Grieve-JA: A Psycholinguistic Dictionary for Grievance-Fuelled Language in Japanese

Emilie Francis
University of Gothenburg, Sweden
Språkbanken Text
emilie.francis@gu.se

Abstract

Since the early 2010s, internet nationalists have been active on Japanese forums, blogs, and social networking services, spreading hate speech and conspiracy. At the same, the involuntary celibate community has gained international notoriety due to acts of mass violence committed by self-identified members. Despite several instances of misogyny motivated violence in Japan, there is little known about Japanese involuntary celibate culture and identity. This paper introduces Grieve-JA, a psycholinguistic dictionary for the analysis of violence, threat, and grievance in Japanese text. The paper outlines the process of developing the dictionary and evaluating it across different corpora representing both neutral and grievance-filled language. As a final measure, the dictionary is used to compare comments from two forums on the Japanese anonymous message board 5channel with comments from Bluesky. Results indicate that Grieve-JA is a useful tool for analyzing extreme language in Japanese as a complement to other psycholinguistic measures.

1 Introduction

Psycholinguistic dictionaries, such as Linguistic Inquiry Word Count (LIWC), assume that emotions and cognitive processes are reflected as measurable features in text (Pennebaker et al., 2015; Boyd et al., 2022; Igarashi et al., 2022). Such psychometric tools have been used to analyze sentiment and psychological processes in text for deception (Rashkin et al., 2017; Almela et al., 2012), far-right populism and radicalization (Breeze, 2020; Araque et al., 2022), religious extremism (Baker and Vessey, 2022), and threat and crises (Etaywe et al., 2024; Choi et al., 2022; Hu et al., 2023). These resources are also employed as features in automatic classification of written deception and threat assessment (Akrami et al., 2018; Rashkin et al., 2017; Glasgow and Schouten, 2014; Chauvin, 2011). The Grievance Dictionary, developed for English by van der Vegt et al. (2021), is one such psychometric tool that has been used to evaluate violent language in lone-actor terrorist manifestos and white supremacy forums.

Threat and risk assessment includes a wide range of violence, such as targeted violence against individuals and mass casualty events. Research has shown that such violence involves grievance-driven planning communicated in advance (Corner et al., 2018; Gill, 2021). The involuntary celibate (incel) community is another subject of study and surveillance using linguistic analysis (Hoffman et al., 2020; Lounela and Murphy, 2024; Jaki et al., 2019; Van Brunt et al., 2021). The community has gained notoriety due to several prolific ideologically motivated attacks committed by self-identified members who communicated violent intentions online (Cecco, 2019; Branson-Potts and Winton, 2018; Morris, 2023).

In Japan, the connection between right-wing ultra-nationalism, 'incel' extremism, and violence has been salient since the early 2000s (Nagayoshi, 2021; Yoshida, 2024; DeCook and Fujioka, 2025). Despite this, there is a lack of research on hateful/threatening language in Japanese. The first step to address this gap is developing linguistic and NLP resources to analyze extremist discourse in Japan.

This paper presents Grieve-JA:[1] a psycholinguistic dictionary for grievance-fuelled violence and threat assessment following van der Vegt et al. (2021), localized for Japanese. The dictionary is tested on eleven different datasets to establish the validity of the measurements for grievance-fuelled language. To conclude, a psycholinguistic analysis of the heretofore unstudied incel and female incel communities of Japan is presented using the dictionary. The contributions of this research are twofold:

[1] github.com/emi-mari/grieve-ja

Proceedings of the Second International Conference on Natural Language Processing and Artificial Intelligence for Cyber Security (NLPAICS'2026)
11–12 June, 2026.

1. a linguistic resource for measuring grievance-fuelled language in Japanese

2. a psycholinguistic analysis of the Japanese involuntary celibate community

The following section presents an overview of the development and use of psycholinguistic dictionaries for various tasks, as well as a brief description of internet nationalism and incel culture in Japan.

2 Background

2.1 LIWC and Psycholinguistic Dictionaries

The Linguistic Inquiry and Word Count (LIWC) was originally created in 1993. Many versions have been released since, with the most commonly used versions being LIWC2015 and LIWC-22 (Pennebaker et al., 2015; Boyd et al., 2022). Each word in a LIWC dictionary is associated with one or more categories. The scores for each category are calculated by reading each word in a text, searching the dictionary for a match, then increasing the category scale for the matched word. The final category scores represent the proportion of total words in the text that match each of the dictionary categories. For example, if a text containing 1,000 words is analyzed and 50 words are related to positive emotion, this number is reported as 5.0% positive emotion. The most recent release, LIWC-22, has over 100 sub-dictionaries that measure various linguistic and psychological processes, such as affect, emotion, part of speech, etc.

J-LIWC is a Japanese translation of the LIWC2015 dictionary, with some adjustments made to account for differences between Japanese and English orthography and grammar (Igarashi et al., 2022). English words that were challenging to translate or culturally specific were generalized, transliterated, or omitted. For example, 'Yiddish' and 'Zoloft' were translated as *idisyugo* "Yiddish" and *kouutuyaku* "antidepressant". As the 'religion' and 'netspeak' categories had a considerable size discrepancy between the Japanese and English versions, Japanese words not present in the initial English version were added to fill the gap. Furthermore, as LIWC counts are based on word segments, it was also necessary to include a tokenization step in the pre-processing stage. The final version J-LIWC includes 11,600 words and 69 categories.

In addition to translations, numerous other dictionaries have been created to supplement LIWC. Among these, several dictionaries specifically measure dehumanization, vulnerability, violence, and threat (Baele et al., 2024; van der Vegt et al., 2021; Choi et al., 2022; Hu et al., 2023).

2.2 The Grievance Dictionary

Based on the methodology of LIWC, van der Vegt et al. (2021) developed the Grievance Dictionary aimed at measuring psychological and social concepts pertinent to grievance-fuelled violence and threat assessment research. The dictionary was designed to be general enough to capture varying types of violent and extremist language, including political/religious extremism and targeted threats. The initial Grievance Dictionary categories were created based on survey responses from experts in terrorism research. Another survey was given to collect suggestions for seed words for each category. This list of terms was extended with nearest neighbours from cosine similarity of GloVe word embeddings. The final un-stemmed dictionary is 5,513 words and has been translated into German, Dutch, and Italian (van der Vegt et al., 2025).

Following the methodology of the Grievance Dictionary, Hu et al. (2023) created a dictionary for studying expression of vulnerability in moments of crisis. The authors examined language use expressing vulnerability, including uncertainty, powerlessness, lack of trust toward authority, and susceptibility to radicalization. Ultimately, they created several lexicons related to grievances, accusations, helplessness, incompetence, out-group speech, etc. which were evaluated using a Twitter dataset of content concerning the January 6th Capitol riot and the COVID-19 pandemic.

2.3 Other Psycholinguistic Dictionaries for Threat

Choi et al. (2022) used word-embedding models to create a threat dictionary that measures threat levels in mass communication, such as news or social media. The dictionary was evaluated by examining patterns in threat related words over time and mapping them to moments in history when the US faced threats, such as violent conflict, natural disaster, and outbreaks of disease. It was also observed that communication that contained more threat terms was more widely shared online.

Incels, typically men who have difficulty forming romantic/sexual relationships with women and hold hostile views towards women and feminism, have been the focus of numerous studies using lin-

guistic analysis (Cottee, 2020; Jaki et al., 2019; Yoder et al., 2023; Johansson Wilén et al., 2024). Baele et al. (2024) used a dictionary based approach to measure violent extremism of online discourse among incels on Reddit, Telegram, forums, and blogs. The "Incel Violent Extremism Dictionary" includes 172 words for acts of violence, weapons, and dehumanizing out-group labels.

2.4 Net Nationalists and Inceldom in Japan

Before forums like 4chan gained notoriety for their role in the rise of the American alt-right, right-wing nationalists called '*netto uyoku*' were frequenting internet forums in Japan (Higuchi et al., 2019; Nagayoshi, 2021; Hack, 2016; DeCook and Fujioka, 2025). Characterized by xenophobia, anti-Korean/Chinese hate speech, and disdain for mass media, the *netto uyoko* community has been a persistent issue in Japanese society and politics (Yoshida, 2024; Yoshida and Demelius, 2025; Fujioka, 2020; DeCook and Fujioka, 2025; Takikawa and Nagayoshi, 2017; Suzuki, 2023; Sakamoto, 2011; Schäfer et al., 2017).

The commonly cited disseminators of such rhetoric are the popular anonymous message boards 5channel and its predecessor 2channel, controversial in Japan for their connection to incidents of mass violence (Sato, 2022; Fujioka, 2020). Many perpetrators of these incidents were motivated by nationalistic discourse online, some explicitly communicating their intention to commit violence on these forums prior to the act (Innami, 2024; Ryall, 2025).

Despite several incidents of misogyny motivated violence in Japan, Japanese incel culture remains obscure (Kihara, 2022; Montgomery, 2021). Two communities on 5channel, *motenai dansei* 'unpopular men' and *motenai onna* 'unpopular women', may be considered part of the global incel community. Like other incel communities, both *motenai dansei* and *motenai onna* have made lack of romantic experience an identity. Both forums refer to members with in-group labels (*moutoko* and *mojo* respectively) and prohibit posts from users who have or have had romantic, sexual, or even platonic relationships with the opposite sex. Although these communities do not use the label 'incel', both exhibit many of the same characteristics as other spaces in the incelosphere (see Section 4).

Due to connections with hate groups and violent extremism, both 2channel and 5channel are consistently monitored by Japanese police (DeCook and Fujioka, 2025; Sato, 2022).

Dataset	Documents	Tokens
news2025	260	91 800
motenai	29 100	3 939 000
wmotenai	575 000	57 182 400
aozora bunko	10 200	110 723 700
japanese-toxic	260	166 800
livedoor	7400	4 936 700
open2-news4vip	56 600	3 336 600
open2-livejupiter	103 900	6 069 100
open2-newsplus	24 400	1 445 700
bluesky	130 600	10 348 900
wrime	4100	272 400
Total	941 800	198 513 000

Table 1: Breakdown of number of documents and tokens per dataset. Numbers have been rounded to the nearest hundred.

3 Data and Methods

The methods used to develop and evaluate Grieve-JA follow the conventional psychometric dictionary creation process (van der Vegt et al., 2021, 2025; Igarashi et al., 2022; Pennebaker et al., 2015).

First, the dictionary is translated and localized following van der Vegt et al. (2025) and Igarashi et al. (2022). Next, the internal consistency and external validity are checked with Cronbach's α and Cohen's d following van der Vegt et al. (2021, 2025); Igarashi et al. (2022). Finally, following van der Vegt et al. (2021) and Igarashi et al. (2022), the Grieve-JA categories are compared with LIWC categories to ensure the measurements of Grieve-JA are distinct from LIWC. Term frequencies are calculated based on the LIWC convention as described in Section (2.1), with analysis carried out using the LIWC-22 software.

3.1 Data

Eleven datasets were used to evaluate Grieve-JA. Five datasets were used by Igarashi et al. (2022) in the development of J-LIWC: the well-established literature corpus AOZORA BUNKO (University, 2023), LIVEDOOR NEWS (Ltd., 2012), and the three OPEN 2CHANNEL datasets (Inaba, 2019).

Six datasets for social media and news were also included: the *unpopular men/women* (MOTENAI/WMOTENAI) 5channel forums, the WRIME social media corpus (Kajiwara et al., 2021), the JAPANESE-TOXIC dataset (Kobayashi, 2019), the BLUESKY-10-MILLION dataset (Ronotalt, 2024), and a small online news dataset.

The Japanese incel forums MOTENAI and WMOTENAI were collected using the Jane API for 5channel. As Jane suspended service to 5channel in July 2023, the dataset contains all forum posts until its suspension.

The NEWS2025 dataset is a small collection of articles from the right-wing nationalist newspaper *Sankei Shimbun* and centre/left-leaning *Mainichi* and *Asashi Shimbun*. The top 25 articles from the society, economy, politics, and international sections of each source were manually collected from October 5th to October 10th, 2025. The motivation for their inclusion was to ensure that terms added to the dictionary to reflect contemporary grievance-filled language, as described in Section 3.2, would not yield low frequencies due to a lack of representation in older datasets.

All data were pre-processed to remove punctuation, markdown, urls, and explicitly personally identifiable information. Additionally, social media datasets were filtered to remove ASCII art and spam. All data were tokenized and processed with MeCab 1.0.10 using the same custom dictionary as J-LIWC[2] and 'Ochasen' output format (Igarashi et al., 2022).

Scores for LIWC2015 are more reliable for comments over 50 tokens van der Vegt et al. (2021), so all datasets were filtered to remove texts under this threshold. In total, eleven corpora with 941,800 documents and 198,513,000 tokens were used for evaluation of Grieve-JA after filtering. The breakdown per dataset is shown in Table (1).

3.2 Dictionary Development

As noted above, the process follows that of van der Vegt et al. (2025) and (Igarashi et al., 2022). Since Japanese presents different challenges in translating from English not considered for translating German, Dutch, or Italian (van der Vegt et al., 2025), the strategies employed for translating LIWC into Japanese explained in Section (2.1) are referenced in such cases.

Translation: While both van der Vegt et al. (2025) and Igarashi et al. (2022) use Google Translate between English and the target language to provide translation candidates, initial translation for Grieve-JA were generated with GPT-4.1. The model was prompted by presenting dictionary terms from the non-stemmed English Grievance Dictionary of 5,513 words in a category-word pair and instructed to provide up to three translation candidates in Japanese for the target word based on the category. For words that require only a single translation, such as proper nouns like 'Putin', the model was instructed to return only one word. An example of the prompt used is provided in Appendix A.

The benefit of using a LLM over traditional machine translation is that synonyms are provided upfront, which eliminated the need to pad vocabulary using semantic similarity with word embeddings to account for large size differences between English and translated dictionaries.

Translations were manually reviewed in tandem by the main author and a professional Japanese translator. The reviewers did not note any mistranslations, but found several instances where the model produced superfluous translations in an effort to provide the maximum candidates when fewer were sufficient (e.g. *zikan wo hakaru mono* 'time measuring thing' for the English word 'clock' when *tokei* 'clock' is sufficient). Superfluous translations were removed and words which the reviewers agreed fit the translation, but were not suggested as one of the tree candidates by the model, were also added. These words were typically regional variations, internet slang, or other informal language (e.g. *satu* 'cops' for 'police'). For words difficult to translate into Japanese from English, the reviewers deferred to one of the approaches employed by the authors of J-LIWC described in Section (2.1).

As the English Grievance Dictionary was developed for the USA and UK, the dictionary needed to be localized for Japanese. For example, the words 'LAPD'[3] and 'CIA'[4] were localized to *keisityô* '(Tokyo) Metropolitan Police Department' and *naityô* 'Cabinet Intelligence and Research Office (CIRO)' respectively. New words were also added to capture current events relevant to grievance categories, such as *gunyôdorôn* 'military drone' in the weapons category. In total, 1,381 words were removed and 686 words were added to the translations.

Finally, translations were tokenized and stemmed with Mecab. Following J-LIWC, words with the same stem and meaning were combined using an asterisk to represent wildcards per LIWC convention (Igarashi et al., 2022). For example,

[2] https://github.com/tasukuigarashi/j-liwc2015

[3] Los Angeles Police Department

[4] Central Intelligence Agency

Category	Example Words	English Translation
fixation	追求, 夢中になる, 集中する	pursue, absorbed in, concentrate
jealousy	羨ましい, 浮気する, 望まれている	jealous, commit adultery, be desired
frustration	緊張する, ストレス, 迷惑	get nervous, stress, nuisance
surveillance	監視している, ウイルス, 盗聴	monitoring, (computer) virus, wiretap
honour	評価する, 礼儀正しい, 勇ましい	esteem, well-mannered, valiant
grievance	差別する, 心配, トラブル	discriminate, concerns, trouble
paranoia	懸念, 不安, 神経質	worry, anxiety, high strung
murder	被害者, 殺し屋, 殺す	victim, hit man, kill
soldier	自衛隊, 攻撃する, 戦争	self defence force, attack, war
god	信仰, 神様, 信じる	faith, god, believe
suicide	未遂, 自殺, うつ病	attempt, suicide, depression
desperation	手も足も出ない, 貧困, ホームレス	at one's wit's end, poverty, homeless
hate	嫌い, 軽蔑する, いじめる	hate, discriminate, bully
help	助ける, 介護士, 福祉	save, nurse, welfare
deadline	時間がなくなる, 期日, 仕事をする	run out of time, due date, do work
loneliness	未亡人, 引きこもり, 別れる	widow, shut-in, break up
weaponry	テロ行為, 武器を持つ, 核兵器	terrorist act, armed, nuclear weapon
relationship	カップル, 結婚している, 恋人	couple, married, lover
impostor	嘘をつく, 詐欺師, 振りをする	lie, swindler, pretend
planning	意図する, 整理する, 目的	intend, organise, objective
threat	リスク, 脅す, ハラスメント	risk, threaten, harassment
violence	傷つける, レイプ, 拉致する	injure, rape, abduct

Table 2: Examples of terms for each category and an English translation. Lists the three most typical words for each category.

all inflected forms of the verb いじめる 'bully', such as いじめた 'bullied', いじめられた 'was bullied', etc., are represented by いじめ*. Alternative forms of words commonly written in two or more of the Japanese writing systems were also included (e.g. *kega* 'wound' can be written as 怪我, けが, or ケガ).

A total of 4,964 terms were included in the final 1.0 version of dictionary, in line with the size of other translations. Table (2) shows a list of example of words for each category and an English translation.

Internal Consistency: A reliable psycholinguistic dictionary must consistently measure the target psychological category, meaning that an accurate tool should identify multiple words within the same category in the same text (Igarashi et al., 2022; Boyd et al., 2022; Pennebaker et al., 2015). If Grieve-JA is used to assess a news article on a political assassination, the dictionary is said to be consistent if it counts multiple words in the murder category within the news article (e.g. 'assassinate', 'gunman', and 'shoot' are terms in the murder category that might be in the article).

Internal consistency is measured using Cronbach's α, treating each category as a factor, dictionary term as an 'item', and individual text as a 'subject' (Igarashi et al., 2022; Pennebaker et al., 2015; Boyd et al., 2022). Cronbach's alpha was calculated for each category in each dataset then averaged across all corpora. As noted by Igarashi et al. (2022); Pennebaker et al. (2015), Cronbach's alpha was designed to measure the internal consistency of surveys and requires adjustment for natural language. Unlike surveys, one does not generally state the same concepts repeatedly in natural language. As a consequence, Cronbach's alpha tends to be underestimated for natural language. Like Igarashi et al. (2022); Pennebaker et al. (2015), the Spearman-Brown prediction formula is employed to counteract this effect under the assumption that the averaged alpha for each category is observed in all eleven datasets. Table (3) shows both the uncorrected and Spearman-Brown corrected Cronbach's alpha for each category.

Overall, the scores reveal that the dictionary is consistent. The highest consistency scores are achieved for the categories 'violence' (0.856), 'murder' (0.852), and 'soldier' (0.849). The lowest scores are observed for the categories 'fixation' (0.587), 'jealousy' (0.601), and 'frustration' (0.618). This result mirrors those of both the original English version and other translations of the Grievance Dictionary (van der Vegt et al., 2025). van der Vegt et al. (2021) suggests that the reason these categories typically score lower across languages is that they are more challenging psy-

chological concepts that are difficult to measure reliably.

Category	Uncorrected α	Corrected α
deadline	0.205	0.739
desperation	0.136	0.634
fixation	0.115	0.587
frustration	0.128	0.618
god	0.244	0.78
grievance	0.172	0.696
hate	0.222	0.758
help	0.149	0.659
honour	0.182	0.711
impostor	0.192	0.724
jealousy	0.12	0.601
loneliness	0.141	0.643
murder	0.344	0.852
paranoia	0.155	0.669
planning	0.29	0.818
relationship	0.292	0.819
soldier	0.338	0.849
suicide	0.211	0.747
surveillance	0.259	0.794
threat	0.252	0.788
violence	0.351	0.856
weaponry	0.326	0.842

Table 3: Average internal consistency per category across datasets represented by Cronbach's (uncorrected) and the Spearman-Brown (corrected) alpha.

Category	Wmotenai	Motenai	WRIME
deadline	0.077	−0.072	0.242
desperation	0.182	0.010	0.179
fixation	0.106	−0.096	0.232
frustration	0.089	−0.063	0.174
god	−0.150	0.010	−0.081
grievance	0.198	0.185	0.136
hate	0.184	0.114	0.104
help	0.211	0.095	0.153
honour	0.119	0.056	0.093
impostor	0.159	0.140	0.026
jealousy	0.283	0.102	0.211
loneliness	0.233	−0.039	0.245
murder	0.124	0.200	0.021
paranoia	0.197	−0.041	0.241
planning	0.097	−0.069	0.272
relationship	0.120	0.015	0.133
soldier	0.079	0.111	0.111
suicide	0.291	0.095	0.252
surveillance	−0.010	−0.065	0.038
threat	0.154	0.182	0.080
violence	0.197	0.199	0.030
weaponry	0.088	0.121	0.020

Table 4: Differences between WMOTENAI, MOTENAI, WRIME, and Bluesky represented with Cohen's d. Highlighted cells show the categories for which the difference is greater than 0.2, where $d < 0.2$ is trivial.

External Validity: Following van der Vegt et al. (2021), Grieve-JA was also tested on four different datasets to assess external validity. This indicates whether dictionary measurements make sense for the text analyzed (i.e. analyzing a children's story should yield negligent results compared to a terrorist manifesto).

The neutral Bluesky dataset was compared to WRIME, MOTENAI, and WMOTENAI. The *motenai* 'unpopular' forums, as part of the 'incelosphere', should score higher on Grieve-JA categories than Bluesky. Particularly, as previous research on the incel community shows, these forums should score higher for violence, hate, and mental health categories (Baele et al., 2024, 2021; Jaki et al., 2019; Balci et al., 2023). The WRIME dataset, collected during the early stages of the COVID-19 pandemic, contains many comments in which authors respond to pandemic policies. As such, it is expected to score higher on categories related to social isolation and anxiety compared to *Bluesky*.

Bluesky and the *motenai* datasets were randomly down-sampled to match the size of WRIME. The mean scores of the *motenai* and WRIME datasets were compared to the means of *Bluesky*, with magnitude of difference represented with Cohen's d. Table (4) reports the Cohen's d effect size between *Bluesky* and the respective in-domain datasets.

Overall, results affirm expectations. For WMOTENAI, the categories of 'jealousy' (0.28) and 'suicide' (0.29) have the largest difference. This observation is in line with previous literature on discourse on the female incel (femcel) community, which has shown hostility expressed at 'attractive' women and a deep impact of loneliness on mental health (Kay, 2024; Bobo, 2023; Francis, 2025a). Further support for this finding is discussed in Section (4). The highest categories for WRIME are 'planning' (0.27) and 'suicide' (0.25), which is indicative of the toll of social isolation from quarantine and social distancing.

With the exception of the 'murder' (0.2) and 'violence' (0.2), the difference between the MOTENAI forum and Bluesky was smaller. While this may suggest MOTENAI users express less grievance-fuelled language compared to WMOTENAI and WRIME, the analysis in Section (4) shows this isn't the case. This may be becaus MOTENAI users use more in-group specific discourse not captured by the dictionary. For example, members frequently engage in self-deprecating threads where they refer to themselves and others as excrement and employ canned expressions.

J-LIWC Correlations: Finally, Pearson's correlation is calculated between J-LIWC and Grieve-JA categories to ensure that the categories of Grieve-JA are distinct from J-LIWC categories (i.e. that the Grieve-JA dictionary is complementary and does not measure the same psycholinguistic traits as J-LIWC). Calculating correlations between Grieve-JA and J-LIWC using the full datasets was too resource intensive, a random sample of 20,000 texts was taken from each dataset. Correlations between Grieve-JA and J-LIWC were measured with Pearson's correlation coefficient $r \geq= 0.2$ and are shown in Appendix B.

Like van der Vegt et al. (2021), no strong correlations were found between Grieve-JA and J-LIWC categories. However, several weak correlations are observed. The largest of these weak correlations is between the category 'loneliness' and 'i' feature of J-LIWC ($r = 0.34$). There is also a weak correlation between the 'hate' category, 'anger' ($r = 0.31$), and 'negative emotion' ($r = 0.29$) features. As Grievance Dictionary categories are designed to supplement LIWC, correlations demonstrate which other psycholinguistic concepts measured through J-LIWC are associated with each Grieve-JA category. The Grieve-JA categories that show any correlation to J-LIWC categories are ones that are expected to be psychologically related, such as 'hate' and 'anger' or 'suicide' and 'death'.

4 An Analysis of the Unpopular

Finally, the dictionary is put to use by analyzing the MOTENAI and WMOTENAI forum datasets with Grieve-JA and comparing the results to the neutral BLUESKY. As shown in Section (3.2), the *motenai* forums share characteristics with English incel and femcel communities; particularly focus on violence and jealousy. These forums also showed aspects that align with *netto uyoku*, such as a higher frequency of hate. To differentiate this analysis from external validity in the previous section, the scope is narrowed to focus on core talking points in incel and *netto uyoku* discourse.

As one of the most important characteristics of the incelosphere is 'anti-feminism' (Ging, 2017; Bauer, 2024), all comments mentioning words related to 'feminism/feminists' are analyzed with Grieve-JA and J-LIWC. Similarly, hatred and violent language directed at 'Korea/Koreans' and 'zaniti' are core characteristics of *netto uyoku* discourse in Japan. The results for the top four categories with the largest difference between the Grieve-JA and J-LIWC analyses for these comments are shown in Figure (1).

Results show there is a considerable difference between the *motenai* forums and Bluesky for these terms. Both have a much higher frequency of language that expresses hate, violence, grievance, threat, and murder for comments mentioning 'feminism/feminists' and 'Korea/Koreans' compared to Bluesky. As shown in Figure (1), one can see that there is a noticeably higher frequency of language expressing negative emotion and anger. Overall, the scores for MOTENAI and WMOTENAI are similar. This is likely a reflection of the general level of extreme language on 5channel and other 'chan' forums, observed in previous research (Sakamoto, 2011; Lombardi, 2025; Colley and Moore, 2022). Despite this, there are notable differences. To explain these differences, netnography was applied to the comments.

The WMOTENAI forum scored much higher for the 'violence' category compared to MOTENAI for the terms 'feminism' and 'feminist'. Upon manual review of the comments, it was observed that users of WMOTENAI frequently discuss men's hatred for and negative stereotypes of feminism/feminists. In particular, users often discuss 'feminism bashing' online and in the media.

On the other hand, a review of comments in the MOTENAI dataset revealed that users portray feminism and feminists as perpetrators of 'violence' and 'hate' against men. The higher score for the 'grievance' category observed for the MOTENAI forum may be attributed to descriptions of feminism/feminists as 'irritating', 'rabid', or 'extreme'. The argument that 'irritating' and 'violent' feminists are oppressing men is prevalent across the incelosphere (Gemelli and Minnema, 2024; Muti et al., 2025; Francis, 2025b; Wiklund, 2020).

For the terms 'Korea/Koreans', the WMOTENAI forum scores much higher than MOTENAI on the 'jealousy' category and 'positive emotion'. Upon reviewing the comments, it was observed that users often discuss Korean pop culture and beauty. Many comments express jealousy toward 'attractive' Korean celebrities, using positive language to describe their physical appearance (often in comparison to Japanese men and women). This observation is consistent with previous research which show femcels focus on physical appearance and compari-

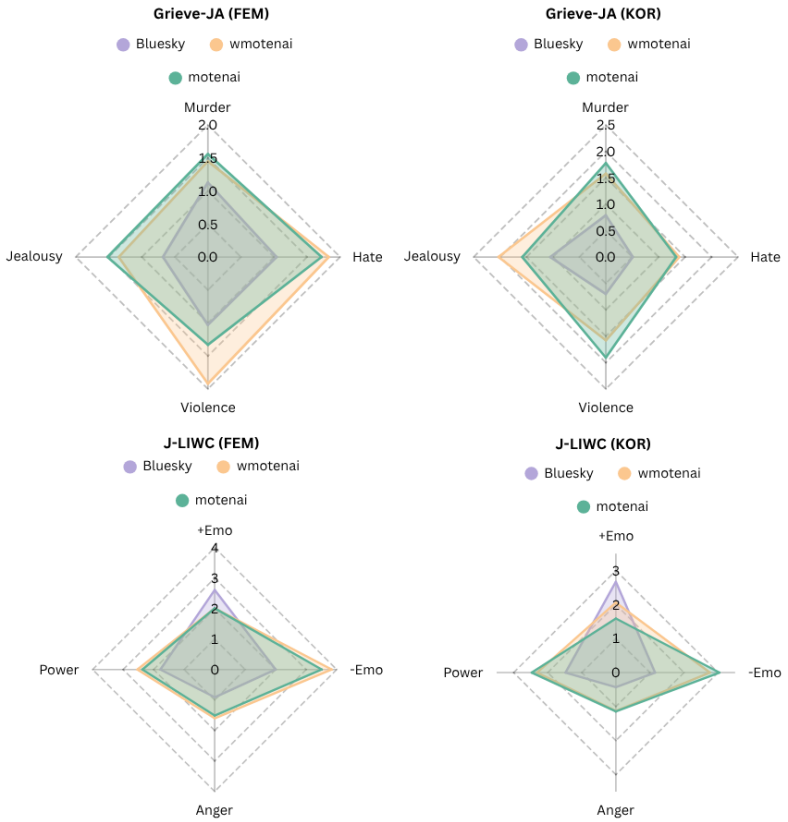

Figure 1: Radar plots showing the scores for Grieve-JA (top) and J-LIWC (bottom) for comments mentioning 'feminism/feminists' (left) and 'Korea/Koreans' (right).

son with 'attractive' women (Evans and Lankford, 2024; Balci et al., 2023; Francis, 2025a).

In comments mentioning *zainiti*, ethnic Korean Japanese citizens who immigrated before 1945, both WMOTENAI and MOTENAI express a high level of hate, violence, murder, and negative emotion. By reviewing the comments, it was observed that users repeat *netto uyoku* rhetoric that *zainiti* Koreans commit violent crimes, destroy Japan, and should be 'sent back' (Sakamoto, 2011; Schäfer et al., 2017). Although not currently a measure of Grieve-JA, it was also observed that both groups employ a high frequency of dehumanizing language.

Another observation in support of the link between MOTENAI and the incelosphere is the similarly sexually explicit and violent language toward women. Users on MOTENAI use sexually demeaning language towards women (e.g. *manko* 'c**t') and express delight at news stories about murdered women. While such threads were less frequent than the previously mentioned self-deprecation threads, they are consistent with the broader incelosphere (Tranchese and Sugiura, 2021; Gemelli and Minnema, 2024).

5 Conclusion and Future Work

This paper introduces Grieve-JA, a Japanese psycholinguistic dictionary for assessing grievance-fuelled language based on the Grievance Dictionary for English. Grieve-JA was applied to two forums from the anonymous message board 5channel *motenai dansei* 'unpopular men' and *motenai onna* 'unpopular women' in a novel analysis of the Japanese involuntary celibate community.

Results show Grieve-JA has potential to be a useful tool in combination with other linguistic measurements such as J-LIWC for analysis of extreme and grievance-fuelled language in Japanese. An analysis of the 'unpopular' forums showed that both communities exhibit similarities to the international incel and femcel communities, as well as rhetoric typical of Japan's *netto uyoku*.

In the future, following Baele et al. (2024), it would be beneficial to include terms specific to the involuntary celibate community in Japan and a category for dehumanization. While this paper has provided a first step, a more dedicated study is needed to explore discourse patterns among the 'unpopular' men and women of Japan.

Limitations

As a translation, Grieve-JA has the same limitation as van der Vegt et al. (2021) and van der Vegt et al. (2025) for English, German, Dutch, and Italian regarding the lower internal consistency score for several dictionary categories. Therefore, like van der Vegt et al. (2021), the authors of Grieve-JA encourage potential users to be cautious in reviewing scores reported for these categories and to support the scores with other measures or evidence from the data.

Additionally, given the fast-changing nature of online language, it is important that the dictionary be regularly updated with new vocabulary in order to effectively measure text from social media and grievances from current events. The dictionary is intended to be a living resource updated regularly with new terms to reflect changes in language over time in subsequent versions.

Finally, only two linguists were involved in the reviewing process. A more community oriented approach with annotators and community analysis may yield more comprehensive results and insights. For this reason, the author of this paper encourages other speakers of Japanese to contribute to future versions of the dictionary.

Ethical Considerations

It is important for potential users to consider whether Grieve-JA is appropriate for the text they intend to analyze. Both the original Grievance Dictionary and Grieve-JA measure the proportion of dictionary terms within a text, they are not measures of writer intent. Therefore, users of such dictionaries should be careful not to make assumptions on the potential for authors of texts within a dataset to commit violent acts based on dictionary scores alone. Furthermore, as both the Grievance Dictionary and Grieve-JA contain proper nouns to reflect discourse on current global conflicts, making value judgments based solely on grievance scores can present biased portrayals of these entities if the dictionary is used in an inappropriate context.

Finally, there is always the risk that such measures can be abused by bad actors to label individuals or groups as violent through misrepresentation of results or lack of context. Given these considerations, research using Greive-JA should be transparent in what is being analyzed, for what purpose, and by whom.

References

Nazar Akrami, Amendra Shrestha, Mathias Berggren, Lisa Kaati, Milan Obaidi, and Katie Cohen. 2018. Profile Risk Assessment Tool (PRAT).

Ángela Almela, Rafael Valencia García, and Pascual Cantos. 2012. Seeing through Deception: A Computational Approach to Deceit Detection in Written Communication. In *Proceedings of the Workshop on Computational Approaches to Deception Detection*, pages 15–22, Avignon, France. Association for Computational Linguistics (ACL).

Óscar Araque, J. Fernando Sánchez-Rada, Álvaro Carrera, Carlos A. Iglesias, Jorge Tardío, Guillermo García-Grao, Santina Musolino, and Francesco Antonelli. 2022. Making Sense of Language Signals for Monitoring Radicalization. *Applied Sciences 2022, Vol. 12, Page 8413*, 12(17):8413.

Stephane Baele, Lewys Brace, and Debbie Ging. 2024. A Diachronic Cross-Platforms Analysis of Violent Extremist Language in the Incel Online Ecosystem. *Terrorism and Political Violence*, 36(3):382–405.

Stephane J. Baele, Lewys Brace, and Travis G. Coan. 2021. From "Incel" to "Saint": Analyzing the violent worldview behind the 2018 Toronto attack. *Terrorism and Political Violence*, 33(8):1667–1691.

Paul Baker and Rachelle Vessey. 2022. A corpus-driven comparison of english and french islamist extremist texts. *International Journal of Corpus Linguistics*, 23:255–278.

Utkucan Balci, Chen Ling, Emiliano De Cristofaro, Megan Squire, Gianluca Stringhini, and Jeremy Blackburn. 2023. Beyond Fish and Bicycles: Exploring the Varieties of Online Women's Ideological Spaces. In *Published in the Proceedings of the 15th ACM Web Science Conference 2023*, pages 43–54. Association for Computing Machinery.

Mareike Fenja Bauer. 2024. Beauty, baby and backlash? Anti-feminist influencers on TikTok. *Feminist Media Studies*, 24(5):1023–1041.

Madeline Bobo. 2023. *Femcels: Where are the Women in the Incelosphere? An Exploratory Content Analysis of Femcel Forums*. Ph.D. thesis, Georgia State University.

Ryan L. Boyd, Ashwini Ashokkumar, Sarah Seraj, and James W. Pennebaker. 2022. The development and psychometric properties of LIWC-22. Technical report, University of Texas at Austin.

Hailey Branson-Potts and Richard Winton. 2018. How Elliot Rodger went from misfit mass murderer to 'saint' for group of misogynists - and suspected Toronto killer. *Los Angeles Times*.

Ruth Breeze. 2020. Angry tweets: A corpus-assisted study of anger in populist political discourse. *Journal of Language Aggression and Conflict*, 8(1):118–145.

Leyland Cecco. 2019. Toronto van attack suspect says he was 'radicalized' online by 'incels' | Toronto van murders. *The Guardian*.

Cherie Chauvin. 2011. Threatening Communications and Behavior: Perspectives on the Pursuit of Public Figures. *Threatening Communications and Behavior: Perspectives on the Pursuit of Public Figures*, pages 1–106.

Virginia K. Choi, Snehesh Shrestha, Xinyue Pan, and Michele J. Gelfand. 2022. When danger strikes: A linguistic tool for tracking America's collective response to threats. *Proceedings of the National Academy of Sciences of the United States of America*, 119(4).

Thomas Colley and Martin Moore. 2022. The challenges of studying 4chan and the alt-right: 'come on in the water's fine'. *New Media and Society*, 24:5–30.

Emily Corner, Paul Gill, Ronald Schouten, and Frank Farnham. 2018. Mental Disorders, Personality Traits, and Grievance-Fueled Targeted Violence: The Evidence Base and Implications for Research and Practice. *Journal of personality assessment*, 100(5):459–470.

Simon Cottee. 2020. Incel (E)motives: Resentment, Shame and Revenge. *Studies in Conflict & Terrorism*, 44(2):93–114.

Julia R. DeCook and Brett J. Fujioka. 2025. Japan's Netto Uyoku and the Crisis of Transnational Digital Surrogate Organizations. *Connective Action and the Rise of the Far-Right: Platforms, Politics, and the Crisis of Democracy*, pages 146–168.

Awni Etaywe, Kate Macfarlane, and Mamoun Alazab. 2024. A cyberterrorist behind the keyboard: An automated text analysis for psycholinguistic profiling and threat assessment. *Journal of Language Aggression and Conflict*.

Hannah Rae Evans and Adam Lankford. 2024. Femcel discussions of sex, frustration, power, and revenge. *Archives of Sexual Behavior*, 53:917–930.

Emilie Francis. 2025a. Between Hetero-Fatalism and Dark Femininity: Discussions of Relationships, Sex, and Men in the Femosphere. In *Proceedings of the The 9th Workshop on Online Abuse and Harms (WOAH)*, pages 325–341, Vienna, Austria. Association for Computational Linguistics.

Emilie Francis. 2025b. Language of the Swedish Manosphere with Swedish FrameNet. In *Proceedings of the Joint 25th Nordic Conference on Computational Linguistics and 11th Baltic Conference on Human Language Technologies (NoDaLiDa/Baltic-HLT 2025)*, pages 170–180, Tallinn, Estoina. University of Tartu Library.

Brett Fujioka. 2020. Toxic internet culture from east to west. *Noema*.

Sara Gemelli and Gosse Minnema. 2024. Manosphrames: exploring an Italian incel community through the lens of NLP and Frame Semantics. In *Proceedings of the First Workshop on Reference, Framing, and Perspective @ LREC-COLING 2024*, pages 28–39, Torino, Italia. ELRA and ICCL.

Paul Gill. 2021. Lone Actor Terrorism. *International Handbook of Threat Assessment*, pages 257–267.

Debbie Ging. 2017. Alphas, Betas, and Incels: Theorizing the Masculinities of the Manosphere. *Men and Masculinities*, 22(4):638–657.

Kimberly Glasgow and Ronald Schouten. 2014. Assessing Violence Risk in Threatening Communications. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 38–45.

Brett Hack. 2016. Subculture as social knowledge: a hopeful reading of otakuculture. *Contemporary Japan*, 28(1):33–57.

Naoto Higuchi, Kikuko Nagayoshi, Mitsuru Matsutani, Kouhei Kurahashi, Fabian Schäfer, and Tomomi Yamaguchi. 2019. *[What are internet right-ringers?]*, volume 97. Seikyusha.

Bruce Hoffman, Jacob Ware, and Ezra Shapiro. 2020. Assessing the Threat of Incel Violence. *Studies in Conflict & Terrorism*, 43(7):565–587.

Wenjia Hu, Zhifei Jin, and Kathleen M. Carley. 2023. Vulnerability dictionary: Language use during times of crisis and uncertainty. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 14161 LNCS:105–114.

Tasuku Igarashi, Shimpei Okuda, and Kazutoshi Sasahara. 2022. Development of the japanese version of the linguistic inquiry and word count dictionary 2015. *Frontiers in Psychology*, 13:841534.

Michimasa Inaba. 2019. Open 2channel dialogue corpus.

Atsushi Innami. 2024. onsi o satugaisita basu zyakkuhan no syônen ni, 5 nengo "turakatta ne"... hontôni atta "saisei" no monogatari [5 years after youth highjacked bus and murdered teacher: "It was hard"... a true account from a survivor]. *Newsweek Japan*.

Sylvia Jaki, Tom De Smedt, Maja Gwóźdź, Rudresh Panchal, Alexander Rossa, and Guy De Pauw. 2019. Online hatred of women in the Incels.me forum: Linguistic analysis and automatic detection. *Journal of Language Aggression and Conflict*, 7(2):240–268.

Evelina Johansson Wilén, Maria Wemrell, and Lena Gunnarsson. 2024. Gender knowledge production and epistemology in the incelosphere. *Journal of Gender Studies*.

Tomoyuki Kajiwara, Chenhui Chu, Noriko Takemura, Yuta Nakashima, and Hajime Nagahara. 2021. WRIME: A New Dataset for Emotional Intensity Estimation with Subjective and Objective Annotations. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2095–2104. ACL.

Jilly Boyce Kay. 2024. The reactionary turn in popular feminism. *Feminist Media Studies*, pages 1–18.

Maki Kihara. 2022. Victims of street massacre in Tokyo's Akihabara area remembered on 14th anniversary of attack. *The Mainichi Shimbun*.

Koga Kobayashi. 2019. Japanese toxic dataset.

Loris Lombardi. 2025. Toxic Talk: Structural Toxicity on a Japanese Anonymous Textboard. *Emerging Media*.

Emilia Lounela and Shane Murphy. 2024. Incel Violence and Victimhood: Negotiating Inceldom in Online Discussions of the Plymouth Shooting. *Terrorism and Political Violence*, 36(3):344–365.

Ronhuit Ltd. 2012. Livedoor news corpus.

Hanako Montgomery. 2021. Knife Attacker on Tokyo Train Says He Wanted to 'Kill Happy Women'. *Vice*.

Steven Morris. 2023. Plymouth shooter fascinated by serial killers and 'incel' culture, inquest hears. *The Guardian*.

Arianna Muti, Sara Gemelli, Emanuele Moscato, Emilie Francis, Amanda Cercas Curry, Flor Miriam Plaza-Del-Arco, and Debora Nozza. 2025. Blue-haired, misandriche, rabiata: Tracing the Connotation of 'Feminist(s)' Across Time, Languages and Domains. In *Proceedings of the The 9th Workshop on Online Abuse and Harms (WOAH)*, pages 299–311, Vienna, Austria. Association for Computational Linguistics (ACL).

Kikuko Nagayoshi. 2021. The Political Orientation of Japanese Online Right-wingers. *Pacific Affairs*, 94(1):5–32.

James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of LIWC2015.

Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking. *EMNLP 2017 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, pages 2931–2937.

Roronotalt. 2024. Bluesky 10-million dataset.

Julian Ryall. 2025. Japan's Nagano on high alert after stabbing spree, copycat threats shake city. *South China Morning Post*.

Rumi Sakamoto. 2011. 'Koreans, Go Home! 'Internet Nationalism in Contemporary Japan as a Digitally Mediated Subculture. *Asia-Pacific Journal*, 9(10).

Yumiko Sato. 2022. Tech Companies Missing a Brutal Wave of Hate Speech in Japan. *Time Magazine*.

Fabian Schäfer, Stefan Evert, and Philipp Heinrich. 2017. Japan's 2014 General Election: Political Bots, Right-Wing Internet Activism, and Prime Minister Shinzō Abe's Hidden Nationalist Agenda. *Big Data*, 5(4):294–309.

Daisuke Suzuki. 2023. netto uyoku ni natta rôhu wa, netouyoka mae kara "Asahi girai" datta... wakai koro no keiken kara mieta "igaina naimen no ugoki" [Before my father became radicalized as an internet right-winger, he was an Asahi-hater...inner workings observed from experience in youth]. *Gendai Media*.

Hiroki Takikawa and Kikuko Nagayoshi. 2017. Political polarization in social media: Analysis of the 'Twitter political field' in Japan. *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017*, 2018-January:3143–3150.

Alessia Tranchese and Lisa Sugiura. 2021. "I Don't Hate All Women, Just Those Stuck-Up Bitches": How Incels and Mainstream Pornography Speak the Same Extreme Language of Misogyny. *https://doi.org/10.1177/1077801221996453*, 27(14):2709–2734.

Globis University. 2023. Aozora bunko corpus.

Brian J. Van Brunt, Bethany S. Van Brunt, Chris Taylor, Nicole Morgan, and Jeffery Solomon. 2021. The Rise of the Incel Mission-Oriented Attacker. *Violence and Gender*, 8(4):163–174.

Isabelle van der Vegt, Bennett Kleinberg, Marilu Miotto, and Jonas Festor. 2025. Translating the Grievance Dictionary: a psychometric evaluation of Dutch, German, and Italian versions. *arXiv preprint: arXiv:2505.07495*, pages 1–10.

Isabelle van der Vegt, Maximilian Mozes, Bennett Kleinberg, and Paul Gill. 2021. The grievance dictionary: Understanding threatening language use. *Behavior Research Methods*, 53:2105–2119.

Maria Wiklund. 2020. *The misogyny within the manosphere. A discourse analysis in a Swedish context.* Ph.D. thesis, Malmö University, Malmö, Sweden.

Michael Miller Yoder, Chloe Perry, David West Brown, Kathleen M. Carley, and Meredith Pruden. 2023. Identity Construction in a Misogynist Incels Forum. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 1–13, Toronto, Canada. Association for Computational Linguistics (ACL).

Yutaka Yoshida. 2024. The Ambivalence of Far-Right Women: Hate, Trauma, Gender, and Neoliberalism in Contemporary Japan. *Journal of Interpersonal Violence*, 39(17-18):3829–3854.

Yutaka Yoshida and Yoko Demelius. 2025. Seduction of far-right actions: A pathway to an authentic self? *Crime, Media, Culture*, 21(2):187–210.

A Prompt Template

```
You are a translator translating a psycholinguistic dictionary of terms for
assessing extreme, threatening, or hateful language.
You will receive a category and a word in English.
Your task is to provide UP TO 3 words in Japanese that BEST align with the
meaning of the English word based on the category.
ONLY translate the English word. DO NOT translate the category.
For some words, such as named entities, there is ONLY one possible translation.
It NOT necessary to provide more than one translation for named entities.

The response MUST be in JSON.

[Example]
Category: deadline
Word: anxiety

Your response:
{{"translation": ["不安", "懸念", "心配"]}}

[Instructions]
Step 1: Identify the BEST translations of the English word in Japanese
- If there is ONLY one possible translation, return it
- Otherwise, provide UP TO 3 Japanese translations for the word
- Return the translations in a list
Step 2: Stop here and output the translations.
Your response MUST be in the following format:

{{"translation": ["", "", ""]}}

[Original]
Category: {cat}
Word: {wrd}

[Your response]
```

B J-LIWC Correlations

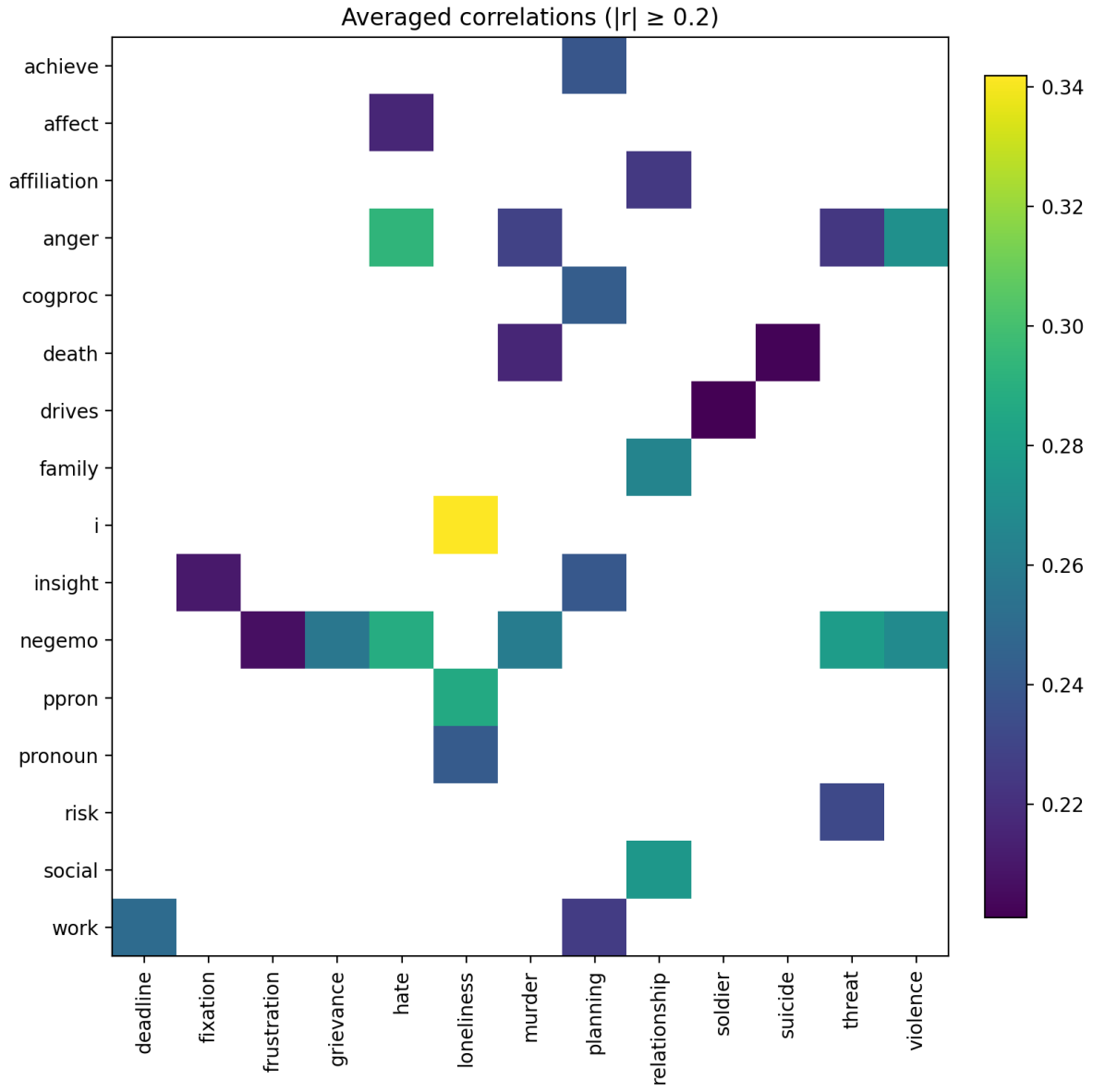

Figure 2: Correlations between Grieve-JA categories (x-axis) and J-LIWC features (y-axis), measured with Pearson's r. Only correlations with a $r \geq 0.2$ are included in the figure, as lower measures are considered insignificant.