottschalk & Gleser (1960) in the pre-mentioned study. Both of these studies agree that it is possible to distinguish between fake and real notes by using this method of analysis, which are dividing the words into categories and counting the number of occurrences in each type, either manually like what Gottschalk & Gleser (1960) have done, or by using the computers like what they have done.

#### **Edelman and Renshaw (1982)**

Edelman & Renshaw (1982), in their article, try to replicate the previous studies that have been done by the previous scholars who have dealt with the language of suicide notes, especially those who aim to distinguish genuine suicide notes from the simulated suicide notes like Osgood and Walker (1959), Gottschalk and Gleser (1960), and Ogilvie et al. (1966). In their study, Edelman & Renshaw (1982) wish to assess the success of these studies in distinguishing the real from fake suicide notes and double-check their results by using a contemporary method, which is a unified theory of discourse analysis to the language of suicide notes. Edelman & Renshaw's (1982) study also aims to make a multivariate model for the suicidal

individual depending on his/her language (Edelman & Renshaw, 1982, p.104). The latter is the main aim of this current paper (to present a simplified model that carries the linguistic features of the online suicide notes language).

In these pre-mentioned studies, Osgood and Walker (1959) and Gottschalk and Gleser (1960), the scholars use different analytical procedures like counting the number of nouns, adjectives, verbs, counting the type-token ratio, counting distress-relief quotient, counting mands, and illness terms. Furthermore, in the study of Ogilvie et al. (1966), the scholars use many other procedures that the computer analysis software (General Inquirer Software) presents. The results of each one of these studies contribute crushingly in distinguishing between the genuine and simulated suicide notes in which the researchers discuss in detail the procedures and the results of these studies). Edelman & Renshaw's (1982) study indicates that there are three problems concerning these studies, which are:

1. There is no serious attempt to present a statistical model including precise numbers and quantitative tools, such as the number of nouns, verbs, type-token ratio, or distress-relief quotient, which helps significantly in distinguishing between the genuine and simulated suicide notes.

2. The analysis of suicide notes depends on a psychological basis or traditional grammar theories, but none of these studies have used unified discourse analysis methods to deal with suicide notes.

3. There is no stable language profile of suicide notes presented in any of these three studies that may be consistently used to describe the suicidal person.

As a consequence of these problems, Edelman & Renshaw (1982) employ a reliable and coherent discourse analysis system, entitled "The Syntactic Language Computer Analysis" (SLCA-111), which has been developed by Cummings and Renshaw (1979) (p. 105).

This system includes several categories into which the words are classified; these categories correlate to the nouns, joint to the verbs, and attribute to adjectives and adverbs. This system also focuses on the different aspects and functions of each category; in other words, it does not deal with specific words only (like counting the type-token ratio, distress-relief quotient), but it aims to understand the function of these categories in the whole text or message (Edelman & Renshaw, 1982, p.106). Thus, this is what distinguishes it from the General Inquirer Software that has been used by Ogilvie et al. (1966) in their study (as illustrated previously). This

computerized discourse analysis system has essentially three basic categories (nouns, verbs, and adjectives/adverbs) and another eight sub-categories that are related to these three basic ones, which are:

1. Social perception: it is usually joint with nouns like proper nouns, pronouns, and references to things (it), indicating the person's ability to perceive the characteristics of animate or inanimate objects.

2. Sensation: it is usually joint with nouns, adjectives/adverbs, referring to what a person perceives as concrete or abstract (like book and truth).

3. Existence: it is usually joint with the three basic categories (nouns, verbs, adjectives/adverbs), indicating the person's ability to perceive something as it exists or not.

4. Motion: it is usually joint with the verbs, referring to what the verb usually expresses, either activity or state.

5. Disposition: it is usually joint with verbs, referring to what the verb usually expresses, either assertion or a condition.

6. Time: it is usually joint with verbs, referring to the tense, whether it is past, present, or future.

7. Symmetry: it is usually joint with verbs, indicating the verb's intention, whether it is transitive or intransitive.

8. Definition: it is usually joint with verbs and nouns, indicating whether these units are modified (Edelman & Renshaw, 1982, pp.106-107).

Utilizing this system, Edelman & Renshaw (1982) apply statistical analysis that depends on "a combination of dictionary "look-up" and positional analysis" (Edelman & Renshaw, 1982, p.107) to the 66 pairs of notes that have been collected by Shneidman and Farberow's (1957). Their findings match the findings of the previous studies. Furthermore, they argue that only 9 of these 36 categories that have been used are highly used to distinguish between genuine and simulated suicide notes, which are: "Total Word Usage," "information Units" (Nouns), "sensation Unit Density" (Concrete-Abstract), "negative Existential Density" (Asserting "not"), "positive Qualification" (Positive modifiers), "negative Authority" (Proper Nouns), "positive generalized Other" (3rd Person Pronouns), "negative Audience" ("you"), "defined Information" (Modified Nouns), and "defined Relational" (Modified Verbs) (p.111).

Thus, the results indicate that genuine suicide notes show a high frequency of words that refer to cognitive states, higher

numbers of words that refer to concrete objects, higher numbers of words that refer to particular people, objects, places, and higher numbers of the occurrences of the pronoun (you). Moreover, and concerning the time references, Edelman & Renshaw's (1982) study reveals a decrease in the future references in genuine suicide notes. At the same time, there is no increase in references to the present and past (p.112). What distinguishes Edelman & Renshaw's (1982) work from the other studies, which have been mentioned is that it emphasizes the existence of negativity and the lower frequency of future tense in genuine suicide notes.

#### **Gregory (1999)**

Many authors have tackled suicide notes, including Adam Gregory (1999). His study aims to formulate a statical model by integrating all the variables in the previous studies. He depends on the same 66 notes that belong to the Shneidman corpus and another 18 suicide notes that he has collected from other studies. His dependence on the previous studies of suicide notes that are related to Osgood and Walker (1959), Edelman and Renshaw (1982), Gottschalk and Gleser (1960), and Ogilvie et al. (1966) helps him to extract nine variables that he uses to distinguish between fake and real notes. These variables or hypotheses are higher usage of nouns, higher usage of verbs, lower usage

of medium sentence length, higher usage of words, lower usage of cognitive verbs, higher usage of positive effects, higher usage of instructions, none or a limited usage of explanations, and external usage of a position of control (Gregory, 1999) cited in (Shapero, 2011, p 65).

After doing such a study, he discovers that the most beneficial oriented variable is content rather than structure variables because it is about either 4 or more than 6, whereas structure variables are about either 1 to 3 or more than 5. After his success in formulating his model, he discovers that 'genuine' attributes, which are the most manifestation for both content and structure variables, have been repeated 15 times, whereas stimulated attributes have been repeated twice (p.65).

Moreover, after adding 18 suicidal notes to his analysis, he also gets the same results: 18 for genuine notes and 2 for simulating notes. He also finds that genuine notes contain higher quantum words, including both effects and itemized instructions, while the variables associated with explanation are regarded as thematic. Furthermore, Gregory conducts his model to analyze Paula's suicide notes cited in Casey in 1993. During his analysis of the mental state of Eddie Gilfoyle, Paula's husband. He recognizes that Gilfoyle killed his wife After he persuades her to write two suicidal letters by claiming it is

related to his future project of suicide (Gregory, 1999) cited in (Shapero, 2011, p. 65). Gregory, a psychologist, formulates such a model focus on distinguishing the killer's mental state from a normal one and distinguishes the fake suicide notes from the real one.

#### **Barak and Miron (2005)**

This work is a combination of three psychological studies that aim to investigate the constancy in the writing of suicidal people in online and offline environments. Two of these studies are not directly related to the current work; the first study investigates the differences in causal attributions in the writing of suicidal people, and the third study investigates the themes of personal distress in the writing of suicidal people. Meanwhile, the second study, which is more relevant to the current work, investigates the differences in the use of words and phrases in the writing of suicidal people (Barak & Miron, 2005, p.510). The data have been selected randomly from online open-access groups and forum-type (about 800 messages have been used). Based on the previous studies about the suicide languages, which indicate that the suicidal person employs within his/her writing more self-focus words, first voice, certain expressions like "me, too," "I also," "me also," and a higher ratio of word between the first and

the last message. Barak and Miron examine their data to check the consistency in these points in the participant's online and offline writings (p.513). Barak and Miron's analysis confirm the previous studies finding and indicate the consistency of the two types of writing of suicidal people. Furthermore, they also demonstrate that the online writings of suicidal people are significantly more self-focus, first voice, and include more expressions like "me, also" than do the online writings of depressed non-suicidal people on the internet.

#### **Malini and Tan (2016)**

This study deals with the suicide notes of Virginia Woolf that she left behind to her husband and sister in 1941. The aim of the study is to prove the genuineness of these notes and make authorship identification to make sure that these suicide notes belong to Virginia Woolf. The method employs semantics forensic and "the Linguistic Inquiry and Word Count" (LIWC) program (p.52). A detailed analysis of the content of suicide notes has been done by examining every single part; the greeting phrase's structure, the first phrase, body, purpose, and the closing phrase. The results confirm the genuine of the two suicide notes and indicate that the suicide notes contain a higher percentage of words that refer to depression and blame for

oneself. Although Virginia Woolf does not directly express her intention to commit suicide, within her letter, she expresses positive emotion to her husband and sister as she indirectly says goodbye (Malini & Tan, 2016).

In another study by "Alfiyan (2018)", in this study, an analysis of one particular suicide note that belongs to Nusadi, a 17 years old boy who has committed suicide in 2016. His suicide has been controversial because all his family members and friends confirm that Nusadi has been a successful boy and has won several competitions in his school; he is far from being depressed. Thus, this study analyzes the content of his suicide notes by using the semantic-pragmatics relevance theory and the "Linguistic Inquiry and Word Count" (LIWC) program. The analysis focuses on the use of the adjectives and sociolinguistics aspects of the suicide note; how the environment influences the language choices. The findings of this study support the validity of the suicide notes, and it indicates Nusadi's depression, which leads him to commit suicide, by the use of the mixture between the positive and negative adjectives (Alfiyan, 2018).

Furthermore, "Hamilton (2013)" also comes from the discipline of psychology. Within his doctoral dissertation, he performs a psychological study of suicide notes. Hamilton deals with suicide notes

from another perspective. He, first, investigates the motives of committing suicide, especially the desire to escape from psychological and physical pain; second, he investigates how each individual commits suicide and the circumstances that have made them do it. In his work, Hamilton (2013) uses a collection of suicide notes from the Metropolitan area of Dayton, Ohio. He makes a detailed analysis of the individual who commits suicide and leaves a note, and after that, he compares the results of this analysis with the results that he has received from the other group, which includes a collection of file cases to an individual who has not left a note. Of course, the details of such analysis are not relevant here but what is relevant is the analysis of the content of suicide notes and the method that has been used.

Hamilton (2013) classifies the suicide notes under investigation into two groups; the first group includes the suicide notes that have been retrieved from suicidal people who commit suicide to escape from psychological pain, while the second group includes the suicide notes that have been retrieved from suicidal people who commit suicide for other motives. The results of such analysis indicate that the suicide notes contain the themes of love, apology, trauma, and pain. Furthermore, the results also show that there are differences in the

content of suicide notes between the two groups (Hamilton, 2013).

Additionally, Roubidoux (2012) investigates the power of suicide note writing in his work. His work has linked the previous studies that have dealt with the characteristics of a powerful person and the way he reflects this power in his writing. Using critical discourse analysis, Roubidoux (2012) examines the 60 pairs of notes that have been presented by Shneidman and Farberow (1957) for the linguistic terms that express power, like the use of personal pronouns. The results of such a study indicate that the use of pronouns can undoubtedly express the power in writing; the results show that the writers tend to use the first-person active pronouns rather than the passive ones, the exclusive pronouns rather than the inclusive, and the singular pronouns rather than the plural. Furthermore, Roubidoux uses these results to distinguish between genuine and simulated suicide notes and concludes that the two types of suicide notes use the pronouns variously (Roubidoux, 2012).

To conclude, different methodologies, tools, and programs have been used to analyze the suicide notes within the studies mentioned above. Different scholars from different disciplines show an interest in this subject; some of them deal with words only and others with the whole message;

some of them are interested in the grammatical and syntactic levels, while others with the phrase and sentence levels. They all have corpora of suicide notes, whether it is written suicide notes or electronic ones. In other words, according to the researchers' knowledge, studies that deal with suicide notes can take four different forms. These include; a) investigating the authorship of a particular note; b) comparing the genuine suicide notes with the simulated ones; c) comparing the suicide notes that have been written by suicidal people with non-suicidal ones, d) and finally, studies focus on analyzing the content of suicide notes.

### **SUICIDAL THOUGHTS POSTS**

#### **SOCIAL MEDIA RELATED STUDIES**

Besides the enormous deaths by suicide, suicidal thoughts, suicidal ideations, and suicide attempts, which are frightening phenomena in the real and virtual world, usually cause severe and dangerous consequences leading in one way or another to death. When dealing with suicide, one must have general information about some terminologies related to suicide to distinguish between the kind of acts each terminology describes, such as suicide thought, non-suicidal self-harm, suicide attempts, and suicidal ideation. So, according to Klonsky et al. (2016), suicidal thoughts are "any suicidal thought or

actions without taking additional steps to distinguish thoughts from plans, from nonfatal attempts, and from attempts that result in death." At the same time, non-suicidal self-harm is "intentional self-injury without intent to die." As to suicide attempt, it can be defined as "a non-fatal, self-directed, potentially injurious behavior with an intent to die as a result of the behavior even if the behavior does not result in injury" finally, concerning suicide ideation can be defined "as thinking about, considering, or planning suicide" (p.14). According to Boyd and Ellison (2007), Social network sites are:

"web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site" (p.1).

Thus, social media are free channels that people use to express what they have in their minds in general, and concerning what this thesis focuses on the most, they also express their thoughts about suicide. Therefore, social media are one of the most significant sources that provide data for these types of studies, and these data can be obtained from different platforms

such as Facebook, Instagram, Twitter, and Reddit. Several studies that have been conducted on online websites have emphasized their values and considered them a window from which we can view people's real emotions. It is also stated that people tend to post their last words on these kinds of websites. Previous studies about human writing state that our writings reflect our psychological states in one way or another; we can learn a lot about a person by studying his/her writing or merely watching his choice of words and his style in writing. Thus, the researcher will list some of these studies that deal with social media data and are still relevant to the current thesis in the following sub-section:

**Brown et al. (2019)**

Brown et al. (2019) work deals with the online posts that have been posted by people with suicidal ideation on Instagram. By applying a qualitative and quantitative linguistic analysis to the data they have received from fifty-two Instagram members, they investigate the suicidal ideation and whether there is a link between acute suicidality and the language that the participants have employed in their Instagram activities. All the participants are teenagers at age 16, and they usually post their suicidal ideations on Instagram. For the qualitative analysis, Brown et al. (2019) interview the participants using

Instagram messenger; they present some questions and ask them to write their answers in the chat (p.4). While for the quantitative analysis, they collect the conversations and the details that the participants add when they post a picture on their accounts ( in picture captions) and utilize "Linguistic Inquiry and Word Count" LIWC software to perform a linguistic analysis (p.5). The results of the qualitative analysis are related to questions that the participants have answered. Meanwhile, the results of the quantitative linguistic analysis, which is relevant to the current study, indicate that there is a slight difference within the language that the participants have used in the interview but no significant differences in languages that the participants have used in their Instagram activities (in picture captions).

Concerning the language that has been used in the interview, the results indicate that the participants with acute suicidality employ more negative emotional words and words that express emotional expressions than the participants with non-acute suicidality, while the other variables are being similar (Brown et al., 2019, p.6).

Brown et al. (2019) conclude that the language that the participants in their Instagram activities have used for about four weeks before the study cannot be used to distinguish between acute suicidality and non-acute suicidality participants since

the Instagram platform is not dedicated to a lengthy discussion platform in contrast Instagram platforms are mostly about sharing photos than expressing their thoughts with long discussion; thus the linguistic analysis in this platform is not very helpful in detecting acute suicidality behavior.

#### **Litvinova et al. (2017)**

Recently, many studies that aim to investigate suicidality have shifted their focus to finding new ways to detect suicidal tendencies automatically. Earlier studies, which have employed suicide notes related to specific authors like Baddeley et al. (2011) and many others, have aimed to create mathematical models, programs, or algorithms that can automatically classify a text as suicidal texts or not. These studies have faced several theoretical and practical criticism; as Litvinova et al. (2017) indicate, they use the data which belong to specific authors that do not necessarily fit with all the writings that belong to this genre. As a result, this work and many others have shifted their focus to using online data that are more diverse and can be obtained as large samples that enable researchers to come up with generalizations easily without any restrictions (p.244).

Litvinova et al.'s (2017) work aim to create a mathematical model that can identify a text, whether it is a suicidal or

non-suicidal text, based on numerical features that have been collected from online writing to young Russian people who have committed suicide; the novelty of this work is that it uses Russian data (p.244). To extract the numerical features of the mathematical model, Litvinova et al. (2017) utilize "the Linguistic Inquiry and Word Count (LIWC)" program to analyze two types of corpora. The first one is related to suicidal texts that have been collected from 45 Russian individuals who have already committed suicide aged from 15-25, while the 26

#### **Kavuluru et al.(2016)**

Kavuluru et al.(2016) conduct a study that aims to recognize the automatically useful comments that have been posted in the Suicidewatch subreddit by employing machine learning and techniques of common language methods (Kavuluru et al., 2016, p.2). The data of this study is about 3000 comments, which are added to Subreddit's API system and nominated firstly by students who are doing the suicide prevention (SP) survey at Kentucky University and Western Kentucky University (p.4). After using the machine-learning model to provide text classification according to three kinds of characteristics: lexical, psychometric, and discourse characteristic, they observe that this model display a high rate for recognizing useful comments, which are

about 90% F-score as well as the existence of contrast and attribution discourses which are regarded as a portion of carefully answers, are seen as very useful with helping comments (p.9). The second corpus is related to non-suicidal texts that have been collected from 1,000 Russian students (Litvinova et al., 2017, p.246). the features of the mathematical model have been selected based on the categories that can be found in the (LIWC) programs, such as “words per sentence (WPS), percent of words longer than six letters (Sixltr), total function words, total pronouns, adverbs, prepositions, etc.” (p.247). The results of applying this mathematical model report 71.5% accuracy in diagnosing a text, whether it is suicidal or non-suicidal once.

#### **Michell (2013)**

In his work, Michell (2013) deals with online communication on Facebook to investigate the authorship of a disputed text by examining its stylistic features compared to several other texts written by known authors. The data in this study consist of eight female participants, their age between 30 to 40, and they share the same demographic categories so that no differences in writing can be found regarding the social, educational, and economic levels. This work aims to explore whether it is possible to identify the author of a particular text, which

usually tends to be short, that has been posted online on social websites or not by using a combination method driven from the frameworks that have been presented by McMenamin (2002) and Olsson (2008). To accomplish his aims, Michell (2013) has stylistically analyzed the disputed text for its style markers (keyness, the frequency of punctuation marks, and the frequency of function words) and has listed them to be compared with each of the eight known authors’ texts to identify the similarities and the differences between them (p.63).

Michell (2013) utilizes Wordsmith tools and the Chi-square test to perform such quantitative analysis. Wordsmith is a software that can be used for the general purposes of text analysis, and its more frequently used with larger numbers of words; different types of corpora. This software has been developed by the British linguist Mike Scott in 1996. It has 12 utilities; within this work, three of its main functions have been used: keyword, wordlist, and Concordance lines.

Meanwhile, the Chi-square test is a statistical test; that can be used to differentiate between two types of texts depending on the frequency of a particular item; it should be noted that if the total frequency of this particular item is less than ten, it cannot be used. The Chi-square test is used within the corpus-based studies

to confirm or refute the null hypothesis, which usually indicates no significant differences in the sets of frequency that a particular item has in the two types of corpora. If the probability of the results of the Chi-square is greater than 0.05, the null hypothesis is accepted, while if the probability of the results of the Chi-square is greater than 0.05, the null hypothesis is rejected (Chaski, 2001, p.9). The qualitative and quantitative analysis results are to come up with a list of the linguistic features of the disputed texts and a list of linguistic features of each of the participant's texts to be compared individually with the disputed text and indicate the similarities and differences. Thereafter, the (writer G), as Michell (2013) refers to him, shares the most relevant linguistic features in his writing with the disputed text (p.131).

#### **Burnap et al. (2015)**

This study depends on a Machine classifier that helps identify the most disturbing content and other themes relating to suicide to provide the outcome of texts relating to suicide that can be found on Twitter. This study also aims to distinguish unsteady references linking to suicide. To achieve these aims, Burnap et al. (2015) employ the following characteristics: psychological, structural, lexical, and emotive characteristics that can be elicited

from social media (Twitter) to establish a collocation of baseline classifiers.

Then, depending on the results of baseline classifiers, which have been developed by establishing another group of classifiers by utilizing the Rotation Forest Algorithm and reporting classifiers with Maximum prospect, they infer the following: Lexical and Grammatical characteristics are almost not found in principal components and are regarded as ineffective in these components, Affective Lexical domains are regarded as very suitable to an emotional and persuasive case which is linked to a particular sort of language and as most relevant features in these components, Wordlists and regular expressions which are elicited from online web linking discussion fora as well as other websites, are seen as very sophisticated patterns and very relevant in principal component by taking the form single-words and wordlist. Unlike the above linguistics characteristics, taken from Twitter, the ideas that act as general semantic domain are seen as ineffective in the components, and Sentiment Scores are also regarded as ineffective and hardly found in each class of component (pp.81-83).

**Greaves and Dykeman (2019)**

Greaves and Dykeman's (2019) study is an example of the types of studies that have been conducted depending on Reddit posts. It depends on "the Linguistic Inquiry and word count (LIWC) software" to investigate the employment of non-suicidal self-injury (NSSI) specific words and the psychological and linguistic characteristics by establishing a corpus that aims to gather general posts that have been posted in Reddit's API System. The data of this study are the public post which is only written in the English language. After establishing a corpus to gather the data and analyzing these data by using LIWC, the researchers infer the following: according to the first object, which is about identifying the reason of include these specific NSSI words in the posts, they discover that reason for posting a particular post which includes NSSI words, is that most participants asking for attention, help, or they feel scared. Concerning the second aim, they discover that the first-person singular is the most occurred pattern like (I, me, and my). Finally, concerning the third aim, which is about the psychological patterns, they discover that negative emotion language is the most occurring type in the texts (Greaves & Dykeman, 2019).

**Jasim and Jaafar (2022) A Corpus Stylistic Study on Suicide notes**

The study was carried out to investigate the language used in a collection of suicide notes and a collection of suicidal thought posts retrieved from Reddit. This study followed an eclectic approach McIntyre and Walker's (2019) approach to conduct a detailed corpus stylistics analysis, and it adapted Leech and Short's (2007) methodology to conduct stylistic analysis based on selected categories that had been chosen from their checklist. The quantitative analysis was performed using WordSmith 8.0 software (Scott, 2021). The objectives of the study are: Firstly, to characterize the linguistic stylistic features of suicide notes posted online and present a simplified model that carries these features, Secondly to characterize the linguistic stylistic features of the second set of data. Thirdly, to examine the similarities and differences between the two types, and finally, the third objective which is about the effectivity of keywords analysis in revealing the style of suicide notes. The results concerning the first objective denoted that the online Suicide notes had distinctive features that could differentiate them from other posts. Further, concerning the second and third objectives, the results also revealed the linguistic stylistic features of the second

type of data and indicated the similarities and differences between the two types.

## FINDINGS

Primitively, all the previous studies that have been mentioned revolve around are “Suicidal people,” “Suicide” and “Suicide note” The first study that has dealt with a collection of suicide notes is De Boismont’s study in 1856 and only after that, many other scholars start to approach the suicide notes from different perspectives (O’Donnell et al., 1993, p.45). Comparing the suicide notes with different other pieces of writing is the most prevalent perspective that the suicide notes studies deal with. The majority of the previous studies are about comparing genuine suicide notes with simulated ones by utilizing different methods and tools, whether manually or by using computer software. The classical study, or it can be considered the basic study within this perspective, is Shneidman and Farberow’s (1957) study in which they present a collection of 721 suicide notes that includes 66 pairs of notes, 33 of them have been written by actual suicidal people, and 33 of them have been written by a control group of non-suicidal people (O’Donnell et al., 1993, p.45). Most of the other pre-mentioned studies have used this collection of notes when they try to analyze the content of suicide notes and

have used the 66 pairs of notes when they try to distinguish between the genuine and simulated suicide notes. So, this can be considered as one of the defects of these previous studies because most of them have used the same data but only different procedures have been applied.

The majority of the studies on suicide are psychological in nature that have been written by psychologists and psychiatrists. Furthermore, these studies include both psychological and linguistic analysis of suicide notes; the researchers have only focused on the linguistic analysis of these studies. Most of them aim to extract a statistical model of the content of suicide notes. Each one of these studies is highly contributing to the history of suicide notes studies; Shneidman and Farberow’s (1957) main contribution is that the dataset that the followed studies depend on. In contrast, Osgood and Walker’s (1959) main contribution is their findings, which reveal that suicide notes tend to have a greater stereotype, disorganized, shorter sentences, simple words, and simple cognitive concepts.

Meanwhile, Gottschalk and Gleser’s (1960) main contribution is when they categorize each word in the texts into a particular grammatical and psychological category; this study considers as the beginning of semantic tagging analysis of suicide notes. Ogilvie et al.’s (1966) main

contribution is the use of computer analysis and semantics tagging by using “the General Inquirer Software” and connecting it with the “Harvard II Psycho-Sociological Dictionary” that helps them to specify singular words inside different classifications in the dictionary. While Edelman & Renshaw’s (1982) main contribution is the new method of analysis using a reliable and coherent discourse analysis system, which is “The Syntactic Language Computer Analysis.”. Finally, Gregory’s (1999) main contribution is how he has collected all the variables from the other studies and formulated a model that investigates them all.

These studies lack the linguistic stylistic interpretation of the statistical results they received from their analyses, especially from a stylistician’s perspective. Furthermore, there are several numbers of studies that deal with suicide notes from different perspectives, and they do include linguistics interpretations like the studies

that deal with authorship identification and those that deal with a specific suicide note that belongs to one particular person.

## CONCLUSION

This paper shows different methodologies and approaches that have been employed in dealing with suicide notes in particular and online posts in general, and it also shows that different stylistic approaches, theories, and corpus tools have been utilized to extract the linguistic stylistic features of a particular text/corpus. Finally, according to the researchers’ knowledge, no previous study has been conducted mainly to perform a stylistics analysis of the online suicide notes and suicidal thought posts found in social media platforms using corpus tools; such analysis can help detect the similarities differences between the types of data.

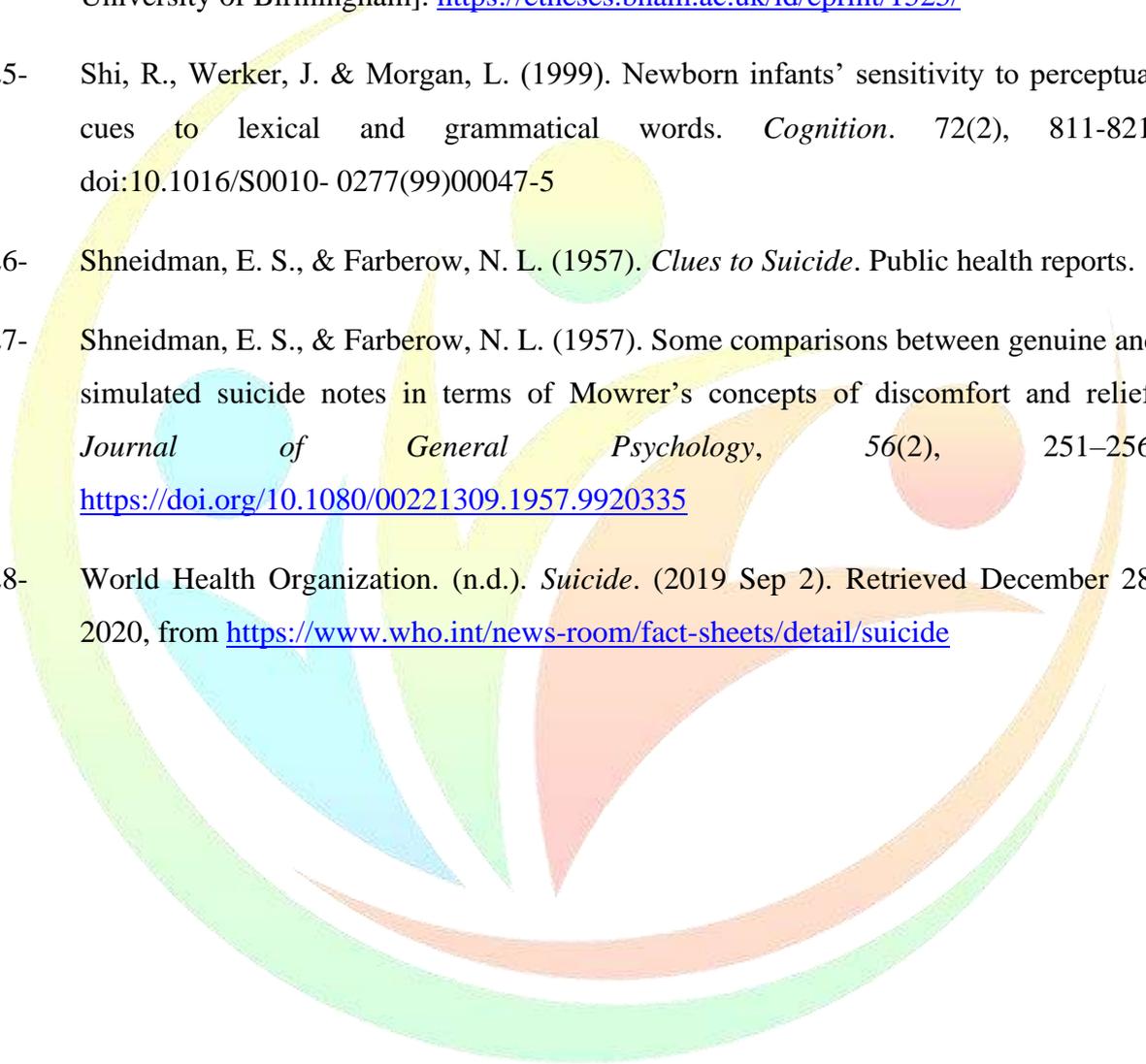
IJRSSH

## REFERENCES

- 1- Baddeley, J. L., Daniel, G. R., & Pennebaker, J. W. (2011). How Henry Hellyer's use of language foretold his suicide. *Crisis*, 32(5), 288-92
- 2- Barak, A., & Miron, O. (2005). Writing characteristics of suicidal people on the Internet: A psychological investigation of emerging social environments. *Suicide and Life-Threatening Behavior*, 35(5), 507-524.
- 3- Boyd, d. m., & Ellison, N. B. (2007). Social networking sites: definition, history and scholarship. *Journal of Computer-Mediated Communications*, 13(1), 210–230.
- 4- Brown, R. C., Bendig, E., Fischer, T., Goldwisch, A. D., Baumeister, H., & Plener, P. L. (2019). Can acute suicidality be predicted by Instagram data? Results from qualitative and quantitative language analyses. *PLoS ONE*, 14(9), 1–12. <https://doi.org/10.1371/journal.pone.0220623>
- 5- Burnap, P., Colombo, W., & Scourfield, J. (2015, August). Machine classification and analysis of suicide-related communication on Twitter. In *Proceedings of the 26th ACM conference on hypertext & social media* (pp. 75-84). Association for Computing Machinery.
- 6- Chaski, C. E. (2001). Empirical evaluations of language-based author identification techniques. *Forensic Linguistics*, 8(1), 1–65. <https://doi.org/10.1558/sll.2001.8.1.1>
- 7- Collins Concise English Dictionary. (n.d). Suicide notes. In *Collins Concise English Dictionary*. HarperCollins. Retrieved December 15, 2020, from <https://www.collinsdictionary.com>
- 8- Edelman, A. M., & Renshaw, S. L. (1982). Genuine versus Simulated Suicide Notes: An Issue Revisited Through Discourse Analysis. *Suicide and Life-Threatening Behavior*, 12(2), 103–113. <https://doi.org/10.1111/j.1943-278X.1982.tb00917.x>
- 9- Gottschalk, L. A., & Gleser, G. C. (1960). An analysis of the verbal content of suicide notes. *British Journal of Medical Psychology*, 33, 195–204.
- 10- Greaves, M. M., & Dykeman, C. (2019). A corpus linguistic analysis of public Reddit blog posts on non-suicidal self-injury. *arXiv preprint arXiv:1902.06689*.

- 11- Hamilton, D. R. (2013). Suicide as an escape from pain: an analysis of suicide notes and case files. [Doctoral dissertation, Wright State University].
- 12- Jasim H.A. and Jaafar, E.A. (2022) A Corpus Stylistic Analysis of Suicide Notes and Suicidal Thought Posts.(Forthcoming)
- 13- Kavuluru, R., Williams, A. G., Ramos-Morales, M., Haye, L., Holaday, T., & Cerel, J. (2016, October). Classification of helpful comments on online suicide watch forums. In *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 32–40. <https://doi.org/10.1145/2975167.2975170.Classification>
- 14- Klonsky, E. D., May, A. M., & Saffer, B. Y. (2016). Suicide, Suicide Attempts, and Suicidal Ideation. *Annual Review of Clinical Psychology*, 12, 307–330.
- 15- Litvinova, T. A. (2016). Corpus studies of speech of individuals who committed suicides. *Russian Linguistic Bulletin*, 3 (7).
- 16- Litvinova, T. A., Seredin, P. V., Litvinova, O. A., & Romanchenko, O. V. (2017). Identification of suicidal tendencies of individuals based on the quantitative analysis of their internet texts. *Computación y Sistemas*, 21(2), 243-252.
- 17- Malini, N., & Tan, V. (2016). Forensic linguistics analysis of Virginia Woolf's suicide notes. *International Journal of Education*, 9(1), 52–57.
- 18- McMenamain, G. R. (2002). *Forensic linguistics: Advances in forensic stylistics*. CRC press.
- 19- Michell, C. S. (2013). Investigating the use of forensic stylistic and stylometric techniques in the analyses of authorship on a publicly accessible social networking site (Facebook) [Doctoral dissertation, University of South Africa].
- 20- O'Donnell, I., Farmer, R., & Catalan, J. (1993). Suicide notes. *The British Journal of Psychiatry*, 163(1), 45–48.
- 21- Ogilvie, D. M., Stone, P. J., & Shneidman, E. S. (2009). Some characteristics of genuine versus simulated suicide notes. In K. Krippendorff & M. Bock (Eds.), *The content analysis reader* (pp. 404–409). Sage.
- 22- 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.  
<https://doi.org/10.1037/h0047078>
- 23- Roubidoux, S. M. (2012). Linguistic manifestations of power in suicide notes: An investigation of personal pronouns [Doctoral dissertation, University of Wisconsin].
- 24- Shapero, J. J. (2011). The language of suicide notes [Doctoral Dissertation, The University of Birmingham]. <https://etheses.bham.ac.uk/id/eprint/1525/>
- 25- Shi, R., Werker, J. & Morgan, L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*. 72(2), 811-821. doi:10.1016/S0010-0277(99)00047-5
- 26- Shneidman, E. S., & Farberow, N. L. (1957). *Clues to Suicide*. Public health reports.
- 27- Shneidman, E. S., & Farberow, N. L. (1957). Some comparisons between genuine and simulated suicide notes in terms of Mowrer's concepts of discomfort and relief. *Journal of General Psychology*, 56(2), 251–256.  
<https://doi.org/10.1080/00221309.1957.9920335>
- 28- World Health Organization. (n.d.). *Suicide*. (2019 Sep 2). Retrieved December 28, 2020, from <https://www.who.int/news-room/fact-sheets/detail/suicide>

The logo for the International Journal of Research in Social Sciences and Humanities (IJRSSH) is a large, stylized graphic. It features a central figure that resembles a person or a flame, composed of several overlapping, curved shapes in shades of blue, green, yellow, and orange. This central figure is set against a background of a large, light-colored circle. Below the graphic, the acronym 'IJRSSH' is written in a bold, orange, sans-serif font.

IJRSSH