Short Communication

COMPUTATIONAL ANALYSIS OF HINDI SUICIDE NOTES USING SNARE

Subhash Chandra[1*], Carole E. Chaski[2]

1. Forensic Psychiatry Fellow (PGY-5), University of Pennsylvania, Perelman school of Medicine, USA

2. Carole E. Chaski, Ph.D., Institute for Linguistic Evidence, 25100 Trinity Drive, Georgetown, DE 19947, USA

*Corresponding author's Email: schandra2@yahoo.com

ABSTRACT

Suicide notes play a pivotal role in death investigation. SNARE (Suicide Note Assessment REsearch) software classifies texts as a suicide note or control text-type with accuracies from 80%-88%, depending on text length, from a database of about 1,000 English sources. Our objective was a pilot study to determine SNARE's reliability among a non-English-speaking sample. Suicide notes were collected in New Delhi from 33 cases of suicide confirmed by autopsy findings, psychological autopsy, inquest paper, and crime-scene investigation. Thirteen out of fifteen legible notes were translated from Hindi into English and run through the SNARE algorithm. The software classified 8 of 13 translated texts as suicide notes (61.5%) and five as control texts (38.5%). All the misclassified notes were longer than 80 words. We concluded that even with the limitations of translation and lower accuracy of SNARE at high word count, the accuracy of 61.5% is greater than that of humans in differentiating between genuine and simulated suicide notes.

Keywords: Suicide note, SNARE, Cross-cultural.

INTRODUCTION

As per the World Health Organization 2012 suicide data, over 800,000 people die due to suicide every year accounting for 1.4% of death worldwide [1]. In India, the rate of suicidal deaths ranged from 15.34 to 21.74 per 100,000 populations, whereas, the global suicide rate is 16 per 100,000 population [2]. Suicides occur cross-culturally, and the incidence of suicide notes varies from 23.4 to 36.2% of suicide victims [3].

In any investigation of death, all circumstances must be taken into account for thoroughness. The assessment of texts left near the dead, "near" in terms of proximity or time, is a best practice for complete death investigation [4]. These documents can contain essential clues in both real suicide and homicide disguised as suicide.

Typically, there are three phases of suicide note assessment during a death investigation. First, the law enforcement officer is given the task of assessing whether a text is a suicide note or not, but there are several reasons that it is problematic for a law enforcement officer to be conducting this assessment [5]. The law enforcement officer's assessment may be influenced by the crime scene or even the reporting person's criminal record [6]. In the next phase of the investigation, the forensic pathologist may be given the task of assessing the note, an assessment that would naturally be influenced by the physical trauma to the body; for instance, a neck injury may be consistent with either suicide or homicide [7]. These skilled professionals get assigned a task in which they have not been trained resulting in misclassification of an important document. A third phase of the investigation may include the services of a forensic psychologist or psychiatrist. Pestian has reported that mental health professionals demonstrate a 50% accuracy at differentiating real suicide notes from simulated suicide notes [8,9].

ALIAS

Automated Linguistic Assessment and Identification System is a suite of text analysis tools built for forensic linguistic analysis [10-12]. One module, SNARE: Suicide Note Assessment Research, classifies text as either a real suicide note or a control document. SNARE was built using a database of more

than four hundred true suicide notes in English ranging from 2 to 1500 words and 500 control documents of other related text types. Most of the documents were American English with a few British English suicide notes, but there were no non-native English suicide notes. On the English database, SNARE has an accuracy rate of 88% [5]. Notes that are shorter than 45 words are more accurately classified by SNARE than longer notes, because longer notes begin to have more elements of the control documents such as apologies or love letters [5]. Considering Pestian's reports, the accuracy of computational analysis supersedes that of humans in differentiating between genuine and simulated suicide notes [8].

Given the universality of suicide deaths and suicide notes, we posed the question: can SNARE work cross-culturally? Then, we decided that we could operationalize this issue by testing how well SNARE works on texts translated into English.

Data

15 suicide notes were collected in India, while one of us (SC), bilingual in Hindi and English, was studying Forensic Medicine and Toxicology at Lady Harding Medical College, New Delhi, India. The notes came from 33 cases in which suicide was confirmed as the manner of death by autopsy findings, psychological autopsy, inquest paper, and crime scene investigation conducted by the first author (SC). Of the 15 suicide notes, one was a picture with a smiling face made on the left palm of the suicide victim and the other one was written on the palm and volar aspect of the wrist of the left arm. Excluding those two suicide notes because of illegibility, there were total 13 suicide notes which were analyzed in this study.

Method

The first author translated the suicide notes into English and input the texts into the SNARE module of web ALIAS, a version of ALIAS that is accessible using a web browser and Internet connection. SNARE uses computational linguistic analysis to identify and count linguistic features in the text. The text analysis results in a quantification of each text; this quantification is used to classify the text as either a suicide note or a control text. SNARE pairs linguistic features with a statistical classifier and gives the result as a suicide note or a control text.

The statistical classifier underlying SNARE is built on approximately 1000 documents which are known to be either real suicide notes or control texts. The particular classifier is Linear Discriminant Function Analysis, using stepwise entry and Mahalanobis distance. Further, the Linear Discriminant Function Analysis is subjected to Leave-One-Out cross-validation for deflating classification rates, and producing the most conservative classification model [13]. The classification model underlying SNARE reaches a cross-validated accuracy of 88% when the suicide note data is less than 45 words, and 86% cross-validated accuracy when the suicide note data is longer. When a new text is placed into SNARE, it is classified according to the statistical model with 86% accuracy. This means that there is at least a 14% error rate, or we can expect SNARE to make 14 errors if it is classifying 100 new texts as either a real suicide note or a control text. We would, therefore, expect at least 2 of the data texts to be misclassified.

RESULTS

Table 1 shows the results of the SNARE classification. Instead of the expected two errors, SNARE misclassified 4 of the real suicide notes as control texts. Interestingly, all of the actual suicide notes that were misclassified as control texts are longer than 80 words, (SN1, SN4, SN90, and SN11) but SNARE was able to accurately assess one very long suicide note (SN12).

Table 1: SNARE results of translated Hindi suicide notes.

Doc ID	Death Characteristics	Word count	SNARE Classification
SN 1	Hanging, suicide note on son's book	108	Control Text
SN 2	Gunshot by a police officer on duty	59	Suicide Note
SN 3	Poisoning by Aluminum Phosphide	17	Suicide Note
SN 4	Hanging	88	Control Text
SN 5	Hanging	17	Suicide Note
SN 6	Hanging	49	Suicide Note
SN 7	Hanging	44	Suicide Note

SN 8	Poisoning	34	Suicide Note
SN 9	Hanging	90	Control Text
SN 10	Hanging	74	Suicide Note
SN 11	Burning	391	Control Text
SN 12	Hanging	428	Suicide Note
SN 13	Hanging	488	Control Text

DISCUSSION

Suicide notes are difficult to classify. Suicide notes have low occurrence and are non-stereotypical or highly inconsistent thus increasing chances of inter-rater and intra-rater variability [14]. With SNARE the authors intend to provide a standard scientific tool for medicolegal purposes which would be acceptable cross-culturally for identifying a given document as a suicide note. The major issue in studying suicide notes is the restriction on the availability of suicide notes. A suicide note is evidence and chain of custody is maintained until the manner of death is confirmed. Sometimes there is 2-3 year time lag before they are available for research or they may not be obtainable at all. This explains the small number of suicide notes in our study. Other researchers have faced similar issues; being able to retrieve only 22 suicide notes over three year period [15].

Given the results of SNARE's classification of these Hindi notes, we conclude that it is indeed possible to use SNARE on translated notes with better than chance accuracy, i.e., better than the accuracy Pestian reported for mental health professionals. But there is clearly a difference in SNARE classification accuracy between native English (80-88%) and translated texts (61.5%). This difference could be attributed to multiple factors: small sample size, SNARE database mostly of American and British suicide notes and the limitation of the software in identifying notes with more than 45 words. We are conducting two supplemental studies. First, we are testing to see if translations produced by a translator with a literary rather than scientific background are classified by SNARE with higher or lower accuracy. Second, we are using TATTLER, multilingual software for forensic text analysis, to develop a Hindi version of SNARE.

In conclusion, this pilot test has shown that translated texts can be processed with at least a fair accuracy with SNARE. Thus, at least some linguistic characteristics of suicide notes appear to be universal and cross-cultural. Further, as with any artificial intelligence tool, rational analysis is not replaced by SNARE, but the human analysis can be supplemented. SNARE provides an objective classification with a known error rate to the mental health professional, forensic pathologist or law enforcement officer tasked with identifying a text as a suicide note or not. Although there were only 13 notes in the current dataset, we think that the dataset provides at least a beginning of an answer to our research question and demonstrates the utility of a computational text analysis tool for suicide note assessment.

REFERENCES

1. Organization WH: Mental health - Suicide Prevention, 2012.

2. Badiye A, Kapoor N, Ahmed S: An empirical analysis of suicidal death trends in India: A 5-year retrospective study. J Forensic Leg Med 27:29-34, 2014.

3. Shioiri T, Nishimura A, Akazawa K, et al.: Incidence of note-leaving remains constant despite increasing suicide rates. Psychiatry and Clinical Neurosciences 59:226-228, 2005.

4. Koehler SA: The role of suicide notes in death investigation. Journal of Forensic Nursing 3:87-88, 92, 2007.

5. Chaski D: Is This a Real Suicide Note? Authentication Using Statistical Classifiers and Computational Linguistics. American Academy of Forensic Sciences, 2011.

6. Chaski CE: Text-typing a Suicide Note and Justice for Diaz-Perez., July 2006.

7. Pusztai A, Bugan A: [Analysis of suicide notes from persons committing completed suicides]. Psychiatria Hungarica: A Magyar Pszichiatriai Tarsasag tudomanyos folyoirata 20:271-80, 2005.

8. Pestian J, Nasrallah H, Matykiewicz P, et al.: Suicide note classification using natural language processing: A content analysis. Biomedical Informatics Insights 2010:19, 2010.

9. John P. Pestian1 PMaML-g: What's In a note: construction of a suicide note corpus. 2012.

10. Chaski CE: Who's At the Keyboard? Recent Results in Authorship Attribution. International Journal of Digital Evidence 4, 2005.

11. Chaski CE: The Keyboard Dilemma and Author Identification. Advances in Digital Forensics III New York, 2007.

12. Chaski CE, Croghan R: Text and Pretext on the Internet: Recognizing Problematic Communications.

13. Jones NJ, Bennell C: The development and validation of statistical prediction rules for discriminating between genuine and simulated suicide notes. Archives of suicide research : official journal of the International Academy for Suicide Research 11:219-33, 2007.

14. Eisenwort B, Berzlanovich A, Heinrich M, et al.: [Suicidology: suicide notes and their themes]. Der Nervenarzt 78:672, 4-8, 2007.

15. Namratha P, Kishor M, Sathyanarayana Rao TS, et al.: Mysore study: A study of suicide notes. Indian J Psychiatry 57:379-82, 2015.