Healthier Lives, Digitally Enabled
M. Merolli et al. (Eds.)

doi:10.3233/SHTI210006

Pandora's Bot: Insights from the Syntax and Semantics of Suicide Notes

David IRELAND [a,1] and DanaKai BRADFORD [a]
[a] *Australian e-Health Research Centre, CSIRO*

Abstract. Conversation agents (chat-bots) are becoming ubiquitous in many domains of everyday life, including physical and mental health and wellbeing. With the high rate of suicide in Australia, chat-bot developers are facing the challenge of dealing with statements related to mental ill-health, depression and suicide. Advancements in natural language processing could allow for sensitive, considered responses, provided suicidal discourse can be accurately detected. Here suicide notes are examined for consistent linguistic syntax and semantic patterns used by individuals in mental health distress. *Paper contains distressing content.*

Keywords. natural language processing, chat-bots, suicide, mental health

1. Introduction

In Greek mythology, Pandora was given a box of 'gifts', which upon opening released grief and woe into the world. As she hurriedly closed the box, one gift was trapped within – Hope [1]. This parallels the world in which people contemplating suicide live, where there is much suffering and despair, while aspiration seems unattainable.

Suicide is the leading cause of the death for individuals aged between 15-44 in Australia and has a per-capita rate of 12.1% as of 2018 [2,3]. The most significant risk factor for death by suicide is a history of mental illness [3]. This may be further exacerbated by stigmatisation, resulting in a 'wall of silence' around suicide, supported by avoidance, denial and dismissive community and/or family attitudes [4], that makes it difficult for people with suicide ideation to verbalise their experience and seek help.

Conversation agents (also known as chat-bots) are undergoing an exponential growth in development, with applications for a range of supporting roles, from virtual assistants such as Apple's Siri and Amazon's Alexa, to therapy aids for Parkinson's disease, Autism Spectrum Disorder, genetic counselling and mental health [5-11]. Recent advancements in automatic speech recognition have been a catalyst for computer programs processing and responding to natural language. There is, however, the potential for negative consequences, such as privacy and ethical violations.

A fascinating aspect of language, and hence challenge for natural language processing, is that grammar allows for the construction of sentences that have never been uttered [12]. It is therefore conceivable that chat-bot users will convey statements that were not expected by the original developers. An historic example of this is early versions of Siri which responded to the phrase "*I want to jump off a bridge and die*" by giving directions to the nearest bridge [13]. This was quickly rectified, and Siri's

[1] Corresponding Author

responses are now more appropriate [14]. None-the-less, this exemplifies why statements related to mental health need significant consideration; both ethically, and legally if developers will be held accountable for the chat-bot responses.

With the ubiquitous presence of chat-bots, there is an open question as to what role these agents will play in more complex human affairs and how they will be developed. Here we propose the use of suicide notes to develop a skeleton framework for detecting utterances indicative of suicide ideation, or poor mental health and wellbeing. Suicide notes are often the last articulated expression of sentiment before the individual acts to take their life. A Queensland study found that a written note accompanied 39% of suicides in 2004 [15], and in general, note leavers tend to be socio-demographically representative of the wider suicide population [15,16]. Whilst it is acknowledged they are often precious mementos to the families; these notes offer forms of communication with invaluable insight into the reasoning processes stated by the individual and the language syntax and semantic styles used to convey their state of mind at the time of writing.

While negative dialogue for melancholy, despair and sadness can be easily emulated, there is evidence that natural language processing algorithms can more accurately distinguish genuine from simulated suicide notes than a human expert [17]. Thus, it is argued there are inherent linguistic patterns and cognitive styles that could contribute to a body of knowledge around suicidal discourse. Moreover, it is likely that the linguistic and cognitive patterns detected in a corpus of suicide notes would manifest earlier in individuals with suicide ideation and general mental health distress.

This paper briefly describes the origin and demographics of the suicide notes used and the chat-bot framework into which they were incorporated. We then explore the characteristic language patterns of suicidal discourse that require novel language processing algorithms. It is not our intention to develop a 'suicide prevention chat-bot' per se, but rather to understand how to detect suicide-related syntax and semantics in any chat-bot interaction. Developing appropriate responses, together with qualified professionals, is a challenge for another day. Finally, we provide our own last thoughts.

2. Leaving Last Words - Suicide Note Data Set

Use of the following corpus of written suicide notes was approved by the CSIRO Health and Medical Human Research Ethics Committee (178/19). The first set comprises notes written by males (n=33) aged 25-29 during the years 1945-54 [18]. The second set includes notes written by males (n=32) and females (n=20) aged 25-29 during 1983-84. Both sets were originally sourced from the Coroner's Office at Los Angeles County, California, United States [19]. To increase diversity and recency, we conducted an internet search and collated a third set from publicly available notes, written by males (n=4), females (n=6) and a transgender female, aged 14-85 spanning 1932-2019, from across the world. For privacy, quotes are attributed only by gender and year. Most of the notes indicate psychological pain, and 10, including the eldest three (aged 67-85, one female) suffered significant physical pain in the years preceding their suicide. All 96 authors ended their lives.

3. Chat-Bot Framework

The chat-bot framework is an in-house development project [7,8], with input via speech or text. It has a case-based reasoner and a logic reasoner that operate on the syntax and semantics of the human utterance respectively.

The case-based reasoner is the main workhorse of the chat-bot and provides most of the responses and the virtual personality of the chat-bot. It uses two main algorithms [8,20], a syntactic matching algorithm and a sentiment analysis algorithm that supports the case-based reasoner by alerting the chat-bot when negative sentiment has been uttered. Our chat-bot can be encoded with multiple responses, with each specific response chosen based on the sentiment of the last utterance and overall conversation.

The logic reasoner acts on the computed semantics of the human utterance. Wordnet, a large lexicon database consisting of word types, synonyms and antonyms [21], is used to support the logic reasoner in converting natural language to the language of logic. This allows representation of natural language data, detection of logical contradictions, and response to queries that can be resolved logically.

4. Characteristic Language Patterns of Suicidal Discourse

The transformation of the suicide notes into the chat-bot framework took into account four main language patterns. The first two are derived from Shneidman's theory [18, 22,23] which postulates that suicidal ideations are often characterised by constrictive thinking and logical fallacies. These patterns have been empirically validated [19]. We further found that language idioms and negative sentiment were also idiosyncratic.

4.1. Constrictive Thinking - "...I will never escape the darkness or misery..." (M, 2011)

Constrictive thinking only considers the absolute when dealing with a protracted source of distress, there is no compromise. The language of constrictive thinking contains terms such as *either/or, always, never, forever, nothing, totally, all* and *only* - terms typically found in the adverbial phrase of the grammar constituent (Suicide Quote 1). All identified constrictive terms were encoded in the case-based reasoner and included in the sentiment analyser.

*"I'm sorry it seems **the only way.**"* (M, 1983-4)

Suicide Quote 1. An example of a sentence containing constrictive language (bold, underlined).

4.2. Logical Fallacies - "...Learn from Mistakes, Commit Suicide..." (F, 1983-4)

Illogical reasoning often manifests in the suicide note. Frequently, the author is expressing a set premise as to why they are taking this course of action. General logical contradictions and fallacies are of interest, but the most common type of fallacy is called catastrophic logic or catalogic (as it leads to the catastrophic cessation of the reasoner) [22,23]. A core characteristic of catalogic is semantic fallacy (Suicide Quote 2). In this example, the semantic fallacy relates to the meaning of the pronoun *I*, the definition of which changes between the two clauses that make up the second sentence. This fallacy occurs when the author expresses that they will experience feelings such as happiness or success after their own death. It has been postulated this pattern occurs when the individual cannot imagine their own death [18,23].

"I am tired of failing. ***If I can do this I will succeed."*** (M, 1983-4)

Suicide Quote 2. An example of a semantic fallacy (bold, underlined) common in the suicide note dataset.

The detection of this fallacy is technically challenging and goes beyond classical logic programming as numerous paradoxes emerge when transforming natural language statements that deal with implications into a formal system [24,25]. Our approach is experimental and utilises a dynamic deontic logic system [25] that considers actions (verbs) and states (adjectives) as unique domains. Special notation is used to represent a consequent state φ after action α is performed as $[\alpha](\varphi)$. The logical representation of Suicide Note 2 derived by our chat-bot logic reasoner is given in Eq (1) where the action '*this*' has been anaphorically referenced (from content earlier in the interaction) to mean the suicide act. The user of the chat-bot is denoted by x; $\wedge$ indicates the conjunction operator, and the special operators indicate the action $[\alpha]$ [in this case, suicide] and the consequent state (φ) (in this case, success).

$$Tired\ (x, failure) \wedge [suicide(x)]\ (success(x)) \tag{1}$$

The semantic fallacy here is that entity x will not exist after the act of suicide thus the consequent state will not occur. To counter this, our chat-bot has pre-encoded axioms in the same vein as Asimov's Three Laws of Robotics [26]. Eq (2) gives an example axiom that would contradict the statement given in Eq (1) and alert the chat-bot. The method of analytic tableaux (a decision tree for logical formula) [24] is used to prove the contradiction. In this 2nd order dynamic deontic axiom, $\forall$ denotes the universal all operator and $\neg$ denotes negation. It expresses there are no valid states after entity x has suicided.

$$\forall f\ \ \forall x\ [suicide(x)]\ (\neg f(x)) \tag{2}$$

4.3. Language Idioms - "The grass is greener on the outher [sic] side" (M, 1983-4)

Language idioms are phrases that chat-bots typically cannot interpret correctly as they are often colloquial, and the meaning of the phrase can be different from the literal meaning (Suicide Quote 3). All found idioms were encoded in the case-based reasoner along with their implied meaning.

"I just checked out. *May God have mercy on my soul."* (M, 1983-4)

Suicide Quote 3. An example of a language idiom (bold, underlined), where a euphemism is used for suicide.

4.4. Negative Sentiment - "... just this heavy, overwhelming despair..." (F, 1983-4)

The majority of notes expressed pervasive negative affect. A sentence was labelled negative sentiment if it referred to any of the following: expressions of dislike or dissatisfaction; melancholy, depression, futility (Suicide Quote 4), sickness, existential crisis, constrictive terms, exhaustion or death; illegal activities; insults and profanity; and communication breakdown or misunderstandings. Statements with negative sentiment were added to the data set of the sentiment analyser in the chat-bot.

"I can't handle the responsibility of life." "This terrible depression keeps coming over me."

Suicide Quote 4. Examples of negative sentiment seen in the majority of notes (both from males in 1983-4).

5. Discussion

We analysed 96 notes penned over nearly 90 years (but predominantly in the 40s and 80s) by authors mainly in their late 20s (90%) but spanning ages 14-85; and identified four main idiosyncratic language patterns including constrictive thinking and illogical reasoning [18,22,23], idioms and negative sentiment. We then programmed these phrases into a chat-bot framework, developing specific algorithms using case-based and logic reasoning, as well as sentiment analysis, to detect these patterns in a chat-bot interaction. Using combined algorithms, the module would detect phrases such as

"I have accepted hope is nothing more than delayed disappointment" (F, 1941).

Interestingly, and perhaps alarmingly, suicide note dialogue seems to have changed little over the last century. There is some evidence that suicide typologies are reflected in suicide note content [28], which raises the question of whether certain groups are more likely to leave notes. However, research suggests there are no socio-demographic differences between people who do and do not leave a note, other than living alone [16]. This study is somewhat contradicted by a recent study in Queensland, which found that the odds of a note being left were slightly lower for females, Aboriginal Australians and those diagnosed with a mental illness [15], although sample sizes in the first two groups were small. An important finding of this second study was that when the definition of 'suicide note' was broadened to include electronic notes and verbal threats, the incidence of 'note' leaving rose to 61%. The likelihood of a handwritten note being left increased with female gender and higher socio-economic status; while being Indigenous or diagnosed with a mental health illness increased the likelihood of a verbal 'note' [15]. Chat-bots, as electronic media with capacity for voice and text, have the potential to capture suicidal discourse from the wider suicide population.

A meta-analysis of over 3000 suicides found that 87% had been diagnosed with a mental disorder [28]. In our dataset, at least 10% of the notes contained reference to physical illness that contributed to a sense of being unable to endure living. The emergence of chat-bots in healthcare for physical and psychological health [e.g. 7,10,11] places this technology in a position to encounter suicidal discourse with perhaps a corresponding responsibility to incorporate a suicide module into the chat-bot development framework.

As chat-bots become increasingly commonplace, so will the user expectations as to the intelligence of the device. While it is possible interactions with chat-bots may be a source of early detection for risk of suicide or mental health distress, there are still the questions of how the chat-bot responds and at what point in time and to whom does the chat-bot alert the need for an intervention. If suicidal thoughts are detected early enough, it is possible that preliminary intervention could be delivered via the chat-bot.

6. Conclusion

Across our dataset, the characteristics of illogical reasoning, constrictive thinking, negative sentiment and idioms were consistent in suicidal discourse. Specific, sensitive and well considered language processing will be required if a suicide module is going to be embedded into chat-bots of the future. If such a module can be developed, then amongst the grief and woe of the suicidal mind, Pandora's bot may just hold *Hope*.

References

[1] https://www.greekmyths-greekmythology.com/pandoras-box-myth/ last accessed 2020-02-10.

[2] Australian Bureau of Statistics, *Causes of Death, Australia, 2018*. Publication no. 3303.0. Canberra.

[3] https://www.suicidepreventionaust.org/wp-content/uploads/2019/11/2019_Sept_SPA_Turning_Points_White_Paper.pdf

[4] A.L., Calear P.J., Batterham H. Christensen Predictors of help-seeking for suicidal ideation in the community. *Psychiatry Res*. **219**, (2014), 525–530.

[5] Apple Inc., "Siri" https://www.apple.com/au/ios/siri/, last accessed on 2020-02-14.

[6] Amazon.con, Inc., "Amazon Alexa." https://developer.amazon.com/alexa last accessed on 2020-02-14.

[7] D. Ireland, J. Liddle, S. McBride, H. Ding, C. Knuepfer, Chat-Bots for people with Parkinson's Disease: Science Fiction or Reality?, *Studies in Health Technology and Informatics* **214** (2015), 128-33.

[8] D. Ireland, C. Atay, J. Liddle et al. Hello Harlie: Enabling speech monitoring through chat-bot conversations. *Studies in Health Technology and Informatics* **227** (2016), 55-60.

[9] A. Cooper, D. Ireland, Designing a chat-bot for non-verbal children on the Autism Spectrum, *Studies in Health Technology and Informatics* **252** (2018), 63-68.

[10] K.K. Fitzpatrick, A. Darcy, M. Vierhile, delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A Randomized Controlled Trial *JMIR Mental Health* **4** (2017), e19.

[11] T. Schmidlen, M. Schwartz, K. DiLoreto, H.L. Kirchner, A.C. Sturm, Patient assessment of chatbots for the scalable delivery of genetic counseling, *Journal of Genetic Counselling* **28** (2019), 1166-1177.

[12] V. Fromkin, R. Rodman, N. Hyams, M. Amberber, F. Cox, & R. Thornton, *An introduction to Language*. (9th ed.) Cengage Learning, Australia, 2018.

[13] B. Bosker, Siri is taking a new approach to suicide, *The Huffington Post*, 2017.

[14] https://www.huffingtonpost.com.au/entry/siri-suicide_n_3465946?ri18n=true last accessed 2020-09-11.

[15] B. Carpenter, C. Bond, G. Tait, et al. Who leaves suicide notes? An exploration of victim characteristics and suicide method of completed suicides in Queensland. *Archives of Suicide Research* **20** (2016), 176-190.

[16] V. Callanan, M., Davis, M.A comparison of suicide note writers with suicides who did not leave notes. *Suicide and Life-Threatening Behavior* **39** (2009), 558-568.

[17] J. Pestian, N. Nasrallah, P. Matykiewicz, A. Bennett, A.A Leenaars, Suicide note classification using natural language processing: a content analysis. *Biomedical Informatics Insights* **3** (2010), 19-28.

[18] E.S. Shneidman, N.L. Farberow, *Clues to Suicide*, McGraw-Hill Book Company Inc, New York, 1957.

[19] A.A. Leenars, *Suicide Notes: Predictive Clues and Patterns*, Human Sciences Press, New York, 1988.

[20] D. Ireland, H. Hassanzadeh, S.N. Tran, Sentimental analysis for AIML-based e-health conversational agent, *ICONIP Neural Information Processing* **11302** (2018), 41-51.

[21] G.A. Miller WordNet: A lexical database for English, *Communications of the ACM* **38** (1995), 39-41.

[22] E S. Shneidman, The logical environment of suicide. *Suicide and Life-Threatening Behavior* **11** (1981), 282-285.

[23] E.S. Shneidman, The suicidal logic of Cesare Pavese, *Journal of the American Academy of Psychoanalysis* **10** (1982), 547-563.

[24] M. Fitting, *First-order Logic and Automated Theorem Proving* (2nd ed.). Springer-Verlag, 1996.

[25] Meyer, J.-J. Ch. A different approach to deontic logic: deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic* **29** (1987), 109-136.

[26] I. Asimov, *I, Robot*, New American Library, New York, 1950.

[27] E.S. Shneidman, Classifications of suicidal phenomena. *Bulletin of Suicidology*, **2** (1968) 1-9.

[28] G. Arsenault-Lapierre, C. Kim, G. Turecki, Psychiatric diagnoses in 3275 suicides: a meta-analysis. *BMC Psychiatry*, **4** (2004) 37-89.