

EMPIRICAL ARTICLE

Using social media to explore the linguistic features in female adults with childhood sexual abuse by Linguistic Inquiry and Word Count

Wenyu Wan¹  | Jiumo Sun² | Jiehan Liu³ | Shu-Wen Yang² | Mingming Liu² | Jia Xue⁴  | Dongdong Jiao⁵ | Xiaoqian Liu⁶

¹Department of Psychology, University of Chinese Academy of Sciences, CAS Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

²Department of Psychology, University of Chinese Academy of Sciences, CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

³Department of Psychology, University of Chinese Academy of Sciences, CAS Key Laboratory of Cognition and Development Psychology, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

⁴Factor-Inwentash Faculty of Social Work & Faculty of Information, University of Toronto, Toronto, Ontario, Canada

⁵Engineer in Basic Technology Laboratory, National Computer System Engineering Research Institute of China, Beijing, China

⁶CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

Correspondence

Xiaoqian Liu, Institute of Psychology, Chinese Academy of Sciences; Address: 16 Lincui Road, Chaoyang District, Beijing 100101, China. Email: liuxiaoqian@psych.ac.cn

Funding information

China Social Science Fund, Grant/Award Number: 17AZD041

Abstract

The adverse impact of childhood sexual abuse experience on a person's physical and mental health is long-lasting. The disadvantageous influence can be reflected in the language expression even if they grow up, especially when the language is not monitored intentionally by the speaker. However, few researchers have focused on the language expression characteristics of this group. This study aims to analyze the message of social media to explore the difference of language expression between adult females with childhood sexual abuse experience (CSA group) and adult women without such experience (control group) by Linguistic Inquiry and Word Count (LIWC). We collected 46 victims (all females) and 46 nonvictims (gender-matched with CSA group) on Sina Weibo, and we applied LIWC to encode and count all the text messages posted on the social platforms. The results of this research suggested that the CSA group differed from the control group in multiple indicators, especially in psychological process words. The victims were less likely to refer to psychological process words, such as body words, sex words, etc. than the nonvictims, however, they preferred to mention human words. Moreover, compared to the control group, the CSA group had published fewer contents and used fewer words that represent the present tense in the social media platforms. The present study provides the research basis for identifying the CSA group in social media platforms in the future.

KEYWORDS

childhood sexual abuse, language expression, LIWC, social media platforms

1 | INTRODUCTION

Childhood sexual abuse (CSA) can be defined as any activity for the sexual satisfaction of an adult with a child before the age of legal consent (Baker & Duncan, 1985; Wright, Fraser, Denman, & Duke, 1987; Johnson, 2004), which is a worldwide problem. It is estimated that 8~31% of women and 3~17% of men have experienced CSA in their

lifetime (Barth, Bermetz, Heim, Trelle, & Tonia, 2013), and the rate for women and men with CSA in China was 10.8 and 4.8%, respectively (Pereda, Guilera, Forns, & Gómez-Benito, 2009).

CSA experience is harmful, grievous, and chronic to survivors, extending from early childhood to adulthood, individual function to social function, physical health to mental health (Fullilove, 2009). In the past two decades, a large number of researchers focused on CSA

group's mental health. They found that CSA experience is linked to increased severe mental disorder including anxiety (Field et al., 2016; Gena, Alessandra, & Danielle, 2013), depression (Chen et al., 2012; Kendler & Aggen, 2014; Musliner & Singer, 2014), post-traumatic stress disorder (PTSD) (Chang, Kaczurkin, Mclean, & Foa, 2018; Meston, Lorenz, & Stephenson, 2013; Polusny & Follette, 1995), and increased higher psychiatric comorbidity including substance abuse, and more suicide attempts (Decou, Lynch, Dehart, & Belknap, 2016; Lee, Lyvers, & Edwards, 2009). It is noteworthy that the number of female CSA victims suffering from depression is twice as many as that of the general population (Molnar, Buka, & Kessler, 2001). Besides, to some extent, this experience leads to different language expression compared with those without CSA (Lorenz & Meston, 2012; Pulverman, Lorenz, & Meston, 2015).

Linguistic expression is related to mental health. The psychological impacts of CSA are imprinted in a person's language (Lorenz & Meston, 2012; Stanton, Meston, & Boyd, 2017). Much research confirmed that early childhood traumatic experiences and consequent mental health problems are reflected in daily language use, particularly in language elements (Pennebaker, Mehl, & Niederhoffer, 2003). Lorenz and Meston (2012) found significant differences between women with and without CSA, such as different rates of using emotional words and words related to facts. As there exists correlation between language use and mental health, in both adults (Boals & Klein, 2005; Hovey, Kim, & Seligman, 2006) and teenagers (Rubens et al., 2013; Simşek & Cerçi, 2013), therefore, it is necessary to investigate the language expression of CSA victims.

The traditional linguistic study method adopts an expressive writing paradigm, that is, the experimenter recruited subjects to the lab and asked them to write a message to describe a certain issue. Recently, Coppersmith, Dredze, Harman, and Hollingshead (2015) examined a broad range of mental health conditions in Twitter data by identifying self-reported statements of diagnosis. It gives us an inspiration that we can access the survivors' language characteristics through texts they published in social media, Sina Weibo, which is one of the most popular social platforms in China. In this platform, users present individual detailed information (e.g., gender, age, relationship status, university, etc.) in their profiles and share what they saw, what they heard, and what they thought by messages or interact with their friends or the public in forms of sentences or pictures, launched by Sina company in China. Compared to the traditional study method, the new approach has many advantages. First, up to June 2018, Sina Weibo's monthly active users has reached 430 million according to the Sina company official report (<http://data.weibo.com/report/>), which means that Sina Weibo has a huge user base. Second, Sina Weibo's data that we obtained from the social media platforms is open-sourced, so we can recruit subjects who meet the experiment requirements easily. Finally, we can gain more ecological validity messages. Psychoanalysis and theory of mindfulness believed that early sexual trauma is reflected in language usage, especially when the language is usually not monitored intentionally by the speaker (Pennebaker & Stone, 2003).

Voluntarily occurring language expression data from social media can eliminate influence originating from social expectation and bias from experimenters in previous laboratory studies (Lorenz & Meston, 2012; Pulverman et al., 2015; Rellini & Meston, 2007). In summary, the social media platform, Sina Weibo, provides a great opportunity for researchers to acquire social media users' linguistic expressions more conveniently and comprehensively. This study aims to assess the characteristics of CSA victims' language expression on social media, and further to explore the language expression differences between CSA group and non-CSA group by using LIWC (Linguistic Inquiry and Word Count) (Pennebaker, Chung, Ireland, Gonzales, & Booth, 2015). This study can help us better understand the impact of CSA on language expression and the psychological process ecologically. And the present study provides the research basis for identifying the CSA group in social media platforms in the future.

2 | METHODS

2.1 | Sample

2.1.1 | CSA group

First, we manually selected 50 victims of sexual assault among over 300 million Weibo users. To verify that each selected subject was a user who had been sexually assaulted in childhood, we used manual searches to select subjects. The main search method as follows: the first approach is that we used the keyword "sexually assaulted" ("性侵" in Chinese) as a search term to fetch microblogs by using Weibo search and checked manually whether the microblog content was about the user's own sexual experience in childhood. The other way is that we found celebrity users who published definitely that she had ever been suffered from sexual abuse in childhood on Sina Weibo and then manually found CSA victims from the comments under the microblogs. The people that published those microblogs or comments (like "when I was 12 years old, I suffered from sexual abuse by my relative") were initially included in our study. Next, we crawled the individual information including gender, age, hometown, etc., Weibo behavior characteristics including the number of fans, mutual fans, followers, etc., and all the posts of these users. Next, we checked whether the CSA case met the inclusion criteria. The inclusion criteria were that (a) these users' original microblogs or comments contained contents that expressed their own CSA experience; (b) female user; (c) older than 18; (d) individual account rather than an organization. Finally, a total of 46 users met all the requirements.

2.1.2 | Control group

The control group members needed to meet the above criteria but had no descriptions of the experience of being sexually assaulted in their microblogs. To identify the differences, we selected 46 nonvictim Weibo users paired with the case group using gender as the control group from the pool of Weibo users. Then we crawled their individual

TABLE 1 Individual profile of the sample of CSA survivors and non-CSA users on Sina Weibo

	Group	N	M	SD	t	p
Verified	CSA	46	1.978	1.474	0.00	1.000
	Non-CSA	46	1.978	1.474		
Gender	CSA	46	2.000	0		
	Non-CSA	46	2.000	0		
Follower	CSA	46	550.783	1,737.433	−0.50	0.617
	Non-CSA	46	688.565	671.111		
Follow	CSA	46	377.717	523.896	−0.54	0.588
	Non-CSA	46	430.935	409.140		
Mutual fans	CSA	46	16.891	56.008	−4.220	0.000
	Non-CSA	46	138.978	188.053		
Weibo number	CSA	46	1105.913	1,735.159	−3.775	0.000
	Non-CSA	46	2646.304	2,155.783		

information, Weibo behavior characteristics, and posts. Table 1 features the demographic profiles of these groups.

2.2 | Measurement

We crawled the 92 Weibo users' data through the Application Programming Interface (API) provided by Sina Weibo. After obtaining the online data of the users, we extracted the features of each user from two aspects: LIWC features (see Gao, Hao, Li, Gao, & Zhu, 2013; Zhao, Jiao, Bai, & Zhu, 2016, for more details) and behavioral features.

LIWC is based on the word count strategy, taking into account both contents and forms of words. Nowadays, a study found that LIWC has good structural validity (Zhao et al., 2016) and was widely used to explore the cognitive, emotional, and structural components behind natural language by psychologists (Pennebaker et al., 2003). Pennebaker and his team devoted themselves to create and develop the LIWC for a long time (Pennebaker et al., 2003; Pennebaker et al., 2015; Pennebaker, Booth, & Francis, 2007; Pennebaker & Chung, 2007; Pennebaker & King, 1999; Pennebaker & Stone, 2003), and developed different language version based on LIWC English version (www.liwc.net). The Chinese version is greatly different from the English version, therefore, Chinese dictionaries are independent. We used a Chinese language psychological analysis system called "Textmind" (Gao et al., 2013) to extract LIWC features. It segments the sentences into words and classifies them into separate dimensions according to the dictionary, which totally has 88 features and was divided into three major categories including general descriptor, linguistic processes, and psychological processes. Those three categories were divided into many subcategories in the meantime (see Supporting Information, Data S1 for more details). After checking all the vocabulary in the text, it calculates the ratio of vocabulary occurrences in each of the 88 categories to the total number of words in the text. These ratios are the LIWC features of the users. Besides, we also calculated 11 behavioral features including the number of words, the average number of words per sentence, number ratio, etc.

Finally, we put all the calculated psychological indices into SPSS is Statistical Product and Service Solutions 15.0 for advanced statistics to compare the case group and the control group by independent sample *t* test.

3 | RESULTS

3.1 | Group differences in general descriptor

There were significant differences in word count ($t = -2.205$, $p = .031 < .05$, $d = -0.46$), rate dictionary cover ($t = -2.708$, $p = .009 < .01$, $d = -0.56$), RateNumeral ($t = -2.967$, $p = .004 < .01$, $d = -0.62$), and punctuations between CSA group and control group. Women in the CSA group used fewer words, Arab numbers, and punctuations than women in the control group (as the Table 2 and Figure 1 show).

3.2 | Group differences in linguistic processes

In linguistic processes, women in CSA group used less present tense markers (e.g., now, at present) than women in control group ($t = -2.110$, $p = .038 < .05$, $d = -0.46$), and there were no significant differences in function words (e.g., maybe, those), pronoun words (e.g., I, you, he, she), and other tense makers (see Table 3 for details).

3.3 | Group differences in psychological processes

As shown in Table 4 and Figure 2, in psychological processes, there were significant differences in the social process, the emotional process, and the biological process. In the social process, women in the CSA group used more human words (e.g., people, the masses) than women in the control group ($t = 2.521$, $p = .015 < .05$, $d = 0.53$). In emotional process, women in CSA group used less negative emotion words (e.g., anxious, jealous) ($t = -2.726$, $p = .008 < .01$, $d = -0.57$), and anxiety words (e.g., nervous, mad) ($t = -3.667$, $p < .000$, $d = -0.75$) than women in the control group. In biological process,

Dimension	Groups	M	SD	t	p	Cohen's d
WordCount	CSA (N = 46)	355.50	851.97	-2.205	0.031	-0.46
	CG (N = 46)	677.30	503.49			
RateDicCover	CSA (N = 46)	0.5751	0.1479	-2.708	0.009	-0.56
	CG (N = 46)	0.6406	0.0711			
RateNumeral	CSA (N = 46)	0.0045	0.0055	-2.967	0.004	-0.62
	CG (N = 46)	0.0073	0.0032			
Period	CSA (N = 46)	0.0089	0.0090	-3.533	0.001	-0.74
	CG (N = 46)	0.0152	0.0080			
Comma	CSA (N = 46)	0.0063	0.0077	-4.603	0.000	-0.96
	CG (N = 46)	0.0134	0.0070			
SemiC	CSA (N = 46)	0.0002	0.0004	-2.368	0.020	-0.53
	CG (N = 46)	0.0005	0.0007			
QMark	CSA (N = 46)	0.0163	0.0021	-2.532	0.014	8.07
	CG (N = 46)	0.0025	0.0012			
Dash	CSA (N = 46)	0.0013	0.0019	-2.974	0.004	-0.67
	CG (N = 46)	0.0025	0.0017			
Parenth	CSA (N = 46)	0.0033	0.0031	-12.163	0.000	-2.54
	CG (N = 46)	0.0367	0.0183			
OtherP	CSA (N = 46)	0.0373	0.0341	-2.843	0.006	-0.59
	CG (N = 46)	0.0551	0.0254			

TABLE 2 Group differences in general descriptor between CSA group and CG group

Abbreviations: CG, control group; CSA, children sexual abuse; OtherP, other punctuation mark; Parenth, parenthesis; QMark, question mark; SemiC, semicolon.

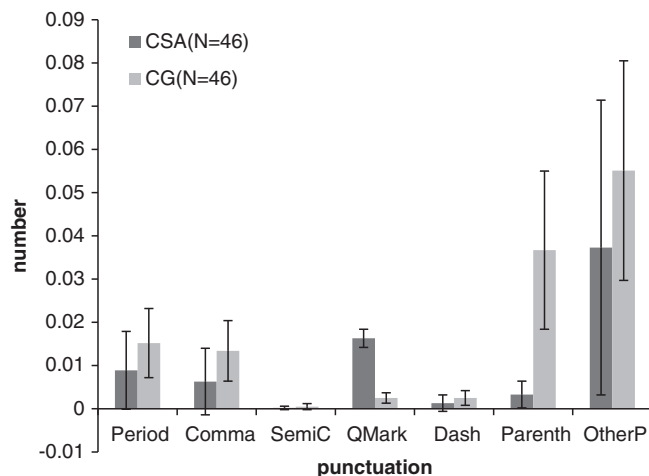


FIGURE 1 The number of different punctuation in the CSA group and the CS group

women in the CSA group used less body words (e.g., neck, skin) ($t = -4.055, p < .000, d = -0.86$), sexual words (e.g., make love, sexuality) ($t = -2.543, p = .013 < .05, d = -0.55$), and ingestion words (e.g., eat, cook) ($t = -2.163, p = .034 < .05, d = -0.45$) than women in

the control group. There were no significant differences in cognitive processes and percept process.

3.4 | Group differences in Weibo behavior characteristics

There are some differences in Weibo behavior characteristics between people in the CSA group and the control group. Women in the CSA group mentioned less other people ($t = -2.816, p < .01$), and URLs than woman in the control group ($t = -2.761, p < .01$). And they also mentioned less emotional words than women in the control group ($t = -7.099, p < .000$) (see more details in Table 5).

4 | DISCUSSION

The present study showed that CSA survivors indeed have different language expressions from those without CSA history. To be specific, there are significant differences in many indicators in the overall description, language process, and especially the psychological process. At the same time, we also found that there were differences in

Dimension	Groups	M	SD	t	p	Cohen's d
PresentM	CSA (N = 46)	0.0033	0.0017	-2.110	0.038	-0.46
	CG (N = 46)	0.0040	0.0013			

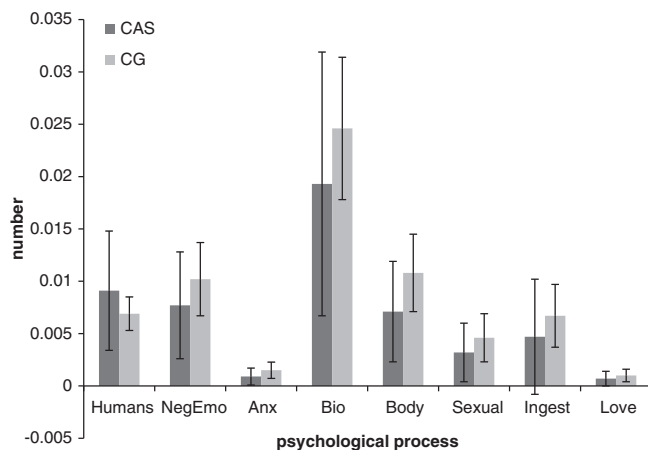
TABLE 3 Group differences in linguistic processes between CSA group and CG group

Abbreviations: CG, control group; CSA, children sexual abuse; PresentM, present tense marker.

TABLE 4 Group differences in psychological processes between CSA group and CG group

Dimension	Groups	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
Humans	CSA (<i>N</i> = 46)	0.0091	0.0057	2.521	0.015	0.53
	CG (<i>N</i> = 46)	0.0069	0.0016			
NegEmo	CSA (<i>N</i> = 46)	0.0077	0.0051	-2.726	0.008	-0.57
	CG (<i>N</i> = 46)	0.0102	0.0035			
Anx	CSA (<i>N</i> = 46)	0.0009	0.00081	-3.667	0.000	-0.75
	CG (<i>N</i> = 46)	0.0015	0.00078			
Bio	CSA (<i>N</i> = 46)	0.0193	0.0126	-2.472	0.016	-0.52
	CG (<i>N</i> = 46)	0.0246	0.0068			
Body	CSA (<i>N</i> = 46)	0.0071	0.0048	-4.055	0.000	-0.86
	CG (<i>N</i> = 46)	0.0108	0.0037			
Sexual	CSA (<i>N</i> = 46)	0.0032	0.0028	-2.543	0.013	-0.55
	CG (<i>N</i> = 46)	0.0046	0.0023			
Ingest	CSA (<i>N</i> = 46)	0.0047	0.0055	-2.163	0.034	-0.45
	CG (<i>N</i> = 46)	0.0067	0.0030			
Love	CSA (<i>N</i> = 46)	0.0007	0.0007	-2.861	0.005	-0.46
	CG (<i>N</i> = 46)	0.0010	0.0006			

Abbreviations: Anx = anxiety; Ingest, ingestion; CG, control group; CSA, children sexual abuse; NegEmo, negative emotion.

**FIGURE 2** The number of words used in the different psychological process in CSA group and CS group

the microblog behavior characteristics between the CSA group and the control group.

TABLE 5 Group differences in Weibo behavior characteristics

Dimension	Groups	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
NumAtMention	CSA (<i>N</i> = 46)	4.9634	13.369	-2.816	0.006	-0.59
	CG (<i>N</i> = 46)	14.6184	19.029			
NumEmotion	CSA (<i>N</i> = 46)	0.3140	0.8461	-7.099	0.000	-1.48
	CG (<i>N</i> = 46)	11.1241	10.2938			
NumURLs	CSA (<i>N</i> = 46)	0.0296	0.0965	-2.761	0.008	-0.57
	CG (<i>N</i> = 46)	1.3039	3.1289			

Abbreviations: CG, control group; CSA, children sexual abuse; NumAtMention, numbers of users mentioned; NumEmotion, emotion words number; NumURLs, URLs number.

In terms of the words count in the overall description, we can see that the CSA group was much smaller than the control group, and the number of emotional words also showed the same result pattern. Previous studies had found that people who had a traumatic experience have significantly fewer words in their language than those who did not have a traumatic experience (Fernandez-Lansac & Crespo, 2015; Huemer et al., 2016). Maybe the traumatic experiences made people have fewer verbal expressions. People with PTSD lack specific, clear, and emotional descriptions of events and therefore show fewer language expressions (Huemer et al., 2016).

In the view of linguistic processes, the CSA group made use of less present tense than the control group, and there were no difference in the past tense and future tense between these two groups. Previous studies showed that the CSA group applied less present tense (Beaudreau, 2007), which is to some extent consistent with the results of our research. Arntz, Hawke, Bamelis, Spinhoven, and Molendijk (2012) found compared with the control group, personality disorders (PDs) applied less present tense in short essays they wrote about their lives. In addition, studies had used the increase usage of present tense

as an indicator of intervention benefits (Burke & Bradley, 2006; Moore & Brody, 2009), because researchers believed that the raising of the current tense means the individuals pay more attention to the moment and reduced complaint about what is going on now.

The present study suggested that there exist significant differences between the CSA group and the control group in the psychological process, such as physiological process words, sex words, body words, etc. People with traumatic experiences are more likely to mention human words than the control group. Currently, scholars were rarely concentrated on "human categories" of the LIWC (Tausczik & Pennebaker, 2010). Some researchers believed that human words are the expression of empathy (Block-Lerner, Adair, Plumb, Rhatigan, & Orsillo, 2007). We explain that people with traumatic experiences have more empathy than the control group because they have suffered misfortunes. Unlike previous studies (Martino, Onorato, & Freda, 2015), our study found the CSA group was less likely to mention negative emotional and anxiety words. And the former also mentioned that the CSA group tended to have less verbal expression including emotional words expression. In terms of physiological process words, such as sex words, body words, etc., the CSA group mentioned them less than the control group did. Although there were a few previous studies whose results were consistent with us, a considerable number of studies whose results were inconsistent with us (Rellini & Meston, 2011). Their results displayed that the CSA group mentioned sexual words more than the control group or at least they have the same tendency (Lorenz & Meston, 2012; Pulverman et al., 2015). The conflicts with some previous results may be caused by cultural difference, which exactly highlight the value of analyzing the CSA victims in China's online social media platforms. It was not difficult to understand that the CSA group was more sensitive to sexuality and therefore intend to suppress their expressions in this regard. This may be a self-regulation and self-defense to limit the recall of traumatic events.

Why do we pay more attention to the linguistic expression difference instead of psychological features directly between female adults with and without CSA history? The reasons followed: although it has been a long time since the traumatic event took place and many of those female victims were not suffering from mental illnesses, these experiences may still affect their cognition, emotion, and behavior unconsciously. These subtle negative effects were not easily detected; however, the analysis of language expression can recognize this delicate difference. More and more researchers realized the importance of the linguistic expressions, so on the one hand, they intentionally instructed the CSA survivors to change their existing expression patterns to promote their mental health, on the other hand, they assessed the intervention effect.

With the development of the Internet and the prevalence of social media, people are willing to share their daily life on social media. Sina Weibo has an enormous user base, which provides a mass potential subjects pool. Although the sample can only represent the group of adult female microblog users with CSA history and maybe it can not represent all the female adults who experienced sexual maltreat during their minority, we think the previous studies still have practical

significance because net citizens are indeed a large group. If we can realize the linguistic expression features of users with CSA history, on this basis we will develop a recognition system to recognize the USA users and intervene on occasion in the future.

This study has some limitations. First of all, although we meet the minimum sample size requirement of the RCT, objectively the sample size is small, beyond that, perhaps, there exists deviation from the overall sample. Second, online analysis is limited on which it is impossible to obtain some of the demographic variables of the subjects, such as the number of years of education and the age suffered from abuse. Therefore, we cannot control the influence of additional variables on language expression. Studies have shown that education level influence language expression (Sarah & Boris, 2015), while studies have shown that language expression is not affected by educational level (Jarrold et al., 2011; Lorenz & Meston, 2012).

Although there exist several limitations, the present research is the first to use social media data to investigate the language expression and behaviors of female CSA victims in China. There is no doubt that using the messages of the Weibo (the self-disclosure of social media paradigm) that CSA users published is of ecological validity rather than using the message that CSA subjects were requested to write following the experimental instructions in a lab (the expressive writing paradigm). To some extent, it avoids the self-reporting bias of the participants, and it avoids the secondary damage to the crowd as well. The victims and survivors were allowed to use nicknames instead of real names to discuss such private topics and conversations in social media platforms. So, it is a good chance for researchers to obtain texts to explore the features of the linguistic expressions of the CSA female adults conveniently and comprehensively. As the results of the present study, the CSA survivors have their own specific linguistic expression feature patterns. It is feasible for us to confirm their identity of CSA survivors in the future and pay more attention to them to provide psychological intervention timely.

ACKNOWLEDGMENTS

This study is supported by China Social Science Fund (17AZD041).

ORCID

Wenyu Wan  <https://orcid.org/0000-0001-6618-1893>

Jia Xue  <https://orcid.org/0000-0002-1668-2531>

REFERENCES

- Arntz, A., Hawke, L. D., Bamelis, L., Spinhoven, P., & Molendijk, M. L. (2012). Changes in natural language use as an indicator of psychotherapeutic change in personality disorders. *Behaviour Research and Therapy*, 50(3), 191–202.
- Baker, A. W., & Duncan, S. P. (1985). Child sexual abuse: A study of prevalence in Great Britain. *Child Abuse & Neglect*, 9(4), 457–467.
- Barth, J., Bermetz, L., Heim, E., Trelle, S., & Tonia, T. (2013). The current prevalence of child sexual abuse worldwide: A systematic

- review and meta-analysis. *International Journal of Public Health*, 58 (3), 469–483.
- Beaudreau, S. A. (2007). Are trauma narratives unique and do they predict psychological adjustment? *Journal of Traumatic Stress*, 20(3), 353–357. <https://doi.org/10.1002/jts.20206>
- Block-Lerner, J., Adair, C., Plumb, J. C., Rhatigan, D. L., & Orsillo, S. M. (2007). The case for mindfulness-based approaches in the cultivation of empathy: Does nonjudgmental, present-moment awareness increase capacity for perspective-taking and empathic concern? *Journal of Marital and Family Therapy*, 33(4), 501–516. <https://doi.org/10.1111/j.1752-0606.2007.00034.x>
- Boals, A., & Klein, K. (2005). Word use in emotional narratives about failed romantic relationships and subsequent mental health. *Journal of Language and Social Psychology*, 24(3), 252–268.
- Burke, P. A., & Bradley, R. G. (2006). Language use in imagined dialogue and narrative disclosures of trauma. *Journal of Traumatic Stress*, 19(1), 141–146. <https://doi.org/10.1002/jts.20102>
- Chang, C., Kaczurkin, A. N., Mclean, C. P., & Foa, E. B. (2018). Emotion regulation is associated with PTSD and depression among female adolescent survivors of childhood sexual abuse. *Psychological Trauma: Theory Research Practice, and Policy*, 10(3), 319.
- Chen, J., Cai, Y., Cong, E., Liu, Y., Gao, J., Li, Y., ... Gao, C. (2012). Childhood sexual abuse and the development of recurrent major depression in Chinese women. *PLoS One*, 42(2), 409–417.
- Coppersmith, G., Dredze, M., Harman, C., & Hollingshead, K. (2015). *From ADHD to SAD: Analyzing the Language of Mental Health on Twitter through Self-Reported Diagnoses*. Paper presented at the The Workshop on Computational Linguistics & Clinical Psychology: From Linguistic Signal To Clinical Reality.
- Decou, C. R., Lynch, S. M., Dehart, D. D., & Belknap, J. (2016). Evaluating the association between childhood sexual abuse and attempted suicide across the lifespan: Findings from a nationwide study of women in jail. *Psychological Services*, 13(3), 254–260.
- Fernandez-Lansac, V., & Crespo, M. (2015). Narrative length and speech rate in battered women. *PLoS One*, 10(11), e0142651. <https://doi.org/10.1371/journal.pone.0142651>
- Field, N. P., Classen, C., Butler, L. D., Koopman, C., Zarcone, J., & Spiegel, D. (2016). Revictimization and information processing in women survivors of childhood sexual abuse. *Journal of Anxiety Disorders*, 15(5), 459–469.
- Fullilove, M. T. (2009). Toxic sequelae of childhood sexual abuse. *American Journal of Psychiatry*, 166(10), 1090–1092. <https://doi.org/10.1176/appi.ajp.2009.09071058>
- Gao, R., Hao, B., Li, H., Gao, Y., & Zhu, T. (2013). Developing Simplified Chinese Psychological Linguistic Analysis Dictionary for Microblog. *International conference on brain and health informatics* (pp. 359–368). Springer, Cham, Switzerland
- Gena, Z., Alessandra, R., & Danielle, D. (2013). The mediating effect of daily stress on the sexual arousal function of women with a history of childhood sexual abuse. *Journal of Sex & Marital Therapy*, 39(2), 176–192.
- Hovey, J. D., Kim, S. E., & Seligman, L. D. (2006). The influences of cultural values, ethnic identity, and language use on the mental health of Korean American college students. *The Journal of Psychology*, 140(5), 499–511.
- Huemer, J., Nelson, K., Karnik, N., Voelkl-Kernstock, S., Seidel, S., Ebner, N., ... Skala, K. (2016). Emotional expressiveness and avoidance in narratives of unaccompanied refugee minors. *European Journal of Psychotraumatology*, 7, 29163. <https://doi.org/10.3402/ejpt.v7.29163>
- Jarrold, W., Javitz, H. S., Krasnow, R., Peintner, B., Yeh, E., Swan, G. E., & Mehl, M. (2011). Depression and self-focused language in structured interviews with older men. *Psychological Reports*, 109(2), 686–700.
- Johnson, C. F. (2004). Child sexual abuse. *Lancet*, 364(9432), 462–470.
- Kendler, K. S., & Aggen, S. H. (2014). Clarifying the causal relationship in women between childhood sexual abuse and lifetime major depression. *Psychological Medicine*, 44(6), 1213–1221.
- Lee, S., Lyvers, M., & Edwards, M. S. (2009). Childhood sexual abuse and substance abuse in relation to depression and coping. *Journal of Substance Use*, 4787(5), 349–360.
- Lorenz, T. A., & Meston, C. M. (2012). Associations among childhood sexual abuse, language use and adult sexual functioning and satisfaction. *Child Abuse & Neglect*, 36(2), 190–199.
- Martino, M. L., Onorato, R., & Freda, M. F. (2015). Linguistic markers of processing trauma experience in Women's written narratives during different breast cancer phases: Implications for clinical interventions. *Europe's Journal of Psychology*, 11(4), 651–663. <https://doi.org/10.5964/ejop.v11i4.991>
- Meston, C. M., Lorenz, T. A., & Stephenson, K. R. (2013). Effects of expressive writing on sexual dysfunction, depression, and PTSD in women with a history of childhood sexual abuse: Results from a randomized clinical trial. *The Journal of Sexual Medicine*, 10(9), 2177–2189.
- Molnar, B. E., Buka, S. L., & Kessler, R. C. (2001). Child sexual abuse and subsequent psychopathology: Results from the National Comorbidity Survey. *American Journal of Public Health*, 91(5), 753.
- Moore, S. D., & Brody, L. R. (2009). Linguistic predictors of mindfulness in written self-disclosure narratives. *Journal of Language and Social Psychology*, 28(3), 281–296. <https://doi.org/10.1177/0261927x09335264>
- Musliner, K. L., & Singer, J. B. (2014). Emotional support and adult depression in survivors of childhood sexual abuse. *Child Abuse & Neglect*, 38 (8), 1331–1340.
- Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Operator's manual: Linguistic inquiry and word count: LIWC2007*. Austin, Texas: LIWC. http://homepage.psy.utexas.edu/HomePage/Faculty/Pennebaker/Reprints/LIWC2007_OperatorManual.pdf.
- Pennebaker, J. W., & Chung, C. K. (2007). Expressive writing, emotional upheavals, and health. In H. Friedman & R. Silver (Eds.), *Handbook of health psychology* (pp. 263–284). New York: Oxford University Press.
- Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2015). The development and psychometric properties of LIWC2007. *Austin*, 29(11), 1020–1025.
- Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77 (6), 1296–1312.
- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54(1), 547–577.
- Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: Language use over the life span. *Journal of Personality and Social Psychology*, 85(2), 291–301.
- Pereda, N., Guilera, G., Forns, M., & Gómez-Benito, J. (2009). The international epidemiology of child sexual abuse: A continuation of Finkelhor (1994). *Child Abuse & Neglect*, 33(6), 331–342. <https://doi.org/10.1016/j.chiabu.2008.07.007>
- Polusny, M. A., & Follette, V. M. (1995). Long-term correlates of child sexual abuse: Theory and review of the empirical literature. *Applied and Preventive Psychology*, 4(3), 143–166.
- Pulverman, C. S., Lorenz, T. A., & Meston, C. M. (2015). Linguistic changes in expressive writing predict psychological outcomes in women with history of childhood sexual abuse and adult sexual dysfunction. *Psychological Trauma*, 7(1), 50–57.
- Rellini, A. H., & Meston, C. M. (2007). Sexual desire and linguistic analysis: A comparison of sexually-abused and non-abused women. *Archives of Sexual Behavior*, 36(1), 67–77.
- Rellini, A. H., & Meston, C. M. (2011). Sexual self-schemas, sexual dysfunction, and the sexual responses of women with a history of childhood

sexual abuse. *Archives of Sexual Behavior*, 40(2), 351–362. <https://doi.org/10.1007/s10508-010-9694-0>

- Rubens, S. L., Fite, P. J., Gabrielli, J., Evans, S. C., Hendrickson, M. L., & Pederson, C. A. (2013). Examining relations between negative life events, time spent in the United States, language use, and mental health outcomes in Latino adolescents. *Child & Youth Care Forum*, 42(5), 389–402.
- Sarah, H., & Boris, E. (2015). Positive emotional language in the final words spoken directly before execution. *Frontiers in Psychology*, 6, 1985.
- Simşek, O. F., & Cerçi, M. (2013). Relationship of the gap between experience and language with mental health in adolescence: The importance of emotion regulation. *Journal of Psychology*, 147(3), 293.
- Stanton, A., Meston, C., & Boyd, R. (2017). Sexual self-schemas in the real world: investigating the ecological validity of language-based markers of childhood sexual abuse. *Cyberpsychology, Behavior, and Social Networking*, 20(6), 382–388.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54. <https://doi.org/10.1177/0261927x09351676>
- Wright, C. M., Fraser, E. M., Denman, M., & Duke, L. (1987). Detection of sexual abuse in children. *LANCET*, 2(8552), 218.
- Zhao, N., Jiao, D., Bai, S., & Zhu, T. (2016). Evaluating the validity of simplified Chinese version of LIWC in detecting psychological expressions in short texts on social network services. *PLoS One*, 11(6), 1–15.

AUTHOR BIOGRAPHIES



Wenyu Wan is a master's student at Chinese Academy of Sciences (CAS) Key Laboratory of Mental Health, Institute of Psychology, CAS and Department of Psychology, University of Chinese Academy of Sciences (UCAS). She graduated with her Bachelor's degree from Beijing Sports University (BSU) in applied psychology. Her research field is focused on cognitive aging and brain plasticity of the elderly. She was also interested in computational cyber psychology and mental health.



Jiumo Sun is a master's student at CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences. Her research field is computational cyber psychology and mental health.



Jiehan Liu, graduating from Dalian University of Technology with a BA in Philosophy, is now studying in Department of Psychology, Chinese Academy of Science, and Institute of Psychology, CAS. Her research interest is conversation skills of preschool children and literacy development in early childhood.



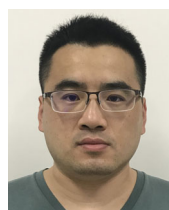
Shu-Wen Yang is a master's student of the CAS Key Laboratory of Behavioural Science of Institute of Psychology of Chinese Academy of Sciences. She has BSc from Tianjin University of Traditional Chinese Medicine in Applied Psychology. Her present research interests are in the area of behavioral decision-making. Her research has mainly appeared (or will appear) in such journals as *Advances in Psychological Science*, *Journal of Dialectics of Nature*, and *Journal of Community Medicine*.



Mingming Liu is a graduate student at CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences (CAS) and Department of Psychology, University of Chinese Academy of Sciences (UCAS) working toward an MA degree in Application Psychology. She holds Bachelor of Science degree in Psychology with honors from Zhejiang University. Her research field is computational cyber psychology and mental health, especially with the application of the new methodology of Ecological Recognition (ER). Her research interests include online behavior analysis and investigating the mental health changing pattern of people who experienced a major life event by ER.



Jia Xue is an assistant professor in the Factor-Inwentash Faculty of Social Work and the Faculty of Information at the University of Toronto. She received her PhD in Social Welfare from School of Social Policy & Practice, along with a dual master's degree in Statistics from Wharton School, University of Pennsylvania. She took her research fellowship at Harvard Kennedy School of Government. Jia applies computational and big data approaches to examine various facets of intimate partner violence and sexual assault. Jia's research agenda extends to the international context. Her article "Rape Myths and the Cross-Cultural Adaptation of the Illinois Rape Myth Acceptance Scale in China," published in the *Journal of Interpersonal Violence*, provides the first psychometric measure to assess attitudes about rape in China. Her research also has been published in scholarly journals such as *Child Abuse & Neglect*, *American Journal of Public Health*, and *British Journal of Social Work*.



Dongdong Jiao, engineer in Basic Technology Laboratory, National Computer System Engineering Research Institute of China. Dongdong Jiao obtained his MS Degree of Computer Science. His study interests include cloud computing, big data analysis, data mining, etc.



Xiaoqian Liu is an assistant researcher at the Institute of Psychology, Chinese Academy of Sciences. She graduated from the University of Chinese Academy of Sciences in 2014 with a PhD in computer science and applied technology. In recent years, Dr. Xiaoqian Liu is engaging in the intersection research work of computer and psychology. Her research fields include computational cyber-psychology, data mining, artificial intelligence applications, and her studies focus on the ecological acquisition methods of daily behaviors such as gait and facial movement, the correlation pattern between offline/online behaviors and psychological indicators such as emotion, personality, mental health, etc., and the automatic identification method of psychological indicators based on offline/online behavior analysis.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Wan W, Sun J, Liu J, et al. Using social media to explore the linguistic features in female adults with childhood sexual abuse by Linguistic Inquiry and Word Count. *Hum Behav & Emerg Tech.* 2019;1:181–189. <https://doi.org/10.1002/hbe2.162>