# Quantifying Changes in Language to Evaluate the Social and Psychological Effect of Participation in a 'Pro-Choice' Suicide Forum

**Jane Dilkes**

University of Birmingham
jdilkes@pm.me

## Abstract

This study quantifies changes in language to evaluate the social and psychological effect of participation in a controversial 'pro-choice' suicide forum. A linguistic frequency approach shows that language becomes increasingly of a style that has been associated with better mental health with increasing participation on the forum, up to around the 100th post made by any participant, at which point language style does not in general continue to change significantly. A subsequent keyword analysis finds significant differences in discourse content between less established and more established participants, with the former focused on suicide method and time references, and the latter using more abstract terms predominantly addressed towards other participants. In addition, it is shown that the forum exhibits established characteristics of a virtual community. It is concluded that the forum may in general help support better mental health of participants, and that this may be related to its accommodation of concerns that are currently not tolerated in the discourse of wider society. There is evidence to suggest that the greater social threat currently may be to ban the forum.

## Introduction

In this study an empirical approach is taken to the pressing issue whether a contested 'pro-choice' suicide forum is the active threat to life it is represented as in worldwide campaigns to shut it down, or a community that in supporting discourse about suicide and related concerns may help to prevent some suicides. This may be related to internet self-harm and suicide content more generally, which as discussed in more detail below are important and complex health issues that require more research.

While suicide is the second leading cause of death globally in those aged 15-29 after road traffic accidents (ONS, 2019), and the leading cause of death in the UK of men aged 20-49, and of women aged 20-34, it is typically a taboo subject, and consequently is often misconstrued. The folk psychology around suicide can be dismissive and unhelpful, for example 'those who talk about suicide are not at risk of suicide'; 'suicidal behavior is motivated by attention-seeking';

'suicide cannot be prevented'; and 'thinking about suicide is rare' (O'Connor, 2021, p. 49). O'Conner (2021) states that the best way non-professionals may help those feeling suicidal is to listen without judgment to their concerns, including suicidal ideation, and without trying to 'fix' their personal existential crisis. While this may feel alien in wider society, it is the natural approach on the forum investigated here, whose participants may typically understand suicidal ideation from personal experience.

## Suicide and the internet

A UK Samaritans report on online self-harm and suicide content states that "The internet can be an invaluable resource for individuals experiencing self-harm and suicidal feelings, however, it can also provide access to content that can be distressing and triggering" (Samaritans, 2020). A systematic review into the relationship between internet use, self-harm, and suicidal behavior in young people (Marchant et al., 2017) similarly found that although there is a risk of harm from online behavior in terms of normalization, triggering, competition, and contagion relating to self-harm, with young people in particular increasingly using social media to communicate their distress and seek support, there are also benefits in terms of crisis support, reduction of social isolation, delivery of therapy, and outreach.

Suler (2005) argues that there is a disinhibition effect of online communities that is a result of dissociative anonymity, invisibility, asynchronicity, solipsistic introjection, dissociative imagination, attenuated status, and authority, but with anonymity, invisibility, and asynchronicity having by far the biggest effect. Grohol (1998) similarly found that individuals experiencing mental health issues talk more openly about their experience in such anonymous environments which remove the fear of negative evaluation; and Gilat and Shahar (2007) found that participants experience an intense sense of affiliation and emotional support through asynchronous communication with the online group, which

also gives the option to 'disappear' from an interaction if it becomes too difficult to deal with (Suler, 2005).

An ethnographic investigation into the harms and benefits of online peer-support relating to self-harm of children and adolescents found that moves to eradicate online self-harm content may cause unintentional harm (Lavis & Winter, 2020). Benefits have been found in having access to others going through the same experience (Stommel, 2009), particularly for an issue that may attract social sanctions and negative judgement, such as suicidal ideation (White & Dorman, 2001). A survey of suicide fora specifically found that "social support in suicide forums is rated as high as support from friends and higher than support from family" (Winkel et al., 2005, p. 714), with support from suicide fora found to correlate with participants' ratings of reduction of their own suicidality. Health professionals must therefore take a more balanced view of self-harm and suicide websites, which have been found to give participants "access to important, socially valued identities, such as being understood, belonging to a community, and coping with their problems" (Baker & Fortune, 2008, p. 118).

## Analysis of social media data relating to suicide

In comparison to the current study, which is concerned with the social and psychological effect of participation in a suicide forum, linguistic analyses of suicide concerns on social media typically have an intervention-motivated focus on locating those who may be having suicidal ideation (Aladağ et al., 2018; Robinson et al., 2016; Westerlund et al., 2015; Zirikly et al., 2019). The 6th Annual Computational Linguistics and Clinical Psychology Workshop shared task (Zirikly et al., 2019) used data from the Reddit forum r/SuicideWatch to identify those at risk of suicide. Data preprocessing for this shared task typically included removing stop words, which is a standard NLP approach. However, typical stop words such as personal pronouns and determiners have been found to be closely linked to people's social and psychological worlds (Pennebaker, 2011). The novel empirical approach of the current study focuses on these psychological correlates of language style for comparative posts relating to level of establishment on the forum. In addition findings are considered against established discourse characteristics hypothesized to indicate the related issue of virtual community (Herring, 2004).

# Data and methods

## Ethics
The British Association for Applied Linguistics good practice guidelines (BAAL, 2021) state that it is acceptable from an ethics viewpoint to use linguistic frequency information from internet discourse, which is the approach of the current study, without specific consent. The British Psychological Society ethics guidelines for internet-mediated research (BPS, 2017) similarly state that consent is not necessary where online data can be considered to be in the public domain. The usage guidelines of the forum investigated in the current study state that the data created by participants is in the public domain. These guidelines in conjunction with concern that the worse harm would be to disturb this participant-led forum led to the decision not to seek specific permission for use of this data.

## Data
Suicide discussion websites have a long internet history. The alt UseNet newsgroup alt.suicide.holiday (ash) started in November 1990 as a place to discuss the relationship between suicide rates and holidays. The 'pro-choice' stance and community language of ash, for example the 'catch the bus' metaphor used to support suicide discourse, transferred to the Reddit discussion website r/sanctionedsuicide subreddit in 2013. This subreddit was banned on 14th March 2018 for breaking Reddit guidelines, specifically it was said to be encouraging violence. The sanctionedsuicide identity restarted on 18th March 2018 at the independent web address sanctionedsuicide.com. This web address was banned by its registrar in January 2021, and the forum consequently moved to the new url sanctioned-suicide.org, also in January 2021, which is where the data used in the current study was collected in January 2022. The sanctioned suicide identity is a reference to the idea that certain forms of suicide are culturally sanctioned, for example Japanese seppuku, a ritual form of suicide used to restore honour; euthanasia; and terrorist martyrdom. These culturally-sanctioned forms of suicide are not regarded as actual suicide in that they are not equatable with depression and other mental illness, and therefore do not carry the shame typically associated with suicide (Pierre, 2015).

### Data collection
There are four major sub-fora on sanctioned-suicide.org: News & Announcements; Suicide Discussion; Recovery; and Offtopic. The data of the current study consists of all posts from the Suicide Discussion sub-forum, which is by far the biggest of these sub-fora and represents the main focus of the forum, and of the current study. Posts collected range from 18th March 2018, the first day of the instantiation of the forum at the independent web address sanctionedsuicide.com, up to 25th January 2022, the date the posts were collected from sanctioned-suicide.org. The Scrapy Python web scraping framework (Scrapy, 2021) was used to write web 'spiders' to collect the following elements of each forum post, each of which is stored as a distinct item:

- the id of the participant who made the post (anonymized)
- the date the post was made
- the text content of the post

This data is referred to in the current study as sanctioned-suicide.

**Sanctioned-suicide participants and posts**

The sanctioned-suicide data is very diverse, consisting of 810,094 posts made by 14,855 participants over 46 months, with participants making 55 posts each on average. 1,766 (12%) of participants have made only one post; 7,594 participants (51%) have made more than 10 posts; and 71 participants (0.5%) have made more than 1,000 posts each. To consider the implications of this pattern of engagement, the proportion of participants whose total posts=1, total posts=2 etc. up to total posts=5, and total posts>10, are listed in table 1 for sanctioned-suicide and two comparative health fora. It can be seen that there is relatively strong repeat engagement on sanctioned-suicide, on which 12% of participants make only one post, compared to 50% for a cancer forum and 31% for a depersonalization forum; the number of active participants on each comparative forum is shown in brackets.

| total posts | sanctioned-suicide (14,855) | cancer (42,311) | depersonalization (7,711) |
|:---:|:---:|:---:|:---:|
| 1 | 12% | 50% | 31% |
| 2 | 8% | 18% | 14% |
| 3 | 6% | 9% | 8% |
| 4 | 5% | 5% | 6% |
| 5 | 4% | 4% | 4% |
| >10 | 51% | 7% | 26% |

Table 1: proportion of participants with specific numbers of posts

Herring (2004) warns that there is often the impulse to categorize online phenomena in broader terms than are warranted by their actual instantiation online, "for example, all groups of people interacting online are 'communities'" (Herring, 2004, p. 339). The discourse characteristics of virtual communities set out by Herring (2004) are used in the current study to help determine whether sanctioned-suicide is a community, which has high relevance in terms of its function for participants. The distribution of participant posts on sanctioned-suicide meets the 'participation' criterion for identification as an online community, in that it demonstrates "Frequent, regular, self-sustaining activity over time" (Herring, 2004, p. 361). In addition there are a number of extended threads on sanctioned-suicide which, in conjunction with distribution of posts by participant, meets the 'interaction' criterion of an online community, which requires "Reciprocity, extended (in-depth) threads, core participants" (Herring, 2004, p. 361).

**Language style and mental health**

A wide range of experimental settings have been used to demonstrate the psychological implications of particular word use (Tausczik & Pennebaker, 2010). Research using social media data specifically has found language patterns relating to a range of emotional and psychological states, including the onset of depression (De Choudhury et al., 2013; Eichstaedt et al., 2018; Guntuku et al., 2019; Yates et al., 2017); and post-traumatic stress disorder (Coppersmith et al., 2014). Dominant findings associate an increased use of first-person singular pronouns, and the pronoun 'I' in particular, with increased depression and self-focus. In comparison increased use of first-person plural pronouns has been associated with better mental health, representing that the speaker is more socially connected to the group, while use of a wider range of pronouns has been associated with awareness of multiple perspectives. Use of emotion words has been associated with emotionality, with positive and negative emotion words associated with better and worse mental health respectively. In terms of suicidal ideation specifically: Egnoto and Griffin (2016) found that the general category 'personal pronouns' is nearly three times as prevalent in suicidal texts in comparison to homicide and neutral texts; O'Dea et al. (2017) found that strongly concerning suicide notes have increased use of first-person singular pronouns, anger words, and an increased focus on the present; Schoene et al. (2021) found an increase in use of the first-person singular pronoun I and a greater use of negations and verbs in genuine suicide note data and depression notes from the Reddit depression data, while pronouns, positive emotions, and an increased focus on the present were the most important linguistic categories for accurate classification of genuine suicide notes; and Wiltsey, Stirman and Pennebaker (2001) found that the poetry of suicidal poets contains throughout their careers more first-person singular pronouns, and fewer first-person plural pronouns.

Van den Nest, Till and Niederkrotenthaler (2019) found that in pro-suicide fora in comparison to anti-suicide and neutral fora there is a decreased rate of emotion words and words relating to social circumstances, and a higher rate of death- and aggression-related words. The current study, which is based entirely in data from a pro-choice suicide forum, focuses on the change over time of the language style of participants, whose persistence on the forum may intrinsically be understood as showing that any suicidal ideation that may have led them to the forum is, or has become, not critical. The focus in the current study, then, is not on the diverse language correlates of suicidal ideation found in the studies discussed above, which are based on quite specific settings: the focus here is on the pronouns and emotion words that have been found to be associated with mental health more generally (Tausczik & Pennebaker, 2010), which are summarized in table 2 .

| lexical category | psychological correlates of increase |
|---|---|
| first-person singular pronouns | worse: depression; self-focus |
| negative words | worse |
| first-person plural pronouns | better: more socially connected to the group |
| third-person pronouns | better: social support; self-distancing |
| range of pronouns | better: considering multiple perspectives |
| positive words | better |
| emotion words | emotionality |

Table 2: psychological correlates of language style

### Natural language processing and linguistic variables

In the current investigation text data is processed using the spaCy natural language processing Python library (Honnibal & Montani, 2021), which uses statistical language models to process text according to its context. Because of the focus in the current study on the psychological correlates of language style, of which pronouns are a key element, the general spaCy pronoun tag -PRON- is analyzed into individual personal pronouns. To determine the range of pronouns used in each post, a count is made of how many of 33 pronoun categories are used at least once. The spaCy-related Textacy Python library is used to calculate the entropy of each post, where entropy is a measure of the average uncertainty associated with a text. And the NRC (Canadian National Research Council) emotion lexicon (Mohammad, 2021), a crowd-sourced resource of words categorized as having positive or negative valence, and belonging in eight different emotion categories (Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, and Trust), is used here to identify and count emotion words. Based on these sources, these key linguistic variables that have been found to be associated with mental health are counted for each separate post.

### Diachronic analysis

Diachronic analysis in the current study is based on participant post index (ppi), where ppi=1, consisting of 14,855 posts from 14,855 participants, is the first post each participant made on sanctioned-suicide; ppi=2 is the second post made by any participant; ppi=3 is the third post and so on up to ppi=200, which is the 200th post made by any participant. Ppi calculations apply across all threads on the forum, and across all 46 months of data.

## Results

In this section the results are reported for two investigations that quantify the change in language of participants with increased participation on sanctioned-suicide. Investigation 1 considers the change in density of use of key elements of language style that have been associated with mental health (table 2). Investigation 2 considers the change in discourse content with increased participation on the forum.

### Investigation 1: changes in language style with increased participation on the forum

The effect on language style of participation on sanctioned-suicide is investigated here with a lexical frequency approach, with the focus, as discussed above, on personal pronouns and word emotion categories, which have consistently been found to be associated with mental health (table 2). The Python SciPy linregress method is used to calculate least squares regression for the average density of each linguistic variable over the range of ppi values investigated, and $r$ is interpreted such that 0.3 denotes a weak relationship; 0.5 denotes a moderate relationship; and 0.7 denotes a strong relationship.

The occurrence in any post of the linguistic variable pronoun_categories (how many of 33 pronoun categories are used), the pronouns I, me, my, we, our, you, your, they, and their, eight emotion categories, and the composite Negative and Positive emotion categories was calculated for each separate ppi, for ppi values from 1 to 200. Except for entropy, which is a value that refers to a whole text (in this case a single post), each variable is considered as a proportion of the total word count of each post – its density in the post. The corresponding $r$ value was calculated for the change of each linguistic variable over the range of ppi values investigated.

As discussed above a comparatively large proportion of sanctioned-suicide participants make more than just 1, 2, or 3 posts (table 1). However, to specifically address the issue that a pattern of change over the ppi range 1-200 towards language of a style that has been associated with better mental health may occur because participants whose language is of a style that has been associated with worse mental health may tend to leave the forum after making a small number of posts, a comparison analysis was made using data from participants who have made more than 10 posts (7,594 participants), which is approximately half of all participants. Those results are in the second column of table 3. It can be seen from table 3 that change in use over time for all of the variables analyzed is virtually the same whether posts for all participants are included, or just posts from participants who have made more than 10 posts.

The results in table 3 show that as ppi increases there is a significant decrease in density of use of the first-person singular pronouns I ($r$=-0.73), me ($r$=-0.48), and my ($r$=-0.69); a moderate increase in density of use of the first-person plural pronoun we ($r$=0.57) and the second-person pronoun you ($r$=0.50) used to address the other; and a weak increase in density of use of the second-person pronoun your ($r$=0.45), the third-person pronouns they ($r$=0.48) and their ($r$=0.47), and the emotion variables Anger ($r$=0.41), Disgust $r$=0.37), Fear ($r$=0.40), Joy ($r$=0.30), Sadness ($r$=0.30), and Trust ($r$=0.35), and the composite emotion variables Positive ($r$=0.39), and Negative ($r$=0.39). The first-person plural possessive pronoun our, the range of pronouns used (pronoun categories = pro. cat.), and the emotion variables Anticipation and Surprise show no significant change by ppi. The whole-post variable entropy shows a significant decrease over time ($r$=-0.61), suggesting that posts become more coherent and uniform. Significant values of $r$ are highlighted bold in table 3.

| ppi 1-200 | all participants | participants with > 10 posts |
|---|---|---|
| | $r$ | $r$ |
| pro. cat. | 0.20 | 0.21 |
| I | **-0.73** | **-0.74** |
| me | **-0.48** | **-0.48** |
| my | **-0.69** | **-0.69** |
| we | **0.57** | **0.57** |
| our | 0.24 | 0.23 |
| you | **0.50** | **0.50** |
| your | **0.45** | **0.45** |
| they | **0.48** | **0.48** |
| their | **0.47** | **0.46** |
| Anger | **0.41** | **0.39** |
| Anticipation | 0.15 | 0.15 |
| Disgust | **0.37** | **0.36** |
| Fear | **0.40** | **0.39** |
| Joy | **0.30** | **0.30** |
| Sadness | **0.30** | 0.29 |
| Surprise | 0.19 | 0.20 |
| Trust | **0.35** | **0.35** |
| Positive | **0.39** | **0.39** |
| Negative | **0.39** | **0.37** |
| entropy | **-0.61** | **-0.62** |

Table 3: change in density of variables from ppi 1 to 200

Figure 1 shows the pattern of change of the mean density per post as ppi increases for the variables pronoun_categories, pronoun_i (I), and the composite emotion variables Positive, and Negative, with data including all participants. It can be seen that for the variables pronoun_categories, pronoun_i, and Positive, the most dramatic change occurs in early posts up to around the 20th post made, while for the Negative emotion variable in comparison there is a more steady increase in density of use as ppi increases.
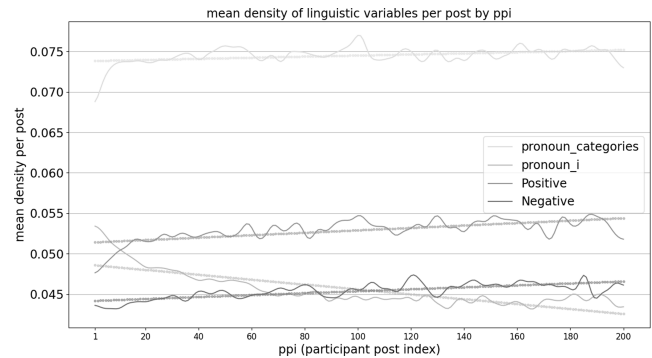


Figure 1: change in density of use of linguistic variables by ppi

To consider the pattern of change in more detail, the analysis was applied to the ppi groups described in table 4.

| ppi range | posts | participants |
|---|---|---|
| 1-5 | 61,297 | 14,854 |
| 6-10 | 43,712 | 9,720 |
| 11-50 | 187,550 | 7,594 |
| 51-100 | 115,993 | 3,097 |
| 101-200 | 126,863 | 1,782 |

Table 4: ppi groups for comparative analysis

The results in table 5 show that the change in density of the linguistic variables investigated is strongest in the ppi range 1-5, representing the newest participation on sanctioned-suicide. Between the first and fifth posts made on the forum there is a significant change in density for all the variables investigated, representing a dramatic change in language style: across this range of posts language becomes increasingly of a style that has been associated with better mental health. The exception to this is use of the first-person plural pronoun our, whose density decreases significantly in those first five posts ($r$=-0.44), and for which thereafter there is no significant change in density of use.

In comparison, for the ppi range 101-200 there is no significant change in the density of any of the linguistic variables: by about the time participants are making their 100th post on the forum, the density of use of these language style elements that have been found to be representative of mental health has stopped changing - it has stabilized at a style that is comparatively representative of better mental health than was found in earlier posts made. In table 5 all significant values of $r$ are highlighted in bold.

| ppi | 1-5 | 6-10 | 11-50 | 51-100 | 101-200 |
|---|---|---|---|---|---|
| pro. cat. | **0.88** | **0.74** | **0.41** | **0.09** | 0.04 |
| I | **-0.94** | **-0.57** | **-0.85** | **-0.56** | -0.03 |

| | | | | | |
|---|---|---|---|---|---|
| me | **-0.81** | **-0.59** | **-0.53** | -0.02 | 0.06 |
| my | **-0.93** | **-0.42** | **-0.57** | **-0.54** | -0.13 |
| we | **0.79** | **0.73** | **0.41** | 0.16 | 0.21 |
| our | **-0.44** | 0.25 | 0.05 | 0.00 | 0.08 |
| you | **0.95** | **0.52** | **0.80** | **0.42** | -0.15 |
| your | **0.93** | **0.83** | **0.56** | 0.15 | 0.14 |
| they | **0.63** | **0.42** | **0.61** | 0.11 | 0.11 |
| their | **0.93** | -0.17 | 0.25 | 0.14 | 0.08 |
| Anger | **-0.90** | -0.09 | 0.23 | 0.12 | 0.04 |
| Anticipation | **0.64** | **0.55** | **0.36** | -0.00 | 0.08 |
| Disgust | **-0.43** | **0.66** | 0.27 | **0.31** | -0.04 |
| Fear | **-0.70** | -0.05 | **0.36** | 0.01 | 0.08 |
| Joy | **0.89** | **0.54** | **0.56** | 0.12 | 0.06 |
| Sadness | **-0.99** | -0.07 | **0.33** | -0.05 | 0.07 |
| Surprise | **0.93** | **0.80** | **0.37** | 0.05 | 0.08 |
| Trust | **0.93** | **0.55** | **0.48** | 0.20 | 0.10 |
| Positive | **0.74** | **0.54** | **0.39** | 0.27 | 0.03 |
| Negative | **-0.82** | -0.05 | **0.49** | 0.11 | -0.00 |
| entropy | **-0.81** | **-0.83** | **-0.83** | **-0.47** | 0.15 |

Table 5: change in density of linguistic variables for ppi groups

The mean entropy of a post also decreases significantly for each set of posts except those in the range 101-200, suggesting that posts become more predictable and uniform with increased participation on the forum. The fact that language style on the forum is established across all participants by around the 100th post they make supports the 'social behavior' criterion for identification as an online community, in that it demonstrates "norms of appropriateness" (Herring, 2004, p. 361).

## Investigation 2: comparison of discourse content between posts with ppi=1 or ppi in range 1-5, against posts with ppi in range 101-200

The keyness of terms is typically used to explore what a corpus is 'about': it relates to the frequency of terms in the corpus of focus compared to their frequency in a reference corpus. To examine the difference in discourse content for new participants in comparison to established participants on sanctioned-suicide, keyword analyses are made here for the posts for which ppi=1 or is in the range 1-5 (less established participants), in comparison to posts for which ppi is in the range 101-200 (more established participants), each with the other as a reference corpus. Results are ranked by log likelihood of the difference in density of each term in each corpus ( p<0.0001). The AntConc software was used for this analysis (Anthony, 2022).

### Keyword analysis with all words included
This first keyword analysis compares all words from posts with ppi in the range 1-5 with all words from posts with ppi in the range 101-200. It can be seen from the results in table 6 that for the posts in which ppi is in the range 1-5 the keywords include four first-person singular pronouns, which are marked with an asterisk. In addition the greetings *hi* and *hello* are keywords for this group, indicating that typical social norms apply when participants first participate in the forum. For the more established participants in comparison (ppi in range 101-200) the characteristic pronouns are second-person, third-person, and first-person plural, also marked with an asterisk. This supports the findings of investigation 1 that language becomes more of a style that has been associated with better mental health with increased participation on the forum.

| | ppi 1-5 | keyness | ppi 101-200 | keyness |
|---|---|---|---|---|
| 1 | *i | 7676.381 | *you | 8679.71 |
| 2 | *my | 4034.669 | *your | 2032.68 |
| 3 | *me | 1709.404 | are | 1582.25 |
| 4 | hi | 1121.186 | *they | 1144.74 |
| 5 | *myself | 807.060 | *we | 1007.98 |
| 6 | hello | 786.103 | people | 897.05 |
| 7 | tried | 704.832 | *their | 578.73 |
| 9 | am | 605.464 | is | 516.07 |
| 10 | been | 545.751 | *us | 504.38 |

Table 6: keywords for posts with ppi in range 1-5 in comparison to posts with ppi in range 101-200

### Focusing on nouns
To focus on the comparative content of discourse, for the following keyword analyses all noun lemmas were extracted from posts.

### Noun lemmas: ppi=1 compared to ppi in range 101-200
The first keyword analysis of noun lemmas is of posts with ppi=1 in comparison to posts with ppi in the range 101-200. The first column in table 7 shows the 20 nouns that are most significantly increased in density where ppi=1 in comparison to the later posts, sorted by log likelihood (p< 0.0001). It can be seen that the most representative nouns for the first posts of all participants on the forum include the time references *year*, *month*, *time*, and *day*, marked *, and references to suicide method, including, *neck, pill, rope, suspension, belt, bag, helium, method, tank, knot*, and *hanging*, marked ! The remaining most representative nouns include *lurker*, which suggests that participants typically read the forum before making their first post.

A second keyword analysis of noun lemmas was made with the posts for which ppi is in the range 101-200 as the corpus of focus, and the set of posts with ppi=1 as the reference corpus. The results, in the second column in table 7, show that the most representative nouns for the more established participants include no time or suicide method references. The keywords for this group are more general and more abstract, including the noun lemmas *peace, luck*, and *journey*.

| | ppi=1 | keyness | ppi 101-200 | keyness |
|---|---|---|---|---|
| 18 | !hanging | 97.273 | patient | 40.997 |
| 19 | family | 96.195 | individual | 39.934 |
| 20 | *day | 93.607 | word | 39.176 |

Table 7: keyword comparison ppi=1 with ppi in range 101-200

The same keyword analyses were run for posts with ppi in the range 1-5 compared with posts with ppi in the range 101-200. The results in table 8 are similar to those in table 7 in which the less established data group consisted of first posts only. There are 12 suicide method references when the less established group consists of posts in the wider range ppi 1-5, and the same time references are present. The keyword *lurker* has disappeared, suggesting this is related to the first post made only. The keyword *school*, which may be likely to relate to specific significant personal experience, has also disappeared, and the potentially positive keywords *friend* and *thank* are now present, with *thank* the 4th most significant keyword; *thank* is likely to be addressed towards responses from other forum participants. For the more established group the keywords still contain more abstract terms including the metonymic *lifer* and metaphoric *journey*.

| | ppi 1-5 | keyness | ppi 101-200 | keyness |
|---|---|---|---|---|
| 1 | *year | 479.261 | people | 751.969 |
| 2 | !neck | 413.715 | hug | 322.737 |
| 3 | !rope | 375.053 | op | 261.677 |
| 4 | thank | 366.329 | peace | 186.235 |
| 5 | !belt | 275.291 | ss | 172.093 |
| 6 | !helium | 246.734 | lifer | 144.399 |
| 7 | family | 233.461 | member | 139.941 |
| 8 | !pill | 225.925 | thread | 106.042 |
| 9 | *month | 225.431 | other | 101.167 |
| 10 | !suspension | 209.937 | man | 93.926 |
| 11 | !bag | 197.861 | luck | 87.296 |
| 12 | !tank | 196.737 | journey | 86.704 |
| 13 | !method | 193.032 | woman | 80.552 |
| 14 | !knot | 192.961 | patient | 80.343 |
| 15 | thought | 175.044 | right | 78.669 |
| 16 | *time | 166.016 | case | 76.255 |
| 17 | *day | 164.054 | user | 74.136 |
| 18 | !hanging | 163.260 | law | 72.818 |
| 19 | !artery | 152.431 | evidence | 66.023 |
| 20 | friend | 150.376 | euthanasia | 64.064 |

Table 8 keyword comparison ppi in range 1-5 with ppi in range101-200

## Discussion

In this investigation it has been shown that in general increased participation on the sanctioned-suicide forum up to around the 100th post made is associated with a change in language increasingly towards a style that has been associated with better mental health. Specifically, increased

On sanctioned-suicide *journey* is metaphoric, and is used in response to the suicidal ideation of other participants, for example "have a safe journey friend", or "good luck on your journey brother" (Dilkes, 2022). This particular metaphoric use of *journey* is related to the 'catch the bus' metaphor which as discussed previously was used to support suicide discourse in the ash UseNet group started in 1990. The presence of this metaphor as a keyword of more established participants in comparison to those participants making their first post on the forum, then, suggests that increased participation leads to a focus on the suicidal ideation of others, and not the self, and this is reflected in the dominant use of the second-person pronouns you and your by this group (table 6).

Five of the keywords for these more established participants are general person references: *people, man, person, member,* and *individual*, compared to the single more personal reference *family* when ppi=1.

The keyword acronym *ss* refers to the sanctioned-suicide forum, the ingroup, while the metonymic term *lifer* represents the outgroup of those who deny the validity of suicide as a choice, and who seek to shut down suicide discourse, and sanctioned-suicide; *lifer* also has connotations of the dogma of extreme right-wing and religious affiliation, including anti-abortion activism. In conjunction with dominant use on sanctioned-suicide of the *journey*-related *ctb* (catch the bus) metaphor used to support suicide discourse, these results demonstrate that sanctioned-suicide meets the 'structure' criterion for identification as an online community, which references "Jargon, references to group, ingroup/out-group language" (Herring, 2004, p. 361). The *ctb* metaphor is not apparent as a keyword in this analysis because it is widely used by both the participant groups compared here.

| | ppi=1 | keyness | ppi 101-200 | keyness |
|---|---|---|---|---|
| 1 | *year | 404.795 | people | 521.055 |
| 2 | !neck | 275.519 | hug | 188.088 |
| 3 | !pill | 192.663 | peace | 136.484 |
| 4 | !rope | 186.918 | op | 122.505 |
| 5 | *month | 166.629 | lifer | 85.730 |
| 6 | !suspension | 164.168 | luck | 68.979 |
| 7 | !belt | 161.883 | system | 66.952 |
| 8 | !bag | 142.766 | right | 63.221 |
| 9 | !helium | 141.904 | man | 56.244 |
| 10 | lurker | 141.564 | ss | 53.862 |
| 11 | !method | 138.721 | other | 53.520 |
| 12 | school | 133.491 | journey | 52.036 |
| 13 | advance | 128.470 | case | 47.931 |
| 14 | !tank | 119.371 | murder | 45.807 |
| 15 | *time | 109.300 | evidence | 44.490 |
| 16 | !knot | 102.550 | person | 44.077 |
| 17 | thought | 97.755 | member | 43.438 |

participation on the forum is associated with decreased use of first-person singular pronouns, and increased use of the first-person plural pronoun we, the second-person pronouns you and your, and the third-person pronouns they and their, which may represent a move in focus away from the self and towards the other, including membership of the group. There is also a significant increase in the range of pronouns used.

While there is increased use of all of the negative emotion variables over time, there is also increased use of the individual positive emotion variables Joy and Trust, and the composite Positive variable. This suggests that sanctioned-suicide supports increased emotion discussion in general, which has been found to be related to increased emotionality. In addition, entropy was shown to decrease with increased participation on the forum, suggesting that participant posts become more coherent and predictable. It is notable that the most dramatic change for all the linguistic variables investigated occurs in the first few posts made on the forum, and that subsequently the rate of change levels off so that language style on the forum is in general established across participants by around the 100th post made.

The keyword analyses in investigation 2 support the findings in investigation 1 that first-person singular pronouns are more associated with early participation on the forum, and pronouns referring to the other are associated with more established participants. The subsequent focus on noun lemma keywords to investigate the content of forum discourse showed that newer participants are focused on more literal concerns including suicide method, which is the dominant concern, and time references. The more established participants in contrast use more abstract terms including peace, luck, the metonymic term lifer, and metaphoric use of journey, and for this group there are no suicide method or time references. While it has been found that severe and low risk of suicide are relatively easy to identify in social media posts (Zirikly et al., 2019) the method and findings of the current investigation may support location of the intermediate experiences of suicidal ideation that have eluded classification tasks.

These findings that language of participants increasingly becomes of a style that has been associated with better mental health are supported by the related finding that sanctioned-suicide meets the participation, interaction, norms of appropriateness, and structure criteria for identification as a virtual community (Herring, 2004). While the diachronic changes in language style found on the forum might perhaps be expected as someone becomes a member of a supportive community, this is of particular relevance for a suicide forum, in which there is ostensibly a dilemma in being both authentically suicidal and belonging to the community (Horne & Wiggins, 2009).

As discussed previously, the best way non-professionals may help those feeling suicidal is to ask people of concern if they are okay, then accept and listen without judgment to their concerns, including the typically taboo topic of suicidal ideation (O'Connor, 2021). The current study suggests that while its existence as a forum for suicide discussion entails that sanctioned-suicide has been associated with actual suicides, in general the forum may help support better mental health of participants precisely by supporting discussion of those shared concerns of participants that are currently typically taboo in wider society. In general participants on the forum appear to join the community of those who share and understand their concerns, where perhaps their previous lack of community exacerbated suicidal ideation. This supports research recommendations described previously that healthcare professionals must take a more balanced view of self-harm and suicide websites, which have been found to give participants "access to important, socially valued identities, such as being understood, belonging to a community, and coping with their problems" (Baker & Fortune, 2008, p. 118). It is also consistent with a report on mental health and human rights by the United Nations High Commissioner for Human Rights, who advises a comprehensive approach to human rights for those with psychosocial disabilities, including protection of their autonomy, agency, and dignity, as well as requiring policy shifts that recognize exclusion and marginalization as the causes and consequences of poor mental health (United Nations High Commissioner for Human Rights, 2017).

## References

Aladağ, A. E., Muderrisoglu, S., Akbas, N. B., Zahmacioglu, O., & Bingol, H. O. (2018). Detecting Suicidal Ideation on Forums: Proof-of-Concept Study. Journal of Medical Internet Research, 20(6), e9840. https://doi.org/10.2196/jmir.9840

Anthony, L. (2022). AntConc (4.0.1.0) [Computer software]. Waseda University. https://www.laurenceanthony.net/software/antconc/

BAAL. (2021). Recommendations on Good Practice in Applied Linguistics. 4th Edition. https://www.baal.org.uk/who-we-are/resources/BAAL-Good-Practice-Guidelines-2021.pdf

Baker, D., & Fortune, S. (2008). Understanding self-harm and suicide websites: A qualitative interview study of young adult website users. Crisis : The Journal of Crisis Intervention and Suicide Prevention, 29(3), 118–122. https://doi.org/10.1027/0227-5910.29.3.118

Coppersmith, G. A., Harman, C. T., & Dredze, M. H. (2014). Measuring Post Traumatic Stress Disorder in Twitter. 4.

De Choudhury, M., Counts, S., & Horvitz, E. (2013). Predicting postpartum changes in emotion and behavior via social media. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 3267–3276. https://doi.org/10.1145/2470654.2466447

Dilkes, J. (2022). The social and psychological work of metaphor: A corpus linguistic investigation [D_ph, University of Birmingham]. https://etheses.bham.ac.uk/id/eprint/12221/

Egnoto, M. J., & Griffin, D. J. (2016). Analyzing Language in Suicide Notes and Legacy Tokens: Investigating Clues to Harm of Self and Harm to Others in Writing. Crisis, 37(2), 140–147. https://doi.org/10.1027/0227-5910/a000363

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preoţiuc-Pietro, D., Asch, D. A., & Schwartz, H. A. (2018). Facebook language predicts depression in medical records. Proceedings of the National Academy of Sciences, 115(44), 11203–11208. https://doi.org/10.1073/pnas.1802331115

Gilat, I., & Shahar, G. (2007). Emotional First Aid for a Suicide Crisis: Comparison between Telephonic Hotline and Internet. Psychiatry: Interpersonal and Biological Processes, 70(1), 12–18. https://doi.org/10.1521/psyc.2007.70.1.12

Grohol, J. M. (1998). Future clinical directions: Professional development, pathology, and psychotherapy on-line. In Psychology and the Internet: Intrapersonal, interpersonal, and transpersonal implications (pp. 111–140). Academic Press.

Guntuku, S. C., Schneider, R., Pelullo, A., Young, J., Wong, V., Ungar, L., Polsky, D., Volpp, K. G., & Merchant, R. (2019). Studying expressions of loneliness in individuals using twitter: An observational study. BMJ Open, 9(11), e030355. https://doi.org/10.1136/bmjopen-2019-030355

Herring, S. C. (2004). Computer-Mediated Discourse Analysis: An Approach to Researching Online Behavior. In S. Barab, R. Kling, & J. H. Gray (Eds.), Designing for Virtual Communities in the Service of Learning (1st ed., pp. 338–376). Cambridge University Press. https://doi.org/10.1017/CBO9780511805080.016

Honnibal, M., & Montani, I. (2021). SpaCy · Industrial-strength Natural Language Processing in Python. SpaCy. https://spacy.io/

Horne, J., & Wiggins, S. (2009). Doing being 'on the edge': Managing the dilemma of being authentically suicidal in an online forum. Sociology of Health & Illness, 31(2), 170–184. https://doi.org/10.1111/j.1467-9566.2008.01130.x

Lavis, A., & Winter, R. (2020). #Online harms or benefits? An ethnographic analysis of the positives and negatives of peer-support around self-harm on social media. https://research.birmingham.ac.uk/portal/en/publications/online-harms-or-benefits-an-ethnographic-analysis-of-the-positives-and-negatives-of-peersupport-around-selfharm-on-social-media(297e63cf-e357-4a09-b4b3-ac51091db945).html

Marchant, A., Hawton, K., Stewart, A., Montgomery, P., Singaravelu, V., Lloyd, K., Purdy, N., Daine, K., & John, A. (2017). A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people::The good, the bad and the unknown. https://research.birmingham.ac.uk/portal/en/publications/a-systematic-review-of-the-relationship-between-internet-use-selfharm-and-suicidal-behaviour-in-young-people(dc610242-b267-400d-920e-1bd57332ecb0).html

Mohammad, S. (2021). NRC Emotion Lexicon. https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

Mother speaks out against suicide forum after son's death. (2021, May 8). BBC News. https://www.bbc.com/news/uk-northern-ireland-56845565

O'Connor, R. (2021). When it is Darkest: Why people die by suicide and what we can do to help. VERMILION.

O'Dea, B., Larsen, M. E., Batterham, P. J., Calear, A. L., & Christensen, H. (2017). A Linguistic Analysis of Suicide-Related Twitter Posts. Crisis, 38(5), 319–329. https://doi.org/10.1027/0227-5910/a000443

ONS. (2019). Suicides in England and Wales—Office for National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/suicidesintheunitedkingdom/2019registrations

Pennebaker, J. (2011). The Secret Life of Pronouns. Bloomsbury Press.

Pierre, J. (2015). Culturally sanctioned suicide: Euthanasia, seppuku, and terrorist martyrdom. World Journal of Psychiatry, 5, 4–14. https://doi.org/10.5498/wjp.v5.i1.4

Robinson, J., Cox, G., Bailey, E., Hetrick, S., Rodrigues, M., Fisher, S., & Herrman, H. (2016). Social media and suicide prevention: A systematic review. Early Intervention in Psychiatry, 10(2), 103–121. https://doi.org/10.1111/eip.12229

Samaritans. (2020). Managing self-harm and suicide content online.

Schoene, A. M., Turner, A., De Mel, G. R., & Dethlefs, N. (2021). Hierarchical Multiscale Recurrent Neural Networks for Detecting Suicide Notes. IEEE Transactions on Affective Computing, 1–1. https://doi.org/10.1109/TAFFC.2021.3057105

Scrapy. (2021). Scrapy | A Fast and Powerful Scraping and Web Crawling Framework. Scrapy. https://scrapy.org/

Stommel, W. (2009). Entering an online support group on eating disorders: A discourse analysis / Wyke Stommel. Rodopi. https://ebookcentral.proquest.com/lib/bham/detail.action?docID=556652

Suler, J. (2005). The online disinhibition effect. International Journal of Applied Psychoanalytic Studies, 2(2), 184–188. https://doi.org/10.1002/aps.42

Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. Journal of Language and Social Psychology, 29(1), 24–54. https://doi.org/10.1177/0261927X09351676

United Nations High Commissioner for Human Rights. (2017). Mental health and human rights.

Van den Nest, M., Till, B., & Niederkrotenthaler, T. (2019). Comparing Indicators of Suicidality Among Users in Different Types of Nonprofessional Suicide Message Boards: A Linguistic Analysis. Crisis, 40(2), 125–133. https://doi.org/10.1027/0227-5910/a000540

Westerlund, M., Hadlaczky, G., & Wasserman, D. (2015). Case study of posts before and after a suicide on a Swedish internet forum. The British Journal of Psychiatry, 207(6), 476–482. https://doi.org/10.1192/bjp.bp.114.154484

White, M., & Dorman, S. M. (2001). Receiving social support online: Implications for health education. Health Education Research, 16(6), 693–707. https://doi.org/10.1093/her/16.6.693

Wiltsey Stirman, S., & Pennebaker, J. W. (2001). Word Use in the Poetry of Suicidal and Nonsuicidal Poets: Psychosomatic Medicine, 63(4), 517–522. https://doi.org/10.1097/00006842-200107000-00001

Winkel, S., Groen, G., & Petermann, F. (2005). [Social support in suicide forums]. Praxis Der Kinderpsychologie Und Kinderpsychiatrie, 54(9), 714–727.

Without suicide forums, Callie might still be alive. (2020, February 10). BBC News. https://www.bbc.com/news/newsbeat-51262240

Yates, A., Cohan, A., & Goharian, N. (2017). Depression and Self-Harm Risk Assessment in Online Forums. ArXiv:1709.01848 [Cs]. http://arxiv.org/abs/1709.01848

Zirikly, A., Resnik, P., Uzuner, Ö., & Hollingshead, K. (2019). CLPsych 2019 Shared Task: Predicting the Degree of Suicide Risk in Reddit Posts. Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology, 24–33. https://doi.org/10.18653/v1/W19-3003