Article

Examining Spanish-Language Pro Non-Suicidal Self-Injury (NSSI) Posts on Tumblr: A Computer-Assisted Text Analysis

Krisy Elrod [*] and Cass Dykeman

Article

Examining Spanish-Language Pro Non-Suicidal Self-Injury (NSSI) Posts on Tumblr: A Computer-Assisted Text Analysis

Krisy Elrod [1,*] and Cass Dykeman [2]

[1] University of Scranton

[2] Oregon State University

* Correspondence: krisy.elrod@scranton.edu; Tel.: 1-570-941-4060

Abstract: Adolescent non-suicidal self-injury (NSSI) often co-occurs with disorders such as depression, anxiety, and PTSD, yet limited research exists on Spanish-language social media contexts. This study employed a computer-assisted text analysis (CATA) approach to examine Spanish-language pro-NSSI Tumblr posts originating from North, Central, and South America. A year's worth of public posts was collected, focusing on captions and hashtags that included NSSI-related terms. Using Linguistic Inquiry and Word Count (LIWC) software, we analyzed linguistic and psychological markers. Log-likelihood ratio tests revealed significantly higher frequencies of words related to negative emotions, sadness, health, and death compared to standard blog norms. Mixed-language posts showed notable code-switching, suggesting a possible emotional distancing mechanism when discussing self-harm. Findings indicate that Spanish-speaking adolescents engaging in pro-NSSI communities exhibit unique linguistic and psychological characteristics, with important implications for clinical assessment and intervention. Mental health counselors and educators can use these insights to develop culturally and linguistically responsive strategies for prevention and support.

Keywords: adolescent; computer-assisted text analysis; CATA; mental health; self-injurious behavior; social media

1. Introduction

Online communities utilize distinctive linguistic practices to define membership terms, foster solidarity among members, share practices internally, and differentiate themselves from outsiders [1,2]. Language is critical in signaling shared identity and is subject to continual adaptation [3]. Linguistic patterns reflect the practices that are specific to virtual communities and the communication codes within groups [4]. In online communities where membership is discouraged through content-banning or blocking, linguistic variations of hashtags grow more popular, complex, and detached from their original spellings, usage, and context [5,6]. This complexity can confuse parents, teachers, and clinicians, thus hindering their ability to provide appropriate care and support.

Research into social media and pro-non-suicidal self-injury (NSSI) focused on five interconnected themes: (a) the prevalence of NSSI, (b) the prevalence of NSSI in Spanish-speaking communities, (c) the widespread use of social media, (d) the presence of pro-NSSI communities on social media platforms, and (e) the lack of education for parents, teachers, and clinicians regarding pro-NSSI terminology in Spanish [2,7–10]. These themes set the stage for the research questions (RQs) that drive this study, which will be examined in greater detail following a comprehensive review of each theme.

Favazza described NSSI as the "deliberate destruction of one's own body tissue in the absence of conscious suicidal intent" [7] (p. 260), and this phenomenon has been recognized for centuries. NSSI behaviors include cutting, burning, picking, scratching, interfering with wound healing, and

hitting an object until injury occurs; they may also encompass more dangerous acts such as bone-breaking or self-poisoning [8]. The prevalence of NSSI continues to rise globally, with studies indicating that 16-22% of adolescents engage in such behaviors, and females more likely to engage in the behaviors [9]. The prevalence of NSSI behaviors peaks at ages 14-15 [10] and is significantly associated with comorbid disorders such as depression, post-traumatic stress disorder (PTSD), and generalized anxiety disorder, justifying its classification as a transdiagnostic issue within the diagnostic and statistical manual of mental disorders [11].

Spanish is spoken by approximately 559 million people worldwide, with 460 million being native speakers. This positions Spanish as the language with the second largest population of native speakers globally, following Mandarin [12]. In the United States (U.S.), 13 percent of the population speaks Spanish at home, making it the country's most widely spoken non-English language. The U.S. has the second-largest population of Spanish speakers, following Mexico. Furthermore, current trends indicate that by 2050, one in three people in the U.S. will speak Spanish, including both monolingual and bilingual individuals [12,13].

Spanish-speaking adolescents and young adults in the U.S. face unique challenges due to acculturation stress and linguistic barriers, which significantly impact their mental health and increase the risk of engaging in behaviors like NSSI [14]. The pressure to assimilate into mainstream western culture while maintaining their own cultural heritage can contribute to emotional distress that leads to NSSI [15]. NSSI among Spanish-speaking adults has been shown to be moderated by ethnic identity, indicating the importance of cultural context in these behaviors [16]. Additionally, a comprehensive review of NSSI among African American and Hispanic adolescents reveals significantly high prevalence rates and unique risk factors within these populations, highlighting the necessity for culturally tailored prevention and intervention strategies that specifically address improving family relationships, addressing acculturation stress and providing coping skills for discrimination and socio-economic challenges [14].

The landscape of NSSI research has shifted with the proliferation of social media, which has fundamentally altered how individuals seek support and interact with others [17]. Online support groups can provide guidance and a sense of community for both positive and negative behaviors. Some groups promote self-harm as a lifestyle, encouraging members to maintain harmful behaviors at a significant personal cost [18]. The extensive use of social media, with adolescents and young adults spending approximately five hours daily on social media, exacerbates these issues [19]. On these platforms, peer and media influences can perpetuate and normalize NSSI behaviors [20].

The COVID-19 pandemic has increased social media usage, which has been associated with worsening mental health outcomes, including heightened anxiety and depression. This connection is particularly evident among adolescents and young adults who frequently use social networking sites [21]. The rise in social media use during the pandemic has also been linked to increased NSSI behaviors. Social media platforms often provide a space where individuals may share and view self-harm content, which can influence and propagate such behaviors [20]. Some individuals with prior mental health difficulties, including those engaging in NSSI, reported an increase in self-injurious urges and behaviors since the onset of COVID-19, exacerbated by social media exposure [21].

Recent research has demonstrated that social media platforms like Instagram can be critical in predicting acute mental health crises, including suicidality and NSSI behaviors [22]. Popular social media platforms among teens and young adults include YouTube and Meta's sites, with other platforms holding smaller shares of the global audience [23]. Despite efforts by these sites to block harmful content, pro-NSSI communities continue to thrive, finding ways to circumvent content bans through the creative use of language and hashtags. The secretive and isolating nature of NSSI makes these online communities both a potential lifeline and a source of reinforcement for harmful behaviors [5,6].

The existing literature that analyzes linguistic patterns has predominantly focused on English-language social media and general populations, leaving a gap in understanding the specific linguistic and psychological characteristics of NSSI among adolescents using the Spanish language on social

media platforms. The linguistic analysis of online pro-NSSI communities can provide invaluable insights. Linguistic Inquiry and Word Count (LIWC) software is particularly suited for this task, analyzing language specific to these communities to better understand how they evolve and communicate [18,24]. Previous studies have utilized computer-assisted text analysis to examine public blog posts on platforms like Tumblr and Reddit, providing a foundation for understanding linguistic markers and psychological processes associated with NSSI [24,25].

This study aims to fill this gap in the literature by utilizing the LIWC software to analyze pro-NSSI content on Tumblr, focusing specifically on public social media posts written by Spanish-speaking Tumblr users in North, Central, and South American countries. The LIWC software was used to analyze the linguistic and psychological processes in pro-NSSI Tumblr posts written in Spanish, English, and mixed languages (Spanish and English).

This research can inform culturally sensitive interventions and support strategies by identifying linguistic markers and psychological processes specific to Spanish-speaking adolescents' communication about NSSI. This is beneficial for clinical mental health counselors working with Spanish-speaking clients, as it highlights potential differences in how these adolescents communicate about NSSI compared to their peers.

To achieve the aims of the study, the following four RQs were addressed.

RQ1: What was the use rate of pro-NSSI-related word categories in public pro-NSSI Tumblr posts written in Spanish?

RQ2: What was the use rate of personal pronoun categories in public pro-NSSI Tumblr posts written in English?

RQ3: Did the use rate of personal pronoun categories differ from norms for such posts in public pro-NSSI Tumblr posts written in English?

RQ4: Did the use rate of specific linguistic and psychological process categories differ from norms for such posts in public pro-NSSI Tumblr posts written in English, Spanish, or mixed language?

2. Materials and Methods

2.1. Design

Our methodology utilized a quantitative computer-assisted text analysis (CATA) approach, which was chosen to capture the linguistic patterns present in a specific time frame, allowing for a detailed analysis of NSSI-related language as it appears in social [26]. Specifically, we used a log-likelihood (LL) ratio, which is considered a quantitative linguistic analysis as it uses statistical probability calculations to compare the frequency of a word or phrase between two different corpora; thereby allowing researchers to identify words that are significantly more prevalent in one corpus compared to another and providing a numerical measure of how key a word is to a specific context or topic. By comparing our corpora to the typical social media blog post corpora, we could identify differences in language use.

The unit of analysis was single words [27]. The variables for the RQ1 included the following word lexicon categories: (a) methods of NSSI, (b) cutting-specific terms, (c) NSSI terms, (d) instruments used, (e) reasons for NSSI, and (f) hidden-hashtag terms. Translation of these published lexicons from English to Spanish for the present study was done by the first author and reviewed for accuracy by a professional translation service. The 15 variables for the RQs 2-4 were selected from over the 90 variables available via the LIWC measure. These 15 variables were selected based upon support for these variables' relationship to NSSI found in the research literature. These 15 variables will be detailed in the measures section.

There were three sub-corpuses in this study: (a) Spanish-only, (b) English-only, and (c) mixed-language. Regarding RQs 3-4, the comparison norm for all corpora was Twitter's set of blog norms in English. No blog norms for Spanish or mixed-language writing might serve as respective baselines for the Spanish-only and mixed-language sub-corpuses. The inferential analysis used in this study

was a LL ratio test, a test that evaluates goodness-of-fit to a hypothesized distribution. Since Pearson's χ^2 can approximate the LL ratio, a priori power analysis for a χ^2 test was employed using G*Power 3.1 [25]. The proper effect size is Cohen's w [26]. Given the number of comparisons planned, the α level was set at .0033. The specific input parameters were: (a) test family = χ^2 tests, (b) statistical test = goodness-of-fit tests: contingency tables, (c) type of analysis = a priori: compute required sample size given α, power, and effect size, (d), w = 0.52, (e) power (1-β error probability) = 0.9, (f) α = .0033, and (g) degrees of freedom = 1. The G*Power 3.1 output included a sample size of 78 and an actual power of 0.90.

2.2. Ethics

The study was reviewed by the Institutional Review Board of Oregon State University (study 8721 on 7/13/2018) and determined that this research does not involve human subjects and thus required no further review.

2.3. Corpus

2.3.1. Data Collection

Tumblr's Application Programming Interface (API) [28] collected public posts between October 23, 2017, and October 23, 2018. The time frame of October 23, 2017, to October 23, 2018, was selected to capture a full year of posts, ensuring a comprehensive dataset. We configured the API to collect posts containing Spanish language in either the captions or the hashtags, used hashtags identifying the post as pro-NSSI (e.g., #autolesiones signified self-harm) [6,18], and that originated in a North, Central, or South American country. The geographical focus on North, Central, and South America was chosen due to the significant Spanish-speaking populations in these regions, as well as the accessibility of free and public data from user accounts located in these regions [28]. The API was instructed to eliminate the following while collecting text: usernames, URLs, location, names, photographs, comments left by other users, and direct quotes. Between October 23, 2017, and October 23, 2018, 1,868 posts and associated hashtags were collected (https://www.tumblr.com/docs/en/api/v2).

Tumblr was chosen as the social media platform as this site is used by individuals in the most likely age range for the onset of NSSI, so adolescents, and allows users to easily gather public data. In addition, Tumblr is a blogging and social media tool that enables users to publish a "tumblelog" or short blog posts. Tumblr's major differences from other social media sites are its free-form nature and users' ability to customize their pages heavily. This allows for more open expressions of thoughts and feelings in posts and the networks users create, which is necessary for accessing writing about NSSI.

2.3.2. Construction

Captions and hashtags associated with posts were collected, and posts were entered into an Excel document. Captions and hashtags were considered two separate parts of the post. For inclusion and exclusion of articles the following criteria were defined.

Inclusion: Posts containing Spanish language anywhere in the post, either caption or hashtag, were included.

Exclusion: First, posts written in languages other than English or Spanish were excluded. Second, posts where both the caption and hashtag were written entirely in English were excluded. Posts entirely in English, where both the hashtag and captions were in English, were excluded to maintain a focus on bilingual and Spanish-language content, which is underrepresented in existing literature. Excluding other languages ensured the analysis remained relevant to the target demographic. Sections that had Spanish language mixed with a language other than English were eliminated.

The corpus contained captions and hashtags for posts in which Spanish was used in either the caption or hashtag or both. Corpus construction continued as follows. If the whole section is in

English, it goes into an English file.   If the whole section is in Spanish, it goes into the Spanish file. If the section contains a mix of English and Spanish, it goes into the mixed file.

The corpus was separated into three sub-corpuses for analysis at the section level. These sub-corpuses were: (a) English-only, including sections of the post, either its caption or hashtags, which were entirely in English, while the rest of the post had Spanish language; (b) Spanish-only, including sections of the post, either the caption or hashtags, entirely in Spanish, and (c) mixed-language, including sections of the post where code-switching occurred in the caption or hashtag.

To illustrate the linguistic context, here are some examples of the corpus in Spanish, English, and mixed languages:

Spanish example: "Siento que no tengo escapatoria. #autolesión #dolor".   Translation: "I feel like I have no escape. #selfharm #pain.

English example: "I can't handle this anymore. #nopuedomas"

Mixed-language example: "Me siento tan sola. I don't know what to do. #autolesión #lost".

These examples highlight the linguistic variations used by individuals in pro-NSSI posts.

2.3.3. Preprocessing

Posts in all sub-corpuses were pre-processed using procedures outlined in the LIWC operator's manual and a supplement to this manual [29,30]. This cleaning involved the following processes: (a) all abbreviations were spelled out (e.g., "max" became maximum), (b) NSSI slang (words used to hide NSSI-related Instagram posts) was not normalized [20], and (c) textese was translated into standard spelling. Once the text was ready for analysis, the file was processed by LIWC software.

2.3.4. Size

The size of the Spanish-only sub-corpus was 42,661 units at word level. The English-only sub-corpus contained 9,814 units and the mixed-language sub-corpus included 4,299 units.

2.4. Measures

2.4.1. Overview

The measures used were selected scales from LIWC, and the Spanish adaptation of the Greaves NSSI Linguistic Scales [31]. Scores for all variables represented a percentage of all words used. The LIWC scales were pre-set scales contained in the LIWC software. Pennebaker [29] reported adequate reliability and validity for these scales. The reference norms used for comparison to the present results were Twitter norms contained in the LIWC psychometric manual [29]. Internal reliability and external validity of LIWC software were well documented [29].

2.4.2. LIWC Measures

The LICW linguistic measures employed were first-person singular, third-person singular, first-person plural, and third-person plural pronouns.   The psychological processes categories used included negative emotion (nasty), anxiety (worried), anger (hate), sadness (crying), cognitive processes (know), insight (consider), body (hands), health (flu), ingestion (dish), and death (coffin).

2.4.3. NSSI-specific Measures

Six specific measures for NSSI were utilized [31]. The first, GNLS_S-Methods of NSSI, evaluates various self-harming behaviors, including biting, burning, erasing, hitting, and picking. The second, GNLS_S-Cutting-Specific Terms, focuses on cutting methods, such as cut and cutting. The third, GNLS_S-NSSI Terms, tracks the usage of terms like non-suicidal, NSSI, suicide, and self-mutil*. The fourth, GNLS_S-Instruments Used, identifies tools used in NSSI, such as blades, bleach, erasers, and fingernails. The fifth, GNLS_S-Reasons for NSSI, examines motives behind NSSI, including feelings of anger or anxiety and desires for attention or control. The sixth, GNLS_S-Hidden Hashtag Terms,

uncovers covert hashtags used by those engaging in NSSI on social media, such as #mifamiliasecreta and #munecas.

2.4.4. Comparison Norms

English Twitter norms contained in the LIWC psychometric manual were used to evaluate the three sample corpora [29]. There are currently no Spanish or mixed-language blog norms available for comparison. While English and Spanish are syntactically dissimilar languages, it is reasonable to assume a roughly one-to-one correspondence between English nouns and Spanish nouns in terms of words [32]. Thus, comparable psychological process categories (nouns) in the second sub-corpus (Spanish-only) were compared to English Twitter norms for psychological process categories (nouns).   Similarly, the third sub-corpus (mixed language) was evaluated using the English Twitter norms for psychological process categories [29].

2.5. Apparatus

The LIWC software has been updated several times, but this study relied on the 2015 English and 2007 Spanish versions. LIWC2015 offered new dictionaries, including more social words and cognitive-process words [29]. The LIWC program can be used to review and analyze many kinds of written text in many languages, including English and Spanish. For the English sub-corpus, the English LIWC2015 dictionary was used; the Spanish LIWC2007 dictionary was used for the Spanish sub-corpus; and, finally, the mixed-language sub-corpus was analyzed using both the English LIWC2015 and Spanish LIWC2007 dictionaries, as recommended by Pennebaker (personal communication, October 30, 2018). LIWC analysis returned results for 90 output variables (i.e., LIWC categories), and LIWC category scores represented the percentage of total words used for the sample. Words could belong to more than one category.

2.6. Data Analysis

For the RQs 1-2, the raw count and percentage of total words used were reported for each variable. For RQs 3-4, the LL ratio test was used to compare the sub-corpuses to [33]. The Bayesian Information Criterion (BIC) was calculated to assess the strength of support for the alternative hypothesis over the null hypothesis [34]. BIC strength descriptors were drawn from [35]. Analyses were completed using R, and the alpha level was set at $p < .05$. Given the large number of hypothesis tests conducted, a Bonferroni correction was used to reduce the chance of a Type I error.

3. Results

In terms of RQ1 (NSSI word-usage rates), the most common NSSI-specific category was Reasons for NSSI (n = 2950) followed by NSSI Instruments Used (n = 652), Cutting-Specific Terms (n = 21), and Hidden-Hashtag Terms (n = 0). For RQ2 (LIWC English pronoun rates), the most frequently appearing pronoun was the 1st person singular (n = 971), followed by 3rd person plural (n = 52), 3rd person singular (n = 49), and 1st person plural (n = 25). With regard to RQ3 (English pronoun rates compared to a norm), the results were as follows: (a) 1st person singular, G2 = X, p = .033, BIC = X, BIC descriptor is "Very Strong"; (b) 3rd person plural,   G2 = X, p = .033, BIC = X, BIC descriptor is "Very Strong", (c)   3rd person singular G2 = X, p = .033, BIC = X, BIC descriptor is "Very Strong"),   and (d) 1st person plural, G2 = X, p = .033, BIC = X, BIC descriptor is "Very Strong". Concerning RQ4 (non-pronoun rates for all three corpuses compared to a norm), the BIC descriptor of "Very Strong" appeared with four categories across all three sub-corpuses: death, health, negative emotion, and sadness. One can review the complete analysis results in Table 1. The raw data is available in a supplemental table that can be found on this research project's Open Science Foundation website.

Table 1. Non-pronoun LIWC Results for All Sub-corpuses (RQ4).

Sub-corpus	Variable	G^2	p	BIC	BIC Descriptor
English	Anger	17.32	0	7.84	Strong
	Anxiety	17.32	0	7.84	Strong
	Body	3.32	0.0683	-6.15	Trivial
	Cognitive	8.14	0.0043	-1.34	Trivial
	Death	28.49	0	19.01	Very Strong
	Health	151.79	0	142.32	Very Strong
	Ingestion	0.7	0.4041	-8.78	Trivial
	Insight	1.72	0.1897	-7.75	Trivial
	Neg. Emo.	95.29	0	85.82	Very Strong
	Sadness	56.75	0	47.27	Very Strong
Spanish	Anger	105.08	0	94.34	Very Strong
	Anxiety	22.53	0	11.8	Very Strong
	Body	16.77	0	6.03	Strong
	Cognitive	171.44	0	160.71	Very Strong
	Death	126.84	0	116.11	Very Strong
	Health	153.38	0	142.65	Very Strong
	Ingestion	2.87	0.0902	-7.86	Trivial
	Insight	18.46	0	7.73	Strong
	Neg. Emo.	261.61	0	250.87	Very Strong
	Sadness	182.03	0	171.29	Very Strong
Mixed	Anger	457.68	0	448.76	Very Strong
	Anxiety	62.75	0	53.82	Very Strong
	Body	6.32	0.0119	-2.6	Trivial
	Cognitive	63.03	0	54.11	Very Strong
	Death	960.63	0	951.7	Very Strong
	Health	554.37	0	545.44	Very Strong
	Ingestion	6.76	0.0093	-2.17	Trivial
	Insight	3.01	0.0828	-5.91	Trivial
	Neg. Emo.	1256.38	0	1247.46	Very Strong
	Sadness	1133.16	0	1124.24	Very Strong

4. Discussion

This study examined the language used in public pro-NSSI blogs using Spanish and posted on Tumblr, focusing specifically on differences in the use of language in Spanish pro-NSSI blogs as compared to blogs overall. The results of the first RQ, which showed that reasons for NSSI were used most across sub-corpuses, can be explained by two possible explanations. The first is that words in the reasons for NSSI category (e.g., anger, anxiety, and sadness) are more associated with words suggestive of negative emotions. These negative-emotion words, particularly during the recall or recollection of traumatic incidents, arise more often when communicants use their dominant language (i.e., Spanish) [32], which is likely the case during posts about self-injury. The alternative explanation is that engaging in this community is a way to belong and feel understood among peers, facilitating discussions about the impulse to harm oneself. Between these two explanations, the

former is the most plausible because past research has demonstrated a relationship between self-injurious behavior and negative emotion-generating experiences such as bullying [36].

The results from the second and third RQs revealed that pronoun use differed from English Twitter norms most frequently and significantly in first-person singular, followed by third-person plural and first-person plural. Using the first-person singular pronoun could indicate depression and perceived lower social status [37]. Adolescents engaging in NSSI are at a higher risk for developing major depressive disorder and exhibiting suicidal behaviors [38], making this an issue of concern for parents and teachers.

The English sub-corpus included language taken from posts where Spanish had been used, which assumes users are bilingual and that they switched languages within the post. Based on the author's experience it is hypothesized that when bilingual individuals use their non-dominant language to discuss emotional issues, it can be a way for them to create distance from or avoid their emotional experience and associated vulnerability. Self-harm is a way to avoid directly confronting intense emotions, so bilingual individuals who engage in self-harm might code-switch when discussing their experience in order to prevent those same intense emotions.

The results from the fourth RQ revealed the psychological process word-use patterns across all three pro-NSSI sub-corpuses. The dominant category in Spanish sub-corpuses was "negative emotion," while in the English sub-corpus, "health" was more prominent. The distinction in these results could be due to the less emotionally charged nature of expressions in a non-dominant language (English) and the greater capacity for emotional expression in romance languages like Spanish. [39]. On a psycho-lexical level, which refers to how language reflects and shapes emotional understanding, these language groups are similar in usage patterns of emotion words, though cultural nuances can affect their connotations and frequency of use [39].

Prior research suggests that people who engage in self-harm are in immense pain and would be expected to use a negative emotional tone to express their sadness. Pennebaker [40] suggested that those who use negative emotional language may not benefit from writing about it and continue to dwell on it. Writing or posting on Tumblr, then, does not help individuals process these emotions; instead, it can contribute to a cycle where they carry on with the same feelings and behaviors. Excessive internet use and engagement with pro-NSSI content online can increase the risk of self-harm and suicidal behavior.

While these data are compelling, a few admitted limitations exist in this study. First, words can be polysemous, and words in LIWC could appear in more than one category, somewhat skewing the data. To control this potential and to avoid confusion, the data were reviewed before being analyzed by LIWC software. Second, only public Tumblr posts from North, Central, and South American countries were considered in this study due to variations in laws regarding free and public access to data amongst additional potential locations. People who post on Tumblr may be doing so through a private account, but they could also be Spanish speakers who do not happen to reside in countries within North, Central, or South America. This combination of factors could fail to account for the entirety of the population. Third, the meanings of words can vary across languages. This study used a Spanish-language corpus, but it also relied on Spanish, English, and mixed-language sub-corpuses because of the phenomenon of code-switching. Also, there were no Spanish or mixed-language blog norms to use for comparison. As such, while the effect size for this study was large, inferences drawn from these data should be colored by the above concerns.

This study reveals several implications for practitioners and researchers. First, this research suggests which lines of inquiry about specific NSSI behaviors and online activities should be pursued when discussing online behaviors with adolescents. These could include examples of posts made online, examples of sites clients visit, or discussions of where clients learn about NSSI or who they are talking to about it. Asking clients about NSSI and online behaviors is a necessary skill for clinical mental health counselors because these may not be topics adolescent clients would bring up on their own, and standard assessment questions might not cover these topics. There is a need for a balanced approach that leverages social media's positive aspects while mitigating its risks.

Additionally, this study points to the need to consider bilingualism with clients with NSSI issues. These clients often feel isolated and might be reaching out to online communities where they are not shamed and where they might feel more understood.   Social media plays a role in both mitigating and exacerbating NSSI risk [41]. The language of their distress could be seen in the frequency of their negative emotional tone or in their sadness. The shortage of bilingual or Spanish-speaking counselors further highlights the urgency of this implication.

A second implication for counselors would involve the Stages of Change (or Transtheoretical) Model [42], specifically allowing practitioners to incorporate better information about linguistic markers in pro-NSSI and recovery-oriented posts. By monitoring clients' language use to indicate stages of change and target interventions for progression more accurately, counselors can provide tailored and individualized treatment to increase recovery rates and lengths of wellness.

Further research must include expanded data collection that includes data from a wider range of social media platforms and at multiple intervals of time.   Longitudinal studies might provide insights into how pro-NSSI related language evolves with cultural and social changes [16].   Further, in-depth linguistic analysis could explore how factors such as age, gender, socioeconomic status, and sexual orientation intersect with language patterns in the pro-NSSI discussions.   Discursive practices in virtual communities form an online identity, express emotions and opinions, and establish communication codes within groups. This indicates the need for adolescents to be involved in co-designing interventions that respect adolescents' preferences and privacy while also addressing NSSI [43].

5. Conclusions

With increased time spent on online platforms following COVID-19 and the growing number of Spanish-speaking adolescents globally, practitioners must continue to explore the discursive practices of pro-NSSI communities and must adapt in new ways to responsibly serve these individuals. Software such as LIWC can reveal psychological processes and patterns of behavior around a variety of mental health issues, and these results can show parents, teachers, and counselors how to intervene, perhaps also shedding insight into the process of recovery. This study suggests that language plays a critical role in expressing NSSI-related emotions and that the interplay of bilingualism introduces complexities that require further investigation. Future studies should explore how bilingual NSSI discourse varies across different emotional contexts and settings.

References

1. Boukhalfi, B., & Amir, M. Interactions discursives et tournures récurrentes dans le discours numérique: Cas des interactions sur Facebook. *Akofena, Varia*, **2022,** 12(3), 127-152.

2. Calvo, N., García-González, S., Perez-Galbarro, C., Regales-Peco, C., Lugo-Marin, J., Ramos-Quiroga, J.A., & Ferrer, M. Psychotherapeutic interventions specifically developed for NSSI in adolescence: A systematic review. *Eur Neuropsychopharmacol*. **2022** 58:86-98. https://www.doi.org/10.1016/j.euroneuro.2022.02.009.

3. Fruttaldo, A. Family portrait: A corpus-based analysis of the discursive construction of traditional families. *de genere: Rivista di studi letterari, postcoloniali e di genere / Journal of Literary, Postcolonial and Gender Studies*. **2024** Retrieved from http://www.degenere-journal.it/

4. Đorđević, J. Digital media discourse in linguistic research. *Faculty of Philosophy in Niš*. **2022** https://doi.org/10.46630/dmd.2022

5. Matsakis, L. How pro-eating disorder posts evade filters on social media **2018**, Retrieved from https://www.wired.com/story/how-pro-eating-disorder-posts-evade-social-media-filters/

6. Moreno, M. A., Ton, A., Selkie, E., & Evans, Y. Secret society 123: Understanding the language of self-harm on Instagram. Journal of Adolescent Health, **2016** 58(1), 78–84. http://dx.doi.org/10.1016/j.jadohealth.2015.09.015

7. Favazza, A. R. The coming of age of self-mutilation. Journal of Nervous and Mental Disease. **1998** 186:259–268.

8. Kerr, P., Muehlenkamp, J., & Turner, J. Nonsuicidal self-injury: A review of current research for family medicine and primary care physicians. *Journal of the American Board of Family Medicine* **2010** 23(2), 240–259.https://doi.org/10.3122/jabfm.2010.02.090110

9. Farkas, B.F., Takacs, Z.K., Kollárovics, N., & Balazs, J. The prevalence of self-injury in adolescence: a systematic review and meta-analysis. *Eur Child Adolesc Psychiatry* **2023** https://doi.org/10.1007/s00787-023-02264-y

10. Esposito, C., Dragone, M., Affuso, G., Amodeo, A. L., & Bacchini, D. Prevalence of engagement and frequency of non-suicidal self-injury behaviors in adolescence: An investigation of the longitudinal course and the role of temperamental effortful control. *European Child & Adolescent Psychiatry*, **2023** 32, 2399-2414. https://doi.org/10.1007/s00787-022-02083-7

11. American Psychiatric Association. Diagnostic and statistical manual of mental disorders **(2022)** (5th ed., rev. ed.). APA. 10.1176/appi.books.9780890425787

12. Instituto Cervantes. Spanish, a growing global language. **2023** Retrieved from [Cervantes Observatorio](https://cervantesobservatorio.fas.harvard.edu/en/about/spanish-in-united-states)

13. Pew Research Center. Future projections of Spanish speakers in the U.S. **2023** Retrieved from [Pew Research Center](https://www.pewresearch.org/hispanic/)

14. Rojas-Velasquez, D.A., Pluhar, E.I., Burns, P.A., Burton, E.T.. Nonsuicidal Self-Injury Among African American and Hispanic Adolescents and Young Adults: a Systematic Review. *Prev Sci.* **2021**, Apr;22(3):367-377. doi: 10.1007/s11121-020-01147-x. PMID: 32671672.

15. Gulbas, L. E., Hausmann-Stabile, C., De Luca, S. M., Tyler, T. R., & Zayas, L. H. An exploratory study of non-suicidal self-injury and suicidal behaviors in adolescent Latinas. *American Journal of Orthopsychiatry*, **2015** 84(4), 302–314. http://dx.doi.org/10.1037/ort0000073

16. Madubata, I. J., Cheref, S., Eades, N. D., Brooks, J. R., Talavera, D. C., & Walker, R. L. Non-Suicidal Self-Injury, Neuroticism, and Ethnic Identity in Young Latina Adults. *Hispanic Journal of Behavioral Sciences* **2020** 42(4), 528-546. https://doi.org/10.1177/0739986320950914

17. Orsolini, L., Reina, S., Longo, G., & Volpe, U. "Swipe & slice": Decoding digital struggles with non-suicidal self-injuries among youngsters. *Frontiers in Psychiatry,* **2024**, *15*. https://doi.org/10.3389/fpsyt.2024.1403445

18. Chancellor, S., Pater, J., Clear, T., Gilbert, E., & De Choudhury, M. #thyghgapp: Instagram content moderation and lexical variation in pro-eating disorder communities. Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, **2016** 1201–1213. https://doi.org/10.1145/2818048.2819963

19. DeAngelis, T. Teens are spending nearly 5 hours daily on social media. Here are the mental health outcomes. **2024,** Retrieved from: https://www.apa.org/monitor/2024/04/teen-social-use-mental-health

20. Lewis, S. P., Kenny, T. E., Pritchard, T. R., & Brewer, A. Self-injury during COVID-19: Views from university students with lived experience. *The Journal of Nervous and Mental Disease*, **2022** 210(11), 849-856. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9616563/

21. John, A., Eyles, E., Webb, R. T., Okolie, C., Schmidt, L., McGuinness, L. A., & Arensman, E. The impact of the COVID-19 pandemic on self-harm and suicidal behaviour: update of living systematic review. *Journal of Affective Disorders*, **2020** 277, 1083-1092. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7871358/

22. Time. AI is turning social media into the next frontier for suicide prevention. **2024**, *Time*. https://time.com/6696703/ai-suicide-prevention-social-media/

23. Pew Research Center. Teens, social media and technology **2022,** Retrieved from https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/

24. Brown, R. C., Fischer, T., Goldwich, A. D., Keller, F., Young, R., & Plener, P. L. #cutting: Non-suicidal self-injury (NSSI) on Instagram. *Psychological Medicine*, **2018** 48(2), 337–346. https://doi.org/10.1017/S0033291717001751

25. Chancellor, S., Lin, Z., Goodman, E. L., Zerwas, S., & De Choudhury, M. Quantifying and predicting mental illness severity in online pro-eating disorder communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* **2016,** (pp. 1171–1184). Association for Computing Machinery. https://doi.org/10.1145/2818048.2819973

26. Weisser, M. Practical corpus linguistics: An introduction to corpus-based language analysis. **2016,** Hoboken, NJ: John Wiley & Sons, Inc.

27. Bjekić, J., Lazarević, L. B., Živanović, M., & Knežević, G. Psychometric evaluation of the Serbian dictionary for automatic text fanalysis-LIWCser. Psihologija, **2014**, 47, 5–32. https://doi.org/10.2298/PSI1401005B

28. Tumblr. Tumblr API documentation. Retrieved July 13, 2024, from https://www.tumblr.com/docs/en/api/v2

29. Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. The development and psychometric properties of LIWC2015. **2015** Austin, TX: Pennebaker Conglomerates, Inc.

30. Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E.. Linguistic Inquiry and Word Count: LIWC2015 (Operator's Manual) **2015** Austin, TX: Pennebaker Conglomerates, Inc.

31. Greaves, M. M., & Dykeman, C. Linguistic Analysis of Nonsuicidal Self-Injury Reddit Posts: Implications for Family Therapy. The Family Journal, **2024,** https://doi.org/10.1177/10664807241248177

32. Bailey, C., McIntyre, E., Arreola, A., & Venta, A. What are we missing? How language impacts trauma narratives. *Journal of Child & Adolescent Trauma*, **2019**, 1–9. https://doi.org/10.1007/s40653-019-00263-3

33. Rayson, P., & Garside, R. Comparing corpora using frequency profiling. Proceedings of the Workshop on Comparing Corpora, **2000**, 9, 1–6. http://dx.doi.org/10.3115/1117729.1117730

34. Neath, A., & Cavanaugh, J. The Bayesian information criterion: Background, derivation, and applications. Wiley Interdisciplinary Reviews: Computational Statistics. **2012**, https://doi.org/10.1002/wics.199

35. Brezina, V. Statistics in corpus linguistics: A practical guide. **2018,** Cambridge, UK: Cambridge University Press.

36. Espelage, D. L., Merrin, G. J., & Hatchel, T. Peer victimization and dating violence among LGBTQ youth: The impact of school violence and crime on mental health outcomes. *Youth Violence and Juvenile Justice,* **2018,** *16*(2), 156-173. https://doi.org/10.1177/1541204016680408

37. Chung, C., & Pennebaker, J. The Psychological Functions of Function Words. In K. Fiedler (Ed.), *Social communication* **2007,** (pp. 343–359). Psychology Press.

38. Turner, B. J., Layden, B. K., & Chapman, A. L. Non-suicidal self-injury, major depressive disorder, and suicidal behavior: A longitudinal study. *Journal of Clinical Psychology*, **2023,** 79 (4), 852-868. https://doi.org/10.1002/jclp.23334

39. Pavlenko, A. Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition, 2008*, 11(2), 147–164. https://doi.org/10.1017/S1366728908003283

40. Pennebaker, J. W. The secret life of pronouns: What our words say about us. **2011,** New York, NY: Bloomsbury Press.

41. Denton, E., & Álvarez, K. The global prevalence of nonsuicidal self-injury among adolescents. *JAMA Network Open*, **2024**, 7(6), e2415406. https://doi.org/10.1001/jamanetworkopen.2024.15406

42. Prochaska, J. O., & DiClemente, C. C. Stages and processes of self-change of smoking: Toward an integrative model of change. Journal of Consulting and Clinical Psychology,   **1983,** 51(3), 390–395.

43. Laestadius, L. I., Craig, K. A., & Campos-Castillo, C. Social media use and mental health: A study of Latinx adolescents. *Journal of Medical Internet Research*, 23, **2021**, (8), e28931. https://doi.org/10.2196/28931