ORIGINAL RESEARCH

What's In a Note: Construction of a Suicide Note Corpus

John P. Pestian[1], Pawel Matykiewicz[1] and Michelle Linn-Gust[2]

[1]Cincinnati Children's Hospital Medical Center, Department of Biomedical Informatics, University of Cincinnati, Cincinnati OH.
[2]American Association of Suicidology, Washington DC. Corresponding author email: john.pestian@cchmc.org

Abstract: This paper reports on the results of an initiative to create and annotate a corpus of suicide notes that can be used for machine learning. Ultimately, the corpus included 1,278 notes that were written by someone who died by suicide. Each note was reviewed by at least three annotators who mapped words or sentences to a schema of emotions. This corpus has already been used for extensive scientific research.

Keywords: natural language processing, computational linguistics, corpus, suicide

Introduction

This research centres on the content of suicide notes. The topic has been studied a great deal, but not to this magnitude. Over three years, suicide notes were collected, digitized, annotated, and analyzed. The sections of this paper present what is known about these notes, how they were collected and annotated, and a discussion of some implications.

Background
Content of notes

Across all age groups, between 10% and 43% of those who end their lives leave a suicide note. What is in a suicide note? Menniger suggested that the wish to die, the wish to kill and the wish to be killed must be present for suicide to occur,[1] but there is a paucity of research about the presence of those three motives in suicide notes. Brevard, Lester, and Yang[2] analyzed notes to determine whether Menniger's three concepts were present in suicide notes. Without controlling for gender, they reported more evidence for the wish to be killed in suicide notes of completers (those who die) than in the notes of non-completers.[2] Leenaars et al[3] revisited Menninger's triad by comparing 22 suicide notes and 22 parasuicide notes that were carefully matched. They concluded that the notes from suicide completers were more likely to have content reflecting anger or revenge and less likely to show escape as a motive. Additionally, although not statistically significant, completers often tended to show self-blame or self-punishment. Another study of 224 suicide notes from 154 subjects characterized note-leavers as young females of non-widowed marital status with no history of previous suicide attempts, no previous psychiatric illness, and religious beliefs. Suicide notes written by young people were found to be longer, rich in emotions, and often begging for forgiveness. Another study noted that genuine notes often included statements such as the experience of adult trauma, expressions of ambivalence, feelings of love, hate and helplessness, constricted perceptions, loss, and self-punishment. One important and consistent finding is the need to control for differences in age and gender.[3]

Using suicide notes for clinical purposes

Of those who attempt suicide for the second time, at least 15% are lethal. As noted by Freedenthal "determining the likelihood of a repeated attempt is an important role of a medical facility's psychiatric intake unit and notoriously difficult because of a patient's denial, intent for secondary gain, ambivalence, memory gaps, and impulsivity".[4] One indicator of the severity and intent is simply the presence of a suicide note. Analysis has shown that patients presenting at an emergency department with non-fatal self-harm in addition to a suicide note are more likely to be at increased risk for completing suicide later.[5] Evidence from a suicide note may illuminate true intentions, but the lack of one does not obviate important questions such as: without a note is the patient's risk less severe? how many patients died by suicide without leaving a note? or is there a difference between the notes of completers and those of attempters? Valente matched notes from 25 completers and 25 attempters and found differences in thematic content surrounding fear, hopelessness, and distress.[6] On the other hand, Leenaars found no significant difference between thematic groups.[3]

The emergence of Natural Language Processing (NLP) and machine learning methods presents the opportunity to re-examine the previous efforts with new analytical tools. Those tools, however, require an annotated corpus of suicide notes.

Methods
Corpus development

A corpus is a collection of written works. An annotated corpus is one that has been reviewed for certain characteristic words, concepts, or sentences, such as: anger, hopelessness or peace. Here we collected 1,319 notes written by people before they died by suicide. The notes were collected between 1950 and 2012 by Drs. Edwin Shneidman, UCLA, and John Pestian, Cincinnati Children's Hospital Medical Center (CCHMC). Database construction began in 2009 and has been approved by the CCHMC's Institutional Review Board (#2009-0664). Each note was scanned into the Suicide Note Module (SNM) of our clinical decision support framework called CHRISTINE.[7] The scanned notes were then transcribed to a text-based version by a professional transcriptionist. Each note was then reviewed for errors by three separate reviewers. Their instructions were to correct transcription errors but to leave indigenous errors like spelling, grammar and so forth alone.

Anonymization

To assure privacy, the notes were anonymized. To retain their value for machine learning purposes, personal identification information was replaced with like values that obscure the identity of the individual.[8] All female names were replaced with "Jane," all male names were replaced with "John," and all surnames were replaced with "Johnson." Dates were randomly shifted within the same year. For example, Nov. 18, 2010, may have been changed to May 12, 2010. All addresses were changed to 3333 Burnet Ave., Cincinnati, OH, 45229, the CCHMC's main campus.

Annotators

It is the role of an annotator to review a note and select those words, phrases, or sentences that represent a particular emotion. Recruiting the most appropriate annotators led us to consider "vested volunteers," or volunteers with an emotional connection to the topic. The emotional connection is what makes this approach different than crowd-sourcing,[9] where there is no known emotional connection. In our case, these vested volunteers who are routinely called *suicide loss survivors*, were generally active in a number of suicide communities. Approximately 1,500 members of several online communities were notified of the study via e-mail or indirectly via Facebook's suicide bereavement resource pages. Of those communities, two groups: Karyl Chastain Beal's online support groups *Families and Friends of Suicides* and *Parents of Suicides*, and the *Suicide Awareness Voices of Education*, directed by Dan Reidenberg, Psy.D., were most active. The notification to potential participants included information about the study, its funding source and what would be expected of a participant. Respondents were vetted in two stages. The first stage was to meet the inclusion criteria (at least 21 years of age, English as a primary language, willingness to read and annotate 50 suicide notes). The second stage included an e-mail sent to participants. In the e-mail, respondents were asked to describe their relationship to the person lost to suicide, the time since the loss, and whether or not the bereaved person had been diagnosed with any mental illness. Demographic information about the vested volunteers is described below. Once fully vetted, they were given access to an automated training site. Training consisted of an online review and annotation of 10 suicide notes. If the annotators agreed with the gold standard at least 50% of the time, they were asked to annotate 50 more notes. They also were reminded that they could opt out of the study at any time if they had any difficulties and were given several options for support.

Emotional assignment

Each note in the shared task's training and test set was annotated by at least three individuals. Annotators were asked to identify the following emotions: abuse, anger, blame, fear, guilt, hopelessness, sorrow, forgiveness, happiness, peacefulness, hopefulness, love, pride, thankfulness, instructions, and information. A special web-based tool was used to collect, monitor, and arbitrate annotator's activity. The tool collects annotation at the word and sentence level. It also allows for different concepts to be assigned to the same word. This feature made it impossible possible to use a simple κ inter-annotator agreement coefficient.[10] Instead, Krippendorff's α[11] with Dice's coincidence index[12] was used. Artstein and Poesio[13] provided excellent explanation of the differences and applicability of variety of agreement measures. There is no need to repeat their discourse; however, it is worth explaining how it applies to the suicide note annotation.

Table 1 shows an example of a single note annotation done by three different coders. At a glance, one can see that the agreement measure has to accommodate multiple coders (a_1, a_2, a_3), missing data, and multi-level agreement ("anger, hate" and "anger, blame" where $d_{Dice} = 1/2$ vs. "hate" and "anger, hate" where $d_{Dice} = 1/3$. Krippendorff's α accommodates all these needs and enables calculations for different spans. Despite the annotators being asked to annotate sentences, they usually annotated clauses and in some cases phrases. For this shared task, the annotation at the token level was merged to create sentence level labels. This is only an approximation to what happens in suicide notes. Many notes do not have typical English grammar structure so none of the known text segmentation tools would work well with this unique corpora. Nevertheless, this crude approximation yields similar inter-annotator agreement (see Table 2). Finally, a single gold standard was created from the three sets of sentence level annotations. There was no reason to adopt any *a priori* preference for one annotator over another, so the democratic

Table 1. Example of a note annotation for different span with corresponding Krippendorff's α and the majority rule.

	I	Hate	You	.	I	Love	You	.	α
Token									
a_1		Hate				Love			
a_2	Anger, hate	Anger, hate			Love	Love			≈0.570
a_3	Anger, blame	Anger, blame	Anger, blame		Love	Love	Love		
Sentence									
a_1		Hate				Love			
a_2		Anger, hate				Love			≈0.577
a_3		Anger, blame				Love			
Majority									
m		Anger, hate				Love			

Table 2. Annotator characteristics.

	Annotators	
Response to call	Directly contacted	1500
	Indirectly contacted via social media	Unknown
	Did not meet inclusion criteria	10
	Completed training	169
	Withdrew before completing annotation	17
	Respondents who fully completed the annotation	64
Annotators performance	Number of notes annotated at least once	1278
	Number of notes annotated at least twice	1225
	Number of notes annotated at least three times	1004
	Mean (SD) annotation time per note	4.4 min (1.3 min)
	Token inter-annotation agreement	0.535
	Sentence inter-annotation agreement	0.546
Gender and age	Males	10%
	Females	90%
	Average age (SD)	47.3 (11.2)
	Age range	23–70
Education level	High school degree	26
	Associates degree	13
	Bachelors	23
	Masters	34
	Professional (Ph.D/MD/JD)	4
Connection to suicide	Survivor of a loss to suicide	70
	Mental health professional	18
	Other	12
Time since loss occurred	0–2 years	27
	3–5 years	25
	6–10 years	14
	11–15 years	13
	16 years or more	12
Relationship to the suicide	Child	31
	Sibling	23
	Spouse or partner	15
	Other relative	9
	Parent	8
	Friend	5

Notes: Some survey questions were not completed. Thus eliminating the ability to sum columns.

principle of assigning a majority annotation was used (see Table 1). This remedy is somewhat similar to the Delphi method, but not as formal.[14] The majority of annotation consists of those codes assigned to the document by two or more of the annotators. Several problems are possible with this approach. For example, it could be that majority of the annotation will be empty. The arbitration phase focused on notes with the lowest inter-annotator agreement where that situation could occur. Annotators were asked to re-review the conflicting notes but, not all of them completed this final stage of the annotation process. Thirty-seven percent of sentences had a concept assigned by only one annotator.

Results

Annotators

The characteristics of the annotators are described in Table 2.

Note content

Selected characteristics of the data are found in Table 3. This table provides an overview of the data using Linguistic Inquiry and Word Count, 2007. This software contains within it a default set of word categories and a default dictionary that defines those words should be counted in the target text files.[15]

Discussion

This paper reports on the results of an initiative to create and annotate a corpus of suicide notes that can be used for machine learning analysis. Sentiment analysis is the process of identifying emotions in text and then evaluating that process. Finding emotional sentiment in text is complex because each annotator brings a different psychological perspective. Agreement between annotators in the range of 0.60–0.80 is considered substantial while a range of 0.40–0.59 is considered moderate.[16] Our moderate performance is what we expected given the amount of notes and annotator differences. In a post-hoc error analysis we found that about 120 sentences were responsible for most of the annotators confusion.

Nevertheless, this corpus provides much opportunity for understanding the language of those who have died by suicide. In particular it creates a vital

Table 3. Frequency and example of assigned emotions.

Description	Frequency	Example
Instructions	609	Careful, cyanide gas in the bathroom
Hopelessness	601	I just didn't want to live anymore
Love	472	I love her
Information	430	I have no debts except for what my wife knows
Guilt	423	Forgive me please
Sorrow	342	Oh, how I suffer
Blame	235	I have been pushed around too much
Hopefulness	216	You will a happy and healthy life
Thankfulness	187	You, John have been so good to me and Jane
Anger	183	Well, Jane I hope this makes you happy!
Fear	154	I am terrified
Happiness/peacefulness	119	I'm ready for the next step with joy and anticipation
Pride	89	We have another sweet little daughter
Forgiveness	61	I do not blame you for anything, my dear
Abuse	53	Life is so cruel when you are persecuted by in-laws and ex-wife

resource for scientists to conduct machine learning and data mining on a large corpus of suicidal language. In one instance it was used as the basis for an international challenge in which 22 teams built machine learning methods designed to identify emotions in the suicide notes.[17] Future uses will no doubt include development of machine learning methods that leads to clinical application. Finally, while this work has focused on the content of emotions and all the challenges of psychological phenomenology that come with this approach, future work should consider how the structural characteristics, including parts of speech and sentence length, should be included.

Acknowledgements

Suicide Loss *Survivors* are those who have lost a loved one to suicide. We would like to acknowledge the roughly 160 suicide loss survivor volunteers who annotated the notes. Without them this research could not be possible. Their desire to help is inspiring and we always will be grateful to each and every one of them.

We would like to acknowledge the efforts of Karyl Chastain Beal's online support groups Families and Friends of Suicides and Parents of Suicides, and the Suicide Awareness Voices of Education, a non-profit organization directed by Dan Reidenberg, Psy.D.

Finally, we acknowledge the extraordinary work of Edwin S. Shneidman, Ph.D., and Antoon A. Leenaars, Ph.D., for their commitment to understanding the complexity of suicide.

Author Contributions

Conceived and designed the experiments: JP. Analyzed the data: JP, PM, MLG. Wrote the first draft of the manuscript: JP. Contributed to the writing of the manuscript: PM, MLG. Agree with manuscript results and conclusions: JP, PM, MLG. Jointly developed the structure and arguments for the paper: JP, PM, MLG. Made critical revisions and approved final version: JP, PM, MLG. All authors reviewed and approved of the final manuscript.

Funding

Competing Interests

Author(s) disclose no potential conflicts of interest.

Disclosures and Ethics

As a requirement of publication author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

References

1. Menninger, K Man Against Himself, Harcourt, Brace & World, 1938.
2. Brevard A, Lester D, Yang BJ. A comparison of suicide notes written by suicide completers and suicide attempters. *Crisis.* 1990;11(7):7–11.
3. Leenaars AA, Lester D, Wenckstern S, Rudzinski D, Breward A. A comparison of suicide notes written by suicide notes and parasuicide notes. *Death Studies.* 1992:16.
4. Freedenthal S. Challenges in assessing intent to die: can suicide attempters be trusted? *Omega (Westport).* 2007;55(1):57–70.
5. Barr W, Leitner M, Thomas J. Self-harm or attempted suicide? Do suicide notes help us decide the level of intent in those who survive? *Accid Emerg Nurs.* 2007;15(3):122–7. Epub Jul 2, 2007.
6. Valente SM. Comparison of suicide attempters and completers. *Med Law.* 2004;23(4):693–714.
7. Pestian JP. Spencer M. Matykiewicz P. Zhang K. Vinks AA. and Glauser T. Personalizing Drug Selection Using Advanced Clinical Decision Support, Biomed Inform Insights. 2009 June 23;2:19–29.
8. Pestian JP, Brew C, Matykiewicz P, et al. A shared task involving multi-label classification of clinical free text. In ACL, editor, *Proceedings of ACL BioNLP.* Prague, Jun 2007. Association of Computational Linguistics.
9. Howe J. The rise of crowdsourcing. *Wired Magazine.* 2006;14(6):1–4.
10. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement.* 1960;20(1):37–46.
11. Krippendorff K. *Content Analysis: An Introduction to its Methodology.* Sage Publications, Beverly Hills, CA; 1980.
12. Dice LR. Measures of the amount of ecologic association between species. *Ecology.* 1945;26(3):297–302.
13. Artstein R, Poesio M. Inter-Coder agreement for computational linguistics. *Computational Linguistics.* 2008;34(4):555–96.
14. Dalkey NC, Rand Corporation. *The Delphi method: An experimental study of group opinion.* Defence Technical Information Center; 1969.
15. Pennebaker JW, Chung CK, Ireland M, Gonzales A, Booth RJ. The development and psychometric properties of liwc2007. *Austin, TX, LIWC. Net;* 2007.
16. Artstein R, Poesio M. Inter-coder agreement for computational linguistics. *Computational Linguistics.* 2008;34(4):555–96.
17. Pestian JP, Matykiewicz P, Linn-Gust M, South B, Uzuner O, Wiebe J, et al. Sentiment Analysis of Suicide Notes: A Shared Task. *Biomed Inform Insights.* 2012;5(Suppl 1):3–16.