Sadness and Anxiety Language in Reddit Messages Before and After Quitting a Job

Molly E. Ireland **Micah Iserman** **Kiki Adams**
Receptiviti
{mireland, miserman, kadams} @receptiviti.com

Abstract

People globally quit their jobs at high rates during the COVID-19 pandemic, yet there is scant research about emotional trajectories surrounding voluntary resignations before or during that era. To explore long-term emotional language patterns before and after quitting a job, we amassed a Reddit sample of people who indicated resigning on a specific day (n = 7,436), each of whom was paired with a comparison user matched on posting history. After excluding people on the basis of low posting frequency and word count, we analyzed 150.3 million words (53.1% from 5,134 target users who indicated quitting) using SALLEE, a dictionary-based syntax-aware tool, and Linguistic Inquiry and Word Count (LIWC) dictionaries. Based on posts in the year before and after quitting, people who had quit their jobs used more sadness and anxiety language than matched comparison users. Lower rates of "I" pronouns and cognitive processing language were associated with less sadness and anxiety surrounding quitting. Emotional trajectories during and before the pandemic were parallel, though pandemic messages were more negative. The results have relevance for strategic self-distancing as a means of regulating negative emotions around major life changes.

1 Introduction

Leaving a job is a major crossroads in adult life. Though in hindsight most people view their choice positively (e.g., with respect to improved career opportunities, pay, and job satisfaction; Parker and Horowitz, 2022), voluntarily resigning is nevertheless a stressful and consequential life transition. Like other major life changes, the social turmoil and financial uncertainty that come with quitting a job are universal, but there are better and worse ways to cope with the emotional fallout.

In the last few years, the psychology of quitting has been complicated by what some have called the Great Resignation. In the United States, voluntary resignation rates have been increasing linearly since the mid-2000s, peaking in 2021, the second year of the pandemic (Gittleman, 2022); indeed, the main difference between pre-pandemic and pandemic-era quitting may be increased likelihood of leaving the labor market entirely (Ferguson, 2022). Whether pandemic-era quitting is the culmination of a decades-long trend or a phenomenon triggered by the pandemic is still being debated (Fuller and Kerr, 2022). What is clearer is that people who chose to leave their jobs in the last few years did so against the backdrop of ongoing social, emotional, and economic upheavals related to the COVID-19 pandemic. A large part of the Great Resignation narrative centers on debates about work culture and work-life balance that are occurring on social media platforms like Reddit (Medlar et al., 2022).

Despite plenty of media dialogue on the subject, there is scarce psychological research on the Great Resignation specifically or the emotions involved in quitting more broadly. Research on quitting or voluntary turnover has historically focused more on antecedents than a broader window considering individuals' psychology before and following resignation (Rubenstein et al., 2018). Earlier research found that quitting results from low or declining job satisfaction (Chen et al., 2011), burnout related to overwork or high-stress responsibilities at work, a lack of work-life balance (especially difficulty disengaging from work when away from the workplace; Sonnentag and Bayer, 2005), and more attractive job opportunities (Rainayee, 2013). Pandemic-era quitting seems to follow a similar course, though burnout related to employers who are perceived as exploitative and COVID-related burnout have become more salient as the pandemic has increased work-related stress for many employees and employers (Jiskrova, 2022).

The following paper introduces the Reddit Job

Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, pages 467–478
July 14, 2023

Change Corpus, a sample that includes all Reddit submissions and comments from individuals who discussed leaving or being fired from a job on a specific day in addition to the same complete history from matched comparison users (i.e., people with similar posting histories as the target users who had never discussed leaving a job). The present paper focuses on a subsample of users who indicated quitting (voluntarily leaving) their jobs. As an initial illustration of how the corpus can be used to answer questions about emotional language over time, we explored how rates of sadness and anxiety words vary as a function of two baseline individual differences relevant to emotion regulation: self-focus and cognitive processing. Finally, we compared sad and anxious language between pandemic-era and earlier quitting.

2 Background

Pronouns are intimately linked with affective language and emotion regulation. In the context of negative affect, first-person singular pronouns (e.g., *I*, *me*, *my*; sometimes called "I"-words) tend to be more closely linked with avoidant emotions such as fear and sadness (Tackman et al., 2019) than approach emotions such as anger (Simmons et al., 2005). In the view that emotions prepare people for pragmatic actions, avoidant emotions such as disgust, fear, and sadness compel people to withdraw from noxious or harmful stimuli, whereas approach emotions such as desire, joy, and anger impel people to engage with the emotional stimulus (Carver and Harmon-Jones, 2009; Corr, 2013). People who use more first-person singular pronouns in everyday conversations and writing tend to be more vulnerable to stress and more prone to affect regulation disorders, such as depression (Baddeley et al., 2013; Tackman et al., 2019). Shifting from a first-person to more distant perspective seems to be a healthy coping strategy when experiencing negative emotions or recounting distressing events. For example, in the expressive writing paradigm, people benefit more from writing about traumatic events when they shift perspectives rather than remaining fixed in a first-person mindset (Holtzman et al., 2017; Pennebaker and Chung, 2007). Self-distancing research similarly shows that people naturally use less "I" when writing or talking about distressing memories, a strategy that reduces negative emotions in the moment in both naturalistic and experimental studies (Park et al., 2016).

Whether self-focus is a product or cause of distress or mental health conditions such as depression has been rigorously debated in psychology. Early theories on depressive realism and aversive self-focus argued that viewing the world and especially the self realistically was distressing; in that view, anyone, regardless of trait negative affectivity or neuroticism, would be disturbed by heightened self-awareness (Wicklund, 1975). Later research qualified those findings, showing that the aversiveness of self-awareness and the tendency towards self-focus after failure but not success were specific to people predisposed to negative affect regulation conditions such as depression (Greenberg and Pyszczynski, 1986; Pyszczynski et al., 1987). In summary, self-focus appears to be harmful when distressed but not otherwise (Pyszczynski et al., 1987), is correlated with trait negative affectivity (Schwartz-Mette and Rose, 2016; Tackman et al., 2019), and can be strategically decreased to help downregulate negative emotions (Kross and Ayduk, 2011).

Cognitive processing language (e.g., *think*, *know*), like self-focus, is not altogether harmful or helpful but can become risky in the context of stress and negative affect. Talking through thought processes can help make sense of emotionally complex issues (e.g., in expressive writing; Kacewicz et al., 2007), and cognitive reappraisal can be a valuable tool for regulating emotions (Riepenhausen et al., 2022). However, chronically high cognitive processing language in conversations (such as in letters or social media messages)—especially in conjunction with negative emotional language—may reflect the kind of rumination (i.e., repetitive, intrusive, inward-focused negative thoughts; Watkins and Roberts, 2020) that characterizes affect regulation disorders such as depression (Dean and Boyd, 2020; Eichstaedt et al., 2018).

Reddit is an increasingly popular resource for social-behavioral scientists interested in analyzing publicly accessible language use surrounding major life events such as romantic breakups (Seraj et al., 2021), community crises such as the COVID-19 pandemic (Ashokkumar and Pennebaker, 2021), and mental health conditions including depression, anxiety, and suicidality (e.g., Matero et al., 2022; Shing et al., 2018). Though Reddit's active user base continues to skew young, American, and male, it is more diverse in terms of ethnic backgrounds, nationalities, and age than typical convenience samples in psychology, such as undergraduate students (Henrich et al., 2010; Sattelberg, 2021). More importantly for collecting conversations about risky and distressing topics, such as ending a relationship or quitting a job, Reddit usernames are typically anonymous, which enables people to discuss negative experiences frankly with few concerns about social or legal risks. Social media analyses also facilitate real-time tracking of changes in social movements. For example, topic analyses of Reddit messages showed that conversations about leaving one's job became more focused on mental health and negative experiences at work after the start of the pandemic (del Rio-Chanona et al., 2022).

Like many social media platforms, Reddit's popularity increased globally during the COVID-19 pandemic. Reddit has also become a hub for social movements related to what's become known as the Great Resignation (e.g., r/antiwork; Medlar et al., 2022). Though, to many, mass resignations appeared to be a zeitgeist triggered by the socioeconomic conditions of COVID, employment data suggest that COVID-era resignations are not unique but are a continuation of linearly increasing voluntary turnover rates dating back to at least a decade before COVID (Gittleman, 2022). Thus, in addition to the primary aim of exploring negative emotional language before and after quitting, a secondary goal of this project was to examine whether negative emotion trajectories differed as a function of quitting era (pre-pandemic or during the COVID-19 pandemic).

3 Method

The sections below first summarize the methods used to assemble the Reddit Job Change Corpus and then discuss the narrower subsample that we focused on in the present analyses. Last, language measures and analytic strategies are described.

3.1 Dataset

In an early phase of this project, we used pushshift.io (which had a searchable archive of Reddit data that is no longer available) to identify a set of users who may have quit their job in two steps: First, we searched for submissions or comments with the query *quit job*, then refined those results by searching for variants of the phrase *I just quit my job*. Second, we collected all submissions and comments from authors in the refined results from between January 2015 and July 2022. This resulted in a sample of 11,391 unique users.

In the current phase of this project (between October 2022 and April 2023), we used the original sample to collect new data, including (a) the full submission and commenting history from 8,797 users with active accounts from the original sample, and (b) any users who were recently active (author of or commenter on up to 102 "hot" submissions at the time of collection) in any of the 1,200 most common subreddits within the original sample. This resulted in a sample of 1,389,763 unique users, which constituted a pool of possible target users (those who may have quit their job) and potential comparison users.

This research analyzes only publicly observable behavior and thus qualifies as exempt under the revised Common Rule in the United States' Federal Policy for the Protection of Human Subjects (Department of Health and Human Services, 2017). In compliance with the Reddit API terms of use, all data analyzed in this research are publicly available and will not be used for commercial purposes.

Target Sample From the new pool of users, we searched for people who may have recently quit their job in two rounds: In the first, we loosely searched for messages (submission title plus body, or comment body) that included (a) a job-related word, such as *job* or *boss*, and (b) a word relating to either quitting or being fired (e.g., *quit*, *resigned*, *fired*, *furloughed*), or (c) a phrase such as *lost my job* or *let me go*. To match, phrases could stand on their own, but quit terms had to be preceded within a sentence by *i*, and fired terms had to be preceded by *i* and *got*, *was*, *was given*, or *have been*. This resulted in a set of 485,005 messages from 271,839 users. In the second round of searching, we lightly cleaned matched messages to remove curly quotes and HTML, then processed them with a dependency tagger (Wijffels, 2023). Once parsed, we used a simple set of dependency-based rules to refine target messages: Each message had to have a self reference (exclusively *i*, *me*, or *my*), job reference (exclusively *job*, *career*, or *position*), and target reference (associated with quitting or being fired). If a message contained all required references, a series of dependency-chain checks were applied to attempt to ensure that the author was talking about their own job, and the target reference applied to that job. If a message passed all dependency checks, it was considered a target message but was additionally checked for hypothetical references (such as *if*, *should*, or *might*) or quotations, which would mark the target message as hypothetical or quoted. See the code for specific checks and criteria: osf.io/p2rt7.

After a refined set of target messages was defined, time references were searched for in the extracted target phrase. If the target phrase included a day reference (such as *yesterday* or *on Monday*), the target sentence included a reference to *minutes*, *hours*, or *days* followed by *ago*, or the target message had no specific time reference but included *just* or *recently*, the message was considered recent. To develop and manually spot-check these criteria, we extracted sentences from target messages along with target phrases and time references: osf.io/xahrc.

The final set of target users were those with messages assigned a target type (quit or fired), not marked as hypothetical or quoted, and marked as recent, and that were not posted in subreddits with names containing the words *meme*, *joke*, *funny*, or *humor*. These criteria resulted in a set of 7,436 users, of whom, 5,357 had only quit messages, 2,016 had only fired messages, and 63 had both.

Comparison Sample To construct a comparison sample, we first removed any users who (a) appeared in the first round of target message identification [i.e., users with any message containing terms loosely relating to a job, and quitting or firing], or (b) made any submissions or comments in subreddits appearing in the second round of target message identification more than once, that also contained the words *work*, *job*, or *career* (such as *r/antiwork* and *r/byebyejob*; which included 71 subreddits). This left 830,960 users to make up the possible comparison pool. To find comparison users, each target user was compared with each user in the comparison pool. The similarity between each user was calculated from inverse Canberra distance between three sets of features: (1) Counts of messages per subreddit [$count_s$; submissions or comments] in which the target user had any messages [$subreddits_t$, where t is the target user; Equation 1], (2) counts of comments and submissions separately across all subreddits [replacing $subreddits_t$ with $\{comments, submissions\}$ for each user in Equation 1], and (3) counts of characters within comments and submissions separately across all subreddits [replacing $subreddits_t$ with $\{nchar(comments), nchar(submissions)\}$ for each user in Equation 1].

$$sim_c = \frac{\sum_{s \in subreddits_t} 1 - \frac{|count_{sc} - count_{st}|}{count_{sc} + count_{st}}}{len(subreddits_t)} \quad (1)$$

These were combined into a weighted average, with subreddit similarity getting 50% weight, message count similarity 30% weight, and message length similarity 20% weight (which helped adjust for differences in similarity distributions between each feature set). These were further weighted by difference in time of first activity (where only users within 1% quantile of the target user were considered) and availability (so each comparison user was only assigned to one target user). After weighting, the comparison user with the highest similarity score was assigned to the given target user.

The comparison users had no target message, so we defined target messages using the paired target user's target message(s): We calculated the position of the target user's target message(s) within their complete message history, then assigned the same target type to the nearest message(s) in similar positions within the comparison user's history. In this way, each comparison user has the same number and type of target message(s).

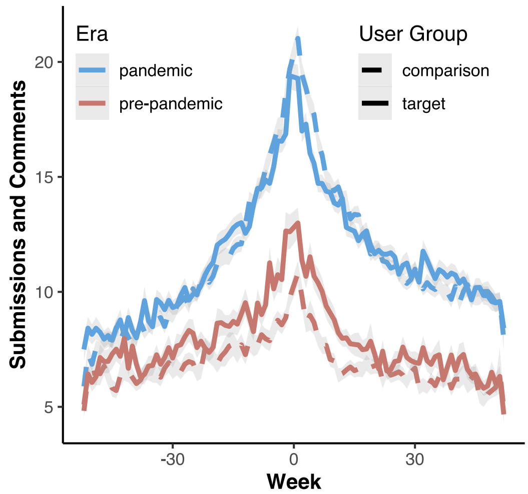

Figure 1: Average message count per week between user groups and eras.

Figure 1 shows the average number of messages (submissions or comments) across users within each week, between the target and comparison samples and between eras. The method of selecting comparison users and assigning them target messages ensured that these distributions would look similar between samples. One remaining difference is the slightly lower number of pre-pandemic messages from comparison users, which is due to the way users were initially sampled: The target sample is more informed by the initial seed sample of users, which includes many users who are no longer active in the most common subreddits within that sample, whereas the comparison sample is primarily made up of users who were recently active at the time of collection.

The fact that there are fewer pre-pandemic messages overall may make comparisons between eras more challenging, as anything that varies by number of messages will also appear to vary by era. A broader challenge comes from the general distribution of messages: All users in this sample must have a 0th week, but they can freely vary in how many surrounding weeks they have. This results in the tent-like distribution of messages around 0, which may make it difficult to identify trends in messages over time, as anything that varies by number of messages will appear to have a strong time association as well.

The scripts used to collect and prepare these samples, along with raw and scored versions of the resulting datasets are available on the Open Science Framework: osf.io/gxbts.

3.2 SALLEE and LIWC

We chose to use SALLEE (Syntax-Aware LexicaL Emotion Engine; Adams, 2022) for measuring emotions. Beyond measuring overall sentiment (i.e., positive or negative emotional tone), SALLEE provides measures of granular emotions that underlie sentiment, such as *fear*, *excitement*, and *gratitude*. SALLEE's lists of emotion words are derived in part from LIWC's affect categories but are applied in a syntax-aware architecture which helps it to perform well on short texts, such as many found on Reddit, and behave reliably in noisy and diverse linguistic contexts. The syntax-aware architecture includes provisions for structures such as intensifiers (e.g., *very*, *so much*), softeners (e.g., *kinda*, *a bit*), negations (e.g., *not*, *never*), punctuation, capitalization, idioms, and words that express emotion and sentiment flexibly depending on context (e.g., swear words). For example, the phrases *pretty darned happy*, *not really happy*, and *NOT happy!* would be scored as moderately positive, somewhat negative, and very negative, respectively. SALLEE outputs a weighted percentage for fourteen specific emotions, three sentiment valences (*goodfeel*, *badfeel*, *ambifeel*), a combined sentiment score reflecting *goodfeel* minus *badfeel*, and combined *emotionality* and *non-emotion* scores.

SALLEE's design makes it particularly valuable for use outside the field of computer science. While more complex models can also offer syntax awareness and a rich array of emotions as output features, they typically do not offer transparency or explainability, do not perform well on casual language, or do not perform well on short texts. Like many other researchers in fields such as sociology, psychology, sociolinguistics, and communication, we found the ability to dissect and explain the way that our data was scored to be vital.

We additionally used the Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2015) anxiety category to measure anxiety; although SALLEE has a fear category that captures worry, anxiety, and fright broadly, it does not have an anxiety-specific category. LIWC is SALLEE's conceptual progenitor and the most commonly used dictionary-based emotional language measure in use in the social-behavioral sciences today (see Eichstaedt et al., 2021). LIWC has fewer words in each of its emotion dictionaries and does not use syntactic context to qualify the weight of individual words, yet it performs well across many social contexts and modalities (Boyd and Schwartz, 2021; Vine et al., 2020). For both tools, we focused on negative emotion words (LIWC anxiety and SALLEE sadness) as the outcomes and first-person singular pronouns ("I"-words) and cognitive processing language, both from LIWC-2015, as moderators.

3.3 Analytic Strategies

Our aims were to explore (1) how individual differences in self-focus (based on first-person singular pronouns, e.g., *I*, *me*, *my*) and cognitive language (based on cognitive words, e.g., *idea*, *think*, *realize*) relate to anxiety and sadness language in social media messages before and after quitting, and (2) how the long-term emotional trajectories associated with quitting a job compare between pre-pandemic and pandemic-era resignations. We focused specifically on sadness and anxiety words as relatively common avoidant emotions that may present barriers to actively coping with major life stressors.

To address these questions, we first concatenated messages by week and analyzed weeks with SALLEE and LIWC (rather than scoring and averaging across individual messages, most of which were too brief for traditional dictionary-based text analysis). Weeks contained about 458 words total on average (*median* = 223) and were highly variable across users (*SD* = 790).

We measured emotional trajectories (linear and quadratic) surrounding quitting as a function of linguistic moderators ("I" pronouns and cognitive processing language), quitting era (resigning before or during the COVID-19 pandemic), and user set (target users who wrote about quitting or comparison users) using linear mixed effects models in R (Bates et al., 2015; R Core Team, 2023). All models included random intercepts for authors nested within dyads (target users and matched pairs). For test statistics, we report F with Kenward-Roger approximated degrees of freedom (Halekoh and Højsgaard, 2014).

Though the full corpus covers individuals' entire submission and comment histories, in some cases for several years, we focused on posts within 52 weeks before or after users' quit messages. We excluded the week centered on a quit message (the quit week) from the dataset used for visualization and statistical tests in order to focus on how people communicate in general—outside of submissions or comments specifically about resigning—and avoid artifacts related to atypical quit weeks at the center of the distribution. Weeks with outliers (>3.5 standard deviations from the mean) for language variables (anxiety, sadness, "I" pronouns, and cognitive processing language) were removed (2.7% of rows from the original sample); we also removed users who were unmatched (lacking a comparison or target users; n = 277) after the word count and posting frequency exclusions.

4 Results

We first regressed emotional language on user set (target versus comparison users) and week in a main effects model. Individuals who discussed quitting their jobs on Reddit used more sadness [$F(1, 3966.2) = 32.93$] and anxiety language [$F(1, 3935.2) = 261.1$] than did matched comparisons across the 2-year time span, and quadratic effects were the best fit for both anxiety [$F(2, 289357) = 32.22$] and sadness [$F(2, 290720.1) = 51.7$], all $p < .0001$.

4.1 First-Person Singular Pronouns

In main effects-only models including user set and week as covariates, baseline "I" rates correlated

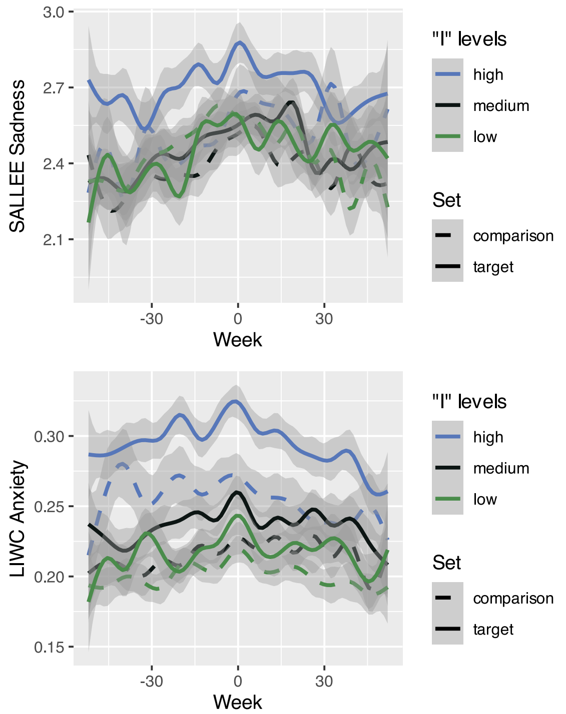

Figure 2: Sadness and anxiety language (% of total words; unweighted for LIWC, weighted for SALLEE) by baseline "I" pronouns. Error bands show 95% confidence intervals. Quit weeks are omitted.

with higher rates of sadness language, $F(1, 8410.4) = 55.50$, and anxiety words, $F(1, 8369.3) = 501.1$.

For the moderator models, we were most interested in interactions with user set, indicating whether moderation by "I" differed between people who discussed quitting on Reddit and comparison users. The two-way interaction between "I"-words and user set, controlling for week as a covariate, was significant for both sadness and anxiety language, both $F > 31$, both $p < .0001$. Figure 2 suggests that the largest differences between target and comparison users occurred before quitting for sadness and after quitting for anxiety.

Simple effect models showed that target users who used high rates of first-person singular pronouns at baseline used more sadness language than comparison users, $F(1, 1905.1) = 28.46$; user set effects were nonsignificant for moderate and low baseline "I"-word usage, both $F < 1$. Users who reported quitting their jobs used more overall anxiety language, relative to comparison users, at all levels of baseline "I"-words, though effects were strongest for people using the most baseline "I" [High $F(1, 1697.90) = 55.53$, Medium $F(1, 1782.12) = 32.48$, Low $F(1, 1768.8) = 18.02$].

The differences between comparison and target users' anxiety and sadness language appeared to

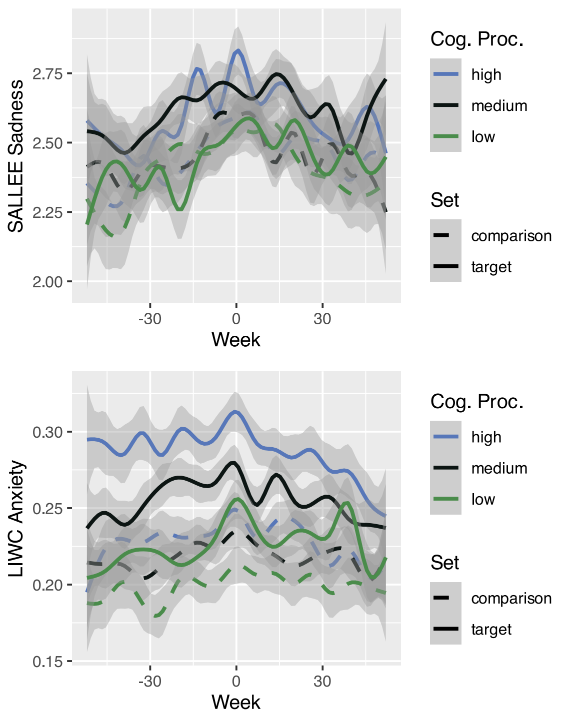

Figure 3: Sadness and anxiety language by baseline cognitive processing language. Error bands show 95% confidence intervals. Quit weeks are omitted.

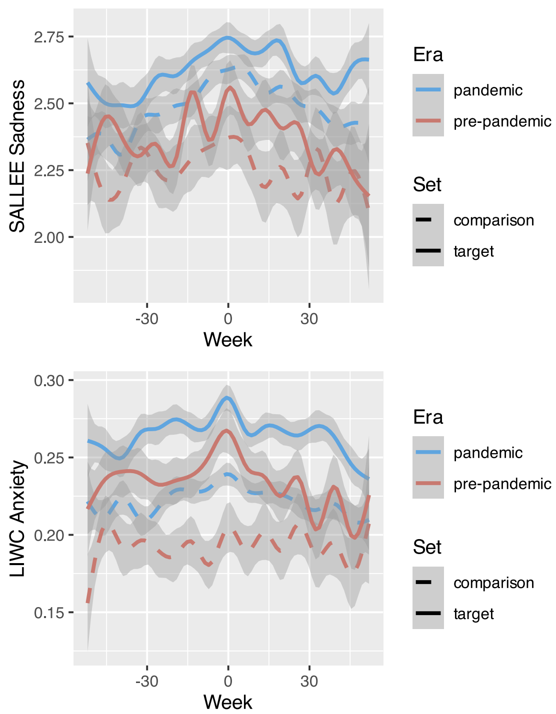

Figure 4: Sadness and anxiety language as a function of era. Error bands show 95% confidence intervals.

be independent of whether they were also talking about work, based on models controlling for work-related language (LIWC's work category, e.g., *boss*, *salary*) as a covariate. Indeed, reading posts from people in the target sample with high baseline "I" and high rates of sadness in the few months before quitting shows that their messages focused primarily on personal grief separate from work (e.g., mourning romantic partners and pets, sadness over family members' serious illnesses). However, anxiety language for high "I" users was more clearly a mix of personal distress and work-related worries, especially after quitting. For example, some messages in the months after quitting reflect the daily life stressors associated with starting a new job (e.g., "My biggest fear is money ... I just started this job so I have no PTO to fall back on.").

4.2 Cognitive Processing

Main effects models including user set and week as covariates showed that baseline cognitive language correlated with higher rates of anxiety [$F(1, 8742.9) = 214.5$] and sadness language [$F(1, 8752.3) = 22.64$] overall, both $p < .0001$.

In the full model regressing anxiety and sadness on the interaction of baseline cognitive processing language, user set, and quadratic effects of week, the strongest effects for both outcome variables were the cognitive processing-by-set interactions. For both, effects were driven by user set effects (i.e., people who quit their jobs using more sadness and anxiety language than comparison users) being greater for those using more baseline cognitive processing language. For anxiety, the effect was significant for all levels [High $F(1, 1735.8) = 146.91$, Moderate $F(1, 1878.1) = 72.79$, low $F(1, 1837.2) = 39.54$, all $p < .0001$]. For sadness, effects were significant only for high and medium cognitive language [High $F(1, 1782.7) = 20.76$, $F(1,1868.8) = 26.199$, Low $F(1,1760.9) = 0.06$]. All models controlled for week as a covariate.

Conclusions were identical when controlling for the degree to which people talked about their job on Reddit by including work-related language as a covariate. Messages with high rates of anxiety written by high baseline cognitive language users tended to focus on general anxiety more than work-related worries. For example, people expressed anxieties about belongings and hobbies ("I recently put a ton of thought into getting either the moto x pure edition ... I am currently using the Galaxy S6 edge, and I swear, I have a panic attack at least once per day worrying about dropping and breaking it") and their own mental health ("So I hope this makes sense. When I am really anxious, [I need] time alone away from people...").

4.3 Quitting Era

Trajectories for sadness and anxiety were roughly parallel between comparison and target groups (Figure 4). There were main effects for pandemic era for both outcomes, indicating that people used more sadness and anxiety language during the pandemic than before regardless of whether they did or did not discuss quitting their jobs on Reddit [Anxiety $F(1, 18892.5) = 85.09$, Sadness $F(1, 19828.1) = 79.92$]. For both emotion variables, there were no significant two- or three-way interactions with user set (all $t < 2$), suggesting that differences between target and comparison users were not limited to the pandemic era. The effect of user set remained significant controlling for era, work-related language, and the quadratic effect of week [Anxiety $F(1, 3998.70 = 313.4$, Sadness $F(1, 4027.1) = 55.53$ both $p < .0001$].

5 Discussion

The results illustrate the potential uses of a new corpus of Reddit messages written by two groups of people: those who indicated on Reddit that they had left their jobs (voluntarily or not) and matched comparison users with similar posting histories who had not discussed a job change. Though the findings are correlational, they have potential relevance for future interventions aimed at helping people cope with career changes and other stressful life events more effectively.

Examining emotional language in the year before and after quitting showed that people who quit their jobs used more anxiety and sadness language than matched comparison users, and these differences were largest for people using high rates of "I" pronouns at baseline. Our results build on self-distancing and expressive writing research to suggest that avoiding self-focus (Kross and Ayduk, 2011) or flexibly regulating perspectives by changing personal pronouns (Seih et al., 2011) may help people experience less distress as they prepare for major life changes.

Anxious language before and in the weeks immediately surrounding quitting was highest for those using high rates of cognitive language at baseline, relative to both comparison users and people with low baseline rates of cognitive language. Words such as *think* and *wonder* reflect self-insight and sense-making in diaries, therapy, or expressive writing (Pennebaker and Chung, 2007) but may be more reflective of hedging (e.g., "I think"), uncertainty (e.g., "I guess"), and rumination in everyday social contexts, such as Reddit (Dean and Boyd, 2020). People with larger negative emotional vocabularies tend to also use more cognitive processing language (Vine et al., 2020). Though our results are preliminary, they support the conjecture that cognitive language may be risky in some social contexts, such as social media.

Together, our findings suggest that predictive models aiming to predict specific life events or linguistic sentiment may produce more accurate or precise results if moderation by pronouns and cognitive language are considered. Exploring possible psychological moderators is especially consequential in transparent models, such as regression or structural equation modeling, where the aim is to model and interpret every feature in depth—understanding each variable's relation to the outcome and other predictors as well as its variance structure—in order to facilitate psychological and behavioral insights (Rudin, 2019).

Methodologically, this project illustrates potential uses of SALLEE, a new sentiment analysis approach used primarily in industry settings (Adams, 2022). SALLEE integrates practical aspects of both traditional dictionary-based emotion measures and syntax-aware techniques that such lexicons typically lack. Like other top-down methods, SALLEE is relatively transparent and face-valid, including only words with explicit emotional content (e.g., *lonely* and *wept* for sadness), in contrast with data-driven approaches that often assign sentiment weights to superficially neutral words. Face validity is not a panacea, and seemingly unambiguous emotion terms may have different implications across different contexts (Chan et al., 2021; Hamilton et al., 2016); however, using explicitly emotional words facilitates straightforward interpretation of results and lowers the adoption barrier for researchers who are new to sentiment analysis. We should note, however, that machine learning or open-vocabulary methods often outperform dictionary approaches in cases with sparse words, novel contexts, and many low-frequency or out-of-vocabulary words (see Eichstaedt et al., 2021).

5.1 Limitations

The sample we collected has limitations shared by most datasets focusing on naturalistic behavior surrounding some event. First, because the starting point in this sample was the quit messages, and not every person posts each week, there are necessarily more target posts than other posts. Quit weeks and weeks immediately surrounding target posts were also more verbose than other weeks and, because work-related concerns were salient at the time, likely included more comments and posts about work or career planning.

Second, the Reddit sample we analyzed is heterogeneous. In most respects, that is a benefit of these data—the conversations covered diverse topics and took place in groups with varying social norms, cohesiveness, and cultures. In that way, these messages are more naturalistic than language from controlled experiments or narrowly focused social media research. Yet there are better options than simply averaging over these differences. For example, emotional expressions are both inflated and suppressed by forum norms regarding emotional self-disclosure (see Balani and De Choudhury, 2015), and the same terms take on different affective meanings across communities (Hamilton et al., 2016). Future research on these or similar data may benefit from clustering forums into psychologically meaningful groups or developing sentiment lexicons tailored to each forum.

Finally, as with any analysis of self-labeled data on social media, we are taking people at their word, accepting the likelihood that some of the messages about quitting in our sample were exaggerated or fabricated (Coppersmith et al., 2015). Despite efforts to stringently filter out hypothetical, satirical, fictional, remembered, or otherwise non-literal references to recent quitting, there are also no doubt some remaining false positives.

5.2 Future Analyses

The corpus we have compiled—including both the messages focused on in the present analyses and those we excluded—is dynamic and growing. In addition to adding new messages as the users in the sample continue to use Reddit, the sample offers a cornucopia of options for studying the psychology of job changes. The sample of excluded users alone is rife for analyses involving sarcasm detection ("Quitting will solve everything!"), advice requests ("What's next if I quit today?"), and counterfactual thinking ("If I'd quit a year ago..."). We are sharing the filtered data as-is but will continue refining it over time. Future analyses may compare SALLEE and LIWC with other language-based emotion measures and experiment with machine learning approaches to forecasting quitting.

The corpus may also be useful for specific workplace applications. Being able to predict voluntary turnover from everyday conversations that are not explicitly about quitting would be invaluable to employers, as organizations lose expertise, social capital, and tangible and intangible investments when employees resign (Rubenstein et al., 2018). However, devising algorithms for predicting the likelihood of leaving a job from language used outside of work introduces ethical quagmires that are beyond the scope of this paper, including questions about the costs of false positives (i.e., being wrongly labeled as a turnover risk by employers).

From the perspective of social-personality psychology, our results add to previous research showing that individual differences in self-focus or self-distancing are relevant to emotional experiences, especially during times of stress or distress (Kross and Ayduk, 2011). We additionally build on the less-established link between cognitive processing and negative emotion (Vine et al., 2020), showing that, independent of self-focus, people who tend to use words referring to thought processes (*think*, *realize*, *wonder*, etc.) at high rates when posting on Reddit use more negative emotional language.

6 Conclusion

Analyses of naturalistic language used in messages on Reddit in the year before and after voluntarily leaving a job showed that people who used the most self-references and cognitive processing language at baseline used more sad and anxious language in the months surrounding quitting. Consistent with research on self-distancing and rumination, low rates of self-referential pronouns and cognitive processing language may be part of a broader pattern of healthy coping with stress and negative emotion. Finally, emotional trajectories for quitting before and after the start of the COVID-19 pandemic were parallel, but pandemic messages were more negative overall. Beyond the psychological implications of this research, methodologically, we have contributed a new publicly-available Reddit corpus and a reliable method for identifying the timing of major life events discussed on social media.

Acknowledgements

Special thanks to Sean Farrell for collecting an initial version of the dataset and generating research ideas during early stages of this project.

References

Kiki Adams. 2022. SALLEE documentation. Online Receptiviti Inc. documentation.

Ashwini Ashokkumar and James W Pennebaker. 2021. Social media conversations reveal large psychological shifts caused by covid-19's onset across us cities. *Science advances*, 7(39):eabg7843.

Jenna L Baddeley, James W Pennebaker, and Christopher G Beevers. 2013. Everyday social behavior during a major depressive episode. *Social Psychological and Personality Science*, 4(4):445–452.

Sairam Balani and Munmun De Choudhury. 2015. Detecting and characterizing mental health related self-disclosure in social media. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1373–1378.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48. R package version 1.1.33.

Ryan L Boyd and H Andrew Schwartz. 2021. Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field. *Journal of Language and Social Psychology*, 40(1):21–41.

Charles S Carver and Eddie Harmon-Jones. 2009. Anger is an approach-related affect: evidence and implications. *Psychological bulletin*, 135(2):183.

Chung-hong Chan, Joseph Bajjalieh, Loretta Auvil, Hartmut Wessler, Scott Althaus, Kasper Welbers, Wouter Van Atteveldt, and Marc Jungblut. 2021. Four best practices for measuring news sentiment using 'off-the-shelf'dictionaries: A large-scale p-hacking experiment. *Computational Communication Research*, 3(1):1–27.

Gilad Chen, Robert E Ployhart, Helena Cooper Thomas, Neil Anderson, and Paul D Bliese. 2011. The power of momentum: A new model of dynamic relationships between job satisfaction change and turnover intentions. *Academy of Management Journal*, 54(1):159–181.

Glen Coppersmith, Mark Dredze, Craig Harman, and Kristy Hollingshead. 2015. From adhd to sad: Analyzing the language of mental health on twitter through self-reported diagnoses. In *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, pages 1–10.

Philip J Corr. 2013. Approach and avoidance behaviour: Multiple systems and their interactions. *Emotion Review*, 5(3):285–290.

Hannah J Dean and Ryan L Boyd. 2020. Deep into that darkness peering: a computational analysis of the role of depression in edgar allan poe's life and death. *Journal of affective disorders*, 266:482–491.

R Maria del Rio-Chanona, Alejandro Hermida-Carrillo, Melody Sepahpour-Fard, Luning Sun, Renata Topinkova, and Ljubica Nedelkoska. 2022. Mental health concerns prelude the great resignation: Evidence from social media. *arXiv preprint arXiv:2208.07926*.

Department of Health and Human Services. 2017. Federal Policy for the Protection of Human Subjects ('Common Rule'). [revised in 2018 and 2019; became effective on July 19, 2018 and delayed until January 21, 2019, and then again delayed until January 21, 2020].

Johannes C Eichstaedt, Margaret L Kern, David B Yaden, H Andrew Schwartz, Salvatore Giorgi, Gregory Park, Courtney A Hagan, Victoria A Tobolsky, Laura K Smith, Anneke Buffone, et al. 2021. Closed- and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychological Methods*, 26(4):398.

Johannes C Eichstaedt, Robert J Smith, Raina M Merchant, Lyle H Ungar, Patrick Crutchley, Daniel Preoţiuc-Pietro, David A Asch, and H Andrew Schwartz. 2018. Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences*, 115(44):11203–11208.

Stephanie Ferguson. 2022. Understanding america's labor shortage: The most impacted industries. *US Chamber of Commerce: Washington, DC, USA*, 7.

Joseph Fuller and William Kerr. 2022. The great resignation didn't start with the pandemic. *Harvard Business Review*.

Maury Gittleman. 2022. The "great resignation" in perspective. *Monthly Labor Review*.

Jeff Greenberg and Tom Pyszczynski. 1986. Persistent high self-focus after failure and low self-focus after success: The depressive self-focusing style. *Journal of Personality and Social Psychology*, 50(5):1039.

Ulrich Halekoh and Søren Højsgaard. 2014. A kenward-roger approximation and parametric bootstrap methods for tests in linear mixed models – the R package pbkrtest. *Journal of Statistical Software*, 59(9):1–30. R package version 0.5.2.

William L Hamilton, Kevin Clark, Jure Leskovec, and Dan Jurafsky. 2016. Inducing domain-specific sentiment lexicons from unlabeled corpora. In *Proceedings of the conference on empirical methods in natural language processing. conference on empirical methods in natural language processing*, volume 2016, page 595. NIH Public Access.

Joseph Henrich, Steven J Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and brain sciences*, 33(2-3):61–83.

Nicholas S Holtzman et al. 2017. A meta-analysis of correlations between depression and first person singular pronoun use. *Journal of Research in Personality*, 68:63–68.

Gabriela Ksinan Jiskrova. 2022. Impact of covid-19 pandemic on the workforce: from psychological distress to the great resignation. *J Epidemiol Community Health*, 76(6):525–526.

Ewa Kacewicz, Richard B Slatcher, and James W Pennebaker. 2007. Expressive writing: An alternative to traditional methods. *Low-cost approaches to promote physical and mental health: Theory, research, and practice*, pages 271–284.

Ethan Kross and Ozlem Ayduk. 2011. Making meaning out of negative experiences by self-distancing. *Current directions in psychological science*, 20(3):187–191.

Matthew Matero, Albert Hung, and H Schwartz. 2022. Understanding BERT's mood: The role of contextual-embeddings as user-representations for depression assessment. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, pages 89–94.

Alan Medlar, Yang Liu, and Dorota Glowacka. 2022. Nobody wants to work anymore: An analysis of r/antiwork and the interplay between social and mainstream media during the great resignation. *arXiv preprint arXiv:2210.07796*.

Jiyoung Park, Özlem Ayduk, and Ethan Kross. 2016. Stepping back to move forward: Expressive writing promotes self-distancing. *Emotion*, 16(3):349.

Kim Parker and Juliana Menasce Horowitz. 2022. Majority of workers who quit a job in 2021 cite low pay, no opportunities for advancement, feeling disrespected. *Pew Research Center*.

James W Pennebaker and Cindy K Chung. 2007. Expressive writing, emotional upheavals, and health. In *Foundations of health psychology*, page 263–284. Oxford University Press.

JW Pennebaker, RJ Booth, RL Boyd, and ME Francis. 2015. Linguistic inquiry and word count: LIWC 2015.

Tom Pyszczynski, Kathleen Holt, and Jeff Greenberg. 1987. Depression, self-focused attention, and expectancies for positive and negative future life events for self and others. *Journal of personality and social psychology*, 52(5):994.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Version 4.3.0.

Riyaz Ahmad Rainayee. 2013. Employee turnover intentions: Job stress or perceived alternative external opportunities. *Business and Management*, 5(1):48–59.

Antje Riepenhausen, Carolin Wackerhagen, Zala C Reppmann, Hans-Christian Deter, Raffael Kalisch, Ilya M Veer, and Henrik Walter. 2022. Positive cognitive reappraisal in stress resilience, mental health, and well-being: A comprehensive systematic review. *Emotion Review*, 14(4):310–331.

Alex L Rubenstein, Marion B Eberly, Thomas W Lee, and Terence R Mitchell. 2018. Surveying the forest: A meta-analysis, moderator investigation, and future-oriented discussion of the antecedents of voluntary employee turnover. *Personnel Psychology*, 71(1):23–65.

Cynthia Rudin. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5):206–215.

William Sattelberg. 2021. The demographics of reddit: Who uses the site. *Alphr*.

Rebecca A Schwartz-Mette and Amanda J Rose. 2016. Depressive symptoms and conversational self-focus in adolescents' friendships. *Journal of abnormal child psychology*, 44:87–100.

Yi-Tai Seih, Cindy K Chung, and James W Pennebaker. 2011. Experimental manipulations of perspective taking and perspective switching in expressive writing. *Cognition & emotion*, 25(5):926–938.

Sarah Seraj, Kate G Blackburn, and James W Pennebaker. 2021. Language left behind on social media exposes the emotional and cognitive costs of a romantic breakup. *Proceedings of the National Academy of Sciences*, 118(7):e2017154118.

Han-Chin Shing, Suraj Nair, Ayah Zirikly, Meir Friedenberg, Hal Daumé III, and Philip Resnik. 2018. Expert, crowdsourced, and machine assessment of suicide risk via online postings. In *Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic*, pages 25–36.

Rachel A Simmons, Peter C Gordon, and Dianne L Chambless. 2005. Pronouns in marital interaction: What do "you" and "i" say about marital health? *Psychological science*, 16(12):932–936.

Sabine Sonnentag and Ute-Vera Bayer. 2005. Switching off mentally: predictors and consequences of psychological detachment from work during off-job time. *Journal of occupational health psychology*, 10(4):393.

Allison M Tackman, David A Sbarra, Angela L Carey, M Brent Donnellan, Andrea B Horn, Nicholas S Holtzman, To'Meisha S Edwards, James W Pennebaker, and Matthias R Mehl. 2019. Depression, negative emotionality, and self-referential language: A multi-lab, multi-measure, and multi-language-task research synthesis. *Journal of personality and social psychology*, 116(5):817.

Vera Vine, Ryan L Boyd, and James W Pennebaker. 2020. Natural emotion vocabularies as windows on distress and well-being. *Nature communications*, 11(1):4525.

Edward R Watkins and Henrietta Roberts. 2020. Reflecting on rumination: Consequences, causes, mechanisms and treatment of rumination. *Behaviour Research and Therapy*, 127:103573.

Robert A Wicklund. 1975. Objective self-awareness. In *Advances in experimental social psychology*, volume 8, pages 233–275. Elsevier.

Jan Wijffels. 2023. *udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'UDPipe' 'NLP' Toolkit*. R package version 0.8.11.