

# Spontaneous Discourse in Response to Advice on Reddit

Neimhin Robinson Gunning  
Trinity College Dublin  
16321701  
nrobinso@tcd.ie

Someone Else  
Another Affiliation  
another@email.org

## Abstract

....

**Keywords:** advice-seeking, spontaneous online discourse, sentiment, affect.

## 1 Introduction

....

## 2 Research Review

Alfandre (2009)

Brookes and Harvey (2016)

Smailhodzic et al. (2016)

Neviarouskaya et al. (2010) developed a rule-based algorithm for recognizing attitude in terms of affect (negative and positive), judgement (negative and positive), and appreciation. The algorithm systematically composes the classifications of constituent sentence parts, such that entire sentences can be classified. Their theory of attitude and the corresponding algorithm are applicable at three levels. At the top level are just three classifications, positive, negative and neutral. At the mid level, positive and negative classifications are further subdivided into, affect, judgement, and appreciation. At the most fine-grained level affect is divided into interest, joy, and surprise (positive affect) and anger, disgust, fear, guilt, sadness, shame (negative affect). The algorithm accounts for negation (by several means), amplification of sentiment, neutralisation of sentiment etc. A corpus of human-annotated sentences is collected and compared to the results of the automated system. The automated system was found to have greater agreement with the human annotations than a baseline system that simply selected the most intense token annotation. Neviarouskaya et al. (2010) discuss some of the failures and failure modes of their system. The accuracy of the automated systems (at predicting human annotations) is highest at the top level of the classification hierarchy (positive, negative, neutral), and the accuracy decreases as we move to finer grained classifications.

Proferes et al. (2021)

Atkins and Harvey (2010)

Miller (2020)

### 3 Research Questions

Miller (2020, 56) suggested investigating the motivations behind engagement with confessional public social media postings, in particular whether it is empathetic concern or sensationalism/“shock value” that drives engagement. We seek to answer similar questions but in a slightly different forum, specifically in public spontaneous advice-seeking discussions with a confessional bent on social media. Also, we take an interest in different types of engagement, empathetic/compassionate/productive engagement, in contrast with sensationalist/antagonistic/trolling engagement.

We narrow our focus to discussions initiated by advice-seeking posts with an authentic, vulnerable tone, avoiding, as far as possible, the inclusion of deceptive, purely comedic, or mocking submissions (although *responses* with comedic/deceptive/mock intention are still of interest to us). In other words, we want to answer questions about cases where posters are facing a genuine moral dilemma and soliciting advice in good faith.

Especially interesting to us are the cases where advice-seekers are given ‘bad news’ (in particular, a negative moral judgement of the advice-seeker’s character or behaviour), and yet show appreciation towards the person who delivers the ‘bad news’. What are the textual markers and schematics of such well-received, negative advice-giving comments?

On the other side of the spectrum, we take an interest in cases where the public discussion devolves into antagonistic and hateful rhetoric. What are the textual markers and schematics advice-seeking posts that coincide with threads developing poorly? In other words, can we extract specific words, phrases, patterns of discourse that one should avoid when genuinely seeking compassionate and nuanced moral advice on public social media fora? Therefore, to answer this question, we are interested in cases in which the original advice-seeking post exhibits an authentic solicitation relating to a genuine moral dilemma.

Finally, can there be a therapeutic benefit to the advice-seeker by submitting their moral dilemma to public scrutiny?

## **4 Methods**

## **5 Results**

## 6 Neimhin's summaries

### 6.1 Recognition of Affect, Judgement, and Appreciation in Text

Neviarouskaya et al. (2010): The authors develop a rule-based algorithm for recognizing attitude in terms of affect (negative and positive), judgement (negative and positive), and appreciation. The algorithm systematically composes the classifications of constituent sentence parts, such that entire sentences can be classified. Their theory and algorithm applies at three levels. At the top level are just three classifications, positive, negative and neutral. At the mid level, positive and negative classifications are further subdivided into, affect, judgement, and appreciation. At the most fine-grained level affect is divided into interest, joy, and surprise (positive affect) and anger, disgust, fear, guilt, sadness, shame (negative affect). The algorithm accounts for negation by several means, amplification of sentiment, neutralisation of sentiment etc. A corpus of human-annotated sentences is collected and compared to the results of the automated system. The automated system was found to have greater agreement with the human annotations than a baseline system that simply selected the most intense token annotation. The authors discuss some of the failures and failure modes of their system. The accuracy of the automated systems (at predicting human annotations) is highest at the top level of the classification hierarchy (positive, negative, neutral), and the accuracy decreases as we move to finer grained classifications.

- If the algorithm proposed in the paper is available as code it would be useful for our textual analyses, but it would too complex to implement from scratch given the time available.

- If an implementation exists it may be possible to make adjustments leveraging other techniques, for instance using word embeddings from a large language model like BERT to make atomic phrasal classifications, and then continuing the authors' algorithm.

- It may be generally true that finer grained analyses/classifications are more difficult than coarser analyses, so the hierarchical approach is useful for validating the methodology at different levels.

- If the corpus described in the paper is available it could be useful to our project.

### 6.2 miller-2020

Miller (2020): Miller (2020) reports a survey of 628 Reddit users. The questionnaire covered Internet usage, Reddit usage (especially in relation to self-disclosure on subreddits), and a set of psychological questions relating to connectedness, social support perception, life satisfaction, narcissism, and sensation seeking.

The sharing of intimate information on Reddit's anonymous-feeling platform was found to be significantly correlated with the sensation seeking tendency, but there was no evidence of an association with narcissism.

The findings are interpreted as evidence that intimate self-disclosure on Reddit is motivated by feeling disconnected and alone.

The research participants were approached with advertisements on Reddit, as well as with posts to related subreddits, e.g. 'r/confessions', 'r/offmychest'.

Miller (2020) suggests that "[f]uture work should investigate whether engagement is driven by empathetic concern, or whether it is spurred by boredom and the tendency of confessions to have a high degree of shock value", which is an interesting line of investigation in the context of spontaneous online advice-seeking. What drives engagement with advice-seeking posts, i.e. what drives Redditors to give advice?

### 6.3 pennebaker-2015

Pennebaker et al. (2015): The LIWC2015 program processes texts and produces linguistic and textual measures, as well as proxy measures of various psychological and developmental attributes.

The proxy measures are essentially relative word counts. For instance the 'negative emotion' proxy is calculated by counting occurrences of words associated with negative emotion.

The 'negative emotion' proxy is subdivided into anxiety, anger, and sadness. Words which contribute to any of these also contribute to the 'negative emotion' score of the text. Interestingly 'positive emotion' is not subdivided, meaning there is a richer analysis of negative emotion. I also saw in the paper "Recognition of Affect, Judgement, and Appreciation in Text" that negative affect was given a richer subcategorization than positive affect.

The authors describe the process by which the LIWC2015 dictionaries and word-categorisations were developed. It involved a significant manual process with collaboration of several judges to create an initial dictionary. The dictionary was then analysed and amended to fix omissions or internal inconsistencies. The set of categories was chosen based on attributes commonly studied in social, health, and personality psychology. The authors give an explanation as to why the internal consistency of linguistic-based psychometrics can be lower or have wider margins than a questionnaire (it comes down to the fact that people rarely repeat themselves in spontaneous natural language, whereas questionnaires make heavy use of repetition, or at least rhyming). Perhaps of interest to our group's analysis is the claim that Spearman-Brown prediction formula is generally a more accurate approximation of a word-category's internal consistency than is Cronbach's alpha.

## 7 Ben's Summaries

### 7.1 Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics

Proferes et al. (2021): The authors present a meta-analysis of 727 manuscripts that used Reddit as a data source (2010-2020). They discuss the manner in which data is retrieved from Reddit, the characteristics of these datasets, the communities being studied, the analytical methods used to process this data, and finally a general ethical discussion of the use of Reddit data in research. This paper presents a systematic review of *how* Reddit is being used in research, the provenance / accuracy of that data, and what consequences arise from using that data.

- Some Reddit communities explicitly state that approval from moderators is necessary before using their subreddit's data. There is also a need to take into account the *size* of a subreddit before performing analysis on their data; smaller communities may host a more intimate and private tone, creating a space for vulnerable conversations, and may not be as receptive to research as larger communities.
- One of the largest indirect known ways of accessing Reddit data is using a platform named Pushshift, ingesting data from Reddit's API for archival purposes. However, the paper states PushShift is not an exact replication of data from Reddit (another paper is cited here; further research needed).
- The use of Reddit data in research was steadily increasing at the time the paper was written. Within their corpus, 2017 saw 102 publications, 2018 saw 146 publications, and 2019 saw 230 publications.
- In terms of discipline usage of Reddit data, the author grouped disciplines into different buckets. Computer Science, Mathematics, and Engineering have the highest count of publications using Reddit data, followed by Medicine & Health, Social Science, and then Humanities.
- 90.7% of publications did not explicitly provide Reddit usernames. 69.1% of papers did not include direct quotes from Reddit.
- The data gathered from Reddit can be inherently biased due to social scoring features; "top posts" tend to receive more replies and receive more attention. Additionally, more broadly agreeable, clever, witty, or inflammatory comments are more likely to be replied to.
- The authors argue that researchers should carefully consider the risks presented to data subjects by including direct quotations or usernames.

### 7.2 The Stories We Tell Influence the Support We Receive: Examining the Reception of Support-Seeking Messages on Reddit

Adelina et al. (2023): This research by Adelina et al. focuses on the interaction between support received on Reddit from advice-seekers and the textual content of their post. The methodology focused on the usage of Latent Profile Analysis. This technique focuses on 'bucketing' similar textual content into various profiles, allowing for the construction of different 'personas' that can be analysed separately. An appropriate number of profiles is essential to conducting this analysis; too many profiles may result in non-convergence, whereas too few may sacrifice accuracy.

Applicable to our use case, their textual preprocessing approach included cleaning and normalising Reddit data. Steps included removing posts under a specific word count (100 words), and retaining only first-level comments, removing nested replies. The authors conducted a classification analysis using several machine learning algorithms rooted on different systematic principles to obtain a holistic overview

of the accuracy of their analysis. A Naïve Bayes approach and a C4.5 decision tree approach offered the higher accuracy for this task, which may be suitable to classification tasks during the implementation of this research project.

Interestingly, the quantity of social support received did not vary between profiles, but the *quality* of the responses did. With sufficient coherence, posters who expressed negativity and low agency in their original posts tended to get more informational (advice, referral) and instrumental support (practical suggestions), whereas posts that were highly incoherent tended to receive less prescriptive support related to other profiles.

## References

- Adelina, N., C. S. Chan, K. Takano, P. H. M. Yu, P. H. T. Wong, and T. J. Barry (2023). The stories we tell influence the support we receive: examining the reception of support-seeking messages on reddit. *Cyberpsychology, Behavior, and Social Networking* 26(11), 823–834.
- Alfandre, D. J. (2009). “I’m going home”: Discharges against medical advice. *Mayo Clinic Proceedings* 84(3), 255–260.
- Atkins, S. and K. Harvey (2010). How to use corpus linguistics in the study of health communication. In *The Routledge handbook of corpus linguistics*, pp. 605–619. Routledge.
- Brookes, G. and K. Harvey (2016). *Examining the Discourse of Mental Illness in a Corpus of Online Advice-Seeking Messages*, pp. 209–234. London: Palgrave Macmillan UK.
- Miller, B. (2020). Investigating reddit self-disclosure and confessions in relation to connectedness, social support, and life satisfaction. *The Journal of Social Media in Society* 9(1), 39–62.
- Neviarouskaya, A., H. Prendinger, and M. Ishizuka (2010). Recognition of affect, judgment, and appreciation in text. In *Proceedings of the 23rd international conference on computational linguistics (Coling 2010)*, pp. 806–814.
- Pennebaker, J. W., R. L. Boyd, K. Jordan, and K. Blackburn (2015). The development and psychometric properties of liwc2015. Technical report.
- Proferes, N., N. Jones, S. Gilbert, C. Fiesler, and M. Zimmer (2021). Studying reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media + Society* 7(2), 20563051211019004.
- Smailhodzic, E., W. Hooijsma, A. Boonstra, and D. J. Langley (2016, Aug). Social media use in health-care: A systematic review of effects on patients and on their relationship with healthcare professionals. *BMC Health Services Research* 16(1), 442.