Universiteit Leiden
The Netherlands

Lexical choices of sharers and non-sharers on child sexual abuse material forums

Boer, M. de; Heeren, W.F.L.; Daser, A.; Gannon, C.; Gnielka, F.; Huikuri, S.; ... ; Blokland, A.

Citation

Boer, M. de, Heeren, W. F. L., Daser, A., Gannon, C., Gnielka, F., Huikuri, S., ... Blokland, A. (2025). Lexical choices of sharers and non-sharers on child sexual abuse material forums. *Applied Corpus Linguistics*, *5*(3). doi:10.1016/j.acorp.2025.100157

Version: Publisher's Version

License: Creative Commons CC BY 4.0 license

Downloaded from: https://hdl.handle.net/1887/4286692

Note: To cite this publication please use the final published version (if applicable).

Applied Corpus Linguistics 5 (2025) 100157

Contents lists available at ScienceDirect

Applied Corpus Linguistics

journal homepage: www.elsevier.com/locate/acorp

Articles

Lexical choices of sharers and non-sharers on child sexual abuse material forums

Meike de Boer [a,b,*], Willemijn Heeren [a,b], Anton Daser [c], Colm Gannon [d], Frederic Gnielka [e], Salla Huikuri [f], Robert Lehmann [e], Rebecca Reichel [e], Thomas Schäfer [e], Alexander F. Schmidt [c], Katarzyna Staciwa [g], Arjan Blokland [a,h]

[a] Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, the Netherlands
[b] Leiden University Centre for Linguistics, Leiden University, Leiden, the Netherlands
[c] Institute of Psychology, Social & Legal Psychology, Johannes Gutenberg University Mainz, Mainz, Germany
[d] School of Computing, Engineering and Mathematical Science, La Trobe University, Melbourne, Australia
[e] Department of Psychology, MSB Medical School Berlin, Berlin, Germany
[f] Child Protection and Technology, ECPAT International, Bangkok, Thailand
[g] Polish Platform for Homeland Security, Poznan, Poland
[h] Leiden Law School, Institute of Criminal Law and Criminology, Leiden University, Leiden, the Netherlands

ARTICLE INFO

Keywords:
Child sexual abuse material
Forensic linguistics
Keyness analysis
Word frequencies
Lexical choices

ABSTRACT

On the dark web, there are forums dedicated to the distribution and discussion of child sexual abuse material (CSAM). Although exchanging material is one of the major purposes of such forums, only a small portion of the users share CSAM themselves. Using keyness analysis, we analyzed word frequencies to see which words were unusually frequent for either CSAM sharers or non-sharers. The language of non-sharing members shows more positivity and rapport-building, which could be a way to compensate for not being able to meet the expectation to contribute material to the forum. In addition, they use more sexually explicit language, potentially to prove that they are a genuine part of the community. Sharers, on the other hand, talk more about the forum and the world outside of the forum where their practices are considered illegal. Hence, many words that are typical for the sharing members are related to the law and law enforcement. Before members start sharing, their language use is situated between non-sharers and sharers. They use positive, rapport-building, and explicit language, although lesser pronounced than non-sharers, and they refer to the forum community but not yet to the world outside the forum. Findings can be used by law enforcement in covert operations, who might want to mimic strategies to compensate for not being able to share CSAM. In addition, the results show that keyness analysis could potentially aid in differentiating between different groups of users on dark web CSAM forums, which could help law enforcement to prioritize target members in large-scale CSAM forums.

1. Introduction

The anonymity of the dark web offers ideal circumstances for people who engage in criminal activities to find each other (Gannon et al., 2023). The dark web can be accessed using The Onion Router (TOR), ensuring the user a high level of anonymity. Besides providing criminal marketplaces for all kinds of illegal products and services, the dark web is home to a global community of individuals who have a sexual interest in children and/or an interest in child sexual abuse material (Europol, 2024). During a period of six months, the daily number of active sites on the dark web was estimated at 80,000 unique .onion addresses, of which 2 % ($N = 900$) contained child sexual abuse material (CSAM). In total, these sites received an estimated amount of 168,152 access requests per day (Owen and Savage, 2015). More recently, Nurmi et al. (2024) estimate that about a fifth of domains on TOR contain some kind of CSAM. Europol (2022) considers CSAM forums as one of the key threats in its fight against the sexual exploitation of children.

Prior studies have shown that the majority of users of CSAM forums do not post any messages or content on the forums. These users access the forum to browse and download CSAM, while not actively

* Corresponding author at: Leiden University Centre for Linguistics, Leiden University, Reuvensplaats 3-4, 2311 BE Leiden, the Netherlands.
E-mail address: meikedeboer@linguaforensica.com (M. de Boer).

https://doi.org/10.1016/j.acorp.2025.100157
Received 1 February 2025; Received in revised form 9 September 2025; Accepted 10 October 2025
Available online 11 October 2025
2666-7991/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

contributing to the discussions or CSAM sharing. This group is frequently labeled as lurkers, both by other forum users and by researchers (e.g., Van der Bruggen et al., 2022). On one particular forum, Van der Bruggen et al. (2022) identified nearly 97 % of the users as lurkers, not contributing any posts to the public part of the forum. Similar results were found by Gnielka et al. (2024), who, on two forums with 936,110 and 592,345 registered members, found less than three percent of the registered members posting public messages. Whereas some CSAM forums have minimum activity requirements such as contributing at least one post before being allowed to enter the forum, the majority of users only contribute just enough to reach these thresholds (Fonhof et al., 2019). In an analysis of a forum with about 15, 000 members and over 400,000 posts, Van der Bruggen et al. (2022) found that only 109 forum members (0.7 %) were responsible for 40 % of the public posts, posting at least 500 posts each.

Of the small minority of members who participate in discussions on the public part of the forums, an even smaller number shares CSAM (Van der Bruggen et al., 2022; Blokland et al., 2024). Note that on the type of community forums discussed in this paper, users who try to sell or trade CSAM are typically banned, so users who share CSAM do not have financial motives (e.g., Chiang, 2024). Rather, there are implicit and explicit expectations regarding sharing: sharers receive more status, especially when sharing new or rare content, and on some forums can earn badges or access special parts of the forum as a benefit (Gannon et al., 2023; Quayle and Taylor, 2002). It is therefore interesting to see how people who do not meet the normative expectations of the forum navigate themselves in conversations with other forum members and what the communicative differences are between the posts of sharers and non-sharers. Hence, provided with a rare opportunity to gain insight into the communication in this community, the current research offers a comparative analysis of what users who share or do not share talk about on an average-scale CSAM forum. Using a mixture of quantitative and qualitative (corpus) linguistic analyses, we investigate lexical choices on the forum. Our research questions are:

1. How do sharing members differ from non-sharing members in terms of word use?
2. How does the language change once a member starts sharing CSAM?
3. How do future sharers differ from non-sharers in terms of word use?

2. Literature review

2.1. Posting behavior on CSAM forums

Based on a quantitative analysis of posting behavior, Van der Bruggen and Blokland (2021) describe six different typical profiles of members of CSAM forums: lurkers, browsers, CSAM interested, escalators, vested members, and managers. For forum members' main goals – accessing CSAM, accessing advice on CSA(M), and finding normalization or validation of their sexual interest (Kloess and van der Bruggen, 2021) – active participation in the conversations is not immediately necessary. Whereas each group spends time on the forum to browse and download CSAM, the activity levels are different (see Table 1).

The least active members are called *Lurkers* and *Browsers* and post no or only a minimal number of messages on the forum. Hence, these groups are not represented in linguistic analyses. As is typical for online communities in general (Herring, 2004), non-active members tend to form the vast majority of members (cf. Gnielka et al., 2024; Van der Bruggen et al., 2022). According to Nonnecke and Preece (2001), forum members may lurk for multiple reasons, including shyness, wanting to remain anonymous, seeking to learn from the community, and feeling unable to contribute. Although one could conclude that the inactive group of members on CSAM forums is less relevant to the police, Van der Bruggen et al. (2022) found that almost all users (> 99 %) download CSAM and are thus in possession of illegal material. In addition, Van der Bruggen and Blokland (2021) show that one-fourth of the limited number of messages browsers post are in the "girls hardcore" section of the forum, thus showing an interest in more sexually explicit content. Hence, these groups, that form the majority of the users, may be equally important to consider as any other group, yet will be necessarily neglected in analyses of public messages on (dark web) forums.

Table 1
Six user profiles on CSAM forums (as identified by Van der Bruggen and Blokland, 2021).

	Group	Characteristics
1	Lurkers	- Least active - Post a low number of messages (or none, if allowed) - Join the forum relatively late
2	Browsers	- Post relatively few messages (but 5× as much as lurkers) - Join the forum relatively late - Post in the more sexually explicit category "girls hardcore"[1]
3	CSAM interested	- Post messages over a longer period - Post in a variety of categories, incl. "boys hardcore" and "girls hardcore" - Over half of this group is a full member (i.e., users who post regularly)
4	Escalators	- Post increasingly often over time - Keep posting actively until the take-down of the forum - 1 out of 10 have a VIP status
5	Vested members	- Join (almost) immediately after the forum emerges - Have a high number of posts - Active in a variety of parts of the forum - Engaged with the social aspect of the community - Almost all users in this group are full members - 1 out of 5 has a VIP status
6	Managers	- Most active posters - 3 out of 4 are involved in forum management (e.g., technical safety) - More than half has Administrator / VIP status

[1] "Hardcore" refers to CSAM in which the nature of the images is explicitly sexual. It opposes to "softcore", which depicts (naked) children without an explicit sexual nature (cf. Anderson, 2024).

The *CSAM interested* and *Escalators* post messages over a longer period of time, with the main difference being that the posting behavior of escalators increases over time (Van der Bruggen and Blokland, 2021). These groups of users represent a group of moderate users who do take part in public discussions but do not post an excessive number of messages. Hence, these groups may be underrepresented in linguistic analyses of CSAM forums.

The majority of forum posts are written by the *Vested members* and *Managers* (Van der Bruggen and Blokland, 2021). These users identify the most with the CSAM community and act as online place managers, branding and maintaining the forum and regulating member behavior (Blokland et al. 2024; Van der Bruggen and Blokland, 2021), such as providing instructions and welcoming new members to the forum (Chiang, 2024). Vested members and Managers typically form a small minority of forum users but are overrepresented in linguistic analyses when sampling on public messages.

Hence, different user groups can be identified on the forums with distinct communication patterns and activity levels. However, distinguishing user groups merely by the number of messages they posted rather than the nature of the communicative content of their posts gives us only a limited insight into this community.

2.2. Public discussions on CSAM forums

Although limited, some linguistic analyses of CSAM forums have been performed. Woodhams et al. (2021) investigated private communications as well as posts on a CSAM forum of 53 suspects under investigation for (online) child sexual abuse offenses. In their communications, 17 suspects mentioned that they accessed CSAM forums to access material, whereas only 11 were already in possession and thus capable of sharing CSAM. In an analysis of the forum posts made by these individuals, of whom 34 had committed hands-on child sexual

abuse, Woodhams et al. (2021) found that common topics were sexual acts (40 %), deviant sexual interests (38 %) such as sadism, bestiality, and incest, security measures to avoid detection by law enforcement (23 %), rules of the forum (21 %), and discussions about sexual interest in different age categories (11 %).

Chiang et al. (2020) looked into the functions that members of a CSAM forum perform with their utterances. Randomly selecting 200 pages of messages on a CSAM dark web forum, they investigated 19,974 conversational turns by 5590 users. Using Move Analysis (Swales, 1981; 1990), they allocated at least one move to each utterance. Move analysis is a text analytical method from genre analysis in which the communicative functions of parts of a text are established (Moreno and Swales, 2018). Chiang et al. (2020) found 20 different moves in the dataset, excluding 2.33 % of the utterances where the function was unclear. Out of these 20 moves, 11 could be further categorized into the following move types: (1) Rapport-building, (2) Image exchange, (3) Image discussion and appreciation, and (4) Assistance (see Table 2). The five most common moves, characterizing over 75 % of the utterances, were: Offering CSAM (30.19 %); Greeting (15.84 %); Image appreciation (12.89 %); General rapport (8.81 %); and Image discussion (7.44 %). Thus, the majority of the moves on CSAM forums are either directly related to offering or discussing CSAM, or to building rapport and being polite (i.e., community-building).

Chiang et al. (2020)'s materials contained 3881 users (69.4 %) who wrote at least two posts. Of these users, 1313 (34 %) offered CSAM at least once. Only 377 users (6.7 %) posted at least 10 messages on the forum. Of them, 314 (83 %) offered CSAM at some point. From this, Chiang et al. (2020) concluded that most highly active users will eventually participate in sharing. Relating this finding to the user typology of Van der Bruggen and Blokland (2021), it is likely that non-sharing members are part of the lesser active subgroups, whereas sharers can be mostly found in the most active user categories. When comparing users with at least two messages who offer CSAM ($N = 1313$) to those who do not ($N = 2568$), Chiang et al. (2020) found that the moves Image Appreciation and Describing Experiences are among the most used moves for non-sharers but are being used by sharers only minimally.

Table 2
Rhetorical moves distinguished by Chiang et al. (2020).

Move Type	Move	Frequency (% of all turns)
Rapport-building	Greeting	15.84
	General Rapport	8.81
	Sign Off	0.86
Image[1] Exchange	Offering CSAM	30.19
	Requesting CSAM	3.40
Image Discussion & Appreciation	Image Discussion	7.44
	Image Appreciation	12.89
	User Appreciation	1.58
Assistance	Seeking Assistance	4.04
	Providing Assistance	2.92
	Denying Assistance	0.06
Other	Expressing Opinion/	2.21
	Preference	2.18
	Describing Experience	2.43
	Exerting Authority	0.26
	Showing Deference	0.38
	Judging Character	1.94
	Seeking User/Interaction Type	0.03
	Law enforcement agent (LEA)	0.04
	Accusation	0.16
	Meeting Planning	
	Law Discussion	
Unclear		2.33

[1] Chiang et al. (2020) use the term Indecent Images of Children (IIOC) for what we call CSAM in the current paper. The terminology in this table corresponds with the current paper.

2.3. Sharing is caring?

An analysis of the posts by users who share CSAM showed that Image Offering can be performed at any time and after any move (although it was less common after the moves Meeting Planning and LEA Accusation). According to Chiang et al. (2020), the fact that sharing CSAM is not restricted to follow after other moves such as CSAM requests, shows that sharing is the main goal of the forum. The most common move preceding or following CSAM sharing was Image Discussion: users often describe or provide commentary on their own or others' shares. The users who shared CSAM performed Requests more often than users who did not share. A potential explanation for this, offered by the authors, is that it is more acceptable to request CSAM for users who participate in sharing themselves. For users who do not share, requesting others to do so may be inappropriate. This shows the covert expectation that users shall not only use the forum to passively consume but also to actively contribute to it. Indeed, requests by non-sharers are often met with the move Exerting Authority: someone reminding them that requesting images is not the purpose of the forum. To lower the risk of being reprimanded, non-sharing members often accompanied their requests by the moves Image appreciation or General rapport. This may be a way to compensate for making a move that is considered inappropriate for them.

Overall, the most common moves used by non-sharers were General Rapport, Image Appreciation, and Describing Experience. According to Chiang et al. (2020), this fits the hypothesis that people who do not contribute CSAM are not adhering to the implicit or explicit expectations of the forum and need to compensate by performing overly positive and social behavior. By explicitly expressing appreciation and building general rapport, they may still be accepted as valuable members of the community, even though they do not share CSAM. In addition, by sharing biographical experiences, these non-sharing members can establish credibility as part of the community and prove that they are engaged with the topic. In the context of exchanging illegal material, the forum users are well aware of the fact that covert law enforcement agents (LEAs) may infiltrate the forum (cf. Chiang, 2021). Since in most jurisdictions, LEAs are not allowed to share CSAM, the move of Image Offering is a way to prove that one is not an undercover officer. In order to avoid suspicion, users who do not share material may need to find other ways to establish credibility on the forum. The assumption that sharing personal experiences can be used as a replacement for CSAM sharing is corroborated by the fact that members who do share CSAM perform this move only rarely.

The study by Chiang et al. (2020) shows that there seem to be major differences in the moves performed by CSAM sharers and non-sharers, and that the latter group shows several strategies to compensate for their lack of CSAM contributions. According to the researchers, this knowledge can be applied by undercover LEAs who are not allowed to share CSAM in their jurisdiction. If they mimic the language of general members without differentiating between sharers and non-sharers, they may come across as impolite, rude, or misusing the forum. However, we still do not know what kind of language members exhibit and which lexical choices they make when talking about CSAM. Studying lexical choices would allow undercover LEAs to further manipulate their linguistic behavior so that they are not perceived as outsiders. In addition, lexical choices may be potentially helpful in distinguishing different user groups and predicting which users would show certain behavior (e.g., sharing CSAM) in the future.

2.4. Lexical choices and aboutness

Although move analysis may give insight into the communicative functions of an utterance, it has some disadvantages (cf. Chiang et al., 2020). First of all, a single utterance may perform different functions, and thus different moves. In addition, the function is not always clear, and the moves can be seen as an interpretation of the most likely

function of an utterance. Finally, the manual turn-by-turn interpretation is described as labor intensive. Hence, a computational approach may be preferred over a manual one, for example by looking at lexical choices.

Lexical choices have been used for differentiating between groups in prior studies for different purposes. For example, Linguistic Inquiry and Word Count (LIWC) has been used to study suicide notes (Handelman and Lester, 2007), right-wing extremism (e.g., Grover and Mark, 2019), terrorism (Pennebaker and Chung, 2009), and threatening letters (Glasgow and Schouten, 2014). LIWC is a text analysis program developed by Pennebaker et al. (2001) that can be used to analyze texts on a word-by-word basis. The software analyzes texts based on over 80 language dimensions, such as emotion, power, metaphysical spaces, and syntactical categories (e.g., first-person pronouns and future tense verbs). Hunter and Grant (2025) however, argue that LIWC is inaccurate and cannot be relied upon without a manual check. Even if LIWC would be reliable, by classifying each word into predefined categories, the exact lexical choices made by the users are still neglected. In addition, although the commercial program is advertised as transparent (Tausczik and Pennebaker, 2010), a full list of language dimensions analyzed could not be found without purchase. Finally, using external or commercial software may be undesirable or even prohibited when working with extremely sensitive data seized by law enforcement, as in the current study.

Another approach in comparing the language in two texts or corpora is keyness analysis. Keyness is a corpus linguistic concept introduced by Scott (1997:236) as a way to establish which words occur "with unusual frequency" in one corpus compared to another or reference corpus. The core principle of keyness analysis is that when two corpora have different (linguistic) characteristics, this will show up in the word frequency lists (Kilgarriff, 1997). Although frequency lists fail to provide meaning or linguistic forms (Gabrielatos, 2018), the lack of information on semantic properties, part of speech, syntactic relations, context, and collocations is compensated when the word lists are considered in their totality (Kilgarriff, 1997). For example, polysemy may lead to the word "bank" not being a key term in a text about banking when compared to a text about river banks. However, the different natures of the texts will show in other key terms such as *money* and *account* vs *river* and *grassy* (Kilgarriff, 1997:233). By identifying key terms, one can get an idea of the concepts that are socially important in that corpus (Scott, 1997). According to Scott (1997), keyness analysis is a way to establish *aboutness*. Aboutness describes the process of understanding the main concepts, topics, or attitudes in a text or corpus, i.e. it gives the reader or analyst an idea of what the texts are about (Phillips, 1989). According to Hutchins (1977), a quick identification of what a document is about is one of the most crucial challenges in information sciences and can help navigate large corpora. According to Gabrielatos (2018), keyness analysis is a useful approach in exploratory research as a way to efficiently see what topics are mentioned more frequently in any of the texts or corpora. An advantage of the approach is that besides a notion of what the texts are about, the exact lexical choices are provided. Thus, keyness analysis not only shows what is written about but also how it is formulated.

2.5. The current study

In this study, using keyness analysis, three comparisons will be made. First, we compare posts written by users who share CSAM to posts of users who do not share CSAM at any time during their stay on the forum. In this comparison, we only use posts written by sharers from the moment that they started sharing CSAM, since their language may not yet be indicative of sharing before this moment. As mentioned in 2.3, insight into the lexical choices of non-sharers may be helpful to undercover LEAs in jurisdictions where they are not allowed to share material, as they may need to compensate for their lack of sharing and avoid inappropriate or unfitting words. In addition, we want to know whether the observations made by Chiang et al. (2020) are reflected in the keywords of the current analysis. Behind the question of how the lexical choices of sharers differ from non-sharers lays the broader question of whether it is possible to group users on such forums based on their lexical choices. Because of the large number of users, LEAs are constantly looking for ways to efficiently differentiate between different user types on dark web forums. When law enforcement can quickly filter out users who are of less interest, e.g. because they do not share material, this helps them to focus their resources on the users of the highest interest. If the techniques used in this paper can differentiate between sharers and non-sharers, they may also be used on other pre-defined user groups such as producers of new CSAM material and people who are known to have committed an offline child sexual abuse offence.

Second, we compare the posts written by CSAM sharers before and after their first CSAM share. This may provide insight into the development users make over time and how the lexical choices of the same group of individuals may change once they start to participate in sharing. In a sense, this is also a sharing vs non-sharing comparison, but with the sharers' prior selves as the reference. Based on Chiang et al. (2020), we expect that non-sharers use more appreciative and rapport-building language as a compensation strategy for not being able to share. If this is the case, these same users are expected to quit these strategies once they have become sharers.

Finally, the posts of future sharers (i.e., sharers before they started sharing) are compared to those of the non-sharers. Any differences between these groups may be used as indications that a certain non-sharing member will later become a sharer. Although we do not know whether non-sharers would have started sharing had the forum not been taken down by law enforcement, this comparison is considered useful. If law enforcement can predict which users are later going to participate in sharing behavior, they may be able to intervene, e.g. by sending them a warning message (cf. Prichard et al., 2024; Wortley et al., 2024), and to allocate their limited resources to those posing the highest risk.

3. Methods

3.1. Data and sample

The analysis was performed using a CSAM forum that law enforcement had seized. Due to privacy as well as legal, ethical, and well-being considerations, neither the name of the forum nor any usernames are provided in this paper. The forum under scrutiny here is the same as that was analyzed in Gnielka et al. (2024). Whereas prior analyses pertained to the frequency of posting, here the focus is on communicative posting content. We cannot give further information regarding the nature of the data, other than that it was a large-scale CSAM forum with material of both male and female victims and 936,110 registered users. The data consisted of 194,551 messages posted on the public part of the forum (i.e., not in private chats).

The language and content of the messages show that English is being used as a lingua franca on the forum and is the dominant language spoken by all users, including non-native English speakers (e.g., "Hello Ím the new one, I don't write very good English but I will do my best."). There are separate subsections where people can communicate in different languages, which are being used sparsely. Hence, the vast majority of the messages on the main parts of the forum are written in English. The few messages in languages other than English were excluded from the analysis.

Table 3
Overview of the number of posts and users included in the analysis.

	Number of posts	Number of words	Median words/post	Number of users
Sharers	77,424	6,192,693	38	5001
Non-Sharers	20,895	863,240	16	10,148
Future Sharers	9238	478,437	21	2441

In our analysis, three sets of posts are distinguished (see Table 3):

1. Sharers (S): Posts written by users who have shared CSAM on the forum. Any messages posted before they actually started sharing are excluded from this set. CSAM sharing was defined as posting a hyperlink at least once. Hyperlinks seem a good proxy for CSAM sharing: using manual coding, Blokland et al. (2024) found that 96 % of posts containing a hyperlink directly referred to CSAM.
2. Non-Sharers (NS): Posts written by users who have not shared CSAM at any point during the period the forum was online.
3. Future Sharers (FS): Posts written by CSAM sharers before they actually started sharing.

Before analyzing the data, data were pre-processed in RStudio (R Core Team, 2024) to exclude hyperlinks and computer code for layout features (e.g., bold text and emojis). An English stop word list from the quanteda package (Benoit et al., 2018) was used to exclude highly frequent function words such as *a* and *the* from the data.

3.2. Method of analysis

The nature of the data requires a versatile approach, as our analysis is both explorative and explanatory in nature. This requires both analyzing the language qualitatively, to see what happens on the forum and how certain activities are being discussed, and quantitatively, to allow the inclusion of as much data from the forum as possible. The following steps were followed in the current analysis.

Step 1: Keyness analysis (quantitative)

As discussed in Section 2.4, keyness analysis is a corpus linguistic technique used to directly compare two sets of texts to evaluate whether a word or other linguistic unit is more typical for one of the sets under comparison (Gabrielatos, 2018; Scott, 1997). For each comparison, one set of texts is used as the study corpus, and another as the reference corpus. The keyness analysis results in a list of items that are unusually frequent in the study corpus relative to the reference corpus. Typically, this can be used to identify large differences in frequencies between the corpora (Gabrielatos, 2018), although it can also be used to establish similarity (Taylor, 2013) or absence (Partington, 2014). Although any type of linguistic unit can be selected as the unit of analysis, word-based analyses are the most common (Gabrielatos, 2018) and were selected for the current analysis.

The keyness analysis was performed in RStudio (v4.4.2) using the quanteda package (Benoit et al., 2018). From the quanteda package, we used the chi-square (χ^2) test to compute the statistical significance of the frequency differences between the study and reference corpus. The χ^2 test was selected because it is a traditional yet robust method for identifying statistically significant differences in word frequencies between corpora. While log-likelihood combined with effect-size measures has gained traction in recent years, notably for its ability to capture nuance in distribution and magnitude of difference, the χ^2 test remains a well-established and widely used statistic for keyness analysis (e.g., Gabrielatos, 2018; Sönning, 2024). The χ^2 test was selected here because it offers a straightforward and interpretable measure of statistical difference that is sufficient for capturing salient lexical contrasts as is the goal of the current study.

Using keyness analysis, the following comparisons were made (see also Section 2.5):

1. Sharers vs Non-Sharers, which is a between-subject comparison.
2. Sharers vs Future Sharers, which is partly a between-subject comparison – most users started sharing from their first post and are only included in the Sharers – and partly a within-subject comparison – for users who posted messages both before and after their first CSAM share.
3. Future Sharers vs Non-Sharers, which is a between-subject comparison.

Each of the three comparisons was run twice: once with set A as the study corpus and set B as the reference corpus, and once in reverse. This resulted in six keyword lists, i.e. two lists per comparison. For each keyword list, the top 100 keywords were selected for further analysis. The top 100 is the average and most common number of keywords used for a follow-up analysis in keyness analysis (Gabrielatos, 2018).

Step 2: Thematic comparison (qualitative)

The lists of 100 keywords resulting from the quantitative analysis were categorized thematically. This was done using open coding, an inductive approach in which the categories are constructed based on the data (Strauss and Corbin, 2004). Using this approach, a list of seven different main categories was created (see also Table 4):

1) Action words, including but not limited to verbs and derivations from verbs. Within this category, subcategories were identified based on sentiment (positive, neutral, negative) and theme (politeness, sexual, CSAM). Words potentially referring to CSAM content, such as "watch", were classified as a separate category rather than being considered neutral.
2) Descriptions, including but not limited to adjectives. Subcategories were again based on sentiment (positive, neutral, negative) and theme (sexual, place/time). The positive and sexual subcategories were merged because in the context of child sexual abuse, the boundaries are often unclear. For example, calling a child sexy, cute, or hot has a sexual connotation.
3) References to individuals, including names. Terms in this category could be referring to children or adults. When it was unclear whether the term referred specifically to a child, it was included in the adult/general category.
4) Body parts. These were split into female and male genitals, other intimate body parts (*ass*), and other body parts (*mouth*). Although other body parts such as the mouth could be considered intimate as well in this context, only buttocks and breasts were classified as general intimate body parts.
5) Clothing, which only consisted of the subcategory underwear.
6) Technical words. These terms refer to CSAM content (excluding action words; see category 1), the forum context (including usernames), and the internet or computers in general.

Table 4
Overview of the categories and subcategories found in the key term lists. For each subcategory, some examples are provided.

Category	Subcategory	Examples
Action words	Negative	reduce, vanned
	Neutral	began, can, said
	Positive	appreciated, attraction
	Politeness	hello, please, thanks, yes
	Sexual	abuse, cum, fucking, licking
	Content (CSAM)	corrupted, download, posted, watch
Descriptions	Negative	annoying, shady, wrong
	Neutral	little, unintelligible, young
	Positive/sexual	amazing, hot, omg, wow
	Place and Time	cabin, ever, public, night
Individuals	Children	age, boys, loli, kids
	Adults/general	anyone, predator, nurse, wife
	Female names	–
	Male names	–
Body parts	Female genitals	pussy, vulva
	Male genitals	glans, penis
	General intimate	ass, buttocks
	Other	arms, finger, mouth
Clothing	Underwear	panties, underwear
Technical	Content	data, link, pornography, video
	Forum	members, messages, posts, threads
	Internet/computers	browser, click, mirror, password
Society	Law/security	criminal, evidence, fbi, justice
	Other	county, government, state, vote
Excluded	Ungrouped	according, cat, information, oracle
	Unknown	droppa, pw, skee, -t

7) Society. A large subcategory within society is law/security.

Finally, some words from the key term lists did not fit into any category and were left ungrouped. In addition, some words were unknown or ambiguous to the researchers, such as abbreviations, slang, or potential technical words. Only if words obviously looked like pieces of computer code, they were classified as Technical (category 6). Because the ungrouped and unknown words cannot contribute to revealing patterns in the data, they were not included in the analysis. For an overview of these excluded words, see Appendix A.

Step 3: Key Word in Context Analysis (mixed)

Without a manual follow-up analysis, keyness analysis may lack context and therefore the meaning of the differences found (e.g., Baker, 2004). To gain some context information for the keywords emerging from the keyness analysis, we performed a Key Word in Context (KWIC) analysis (Luhn, 1959) using the quanteda package in RStudio. In the KWIC analysis, for a selection of keywords for which the context was not immediately clear, their occurrences were analyzed with 10 words to the left and right of the keyword. Subsequently, a number of occurrences were sampled (as described in Appendix C) and their context was labeled thematically to provide information regarding the contexts of the keywords. The results of the KWIC analysis are provided in Appendix C and are described concisely in the next section.

4. Analysis

This section presents the results of the thematic analysis of the keyness lists and the subsequent KWIC analysis. Appendix B gives an overview of the number of keywords in the thematic categories for each of the six comparisons. Since the results for the Sharers were highly similar regardless of the reference corpus, the two comparisons involving the Sharers are presented together in Section 4.1. Despite their similarities when compared to the Sharers, there may be differences between the Future Sharers and Non-Sharers that can be used to predict sharing behavior in the future. Hence, the FS and NS sets are compared in 4.2.

4.1. The language of sharers and non-sharers

When comparing the language of Sharers to the other two sets, we see that the lexical choices in posts written by Sharers are quite different from the others. This section discusses the results for the two comparisons involving the Sharers from the moment they started sharing CSAM to the two non-sharing (or not-yet-sharing) sets.

4.1.1. Action words used by sharers and non-sharers

Action words contributed to 10–12 % of the top 100 key term lists for the Sharers and 29–39 % in the lists for the two non-sharing sets (i.e., NS and FS). The action keywords are provided in Table 5 The main differences can be found in the positive and politeness action words, which were absent in the Sharers lists. Apparently, people who are not sharing CSAM (at that moment) are more positive in their choice of words and, on top of that, more polite. In addition to these differences, the keywords for the non-sharing sets of posts included more sexual action words which were also more explicit, such as *fucking* and *sex*. In contrast, sexual action keywords for the sharing set were limited to *abuse* and *exploitation*. It is interesting to explore the context of these terms, as this does not seem to match the forum users' own worldviews and perspectives (Merdian et al., 2014; Van der Bruggen and Blokland, 2021). The KWIC analysis showed that when using the word *abuse*, Sharers referred to society's point of view rather than their own perspective in more than half of the occurrences. Non-Sharers, on the other hand, take the most critical view, using the word *abuse* deliberately to refer to potential harm (see Appendix C, Table C1).

The distribution of keywords for the other action word subcategories is more similar. Negative action words are almost absent in the keyword lists for any of the three groups. This could either mean that negativity is rare on the forum or that it is shown in similar frequencies and terms in all sets. All keyword lists contain neutral action words, but their neutral nature makes it difficult to compare them to each other. Finally, the three sets have keywords potentially related to (CSAM) content. Here, the non-sharing sets (i.e., NS and FS) include more terms related to passive actions such as *watching, looking*, and *re-uploading* (requests to re-upload inaccessible CSAM), whereas the sharing set contains more active actions such as *added* and *posted*. Indeed, the KWIC analysis shows that the word *watch* refers to watching CSAM in the majority of the cases, especially for the NS set (see Appendix C, Table C3). For Sharers, the word is used in an additional context and often refers to law enforcement and security. Interestingly, *watching* was not always a passive act: in a quarter of the occurrences in the NS group, it refers to offline CSA situations, including children watching the adult masturbate. This demonstrates that Non-Sharers cannot necessarily be regarded as less severe abusers than Sharers. The keyness of *re-uploading* in the Non-Sharing set is interesting because it shows that these users make re-uploading requests to address non-working links when they are trying to consume CSAM. Whereas Chiang et al. (2020) found that requests were more typical for sharing members, re-uploading requests may be more typical for members who do not share. Potentially, re-uploading requests are considered less inappropriate because the sharing member has already made their own decision to upload CSAM and is not requested to upload more.

4.1.2. Descriptions used by sharers and non-sharers

In the keyness lists for the Sharers, descriptions such as adjectives are almost absent. In the other lists, however, descriptions take up 22 to 25 % of the list. As with the action words, negative descriptions were not typical for either set (see Table 6). The non-sharing sets (FS and NS) included neutral descriptions potentially used to describe a child (e.g., little, naked, younger) or the content (e.g., unintelligible). Indeed, the KWIC analysis showed that the vast majority of these terms were used to refer to children (see Appendix C, Table C4). Of these descriptions, *naked* could be considered sexual rather than neutral but was so different from the other positive/sexual descriptions that it was considered neutral. The majority of descriptions were overly positive (e.g., amazing, awesome, best, great, perfect) or even sexualized (e.g., cute, hot, sexy). Finally, some words in the lists related to place and time, which were considered most fitting under the higher category of descriptions. Here, the words *ever* and *never* were key to the non-sharing users, possibly because they were fantasizing or anticipating future opportunities for committing or sharing images of child sexual abuse. The KWIC analysis supported this hypothesis, showing descriptions referring to both watching CSAM as fantasizing hands-on abuse – although it also revealed a different category of conversations in which members reported feelings related to pedophilia, such as sexual attraction to children and loneliness. For the sharers, the only key term potentially related to place and time was *public*, which could refer to a location where abuse takes place, to society, or to the digital world inside or outside the forum. The KWIC analysis showed that for the Sharers, the vast majority of uses of the word were for *public key* and other digital spaces, whereas for the Non-Sharers, it was used relatively often to refer to places of abuse (see Appendix C, Table C5).

4.1.3. References to individuals, body parts, and clothing by sharers and non-sharers

With regards to references to individuals (1–6 %) and their body parts (0–6 %) or clothing (0–1 %), there are some differences in the keyness lists for sharing and non-sharing users. Posts in the non-sharing sets (FS and NS) included more references to children and individuals in general, including more names (see Table 7). On top of this, terms referring to body parts (including genitals) and clothing can only be found in these sets and are absent from the Sharers' key term lists. This shows that despite the lack of CSAM sharing in these posts, the victims

Table 5
Comparison of Action words from the keyword lists for Non-Sharers vs Sharers (left) and Future Sharers vs Sharers (right).

Category	Subcategory	Non-Sharers	Sharers	Future Sharers	Sharers
Action words	Negative	–	–	vanned	–
	Neutral	can, cant express extract find get, got im started tried work, working	closing completed said signed	began get feel	closing completed counts said signed
	Positive	appreciated attracted like love, loved wish	–	attracted attraction like love	–
	Politeness	hello hi please, pls, plz thank, thanks, thankyou, thx ty	–	gracias gratz hi okay please thank, thanks, thankyou, thx yeh	–
	Sexual	cum fuck, fucking suck, sucking	abuse exploitation	busted fucking imitating laid lick masturbated sex suck	abuse exploitation
	Content	looking reup, reupload, re-up, re-upload watch	added posted request, requests	looked reup, re-upload	added posted received request, requests

Table 6
Comparison of Descriptions from the keyword lists for Non-Sharers vs Sharers (left) and Future Sharers vs Sharers (right).

Category	Subcategory	Non-Sharers	Sharers	Future Sharers	Sharers
Descriptions	Negative	shahdy [sic] wrong	–	foul shady, shadys	–
	Neutral	little much old tiny younger	–	big little much naked old slim unintelligible younger	–
	Positive/ sexual	amazing awesome beautiful best cute good great hot nice omg perfect sexy wow	–	amazing awesome beautiful best cute divine good great hot lovely nice sexy wow	thick
	Place and Time	ever never	public	never	public

Table 7
Comparison of Person references from the keyword lists for Non-Sharers vs Sharers (left) and Future Sharers vs Sharers (right).

Category	Subcategory	Non-Sharers	Sharers	Future Sharers	Sharers
Persons	Children	age boys girl, girls kids niece	child -year- old	age girl, girls kids loli	year- old
	Adults/general	anyone mistress shes someone	team	pedo nurse	–
	Female names	6[*name*]	–	9[*name*]	1 [*name*]
	Male names	3[*name*]	–	7[*name*]	2 [*name*]
Body parts	Female genitals	cunt pussy	–	pussy vulva	–
	Male genitals	cock, cocks	–	glans	–
	General intimate	ass	–	–	–
	Other	mouth	–	mouth	–
Clothing	Underwear	panties	–	–	–

depicted in the images are being described. Besides describing what can be seen on CSAM posted by other users, these users may discuss their own experiences and fantasies related to child sexual abuse. On the other hand, users who share CSAM seem to adhere to a "show, don't tell" principle where descriptions of who and what can be seen in the material is not explicitly mentioned.

4.1.4. *Technical words used by sharers and non-sharers*

Whereas the categories above were more typical for the non-sharing sets, 37–39 % of the key term lists for Sharers are related to technology,

i.e. (CSAM) content, the (CSAM) forum, and the internet or computers in general (cf., 6–8 % in the NS and FS sets).

Regarding (CSAM) content, the majority of key terms for the non-sharing sets (NS and FS) refer to videos, whereas images are more typical for Sharers (see Table 8). Perhaps images are often uploaded by Sharers but invoke less responses and discussion from non-sharing members than videos. In addition, spelling varieties for *previews* are typical for Sharers. Whereas above we saw that Sharers do not necessarily seem to describe the content of their materials in depth, they do explain the technical aspects. The shift from the abbreviation *porn* (Future Sharers) to the full word *pornography* (Sharing) may be an artifact from the analysis.

References to the forum and its structure seem typical for sharing members. This could be explained by the fact that they have spent a longer period of time on the forum and are more invested in the current forum and the larger CSAM community as a whole. On top of forum references, the key terms associated with Sharers show abbreviations such as mods (moderators) and mvp (most valuable player). Such terminology is in-group language typical for forum speech and partly forum-specific. These terms show a high familiarity with the forum subculture. This fits the category of Vested members as described by Van der Bruggen and Blokland (2021).

Finally, Sharers use terms related to the internet or computers. For example, key terms are *back-up, bug*, and *mirror*, which are related to cybersecurity and data management. This involvement in technical safety fits Van der Bruggen and Blokland (2021)'s description of forum managers. Based on their lexical choices, an image emerges where Sharers are more invested in the community and more involved with the management and security of the forum.

4.1.5. References to society used by sharers and non-sharers

Besides references to the forum and the online world, references to the world outside of the forum are typical for Sharers. In the key term lists of non-sharing sets of posts (NS and FS), none of the terms referred to the law, politics, or society in general. For the Sharers, 27–31 % of the key terms refer to law, law enforcement, or security, and another 5–7 % to other societal aspects (see Table 9).

4.1.6. Sharers vs non-sharers

To summarize, once they start sharing, members seem to be more invested in the forum community and the technical aspects of it. In addition, they are concerned with security and legal issues, and with the society outside of their online community. Before members start sharing, they closely resemble non-sharers. Since we know that the most active forum members may potentially all start to share if they remain active long enough (Chiang et al., 2020), there may not be a difference between the pre-sharing and non-sharing sets. Still, in the next section, we compare the future sharers to the non-sharers using keyness analysis to zoom in on any differences.

4.2. Pre-sharing indications of future sharing behavior

In the comparisons with posts from Sharers, the Future Sharers and Non-Sharers sets led to highly similar key terms. This shows that the sets are similar, and that further comparison may not lead to meaningful differences. Using keyness analysis, differences between any sets of texts will lead to a top 100 – even if, in fact, the posts are of similar nature. Still, it is interesting to see which differences emerge from further comparison and if it is possible to predict which non-sharing members will later start to share, based on their lexical choices.

Even when just looking at the distribution in the first two columns of Appendix B, most categories seem rather similar. Disregarding the categories where the number is (nearly) identical (±1 word), the following categories are more typical for Non-Sharers: neutral action words, positive action words, action words related to politeness, sexual action words, action words related to (CSAM) content, positive/sexual adjectives, references to children and adults, female genitals, underwear, and technical terms referring to content. Words more typical for Future

Table 8
Comparison of Technical words from the keyword lists for Non-Sharers vs Sharers (left) and Future Sharers vs Sharers (right).

Category	Subcategory	Non-Sharers	Sharers	Future Sharers	Sharers
Technical	Content	filename.ext	data	2[filename].rar porn	file
		re	image, images	vid	image, images
		vid, vids	mb		key
		video, videos	pornography		link
			preview, previews		mb
			size		pornography
					prev, preview, previews
					size
	Forum	[*username*]	board	re	members
			following	[*username*]	message, messages
			2[*forum name*]		mvp
			hosts		name
			member, members		3[*other forum*]
			mods		posts
			mvp		thread, threads
			posts		4[*username*]
			report, reports		users
			section		
			staff		
			thread, threads		
			3[*username*]		
			users		
			vip		
	Internet/computers	password	application	–	archive
			backup		backup
			bug		browser
			click		bug
			dateline		click
			key		dateline
			mirror, mirrors		mirror, mirrors
			pgp		network
			tor		online

Table 9
Comparison of references to Society from the keyword lists for Non-Sharers vs Sharers (left) and Future Sharers vs Sharers (right).

Category	Subcategory	Non-Sharers	Sharers	Future Sharers	Sharers
Society	Law/ security	–	alleged, allegedly arrested attorney case charged, charges council court crimes criminal enforcement evidence fbi federal investigation investigators judge justice law officers operation order police security sentenced victims	–	Accused allegedly arrested attorney authorities charged, charges court crimes criminal enforcement evidence fbi federal investigation investigators judge law officers officials operation police report, reports reported security sentenced trial victim, victims warrant
	Other	–	county department district government national state vote	–	council county department district government

Sharers are: negative action words, negative descriptions, descriptions of place and time, male names, forum-related words, and ungrouped or unknown words.

Overall, wherever there is a difference, Non-Sharers show the typical picture for non-sharing posts that was painted in 4.1, whereas Future Sharers – when compared to Non-Sharers – resemble the Sharers (see Table 10). Future Sharers are less positive and polite than Non-Sharers, are less sexually explicit in terms of action words and descriptions, and refer less to individuals, body parts, and underwear.[1] Future Sharers use much more references to the forum, showing a higher level of investment in the community. In addition, they use more ungrouped and unknown words, which may be an indication that they talk about a larger variety of unrelated topics and use more community-specific language and abbreviations not known to the researchers. The results indicate a development from Future Sharing members becoming more invested to the point where they start sharing CSAM.

Whereas for most categories, Future Sharers seem to behave already more similar to Sharers in comparison to Non-Sharers, there are some differences. First, Future Sharers use more words related to place and time – to such an extent that this category was not yet included as a category before the analysis of this keyword list. Second, just like the Non-Sharers, Future Sharers typically do not talk about the society outside of the forum such as politics, law, and security. Apparently, these topics become more relevant to the forum members once they start to actively share abuse material to the forum. Table 10 shows the categorized key terms for the NS-FS comparisons.

5. Conclusion

This paper used keyness analysis to compare sharing and non-sharing members of a CSAM forum. First, we aimed to answer the question of how sharing members and non-sharing members differ from each other in terms of keywords. The results indicated that sharing members are more invested in the forum and talk more about technical aspects of the forum, cybersecurity, and the outside world such as laws and politics. These types of terms seem to reflect the *Vested members* and *Managers* categories as described by Van der Bruggen and Blokland (2021). Indeed, Chiang et al. (2020) found that the most active users are most likely to participate in illegal activities on the forum such as uploading material. Non-sharing members, on the other hand, use more positive, polite, and explicit words. This fits the findings by Chiang et al. (2020), who found that image appreciation, rapport-building, and describing experience were typical moves for non-sharers and may be used to compensate for the lack of sharing.

By analyzing the posts of sharers before they actually started sharing, we were able to show that future sharers resemble non-sharers in the sense that they seem to use the same compensation strategies and do not yet talk about issues related to society and (cyber-) security. Hence, although the future sharers partly consist of the same users as the sharers group, their lexical choices more closely resemble the non-sharers. Because many of the differences can be related to the compensation strategies as described by Chiang et al. (2020), it makes sense that the same user group shows a development from pre- to post-sharing. Once they start sharing, the compensations are redundant as they have already proven themselves to be a valuable and authentic part of the community (cf. Chiang, 2024; Grant and MacLeod, 2020).

Finally, to see if future sharers already show indications that they may start to share in the future, the pre-sharing posts were compared to non-sharers. This comparison showed that future sharers are in the middle between non-sharers and sharers, showing some indications of future sharing behavior. Before sharers start sharing, their investment in the forum seems to be higher than for non-sharers because they use more terminology related to the forum. In addition, the compensation strategies are less visible in future sharers than non-sharers, potentially because they are already anticipating on the fact that they will be sharing material in the future.

An important difference between sharers and non-sharers, also absent for future sharers, is the sharers' focus on law and security. Apparently, this only becomes an important topic for forum users once they are sharing material themselves. This is an interesting finding since we know that almost all members of a CSAM forum download illegal material (Van der Bruggen et al., 2022). Perhaps, users who "merely" download feel safe because of the anonymity of the TOR browser or because they perceive themselves as one of too many forum users to be detected by law enforcement. Alternatively, non-sharing members may be less conscious of the illegal nature and trivialize their impact because their crimes are perceived as victimless (cf. Merdian et al., 2014). However, another possibility is that non-sharing members simply are newer to the forum and have less technical and legal knowledge allowing them to participate in such discussions (cf. Chiang, 2024; Nonnecke and Preece, 2001).

Summarizing, similar to Chiang et al. (2020), we found differences between sharing and non-sharing members on a CSAM forum, which gave more insight into the community. By including the lexical choices made by the forum users, we provided more tools for covert operations by law enforcement (cf. Chiang, 2021; Grant and MacLeod, 2016). In addition, the findings can be potentially used to find indications of escalating behavior from non-sharer to sharer. This could help in "triage" when law enforcement aims to focus on certain high-risk individuals, or to be able to intervene before the escalation has taken

[1] With the exception of male names.

Table 10
Comparison of keyness lists for Non-Sharers vs Future Sharers.

Category	Subcategory	Non-Sharers	Future Sharers	Subcategory	Non-Sharers	Future Sharers
Action words	Negative	–	reduce vanned	Politeness	hello please, pls, plz thank, thanks	gracias gratz yeah yup
	Neutral	can, cant extract help im open whats work, working	began hit means tend	Sexual	cum licking sucking	busted
	Positive	appreciated love, loved loving smiled	desire	Content	corrupted download looking reup, reupload upfile upload watched	–
Descriptions	Negative	shahdy	annoying foul shady, shadys	Positive/ sexual	gently great	–
	Neutral	directly normally tiny	limited ripe unintelligable	Place and Time	night	cabin moment positioned season side wall
Persons	Children	cousin niece son, son's	child loli	Adults/ general	anyone aunt mistress someone uncles wife	daddy mr nurse predator
	Female names	11[*name*]	10[*name*]	Male names	4[*name*]	7[*name*]
Body parts	Female genitals	clit cunt slit	vulva	General intimate	ass	buttocks
	Male genitals	cock, cocks	glans penis	Other	finger	arms
Clothing	Underwear	panties underwear	–			
Technical	Content	copy file filename,ext link, links re re-up video, videos	childporn hurtcore pornography	Forum	[*forumname*] messages pass password [*username*]	admins captain dee,rar definition forum hosts layout members nickname section, sections staff torchat troll 6[*username*] vip
	Internet/ computers	virus	e-mail			
Society	Law/security	warrant	–			

place. Similarly, keyness analysis may be used to distinguish other user groups such as producers of new CSAM material and people who are known to have committed an offline child sexual abuse offence. This way, linguistic analyses may aid in stopping undesirable behavior and prevent children from being sexually abused or exploited in the future.

Funding information

This project is funded by the European Union's Internal Security Fund - Police under Grant Agreement 101084525.

CRediT authorship contribution statement

Meike de Boer: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Willemijn Heeren:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Anton Daser:** Writing – review & editing. **Colm Gannon:** Software, Resources, Data curation. **Frederic Gnielka:** Writing – review & editing. **Salla Huikuri:** Project administration, Funding acquisition. **Robert Lehmann:** Writing – review & editing, Supervision. **Rebecca Reichel:** Writing – review & editing. **Thomas Schäfer:** Data curation. **Alexander F. Schmidt:** Writing – review & editing, Supervision. **Katarzyna Staciwa:** Writing – review & editing. **Arjan Blokland:** Writing – review & editing,

Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Key terms excluded from the analysis

Table A

Table A
Ungrouped and Unknown words from the comparisons of keyness lists.

Comparison	Non-Sharers vs Sharers		Sharers pre vs post Sharing		Non-Sharers vs Pre-Sharers	
	Non-Sharers	**Sharers**	**Pre-Sharing**	**Sharers**	**Non-Sharers**	**Sharers Pre-Sharing**
Ungrouped	advance	according	bely	according	advance	darkness
	chance	canary	chance	case	button	giftbox
	directly	former	darkside	information	cat	initially
	really	including	just	note	really	mean
	whats	information	milk	signature	specially	metal
		list	really	text		milk
		name	rhythm			oracle
		oracle	stuff			pen
		project				rather
		signature				square
						well
Unknown	–	p	enh	mo	b	bd
		-t	nk	pw	d	dee
		u.s	r's	-t	m	droppa
			tddc		pm	enh
			x		pw	mk
			yates		v	nk
					vd	r's
						shah
						skee
						tc
						tlz

Appendix B. Overview of the distribution of keywords from the keyness analyses

The distribution of the keywords over the seven themes distinguished in the data can be found in Table B. Each of the three sets was compared to the other two groups. The top row in Table 5 shows the study corpus (in **bold**), i.e., the set for which the keywords were more frequent in comparison to the reference corpus. The set used as the reference corpus is shown in the second row (in *italics*). The first two numerical columns (in grey) show the number of keywords for the Non-Sharers and the Future Sharers in comparison to each other. The middle columns (in white) show the number of keywords for the Future Sharers vs the Sharers. The final two columns (in grey) show the comparisons between the Sharers vs the Non-Sharers.

Table B
Thematical distribution of the top 100 keyword lists for Non-Sharers (NS) in comparison to Future Sharers (FS) and Sharers (S); FS in comparison to NS and S; S in comparison to FS and NS.

	Study corpus	Non-Sharers	Future Sharers		Sharers		Non-Sharers
	Reference corpus	*FS*	*NS*	*S*	*FS*	*NS*	*S*
Category	**Subcategory**						
Action words	Negative	0	2	1	0	0	0
	Neutral	9	4	3	5	4	12
	Positive	5	1	4	0	0	6
	Politeness	6	4	10	0	0	10
	Sexual	3	1	8	2	2	5
	Content (CSAM)	8	0	3	5	4	6
Descriptions	Negative	1	4	3	0	0	2
	Neutral	3	3	8	0	0	5
	Positive/sexual	2	0	13	1	0	13
	Place and Time	1	6	1	1	1	2
Individuals	Children	4	2	5	1	2	6
	Adults/general	6	4	2	0	1	4
	Female names	11	10	9	1	0	6
	Male names	4	7	7	2	0	3

(*continued on next page*)

Table B (*continued*)

	Study corpus	Non-Sharers	Future Sharers		Sharers		Non-Sharers
Body parts	Female genitals	3	1	2	0	0	2
	Male genitals	2	2	1	0	0	2
	General intimate	1	1	0	0	0	1
	Other	1	1	1	0	0	1
Clothing	Underwear	2	0	0	0	0	1
Technical	Content	9	3	4	11	8	6
	Forum	5	21	2	16	21	1
	Internet/computers	1	1	0	10	10	1
Society	Law/security	1	0	0	31	27	0
	Other	0	0	0	5	7	0
Excluded	Ungrouped	5	11	8	6	10	5
	Unknown	7	11	6	3	3	1
TOTAL		100	100	100	100	100	100

The exact keywords in the six lists are presented and discussed in Sections 4.1 and 4.2.

Appendix C. Results from the KWIC analysis

Our data agreement does not allow us to provide all concordances from the KWIC analysis. This appendix shows the distribution of themes identified in the contexts of the keywords "abuse" (Table C1), "watch" (Table C2), (children. For each theme, a short description and some examples are provided. In the examples, the keyword is provided in bold, and the 10 words to the left and right are provided as long as they were part of the same message. Note that in the concordance analysis, punctuation was removed. Occurrences that were either in languages other than English or duplicates were excluded.

Table C1
Themes in the KWIC concordances of the word abuse. In brackets, the total number of occurrences is provided, as well as the number of occurrences included in the thematic coding (i.e., 10 % for the study corpus and 50 for each of the reference corpora).

Theme	Sharers (n = 1690; 169)		Non-Sharers (n = 118; 50)		Future Sharers (n = 72; 50)	
	#	%	#	%	#	%
Society	92	54.4	13	26	15	30
Extreme	25	14.8	5	10	5	10
Neutral	18	10.7	11	22	10	20
Denial	14	8.3	6	12	12	24
Critical	0	0	7	14	1	2
Other	10	5.9	5	10	5	10
Excluded	10	5.9	3	6	2	4
TOTAL	169	100	50	100	50	100

Society = CSA(M) described from the perspective of society outside of the forum, including the victim's perspective, law enforcement, accusations, and quoted news articles.

(1a) NSW government does not know how many reports of sexual **abuse** against children in state care it receives It decided that
(1b) But the Royal Commission into Institutional Responses to Child Sexual **Abuse** last year exposed deficiencies in record keeping by state governments
(1c) identified five different boys who all made allegations of sexual **abuse** There's not much question in my mind that Michael was
(1d) said Recent events brought to light the horrific reality of **abuse** of minors and vulnerable people by members of the church
Extreme = descriptions of extreme forms of abuse, including warnings. Indirectly, this shows that the other forms of abuse on the forum may not be considered as abuse, or at least as less severe.
(2a) or stabbing or needling Serious whipping or hitting violent Verbal **abuse** while fucking the kid Rugged and rough penetration especially with
(2b) I do say calmly that those who rape hurt and **abuse** children should be tortured to the brink of death and
(2c) bad for her do not watch if you can't stand **abuse** quote She does seem simple but no reason to deny
(2d) children some of them very young to the most horrifying **abuse** imaginable He deliberately targeted and groomed vulnerable communities abroad to
(2e) and prosecutors and should look at it from a sexual **abuse** and human trafficking point of view In addition he said

Neutral = occurrences in which the word abuse seems to be used as a "neutral", non-judgmental description of CSA (i.e., the members seem to use society's labels without discussing it).

(3a) with friends or work colleagues and the topic of child **abuse** and pedophiles arises How do you deal with it Most
(3b) instructions on how to produce child pornography and how to **abuse** children You were at the very top of the tree
(3c) two manuals one giving instructions on how to groom and **abuse** a child and the other entitled Producing Kiddie Porn for
(3d) Frank seeking revenge on his x-wife conceives a plan to **abuse** his loving step daughter
(3e) Amazing little girl **abuse** so perverse I love it

Denial = occurrences in which members are explicitly denying that their actions can be considered abuse.

(4a) known fact about us All Pedophiles want to harm and **abuse** children Which in most cases it is the exact opposite
(4b) be produced and you personally are still not causing any **abuse** Yes many children are not pleased to be used in
(4c) daughter and her BFF has to do no rape no **abuse** and no force needed
(4d) this type of show for ppl This definitely is not **abuse** here.Take a look for yourself I for one like this
(4e) had these series for over years who said anything about **abuse** Or whether they are happy or *not I* am providing

Critical = occurrences in which members are critical of the overall abuse taking place on the forum (i.e., not specifically discussing extreme abuse forms).

(5a) seen I gather there's issues of anxiety and cases of **abuse** featured in CP and not just in hurtcore videos But
(5b) porn which makes us believe over time it's ok to **abuse** children Being **abused** as a child which tells us its
(5c) fact say that some children did have enjoyable experiences being **abused** but when they grow up it fucks them up thinking
(5d) the person is harmed when photos of their child sex **abuse** are viewed So the way to solve this is to

Other = references to other forms of abuse, i.e., not sexual abuse, including abuse of power, violations of the rules, and substance abuse (Table C2).

Table C2
Themes in the KWIC concordances of the word watch. In brackets, the total number of occurrences is provided. Because the main purpose of the KWIC analysis was to verify whether the action word referred to the passive act of consuming CSAM, only the two most different groups were compared. The word was key for the NS group, of which 10 % of the occurrences were analyzed, in comparison to the S group, of which the first 50 occurrences were sampled.

Theme	Sharers (n = 1468)		Non-Sharers (n = 526)	
	#	%	#	%
CSAM	22	44	31	58.5
Offline	2	4	14	26.4
Security	9	18	0	0
Other	6	12	8	15.1
Excluded	11	22	0	0
TOTAL	50	100	53	100

CSAM = The action word *watch* refers to watching CSAM or watching a child.

(6a) or not be tortured and beaten I for one cannot **watch** nor enjoy videos that show someone being forced into sex
(6b) being caught i'll just hang out on pedo boards and **watch** and fap i'll add my big thank you to the
(6c) ooh i could **watch** her all day with that suck mouth thanks for posting
(6d) The link isnt working for me so i cant **watch** or download this video

Offline = The action word *watch* refers to watching children in real life, or children watching the adult masturbating in real life.

(7a) a seedy pervert getting them to jack me off or **watch** me while I jacked myself off Cant beat the good
(7b) planned to get them into the house so I could **watch** them closer I made snacks and drinks for the girls
(7c) on behind her I was too scared to have her **watch** me cum in front of her and really didnt want
(7d) brought her sister to a crack in my door to **watch** me masturbate and cum to porn all the time I

Security = The members are warning each other for law enforcement agencies on the forum, or are referring to court and punishment.

(8a) LEA are there fishing around but keep it tight and **watch** the files coming in I have seen a flood of
(8b) its LEA but I do think its a sinking ship **Watch** yourself out there friends or you may find yourself swimming
(8c) Barr a senior researcher on women's rights with Human Rights **Watch** the New York-based organization Protecting children from sexual abuse requires
(8d) other social medias you can see him wearing the same **watch** Another big mistake he made was putting pictures of some

Other = The action word *watch* refers to something other than CSA(M) (Table C3).

Table C3
Themes in the KWIC concordances of the descriptions potentially describing children. In brackets, the total number of occurrences is provided. Because the main purpose of the KWIC analysis was to verify whether the descriptions referred to children, only the first 50 occurrences from the Non-Sharing set were analyzed.

	little (n = 2277)		*younger* (n = 322)	
	#	%	#	%

(*continued on next page*)

Table C3 (*continued*)

	little (*n* = 2277)			*younger* (*n* = 322)	
Child	30	60	Child	43	86
Child's body	13	26	Adult	5	10
Other (sexual)	4	8	Excluded	2	4
Other (non-sexual)	3	6			
TOTAL	50	100		50	100

Child = reference to the child as a whole, including *girl, boy, child*, and derogatory terms.
Child's body = references to the body or body parts of a child, including their genitals.
Other (sexual) = uses of *little* not referring to children or their bodies, but used in a sexual context.
Other (non-sexual) = non-sexual uses of the term *little*.
Adult = references to the *younger* self of the poster.
NB Examples are not provided due to their graphic content and lack of additional information (Table C4).

Table C4
Themes in the KWIC concordances of ever and never. In brackets, the total number of occurrences is provided. Because the main purpose of the KWIC analysis was to verify whether the descriptions referred to hands-on CSA, only the first 50 occurrences from the Non-Sharing set were analyzed.

	ever (*n* = 927)		*never* (*n* = 1486)	
	#	%	#	%
Hands-on	19	38	6	12
CSAM	14	28	18	36
Loneliness	6	12	4	8
Pedophilia	0	0	10	20
Law enforcement	4	8	3	6
Technical/forum	4	8	8	16
Other	2	4	1	2
Excluded	1	2	0	0
TOTAL	50	100	50	100

Hands-on = references to (hypothetical) hands-on CSA.

(9a) with a nice ass shes really the only preteen im **ever** gonna be interested in fucking and guess i just needed
(9b) her that it was the most amazing thing I had **ever** felt and that she was really good at it She
(9c) I for one believe it is better to **never** touch a child
(9d) a promise to myself that no matter what i would **never** hurt a kid never do anything that they clearly don't

CSAM = references to watching CSAM.

(10a) This is one of the hottest things I have **ever** seen D Also the password works fine Thank you for
(10b) one of the best i have **ever** seen wish there more of her
(10c) yes i have **never** seen things like this on [XXX] where did you find

Loneliness = references to feeling isolated and lonely in the outside world.

(11a) of I guess I am the loneliest person I have **ever** met or heard of Also my family sees me as
(11b) as I'm concerned he will be the only one I'll **ever** tell that I'm a pedo
(11c) No and nobody **ever** will or I would probably commit suicide I wouldn't let
(11d) how my life would be if a opened I have **never** ever met a single person who usually speaks about important
(11e) who say sexuality is a choice are quite wrong I **never** chose to be a pedo it's just what got me

Pedophilia = references to the member's identity a pedophile

(12a) young to teen and from teen to adult I was **never** born a pedophile I went to church til well my
(12b) are repulsed by adults bodies of both sexes I have **never** found anything remotely interesting about females I never saw the
(12c) as I can remember I feel no remorse for it **Never** had and never will I know some of you are (Table C4)

Table C5
Themes in the KWIC concordances of public. In brackets, the total number of occurrences is provided. For the study corpus, the first 100 occurrences were sampled.

	Sharers (*n* = 2202)	*Non-Sharers* (*n* = 136)

(*continued on next page*)

Table C5 (*continued*)

	Sharers (n = 2202)		*Non-Sharers (n = 136)*	
	#	%	#	%
Internet	85	85	24	48
General	9	9	7	14
Location	2	2	12	24
Information	2	2	1	2
Security	0	0	4	8
Excluded	2	2	2	4
TOTAL	100	100	50	100

Internet = references to public websites, public parts of the forum, or other "public" places on the internet. E.g., 64 occurrences in the Sharers set were part of the word group *public key*.

General = reference to the general public, including public opinion.

Location = reference to CSA happening in a public space, such as a public bathroom.

Information = information (made) known to the public, including because of LEA detection.

Security = references to not watching CSAM in public to avoid detection.

References

Anderson, S. (2024). Universal classification schema: harmonizing terminology for child sexual exploitative material – an application-level ontology (Version 2.0). retrieved from INHOPE on 21 February 2024. [Classified document.].

Baker, P., 2004. Querying keywords: questions of difference, frequency, and sense in keywords analysis. J. Eng. Linguist. 32 (4), 346–359. https://doi.org/10.1177/0075424204269894.

Benoit, K.h, Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., Matsuo, A., 2018. quanteda: an R package for the quantitative analysis of textual data. J. Open Source Softw. 3 (30), 774. https://doi.org/10.21105/joss.00774.

Blokland, A., Daser, A., de Boer, M., Gannon, C., Gnielka, F., Huikuri, S., Reichel, R., Shäfer, T., Schmidt, A.F., Staciwa, K., Lehmann, R., 2024. Why do users continue to contribute to darknet child sexual abuse material forums? examining social exchange, social capital, and social learning explanations using digital forensic artifacts. Child Abuse Negl. 153, 106815. https://doi.org/10.1016/j.chiabu.2024.106815.

Chiang, E., Nguyen, D., Towler, A., Haas, M., Grieve, J., 2020. Linguistic analysis of suspected child sexual offenders' interactions in a darknet image exchange chatroom. Int. J. Speech Lang. Law 27 (2), 129–161. https://doi.org/10.1558/ijsll.41446.

Chiang, E., 2021. Send me some pics': performing the offender identity in online undercover child abuse investigations. Polic.: J. Policy Pract. 15 (2), 1173–1187. https://doi.org/10.1093/police/paaa065.

Chiang, E., 2024. I read the rules and know what is expected of me": the performance of competence and expertise in 'newbie' offenders' membership requests to dark web child abuse communities. Discourse Context. Media 57, 100744. https://doi.org/10.1016/j.dcm.2023.100744.

Europol (2022). *Child sexual exploitation*. Retrieved from https://www.europol.europa.eu/crime-areas/child-sexual-exploitation.

Europol, 2024. Internet Organised Crime Threat Assessment (IOCTA) 2024. Publications Office of the European Union, Luxembourg.

Fonhof, A.M., van der Bruggen, M., Takes, F.W., 2019. Characterizing key players in child exploitation networks on the dark net. In: Complex Networks and Their Applications VII: Volume 2 Proceedings The 7th International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2018 7. New York, NY. Springer International Publishing, pp. 412–423.

Gabrielatos, C., 2018. Keyness analysis: nature, metrics and techniques. Eds. In: Taylor, C., Marchi, A. (Eds.), Corpus Approaches to discourse: A critical Review. Routledge, pp. 225–258. https://doi.org/10.4324/9781315179346.

Gannon, C., Blokland, A.A.J., Huikuri, S., Babchishin, K.M., Lehmann, R.J.B., 2023. Child sexual abuse material on the darknet. Forens. Psychiatr. Psychol. Kriminol. https://doi.org/10.1007/s11757-023-00790-8.

Glasgow, K., Schouten, R., 2014. Assessing violence risk in threatening communications. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. https://doi.org/10.3115/v1/W14-3205.

Gnielka, F.M., Reichel, R., Blokland, A., Daser, A., De Boer, M., Gannon, C., Schmidt, A. F., Schäfer, T., Huikuri, S., Staciwa, K., Lehmann, R.J.B., 2024. Missing the mark? Identifying child sexual abuse material forum structure and key-players based on public replies and private messaging networks. Humanit. Soc. Sci. Commun. 11 (1), 1459. https://doi.org/10.1057/s41599-024-03954-x.

Grant, T., MacLeod, N., 2016. Assuming identities online: experimental linguistics applied to the policing of online paedophile activity. Appl. Linguist. 37 (1), 50–70.

Grant, T., MacLeod, N., 2020. Language and Online identities: The undercover Policing of Internet Sexual Crime. Cambridge University Press.

Grover, T., Mark, G., 2019. Detecting potential warning behaviors of ideological radicalization in an alt-right subreddit. In: Thirteenth International AAAI Conference on Web and Social Media. https://doi.org/10.1609/icwsm.v13i01.3221.

Handelman, L.D., Lester, D., 2007. The content of suicide notes from attempters and completers. Crisis 28 (2), 102–104. https://doi.org/10.1027/0227-5910.28.2.102.

Herring, S.C., 2004. Computer-mediated discourse analysis: an approach to researching online behavior. Eds. In: Barab, S.A., Kling, R., Gray, J.H. (Eds.), Designing For Virtual Communities in the Service of Learning, pp. 338–376.

Hunter, M., Grant, T., 2025. Is LIWC reliable, efficient, and effective for the analysis of large online datasets in forensic and security contexts? Appl. Corpus Linguist. 5 (1), 100118. https://doi.org/10.1016/j.acorp.2025.100118.

Hutchins, W.J., 1977. On the problem of 'aboutness' in document analysis. J. Inform. 1 (1), 17–35.

Kilgarriff, A., 1997. Using word frequency lists to measure corpus homogeneity and similarity between corpora. In: Proceedings 5th ACL workshop on very large corpora. Beijing and Hong Kong, pp. 231–245.

Kloess, J.A., van der Bruggen, M., 2021. Trust and relationship development among users in Dark Web child sexual exploitation and abuse networks: a literature review from a psychological and criminological perspective. Trauma Violence Abuse 24 (3), 1220–1237. https://doi.org/10.1177/15248380211057274.

Luhn, H.P., 1959. Keyword-in-Context Index for Technical Literature (KWIC Index). IBM Advanced Systems Development Division, Yorktown Heights, New York.

Merdian, H.L., Curtis, C., Thakker, J., Wilson, N., Boer, D.P., 2014. The endorsement of cognitive distortions: comparing child pornography offenders and contact sex offenders. Psychol. Crime Law 20 (10), 971–993. https://doi.org/10.1080/1068316X.2014.902454.

Moreno, A.I., Swales, J.M., 2018. Strengthening move analysis methodology towards bridging the function-form gap. Engl. Specif. Purp. 50, 40–63. https://doi.org/10.1016/j.esp.2017.11.006.

Nonnecke, B., Preece, J., 2001. Why lurkers lurk. In: Proceedings from Americas Conference on Information Systems. Boston, USA.

Nurmi, J., Paju, A., Brumley, B.B., Insoll, T., Ovaska, A.K., Soloveva, V., Vaaranen-Valkonen, N., Aaltonen, M., Arroyo, D., 2024. Investigating child sexual abuse material availability, searches, and users on the anonymous Tor network for a public health intervention strategy. Sci. Rep. 14 (1), 7849. https://doi.org/10.1038/s41598-024-58346-7.

Owen, G., Savage, O., 2015. The TOR Darknet. Global Commission on Internet Governance, London. Retrieved from. https://www.cigionline.org/static/documents/no20_0.pdf.

Partington, A., 2014. Mind the gaps: the role of corpus linguistics in researching absences. Int. J. Corpus Linguist. 19 (1), 118–146. https://doi.org/10.1075/ijcl.19.1.05par.

Pennebaker, J.W., Chung, C.K., 2009. Computerized text analysis of Al-Qaeda transcripts. In: Krippendorff, K., Bock, M.A. (Eds.), The Content Analysis Reader. Sage.

Pennebaker, J.W., Francis, M.E., Booth, R.J., 2001. Linguistic Inquiry and Word Count: LIWC 2001. Erlbaum, Mahwah, NJ.

Phillips, M., 1989. *Lexical Structure of text*. Discourse Analysis Monograph no. 12. English Language Research, University of Birmingham.

Prichard, J., Wortley, R., Watters, P., Spiranovic, C., Scanlan, J., 2024. The effect of therapeutic and deterrent messages on internet users attempting to access 'barely legal' pornography. Child Abuse Negl. 155, 106955. https://doi.org/10.1016/j.chiabu.2024.106955.

Quayle, E., Taylor, M., 2002. Child pornography and the internet: perpetuating a cycle of abuse. Deviant Behav.: Interdiscip. J. 23, 365–395. https://doi.org/10.1080/01639620290086413.

R Core Team, 2024. R: A language and Environment For Statistical Computing (v4.4.2). R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Sönning, L., 2024. Evaluation of keyness metrics: performance and reliability. Corpus Linguist. Linguist. Theory 20 (2), 263–288. https://doi.org/10.1515/cllt-2022-0116.

Scott, M., 1997. PC analysis of key words - and key key words. System 25 (2), 233–245. https://doi.org/10.1016/S0346-251X(97)00011-0.

Strauss, A.L., Corbin, J., 2004. Open coding. Ed. In: Seale, C. (Ed.), Social Research methods: A reader. Psychology Press, pp. 303–306.
Swales, J. (1981). *Aspects of article introductions: Aston ESP Research Reports No. 1*. Language Studies Unit, Aston University, Birmingham.
Swales, J., 1990. Genre Analysis: English in Academic and Research Settings. Cambridge University Press, Cambridge.
Tausczik, Y.R., Pennebaker, J.W., 2010. The psychological meaning of words: LIWC and computerized text analysis methods. J. Lang. Soc. Psychol. 29 (1), 24–54. https://doi.org/10.1177/0261927×09351676.
Taylor, C., 2013. Searching for similarity using corpus-assisted discourse studies. Corpora 8 (1), 81–113. https://doi.org/10.3366/cor.2013.0035.
Van der Bruggen, M., Blokland, A., 2021. A crime script analysis of child sexual exploitation material fora on the darkweb. Sex Abuse 33 (8), 950–974. https://doi.org/10.1177/1079063220981063.
Van der Bruggen, M., van Balen, I., van Bunningen, A., Talens, P., Owens, J.N., Clapp, K., 2022. Even "lurkers" download: the behavior and illegal activities of members on a child sexual exploitation TOR hidden service. Aggress. Violent. Behav. 67, 101793. https://doi.org/10.1016/j.avb.2022.101793.
Woodhams, J., Kloess, J.A., Jose, B., Hamilton-Giachritsis, C.E., 2021. Characteristics and behaviors of anonymous users of dark web platforms suspected of child sexual offenses. Front. Psychol. 12, 623668. https://doi.org/10.3389/fpsyg.2021.623668.
Wortley, R., Findlater, D., Bailey, A., Zuhair, D., 2024. Accessing child sexual abuse material: pathways to offending and online behaviour. Child Abuse Negl. 154, 106936. https://doi.org/10.1016/j.chiabu.2024.106936.