


Articles

A corpus-assisted discourse analysis of children's and groomers' talk in online grooming interactions

Craig Evans^a, Nuria Lorenzo-Dus^{b,c,*} ^a Independent^b Swansea University, Singleton Park, Swansea, SA2 8PP, Wales, UK^c Universitat Politècnica de Valencia, Spain

ARTICLE INFO

Keywords:

Online grooming
Discourse
Children's online communication
Manipulation discourse
Corpus Assisted Discourse Studies
Online Child Sexual Exploitation and Abuse

ABSTRACT

Harmful communication may not always be recognisable as such, especially when it is manipulative and deceptive and appears to be indistinguishable from innocuous communication. This is the case with online child sexual grooming, where talk from interactions between groomers and children may resemble that seen between friends or consenting adults chatting. However, recognising that online grooming may be taking place is not simply a matter of spotting tell-tale words or phrases. It requires engaging with ways that online grooming is discursive: involving groomers and children using language to perform particular functions as they pursue different goals through a dynamic exchange. We address this need in this study by providing the first ever *complete* account of online grooming discourse, one that identifies features not only of groomers' talk but also of children's, using collocates of the most frequent content words in a corpus of each. Comparing findings between the two highlights distinctiveness that helps make online grooming communication more identifiable. It also reveals strong similarity, perhaps reflecting groomers' efforts to minimise perpetrator/victim contrast for deception purposes. An advantage of using a corpus-assisted discourse studies approach, as found in our study, is that it can uncover subtle, non-obvious patterns that may serve as indicators of online grooming despite such deception.

1. Introduction

Online grooming is a pervasive and escalating form of online child sexual exploitation and abuse (OCSEA). According to the National Society for the Prevention of Cruelty against Children (NSPCC, 2024), the number of *police-recorded* cases of online grooming in the United Kingdom (UK), for instance, has increased by 89 % since 2017/2018, when the offence of online grooming first came into force in the country. The actual number of online grooming offences is likely to be considerably higher, given that OCSEA in general, and online grooming in particular, are known to be significantly under-reported and therefore under-recorded officially. More positively, stakeholder interest in preventing OCSEA, specifically online grooming, has also increased in recent years, with emerging legislation – such as the UK's Online Safety Act¹ (2023) – addressing online grooming more centrally than hitherto. From an academic perspective, novel conceptualisations of online grooming as a *communicative* practice of manipulation have also been

developed, which deploy linguistics methods to identify patterns in language use within such practice. Most of this research, however, has exclusively examined the language of offenders. While this new linguistic lens on what is essentially a linguistic practice is welcome and has yielded significant results (see Section 2), it overlooks the dynamic, two-way (offender–child) nature of the communicative practice of online grooming. As a communicative practice of manipulation, moreover, online grooming is far from a level playing field: during online grooming interactions, offenders take advantage of what is already an imbalanced power relation with the children they abuse. Children's communicative engagement during online grooming interactions is thus severely constrained. The comparative dearth of research into children's language use during grooming interactions, including the ways in which their ability to exercise communicative agency is curtailed – that is, groomer–child interactional dynamics – thus constitutes an important gap in knowledge which the study reported in this article seeks to fill. To this end, a primary aim of the present study is to provide a *complete* account

* Corresponding author.

E-mail addresses: n.lorenzo-dus@swansea.ac.uk, nulodu@idm.upv.es (N. Lorenzo-Dus).¹ <https://www.gov.uk/government/publications/online-safety-act-explainer/online-safety-act-explainer>

of online grooming discourse, one that identifies features both of groomers' and children's talk.

The article is structured as follows. Section 2 provides a necessarily brief overview of extant work on online grooming as a practice of communicative manipulation, including groomers' tactical language use and incipient research into children's communication during online grooming. This is followed, in Section 3, by a description of the methodology adopted in the study, with a focus on the particular techniques deployed within the corpus-assisted discourse studies (CADS) approach followed. Section 4 firstly identifies a set of features representative of the discourse of children, and a set representative of the discourse of groomers, and then discusses each feature comparatively, exploring both similarities and differences. Section 5 discusses key findings of the study.

2. Online grooming as two-way communication

Online grooming is a digital practice of communicative manipulation, whereby an adult uses discourse (i.e., text, images, etc.) to lure a child into partaking in sexual activity online and/or offline (Lorenzo-Dus et al., 2016, 2020; Lorenzo-Dus, 2022). Foundational work on this issue was produced within the fields of, primarily, psychology, criminology and the computer sciences (see, e.g., Whittle et al., 2013; Babchishin et al., 2015). Across these studies, the importance of looking at language use was acknowledged, albeit that generally in an implicit manner. Within the computer sciences, for instance, early work on online grooming detection called for micro-level, contextualised description of the linguistic means via which groomers seek to fulfil their abusive goals online (see, e.g., Kontostathis et al., 2009; Cano Basave et al., 2014; Bogdanova et al., 2014; Vartapetian and Gillam, 2014).

A growing number of studies of online grooming discourse have emerged since the mid-2010s. These differ from previous language-based studies that had adopted decontextualised language analysis methods, often through implementation of psycholinguistic software such as Linguistic Inquiry Word Count (LIWC – Pennebaker et al., 2015; see Lorenzo-Dus and Kinzel, 2019, for a critical review of this work). The majority of studies within what may be seen as a 'second wave' of language-based research, with a focus on discourse, have nevertheless continued to assign primacy to groomers' input and the tactics they deploy to groom children online (see, e.g., Lorenzo-Dus et al., 2016, 2020, 2023; Chiang and Grant, 2018; Schneevogt et al., 2018; Lorenzo-Dus, 2022; Pérez-Sabater et al., 2024). Examples of these tactics include trust building for deceptive purposes via the use of compliments (Lorenzo-Dus and Izura, 2017); questions about relations, hobbies and so on; self-disclosure about 'vulnerabilities' (e.g., feeling lonely); use of implicit (Lorenzo-Dus and Kinzel, 2021) and explicit sexual terms (e.g., Grant and Macleod, 2020); use of requests to meet offline and/or online (Kinzel, 2021; Lorenzo-Dus, 2022); use of threats (Schneevogt et al., 2018); and use of criticism and othering of children's social support networks, principally family and peers, as a way to create distance from them and a consequent sense of emotional isolation in the child (Lorenzo-Dus, 2022).

The focus on groomers, and their discourse, likely owes to two practical reasons. One of them is the applied research goals of most studies of online grooming and other forms of OCSEA, which have prioritised detection-orientated tasks – that is, identifying language signals of grooming, which have been equated with groomers' language signals. Such research has sought to derive results that, for example, may 'constitute a scientific basis for legal arguments' (Drouin et al., 2017, p. 453) or help 'refine the strategies used by law-enforcement bodies' (Chiang and Grant, 2017, p. 124) to counter the bad actors involved (i.e., the perpetrators).

Another reason for the focus on groomers lies in restrictions on access to authentic online grooming data; that is, online grooming chatlogs that involve offenders and actual children. Most research into online grooming has examined a dataset containing c. 600 transcripts of US-

based decoys posing as children as they interacted with offenders online: the Perverted Justice dataset. The advantages and disadvantages of using this dataset for language-based analysis have been debated. The former include the ability to derive valuable information about groomers' strategic language use (Chiang and Grant, 2017) – something that, to our knowledge, only one study (Lorenzo-Dus et al., 2023) has empirically tested. The authors used the same taxonomy to annotate groomer discourse in the entirety of the Perverted Justice dataset and in a dataset shared by UK law enforcement entailing online grooming chatlogs between offenders and actual children. The results showed high similarity levels regarding the comparative frequency of key tactics (Lorenzo-Dus et al., 2023). As for the disadvantages, these include the possibility that decoys use language differently to children. This seems a logical assumption to make. Studies have been conducted that show differences in the language of decoys when compared to groomers (Drouin et al., 2017), as well as in the language of groomers interacting with decoys posing as boys as compared to girls (van Gijn-Grosvenor and Lamb, 2016). However, to our knowledge, no study to date has compared decoy and child language in online grooming contexts.

The marked research focus on groomers' discourse has led to conflating groomers' communicative goals with the goals of online grooming communication, without sufficiently explicit acknowledgement of the fact that the children in online grooming interactions also have communicative goals. In doing so, the practice of online grooming has been equated to groomers' practice and, therefore, only partially accounted for. Yet, online grooming entails two-way interaction – between an offending adult and a child.² Just like most other one-to-one interactional contexts, it is only possible to understand what groomers are doing through their talk by seeing this in relation to what children are doing through their talk. Interpersonal dynamics are, thus, central to online grooming discourse. Within such dynamics, moreover, facework is crucial, as a relatively recent body of online grooming research shows. This has focused primarily on the importance of groomers' strategic facework (e.g., Grant and Macleod, 2020; Lorenzo-Dus, 2022; Pérez-Sabater et al., 2024; Schneevogt et al., 2018). It has also started to be examined in relation to children's language use, especially when communicatively resisting groomers' tactical discourse (e.g., Kloess et al., 2017; Chiang and Grant, 2018; Thomas et al., 2023; Lorenzo-Dus and Evans, 2025). In these studies, and our work, facework is understood as being a fundamental, structural feature of social interaction that involves all participants conducting themselves according to a social code that is governed by a set of unwritten rules and context-specific, standardised practices (Goffman, 1967; 1981). These serve to uphold the positive public image – or face – of every individual taking part in any given encounter, including through discourse 'politeness' that seeks to attend to one's interactant face needs. Face needs, in turn, may be further broken down, in different ways. They may be conceptualised, for example, as a general orientation to positive or negative politeness, respectively understood as the desire to be liked and appreciated or not to be imposed upon (Brown and Levinson, 1987). Alternatively, facework may be seen in terms of managing identity and social rapport in interaction (Spencer-Oatey, 2000; 2008), whereby a distinction is made between quality face (positive self-image from personal qualities) and social identity face (derived from group memberships/roles). Face needs may also be threatened, a discourse practice that may be examined through the notion of impoliteness and operationalised in terms of positive / negative face attack orientation (e.g., Culpeper 1996, 2005), or face/sociality rights/goal attack orientation (Culpeper, 2011), amongst other.

Indeed, over the past decades, the concepts of facework and politeness/impoliteness have been extensively reformulated conceptually,

² There are contexts in which online grooming occurs within multi-participant interactions, involving, for example, more than one offender. These are not within the scope of the present study.

methodologically and empirically within the field of linguistics. Facework has for instance been reconceptualised as relational work (e.g., Locher and Watts, 2005), a tripartite distinction having been drawn between polite (marked, positively evaluated behaviour), politic (socially appropriate or unmarked) and impolite (marked, negatively evaluated behaviour) in context. Politeness and impoliteness approaches have also been scrutinised in terms of, for instance, cross-cultural/linguistic relativity, intentionality, and emic/etic validity (for a critical review see, e.g., Garcés-Conejos Blitvich and Sifianou, 2019). Moreover, most of these approaches have been operationalised via taxonomies – such as lists of politeness and impoliteness strategies, triggers and formulae. As Garcés-Conejos Blitvich (2010) and Garcés-Conejos Blitvich et al. (2011) and Culpeper (2016), amongst other scholars, argue, if handled critically, such taxonomies constitute helpful analytic categories that expose behavioural patterns. Such critical handling includes an understanding that the use of taxonomies does not entail adopting a top-down, analyst approach. As advocated within the so-called third wave of (im)politeness research (see, e.g., Kádár and Haugh, 2013), analysts (here, the study authors) need to consider how interactants (here, groomers and children's) produce and take up each other's facework on a moment-by-moment basis.

Overall, then, facework – and the use of discourse politeness and impoliteness – is particularly salient in interactional contexts in which meaning negotiation is sensitive, indeed illegal, as is the case in cyber-crime communication, including online grooming.³ A case in point is groomers' manipulative use of reciprocated disclosure as facework to foster closeness with and trust in the children they prey on (Lorenzo-Dus, 2022). Groomers perform this type of disclosure, amongst other ways, through mirroring children's discourse – strategically deploying linguistic accommodation (Giles et al., 1991). Similarly, power imbalance and other important contextual constraints, may result in children's use of facework that seeks to minimise conflict through, for instance, the use of indirectness when trying to 'negotiate' image sending or meeting offline (e.g. Kloess et al., 2017; Lorenzo-Dus and Evans, 2025) with groomers. It is only by examining the dynamics of groomer–child interaction, therefore, that we can enhance our current knowledge of online grooming as a communicative practice of manipulation.

The value of looking at children's talk in research on online communication is not limited to how this contributes to knowledge and understanding about groomers' talk as part of a dynamic interactional process. Children's talk is important in its own right. In this sense, it is worth reminding ourselves of the fact that children have (digital) communicative agency. The term agency is here understood as engagement – displaying their ability to communicate, including in digital environments. Agency does not assign any responsibility – let alone blame – to children for the abuse that online grooming communication carries, which relies exclusively and always with the offending adult: the groomer. Critically, in the context of online grooming, children's communicative agency is seriously constrained by power imbalance and other factors, including digital affordances that enable, for example, unsolicited access to them by bad actors.⁴

Despite the importance of children's digital agency, which is realised via discourse, there is a comparative dearth of research into the characteristics of children's language use in digital environments in general (but see Christakis and Hale, 2025) and in relation to sexuality in particular. Regarding the latter, the focus has been on sexting – primarily consensual sexting practices between children (Barroso et al., 2023). The limited research into children's digital discourse around

sexuality has paradoxically often been conducted using non-linguistic tools, favouring self-reported methods, primarily surveys. This research has revealed that children regularly use digital media as a 'developmental forum' for sexual experimentation (see, e.g., Quayle et al., 2012; Cooper et al., 2016). Their natural curiosity and pushing of normative boundaries around sex and relationships are well attested (see, e.g., Ningrum and Kusbaryanto, 2021). These must be factored in when examining their digital communicative practices.

A small number of studies have examined child–groomer interactional dynamics in relation to specific 'activities', mainly children's attempts at resisting groomers' advances, as expressed through their digital interactions, alongside groomers' reactions to such resistance (Kloess et al., 2017; Thomas et al., 2023; Lorenzo-Dus and Evans, 2025), and groomers' use of threats alongside children's attempts at managing them (Seymour-Smith and Kloess, 2021). These studies coincide in identifying a wide range of discourse strategies – from excuse and plea making through to explicit statements of intention to report abuse – that children deploy in their interactions with groomers online. To our knowledge, however, no research has produced a holistic account of children's discourse during online grooming alongside groomers' language use. This study contributes to filling this important gap in knowledge. It adopts a CADS approach to investigate child and groomer patterns of language use in online grooming interactions, as next described.

3. Methodology

3.1. Data

The corpus used in this study comes from a large dataset of chatlogs that reproduce interactions between groomers and children via a variety of social media platforms over a five-year period (2014–2019). The chatlogs were shared by UK law enforcement for the purposes of developing research-informed, anti-online grooming solutions. The dataset was safely stored and anonymised by the research team (which includes the study authors) prior to analysis. Corpus pre-processing, including spelling standardisation and transliteration of emojis and other graphicons, was also undertaken prior to analysis.

From this dataset, we built a specialised corpus comprising 75 chatlogs. Specialised corpora tend to be the norm in studies that concern cyber-crime, where the text and communicative contexts feature at large socially – as attested by escalating figures for online grooming (see Section 1) – but access to texts by researchers is particularly challenging (for the reasons discussed in Section 2). This makes it extremely difficult to collect large and balanced corpora (see Nini, 2020; Wright, 2025). In our case, we adhered as best as possible to established principles of corpus building, including representativeness, when building our specialised corpus on online grooming interactions. The data-sharing agreement in place prevents disclosure of specific demographics regarding the groomers' and children's gender, age and other social features. However, the groomers in the specialised corpus were all adults and over two-thirds of them were males; over 80 % of the children were girls aged 13–17. We also selected chatlogs that were diverse in length and duration, excluding those that had fewer than 200 words. In all cases, the chatlogs entailed one-to-one interactions – instances of multi-party conversation were excluded for this study.

For the purposes of producing an account of children's talk, the specialised corpus was divided into two sub-corpora using the web-based corpus software CQPWeb:⁵ one consisting of the children's turns

³ Given the richness and complexity of this field of research, the well-known terms of facework and politeness/impoliteness are used as shorthand from this point onwards.

⁴ For a discussion of the term agency in the context of child sexual exploitation and abuse, see, e.g., Beckett (2019).

⁵ <https://cqpweb.lancs.ac.uk>. We deployed a local installation of the software within a secure digital platform made available at the authors' institution, such that the anonymised dataset was never shared beyond said platform. Access to the data was restricted to the research team under strict security measures and as per the terms of the data sharing agreement with law enforcement.

from the chatlogs (hereafter C—Corpus) and the other consisting of the groomers' turns (hereafter G—Corpus). C—Corpus consists of 42,715 words and G—Corpus of 69,191 words.

3.2. Framework and procedure

Full ethical approval was obtained from the authors' institution prior to commencing the study. A robust researcher well-being protocol was also devised and implemented from the outset – and remains active beyond the 'analysis phase' of the project.

As stated in Section 1, a central aim of this paper is to provide an account of children's and groomers' talk, that is, an overall impression of what their talk is made up of/concerned with. This may include both representational aspects (e.g., meaning, topics) and performative aspects (e.g., interpersonal actions) of language use. By 'account', we mean a cumulation of the most salient language-based features. This must consider distinctness as well as similarity – an account of a specified discourse should be just that; that is, not made up of features that could also constitute an account of another discourse that may be very similar.⁶ Also, distinctness is necessary for achieving our study's applied research purpose of making online grooming more identifiable/recognisable.

CADS is an especially effective approach for highlighting 'the features of particular discourse types' (Partington, 2010, p. 88) by combining quantitative results that are semi-automatically produced by corpus techniques with 'close reading approaches' (Egbert and Baker, 2020, p. 8). This is achieved in a principled way, using frequency and statistical measures to identify language-based patterns in a corpus and, therefore, a discourse the corpus represents.

Our approach combines two well-established corpus linguistics procedures for analysing discourse: word frequency and collocation. We use these as part of a two-staged process, starting by generating a list of the most frequent content words in each corpus. We then looked at the collocates of these to help identify features of each corpus and the children's and groomers' talk they represent. Collocate patterns are also likely to constitute distinct elements. This is because collocates are words that regularly co-occur in near proximity where 'the relationship' between them 'is statistically significant in some way' (Baker, 2006, pp. 95–96) and they are therefore likely to reflect context-specific choices that may be features of particular discourses.

Following data redaction (see Section 3.1), we created lists of the most frequent words in LancsBox⁷ (factoring in a normalised dispersion using DPNorm score <0.6), one each for G—Corpus and C—Corpus.⁸ We manually filtered out function words to leave only content words, as function words are highly productive and, individually, can represent multiple meanings/functions (and are therefore not reliable for identifying meaning/function patterns). We next manually reviewed the wordlists and selected the top 20 most frequent content words in each list for further analysis. We judged that the top 20 words in both lists included enough lexical variety to allow for consideration of a range of features. Limiting to 20 each (40 in total) represented a balance of recall and precision, where 40 represented a manageable figure for analysing the collocates of each. Collocates for each of the 20 most frequent content words in both corpora (the 10 strongest or all collocates if there were fewer than 10) were then generated using CQPWeb. We used the statistical measure Mutual Information (MI) and a collocation window

of three words to the left and three to the right, with the requirement for collocates to occur a minimum of five times.⁹

Next, we compiled a list of 'collocate pairs' (each one of the content words in the list of 20 most frequent combined with each of its collocates). We examined them manually to identify thematic categories from each of them and then allocated each pair to one or more category. In order to check the accurate inference of meaning and/or function of the collocates when they occur with the highly frequent words, we reviewed all concordances (or up to 50 where there were more than 50) for each collocate pair. The concordance review process involved reading expanded extracts to ensure meaning disambiguation (see Gillings and Mautner, 2024). We repeated this process for both G—Corpus and C—Corpus, further identifying the proportions of collocate pairs for each category, and factoring this into the analysis. We compared the findings between both corpora to provide an account of groomers' and children's talk in online grooming interactions. In doing so, and whenever relevant, we drew upon facework and discourse politeness/impoliteness (see Section 2). This was the case when considering results in terms of groomer-child interpersonal dynamics, especially their navigating their own and their interactant's face needs.

4. Results

4.1. Overview

Three main aspects of language use emerged as relevant in the analysis: particular Topics, Conversation Management and Interpersonal Dynamics. Quantifying the tendency for the identified collocate pairs to fit one or more of these aspects of language use provides an initial overall impression of what the talk represented in C—Corpus and G—Corpus is *made up of*, as illustrated in Fig. 1.

The most common aspect of language use representing children's and groomers' talk, in the C—Corpus and G—Corpus, respectively, is the way it is about certain topics. Just over two fifths of the collocate pairs identified, 41 % in C—Corpus and 44 % in G—Corpus, were classified as Topics-Based. The next most common is Conversation Management, which counts for 27 % of the collocate pairs from C—Corpus and 28 % of those from G—Corpus. This is followed by Interpersonal Dynamics, which represents 21 % of the C—Corpus and 20 % of the G—Corpus collocate pairs. A number of collocate pairs were classified as Miscellaneous – 11 % in C—Corpus and 5 % in G—Corpus – where an examination of concordances revealed them to be 'genuinely uninterpretable' (Gillings and Mautner, 2024, p. 54); they are therefore not included in the analysis that follows.

The three aspects of language use that represent children's and groomers' talk in online grooming interactions each consist of a number of categories. These are listed in the right-hand boxes of Fig. 2.

As shown in Fig. 2, Topics-Based language use is represented by three collocate pair categories. These include Sex, where a collocate pair tends to be part of language use that primarily refers to sex in some way; Relationship, where its use tends to refer to the idea of a romantic or other kind of personal relationship between child and groomer; and Past, Present, Future, where it is primarily used to remember actions or events, talking about current activities or referring to future plans.

Conversation Management is represented by two collocate pair categories: Established Conversation Routines, where a collocate pair tends to be part of conventionalised linguistic forms used to start or end an interaction; and Reported Speech/Clarification, where a collocate pair tends to be part of language used to manage conversational flow with respect to establishing meaning and understanding.

The third aspect of language use that represents the children's and

⁶ For a discussion see various chapters in Taylor & Marchi (2018).

⁷ <https://lancsbox.lancs.ac.uk>

⁸ <0.6 was selected because it has been shown to capture the required level of dispersion in a similar specialised corpus (see Lorenzo-Dus et al., 2020). For a review of corpus statistics measures, including DPNorm, see Gries (2024).

⁹ For a detailed discussion of the use of MI in applied linguistics research see Ballance (2022, p. 8), including the recommendation that its use be complemented with 'robust dispersion measures'.

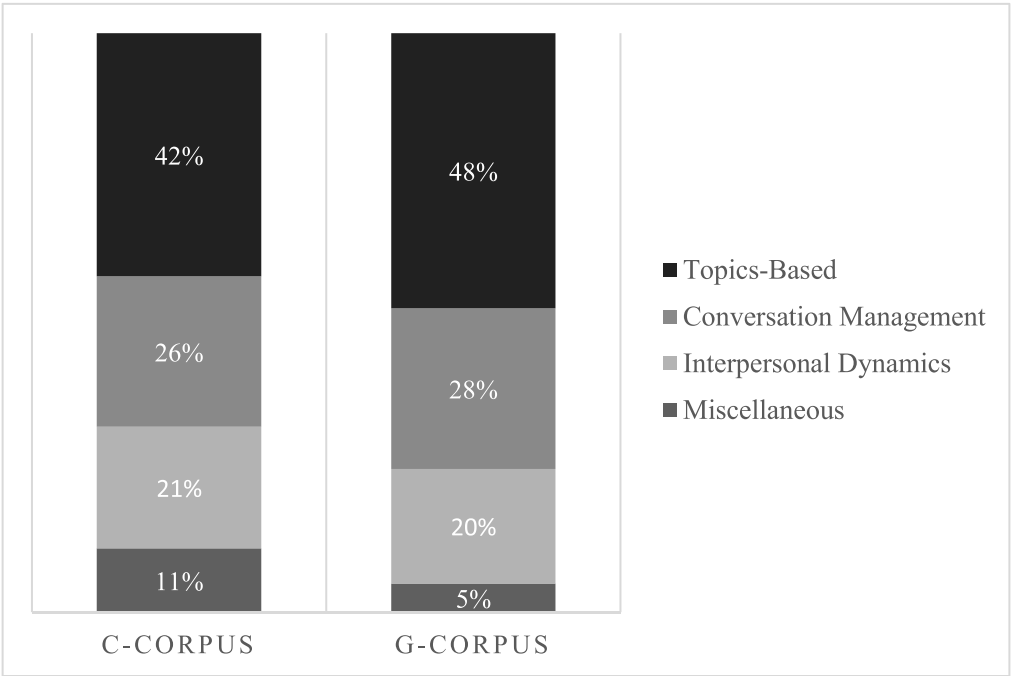


Fig. 1. Proportions of collocates representing aspects of language use that constitute children’s and groomers’ talk.

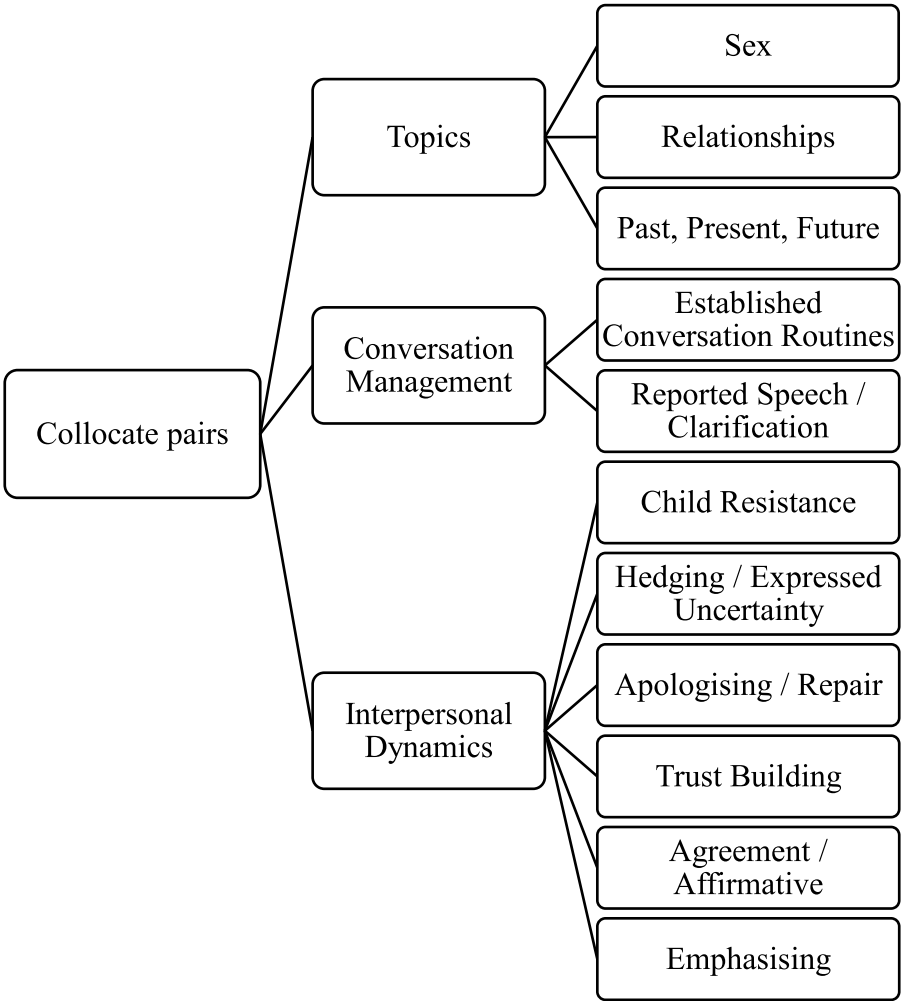


Fig. 2. Breakdown of aspects of language use into collocates pair categories.

groomers' talk examined in this study, Interpersonal Dynamics, consists of six collocate pair categories. These are Child Resistance, where a collocate pair is part of children expressing resistance (or reflecting this in the case of collocate pairs in G-Corpus); Hedging/Expressed Uncertainty, where a collocate pair is part of minimising the imposition of something said or to represent not knowing something; Apologising/Repair Work, where a collocate pair is part of mitigating an offence/potential offence or to address a breakdown in relations; Trust Building, where a collocate pair is part of language used to foster a sense of closeness between child and groomer (and in a way not represented by any of these other categories); Agreement/Affirmative, where a collocate pair is used to establish accord between groomer and child; and Emphasising, where a collocate pair is used to express the strength of a feeling or attitude.

4.2. Topics-Based

The three categories representing collocate pairs that tend to occur as part of Topics-Based language use (Sex; Relationship; Past, Present, Future) each consist of a number of subcategories. These are presented in Table 1 along with concordance examples and a comparison of the proportion of the collocate pairs that make up the (sub)categories across the two corpora.

Sex is the largest Topics-Based category for collocate pairs in both corpora, counting for over three-fifths (62 %) in C-Corpus and over two-thirds in (70 %) in G-Corpus. As several of the subcategories show, the majority of the Sex collocate pairs do not represent simply talking about sex but action through talk where sex is the object of a request, offer and so on (e.g., 'I really wanna see you blow a load'; 'would you like to have a sexy web call with me'). The collocate pair findings reveals a high degree of similarity in the sex talk of children and groomers in online grooming interactions; for example, the subcategories representing sexual offers and requests count for 42 % of Topics-Based collocate pairs in C-Corpus and 41 % in G-Corpus. This shows children's communicative agency in sex talk when they interact with groomers online (see Section 2 for an explanation of 'agency' meaning engagement but not blame).

Strong similarity across the two corpora in the Topics-Based category Sex not only occurs with respect to what children and groomers are doing with their talk but also with respect to the style of their language. This is revealed by looking at the concordances of the Sex collocate pairs. For example, in both C-Corpus and G-Corpus, sex references can be explicit ('make you fuck your ass and do something at the same time') or implicit ('want a show'); the language often reflects hyper-confidence, as suggested by the provocative style of flirtation routines ('bet you can't make me'); and sexual roleplay language is often used, realised as the form of certain terms of address (e.g., 'that's it slave good boy') or when adopting the behaviour and style of a sexual stereotype such as someone who is domineering (e.g., 'I feel like I'm going to have to punish you for this'). As these examples show, sex talk in online grooming interactions is highly performative, characterised by exaggeration and knowingness that suggest a child's perspective of engaging in harmless, in-the-moment fun. Children's sense of online sex talk being part of experimentation and play without consequences supports the groomer's purpose of getting children to drop their guard as they seek to manipulate and control them.

Although the majority of the findings reveal strong similarity in the sex talk of groomers and children (a likely effect of groomers' tactical purpose, suggested above), there is a striking difference in one subcategory, Other Sexual References, which counts for 23 % of Topics-Based collocate pairs in G-Corpus compared to 11 % in C-Corpus. This refers to sex talk in general and suggests that it is more a normalised part of groomers' talk than children's, where the frequency difference may represent a feature that distinguishes the two.

Another Topics-Based category is Relationship; that is, where the talk in online grooming interactions is about the groomer's and child's

Table 1

Examples to illustrate Topics-Based collocate pair categories/subcategories and proportions of these in C-Corpus and G-Corpus.

| TOPICS-BASED | | % of collocate pairs | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|------------------|
| C-Corpus | G-Corpus | C-Corpus 62 % | G-Corpus 70 % |
| 1. SEX | | | |
| 1a. Requests | | | |
| e.g. <u>make+me</u> (<u>make</u> out with <u>me</u> like now); <u>wanna+see</u> (i really <u>wanna see</u> you blow a load); <u>go+cam</u> (want to <u>go</u> on <u>cam</u> for me?) | e.g. <u>fuck+daddy</u> (show <u>daddy</u> how you'd <u>fuck</u> his cock); <u>come+join</u> (<u>come</u> <u>join</u> me?) | 33 % | 29 % |
| 1b Offers | | | |
| e.g. <u>want+show</u> (<u>want</u> a <u>show</u> on tuesday?); <u>want+if</u> (you can finger me <u>if</u> you <u>want</u>) | e.g. <u>like+would</u> (<u>would</u> you <u>like</u> to have a sexy web call with me); <u>want+watch</u> (having a wank <u>want</u> to <u>watch</u>) | 9 % | 12 % |
| 1c. Challenges / Urging and Expressions of Influence or Desire | | | |
| e.g. <u>make+me</u> (i bet you cant <u>make me</u>); <u>like+feel</u> (i <u>feel</u> <u>like</u> I'm going to have to punish you for this!) | e.g. <u>just+wanted</u> (<u>just</u> <u>wanted</u> to kiss those lips); <u>see+love</u> (Would <u>love</u> to <u>see</u> a video of you using it) | 9 % | 6 % |
| 1d Other Sexual References | | | |
| e.g. <u>good+boy</u> (oh that's it slave <u>good boy</u>); <u>think+of</u> (pool, can't really <u>think of</u> anywhere else i want to do it); <u>time+at</u> (make you fuck your ass and do something <u>at</u> the same <u>time</u>) | e.g. <u>think+of</u> (I can <u>think of</u> more); <u>good+girl</u> (mm <u>good girl</u>); <u>right+I'm</u> (I'm hard <u>right</u> now) | 11 % | 23 % |
| 2. RELATIONSHIP (i.e. personal or romantic) | | | |
| 2a. Status / Nature | | | |
| e.g. <u>really+love</u> (I <u>really</u> do <u>love</u> you); <u>like+feel</u> (i <u>feel</u> <u>like</u> I know you very well); <u>know+each</u> (we don't really <u>know</u> anything about <u>each</u> other) | e.g. <u>like+feel</u> (Why do it <u>feel</u> <u>like</u> your annoyed with me); <u>know+each</u> (If we get to <u>know</u> <u>each</u> other better and meet); <u>think+I'm</u> (Do u <u>think</u> I'm too old) | 8 % | 6 % |
| 2b Requests, Offers, or Expressions of Desire | | | |
| e.g. <u>want+talk</u> (I still <u>want</u> to <u>talk</u>); <u>please+tell</u> (you wont ruin anything <u>please</u> just <u>tell</u> me) | e.g. <u>want+talk</u> (I <u>want</u> to <u>talk</u> as long as possible); <u>come+back</u> (please <u>come</u> <u>back</u> on) | 13 % | 7 % |
| 3. PAST, PRESENT, FUTURE | | | |
| 3a. Looking Back | | | |
| e.g. <u>time+last</u> (I remember <u>last</u> <u>time</u> we got no where); <u>going+was</u> (I <u>was</u> <u>going</u> to tell my sister that you liked me) | e.g. <u>time+last</u> (<u>last</u> <u>time</u> we spoke we got a little naughty) | 4 % | 2 % |
| 3b Looking Forward | | | |
| e.g. <u>go+where</u> (let me know if you find a day <u>where</u> we can <u>go</u> out); <u>time+have</u> (I do it better tomorrow because I <u>have</u> more <u>time</u>); <u>time+first</u> (you know the <u>first</u> <u>time</u> we meet) | e.g. <u>time+next</u> (<u>next</u> <u>time</u> we cam you're the boss); <u>go+we</u> (I'll learn about you as <u>we</u> <u>go</u>) | 13 % | 9 % |
| 3c. Current or Routine Activities | | | |
| N/A | e.g. <u>just+woke</u> (in bed naked as <u>just</u> <u>woke</u> up); <u>right+now</u> (Are you sat in the car <u>right</u> <u>now</u>) | 0 | 6 % |

The two words that represent a collocate pair are underlined in all concordances presented in this paper.

personal or romantic relationship. In C-Corpus, 21 % of Topics-Based collocate pairs were categorised as Relationship, about a third of those for Sex, and even less in G-Corpus, 13 %, which represents less than a fifth of the Sex collocate pairs in that corpus. The Relationship collocate pairs were divided into two subcategories, Status/Nature, where the

idea of the child–groomer relationship is the focus of the talk represented by a collocate pair (e.g., ‘I really do love you’), and Requests, Offers or Expressions of Desire, where the relationship is indirectly referred to through non-sexual talk as action (e.g. ‘you won’t ruin anything please just tell me’).

The higher proportion of Relationship collocate pairs identified in C–Corpus compared to G–Corpus suggests the idea of a personal or romantic relationship is more of a concern in the talk of children than groomers. Differences between C–Corpus and G–Corpus are also evident in the language around collocate pairs in this category. Concordance examples in the subcategory Nature/Status reveal that children tend to be concerned with the question of whether their relationship with the groomer exists (e.g., ‘I feel like I know you very well’; ‘we don’t really know anything about each other’). Groomers, on the other hand, tend to refer to their relationship with the child as part of a negotiation of their romantic relationship status (e.g., ‘Do u think I’m too old’; ‘If we get to know each other better and meet’), which is linked to their tactical purpose of deceptive trust development (Lorenzo-Dus et al., 2020).

The final Topics-Based category, Past, Present, Future, counts for the same proportion in C–Corpus and G–Corpus, representing 17 % of Topics-Based collocate pairs in both. There is also a similarity at the subcategory level, where Looking Forward collocate pairs (13 % and 9 % of Topics-Based in C–Corpus and G–Corpus, respectively) are noticeably more common than Looking Back collocate pairs (4 % and 2 %, respectively). However, a consideration of the concordances of the Looking Forward collocate pairs reveals differences in the way they use language.

In C–Corpus, these highlight how children tend to have a romantic attitude about the future, as illustrated by their talk about imagined future scenarios, for example: ‘if you could go anywhere in the world... where would you go’. The collocate pair *time+first* also suggests this tendency, as occurs when children talk about the imagined situation of their first in-person meeting with a groomer in idealised terms of them being in control, for instance when stating a *no sex* condition (e.g., ‘not the first time we meet’). In reality, manipulative, sexually motivated adult groomers are more likely to dictate the terms of any in-person meet-up. In G–Corpus, Looking Forward is more concerned with making definite plans (e.g., ‘when your free I’ll drive up to where u are and we can go for a drive’) or strong assertions about how things are going to be (e.g., ‘next time we cam you’re the boss’). This use of language supports the groomer tactic Further Contact identified in previous research (Lorenzo-Dus et al., 2020).

The most striking difference between C–Corpus and G–Corpus is with the Past, Present, Future subcategory Current or Routine Activities, which includes collocate pairs that occur as part of references to or enquiries about what the speaker or addressee is currently up to or does as a matter of routine. This subcategory was only identified for the findings from G–Corpus, where it represents 7 % of the Topics-Based collocate pairs; it reflects groomers’ tactical work as they attempt to involve themselves in a child’s life by keeping tabs on what the child is up to (e.g., ‘Are you sat in the car right now’) and updating them about their own current situation (e.g., ‘in bed naked as just woke up’).

4.3. Conversation management

The two collocate pair categories that represent language used to manage conversational flow are presented in Table 2; one of these (Established Conversation Routines) is divided into two subcategories. Like Table 1, Table 2 presents concordance examples and a comparison of proportion differences represented by the (sub)categories across the two corpora.

In both corpora, collocate pairs are fairly evenly distributed across the two Conversation Management categories: 53 % of the Conversation Management collocate pairs in C–Corpus and 57 % in G–Corpus represent Established Conversation Routines, while 47 % in C–Corpus and 43 % in G–Corpus represent Reported Speech and Clarification. A

Table 2

Examples to illustrate Conversation Management collocate pair categories/subcategories and proportions of these in C–Corpus and G–Corpus.

| CONVERSATION MANAGEMENT | | % of collocate pairs | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|--------|
| C–Corpus | G–Corpus | C–Corp | G–Corp |
| 4. ESTABLISHED CONVERSATION ROUTINES | | 53 % | 57 % |
| 4a. Openings | | | |
| e.g. <i>like+add</i> (I’d <u>like</u> to <u>add</u> you as a contact); <i>good+thanks</i> (I’m <u>good</u> <u>thanks</u> you?) | e.g. <i>like+add</i> (I’d <u>like</u> to <u>add</u> you as a contact); <i>good+you</i> (I’m <u>good</u> <u>you</u> ?) | 11 % | 25 % |
| 4b Closings | | | |
| e.g. <i>back+get</i> (I’ll <u>get</u> <u>back</u> around 1230); <i>go+bed</i> (I’m gonna <u>go</u> <u>bed</u> now anyway absolutely shattered); <i>right+back</i> (sorry i have to go for 20 mins, be <u>right</u> <u>back</u>) | e.g. <i>go+sleep</i> (I think I should <u>go</u> <u>sleep</u>); <i>right+back</i> (be <u>right</u> <u>back</u> going to the toilet); <i>go+out</i> (I have to <u>go</u> <u>out</u> for 30 min) | 42 % | 32 % |
| 5. REPORTED SPEECH AND CLARIFICATION | | 47 % | 43 % |
| e.g. <i>mean+what</i> (What do you <u>mean</u> challenging); <i>said+I</i> (I never <u>said</u> I didn’t want to talk); <i>said+what</i> (I take <u>what</u> I <u>said</u> back) | e.g. <i>said+I</i> (I <u>said</u> I’m not a crazy desperate fool that begs); <i>mean+what</i> (I don’t get <u>what</u> you <u>mean</u>); <i>say+something</i> (Did I <u>say</u> <u>something</u> wrong :)) | 47 % | 43 % |

more notable difference between the corpora is evident in the proportions of the two Established Conversation Routines subcategories. In C–Corpus, Openings count for 11 % and Closings 42 % of the Established Conversation Routines category, while in G–Corpus, Openings count for 25 % and Closings 32 % of this category.

The need to perform different kinds of facework when withdrawing from an interaction, such as giving reasons for leaving a conversation (e.g., the need to go to bed) or managing expectations about future contact (e.g., ‘be right back’; ‘I’ll get back around 12.30’), may explain the greater tendency for collocate pairs to represent Closings than Openings. The latter is more likely to involve highly conventional conversation-starting routines (e.g., requests to ‘add’; *how are you*-type greetings) and therefore produces less variety in collocate pairs. Overall, though, C–Corpus was found to have notably more Closings and fewer Openings than G–Corpus, reflecting the fact that groomers are more likely to initiate contact when pursuing children and children are more likely to withdraw when resisting groomers.

With Reported Speech and Clarification, the fairly similar proportion of Conversation Management collocate pairs representing this category in both corpora corresponds to general similarity in the way language is used. For example, as a review of concordances in C–Corpus and G–Corpus reveals, the collocate pairs represent a variety of actions around what was said and meant in a conversation, such as asserting (‘does not mean I always am’), questioning (‘What do you mean’) and so forth.

However, the collocate pairs also help to highlight evidence of difference in some general tendencies, as is the case with the collocate pair *said+I*. In G–Corpus, the words ‘I’ and ‘said’ tend to co-occur as part of groomers’ tendency to report speech when asserting or reaffirming what they have said, especially in a manner that expresses the certainty and consistency of their position (e.g., ‘like I said’). In C–Corpus, ‘I’ and ‘said’ tend to co-occur as part of a variety of actions that children perform to maintain their on-the-record position when responding to groomers’ challenge of this. These include denying (e.g., ‘I never said I didn’t want to talk’), withdrawing (e.g., ‘forget I said anything’) and withholding (e.g., ‘no I said tomorrow night’; ‘I said I wasn’t doing anything’). This suggests a tendency for children’s meaning clarification to involve them needing to fend off the groomer’s attempt to get them to do something.

4.4. Interpersonal dynamics

The six collocate pair categories that represent language associated with the Interpersonal Dynamics aspects of interactional talk are presented in Table 3, illustrated by concordance examples and comparison of category size across the two corpora.

The relative proportions of the categories presented in Table 3, which are either fairly similar or significantly different, suggest that children's and groomers' talk may be more closely aligned with respect to some Interpersonal Dynamics features than others. However, while the proportions may be similar, the use of language represented by these categories can be different. This is the case with Apologising/Repair Work, which represents a similar proportion of Interpersonal Dynamics collocate pairs in C—Corpus and G—Corpus, 22 % and 25 %, respectively. While Apologising/Repair Work often serve the same purpose of helping to maintain good relations in C—Corpus and G—Corpus, children's apologising more often tends to take the form of self-blame (e.g., 'I'm so paranoid today I'm sorry I'm so sorry') and efforts to reassure the groomer (e.g., 'I'm sorry wish I could stay up longer'). Groomers' apologising, on the other hand, tends more often to take the form of appealing for forgiveness or expressing a regretful attitude (e.g., 'Come on I'm sorry'; 'I didn't mean it').

Differences in children's and groomers' talk are also revealed by the collocate pairs of the categories Hedging/Expressing Uncertainty and Trust Building, even though the relative proportions of these categories between C—Corpus and G—Corpus are similar. Hedging/Expressing Uncertainty counts for 20 % of the Interpersonal Dynamics collocate pairs in C—Corpus and 18 % in G—Corpus. In C—Corpus, children tend to use hedging or expressions of uncertainty to mitigate face threats when

Table 3

Examples to illustrate Interpersonal Dynamics collocate pair categories/sub-categories and proportions of these in C—Corpus and G—Corpus.

| INTERPERSONAL DYNAMICS | | % of collocate pairs | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|--------|
| C—Corpus | G—Corpus | C—Corp | G—Corp |
| 6. AGREEMENT / AFFIRMATIVE | | | |
| e.g. <i>know+I</i> (Hah. I <u>know</u> what you mean); <i>please+yes</i> (<u>yes</u> please); <i>good+sounds</i> (sounds <u>good</u> to me) | e.g. <i>know+I</i> (yeah I <u>know</u>); <i>mean+know</i> (I <u>know</u> what you <u>mean</u>) | 12 % | 21 % |
| 7. CHILD RESISTANCE (or reflected resistance) | | | |
| e.g. <i>right+now</i> (Just not really feeling it <u>right now</u>); <i>think+don't</i> (I <u>don't</u> think its working); <i>time+another</i> (maybe we should do this <u>another time</u>) | e.g. <i>come+didn't</i> (how <u>come</u> you <u>didn't</u> come on last night) | 24 % | 5 % |
| 8. TRUST BUILDING | | | |
| e.g. <i>mind+don't</i> (hope you <u>don't mind</u> i added you); <i>like+look</i> (I look <u>like</u> shit in the morning) | e.g. <i>tell+can</i> (You <u>can</u> tell me anything); <i>like+looks</i> (hair looks nice I <u>like</u> that naturally curly) | 14 % | 13 % |
| 9. APOLOGISING / REPAIR WORK | | | |
| e.g. <i>make+you</i> (I'm not trying to <u>make</u> you feel bad); <i>sorry+I'm</i> (I'm <u>sorry</u> just not myself) | e.g. <i>sorry+I'm</i> (Come on I'm <u>sorry</u>); <i>mean+didn't</i> (I <u>didn't</u> mean it) | 22 % | 25 % |
| 10. EMPHASISING | | | |
| e.g. <i>really+is</i> (love <u>is</u> a <u>really</u> meaningful word); <i>like+actually</i> (I <u>like</u> someone to <u>actually</u> show interest in me) | e.g. <i>fuck+as</i> (I get one word answers and your dry <u>as</u> fuck); <i>really+was</i> (I <u>was</u> <u>really</u> excited earlier but you went off) | 8 % | 18 % |
| 11. HEDGING / EXPRESSING UNCERTAINTY | | | |
| e.g. <i>really+not</i> (Just <u>not</u> really feeling it right now); <i>think+need</i> (i <u>think</u> I <u>need</u> to sleep); <i>know+don't</i> (I <u>don't</u> <u>know</u> where to meet) | e.g. <i>know+don't</i> (I <u>don't</u> <u>know</u> what to say); <i>know+where</i> (I just want to <u>know</u> where we stand) | 20 % | 18 % |

resisting groomers (e.g., 'I think I need to sleep'; 'I don't know I'm 12'). Similar use of hedges to minimise the risk of causing offence was also found in G—Corpus, but the collocate pairs in this category for G—Corpus also reveal how groomers use hedges to minimise the idea of their imposition rather than the effect of it; for example, when groomers use the collocate pair just+asking to insinuate that the child is being unreasonable in their reluctant response to the groomer's request or query about something. In doing so, groomers are arguably gaslighting the child by encouraging them to question whether their resistance is justified (which, objectively speaking, it always is, given the illicit context).

An examination of the language for the category Trust Building, which counts for 14 % of Interpersonal Dynamics collocate pairs in C—Corpus and 13 % in G—Corpus, reveals considerable distinctions in the ways in which children and groomers attempt to build trust through their talk. Children tend to do this by expressing concern about not wanting to offend or disappoint the groomer, as revealed by the collocate pair mind+don't (e.g., 'I don't mind'; 'hope you don't mind'), where they use language in an accommodating manner, and the collocate pair really+look (e.g., 'I really do look ugly right now'; 'I look really shit lol'), where they tend to be concerned about how the groomer views their appearance. The Trust Building collocate pairs in G—Corpus, on the other hand, highlight groomers' tendency to invite children to trust them (e.g., 'It's okay take your time I won't bite') or to pay children compliments (as found with the collocate pair really+you're, e.g., 'You're really hot!'; 'You're beautiful you really are'; 'Well I think you're really attractive'). The category suggests the tendency for children to employ facework that seeks to minimise imposition on one's interactant's perceived face needs (i.e., negative politeness strategies in Brown and Levinson's (1987) framework) and groomers to employ facework that sought to make one's interactant feel liked or appreciated (i.e., positive politeness strategies in Brown and Levinson's (1987) framework), when engaging in trust building through their talk.

Distinct features of children's and groomers' talk are particularly highlighted by the language of Interpersonal Dynamics categories where there is significant difference. This is obviously the case with Child Resistance, which represents 24 % of Interpersonal Dynamics collocate pairs in C—Corpus compared to 5 % in G—Corpus, the latter only reflecting child resistance. Child resistance entails the use of facework / relational work that includes politeness and, to a lesser extent, impoliteness (for a detailed analysis see Lorenzo-Dus & Evans 2025).

With the category Emphasising, the proportion of collocate pairs in G—Corpus is over twice that in C—Corpus, 18 % compared to 8 %. Patterns in language use around the collocate pairs reveal a striking distinction between how groomers and children use emphasis. Groomers tend to produce emphatic language in relation to something sexual, as highlighted by collocate pairs containing the word 'really' (e.g., 'get really wild for daddy now'; 'I was really excited earlier'; 'I really wanna fuck you'). Children tend to use emphasis in relation to something mental, particularly emotion (e.g., 'Love is a really meaningful word'; 'I actually like you genuinely').

Finally, the Interpersonal Dynamics category Agreement/Affirmative contains collocate pairs that tend to occur as part of language used to express agreement and a sense of solidarity and common purpose (e.g., 'I know what you mean' – from the know+I collocate pair, with this example occurring in both corpora). They also occur as part of language used pragmatically to positively reinforce certain kinds of desired behaviours or situations by expressing acceptance of these. This is especially a feature of groomers' talk, as highlighted by collocate pairs containing the highly frequent word 'good' (e.g., 'C: I'm not being moody I swear G: Okay good'; 'C: Alone G: that's always good'). This can represent a subtle groomer tactic that involves conditioning children to act in ways that support groomers' manipulative and abusive purposes. It also helps to account for why the proportion of Interpersonal Dynamics collocate pairs represented by the category Agreement/Affirmative is greater in G—Corpus than C—Corpus, 21 % compared to 12 %.

5. Discussion and conclusion

A notable overall feature of online grooming discourse revealed by the findings in our study is the high similarity between groomers' and children's talk. All but one of the collocate pair (sub)categories are the same, and the proportion of collocate pairs represented by these is often similar across the two corpora. To some extent, this is to be expected of people participating in the same conversation and therefore talking about the same things. However, there is high similarity that might not be expected, such as the finding that the nature of children's sex talk, like groomers', is largely action-based (e.g., it involves offers, requests), and that communicative agency is as much a feature of children's talk as groomers'. This challenges expectations that online grooming involves obviously dominant, controlling abusers and passive victims.

The high similarity between the talk of groomers and children actually creates a false impression of online grooming interactions, which in reality constitute a distinct difference between goal-driven adult perpetrators using manipulation tactics and children engaging in exploratory and experimental behaviour while at a formative stage of life. One explanation for why groomers' and children's talk is highly similar is the influence of norms and adoption of features of a certain language style associated with sex talk; for example, a hyper-performative style involving the exaggerated expression of desire and confidence as part of a flirtation routine, which is especially likely to occur when groomers and children adopt personas and engage in sexual roleplay. This kind of behaviour may be facilitated by the digital affordances of online communication, such as remoteness and anonymity, which can allow for identities to be performed through digital texts (Papacharissi, 2013; Bou-Franch and Blitvich, 2019; Sardá et al., 2019).

Similarity is problematic because it hides the unequal, abusive relationship behind the impression of equality and means contextual knowledge (e.g., speakers' age) is needed to recognise online grooming is taking place. However, using CADS in this study has helped to reveal subtle but important differences in the language that make online grooming communication more identifiable. This approach is particularly effective when looking at deceptive communication, as it helps reveal implicit features that may be based on unconscious tendencies that groomers cannot easily disguise; for example, in the present study, the apparently greater tendency for sex to be a normalised part of groomers' talk, or for children's use of emphasis to be in relation to personal or romantic relationships and groomers' use of emphasis to be in relation to sex.

The value of using CADS lies not only in drawing out non-obvious distinct features that might otherwise be drowned out by dominant surface similarities but also in the way that it provides a richer account of online grooming communication than would be achieved with a content analysis or topic modelling techniques (see Brookes and McEnery, 2019), where identifying topic categories can represent the end of the analysis. This has proven to be the case in our study where examining patterns across categories has provided insight about children's and groomers' language use tendencies that might not have been apparent from looking at the categories or individual linguistic features alone.

In summary, generating findings on the features of children's and groomers' talk helps provide an account of online grooming communication overall. This is because online grooming communication is relational – it constitutes a two-way process. As the research is applied, it is also important for it to produce insights that may help practitioners in tackling the social problem online grooming represents. In our study, we have strived to do this by identifying features of online grooming discourse to help improve recognition of when it occurs as well as understanding of how it works as a communication practice. As the relational aspects of language use tend to be most evident in particular instances of interaction, while discourse features that distinguish online grooming as a particular kind of communicative activity tend to be most

evident as patterns across texts, CADS, which allows for the analysis of both, has proven to be an effective approach for researching online grooming communication in this study.

Data availability

Due to the sensitive nature of the data analysed in the current study, the data cannot be deposited in a repository.

Funding statement

The research for this study has been supported by Safe Online, Tech Coalition, and Spanish Ministry of Science and Innovation (MICN/AEIO/10.13039/501100011033).

CRediT authorship contribution statement

Craig Evans: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Nuria Lorenzo-Dus:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Nuria Lorenzo-Dus reports financial support was provided by Safe Online, Tech Coalition and the Government of Spain. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to express their gratitude and appreciation to the police forces involved for their assistance, time and effort in providing access to the data examined.

References

- Babchishin, K.M., Hanson, R.K., VanZuylen, H., 2015. Online child pornography offenders are different: a meta-analysis of the characteristics of online and offline sex offenders against children. *Arch. Sex. Behav.* 44 (1), 45–66.
- Baker, P., 2006. *Using Corpora in Discourse Analysis*. Bloomsbury.
- Ballance, O., 2022. Methodological considerations for the use of mutual information: examining the role of context in collocation research. *Res. Methods Appl. Linguist.* 1 (3), 1–12. <https://doi.org/10.1016/j.rmal.2022.100024>.
- Barroso, R., Marinho, A.R., Figueiredo, P., Ramião, E., Silva, A.S., 2023. Consensual and non-consensual sexting behaviors in adolescence: a systematic review. *Adolesc. Res. Rev.* 8 (1), 1–20.
- Beckett, H., 2019. Moving beyond discourses of agency, gain and blame: reconceptualising young people's experiences of sexual exploitation. In: Pearce, J. (Ed.), *Child Sexual exploitation: Why theory Matters*. Policy Press, pp. 23–42.
- Bogdanova, D., Rosso, P., Solorio, T., 2014. Exploring high-level features for detecting cyberpedophilia. *Comput. Speech. Lang.* 28 (1), 108–120.
- Bou-Franch, P., Blitvich, Garcés-Conejos, 2019. *Analyzing Digital discourse: New insights and Future Directions*. Palgrave Macmillan.
- Brookes, G., McEnery, T., 2019. The utility of topic modelling for discourse studies: a critical evaluation. *Discourse Stud.* 21 (1), 3–21.
- Brown, P., Levinson, S.C., 1987. *Politeness: Some universals in Language Usage*. Cambridge University Press.
- Cano Basave, A., Fernandez, M., Alani, H., 2014. Detecting child grooming behaviour patterns on social media. In: *SocInfo 2014: The 6th International Conference on Social Informatics*. <http://oro.open.ac.uk/41394/1/predatorBehaviour.pdf>.
- Chiang, E., Grant, T., 2017. Online grooming: moves and strategies. *Lang. Law* 4 (1), 103–141.
- Chiang, E., Grant, T., 2018. Deceptive identity performance: offender moves and multiple identities in online child abuse conversations. *Appl. Linguist.* 40 (4), 675–698.
- Christakis, D.A., Hale, L., 2025. *Handbook of Children and screens: Digital media, development, and Well-Being from Birth Through Adolescence*. Springer.

- Cooper, K., Quayle, E., Jonsson, L., Svedin, C., 2016. Adolescents and self-taken sexual images: a review of the literature. *Comput Hum. Behav* 55, 706–716.
- Drouin, M., Boyd, R., Hancock, J., James, A., 2017. Linguistic analysis of chat transcripts from child predator undercover sex stings. *J. Forens. Psychiatry Psychol.* 28 (4), 437–457.
- Egbert, J., Baker, P., 2020. *Using Corpus Methods to Triangulate Linguistic Analysis*. Routledge.
- Giles, H., Coupland, N., Coupland, J., 1991. Accommodation theory: communication, context, and consequence. In: Giles, H., Coupland, J., Coupland, N. (Eds.), *Contexts of Accommodation*. Cambridge University Press, pp. 1–68.
- Gillings, M., Mautner, G., 2024. Concordancing for CADS: practical challenges and theoretical implications. *Int. J. Corpus Linguist.* 29 (1), 34–58.
- Goffman, E., 1967. *Interaction Ritual: Essays on Face-to-Face Behavior*. Doubleday, Garden City, NY.
- Goffman, E., 1981. *Forms of Talk*. University of Pennsylvania Press, Philadelphia.
- Grant, T., Macleod, N., 2020. *Language and Online identities: The undercover Policing of Internet Sexual Crime*. Cambridge University Press.
- Gries, S.T., 2024. A review: corpus statistics, the ‘usual’ approaches. S. T. Gries, *Frequency, dispersion, association, and keyness: Revising and Tupleizing Corpus-Linguistic Measures*, pp. 12–79. John Benjamins.
- Kinzel, A., 2021. *The Language of Online Child Sexual groomers: A corpus Assisted Discourse Study of intentions, Requests and Grooming Duration*. Swansea University. <https://doi.org/10.23889/SUthesis.59027>. Doctoral thesis.
- Kloess, J.A., Hamilton-Giachritsis, C.E., Beech, A.R., 2017. A descriptive account of victims’ behaviour and responses in sexually exploitative interactions with offenders. *Psychol. Crime Law* 23 (7), 621–632.
- Kontostathis, A., Edwards, L., Leatherman, A., 2009. ChatCoder: toward the tracking and categorization of internet predators. In: Workshop held in conjunction with the Ninth SIAM International Conference on Data Mining (2009).
- Lorenzo-Dus, N., Kinzel, A., 2019. So is your mom as cute as you?’: examining patterns of language use by online sexual groomers. *J. Corpora Discourse Stud.* 2, 14–39.
- Lorenzo-Dus, N., & Kinzel, A. (2021). We’ll watch TV and do other stuff’: a corpus-assisted discourse study of vague language use in online child sexual grooming. *Exploring discourse and ideology through corpora*, 189–210.
- Lorenzo-Dus, N., Izura, C., Pérez-Tattam, R., 2016. Understanding grooming discourse in computer-mediated environments. *Discourse Context. Media* 12, 40–50. <https://doi.org/10.1016/j.dcm.2016.02.004>.
- Lorenzo-Dus, N., Kinzel, A., Di Cristofaro, M., 2020. The communicative modus operandi of online child sexual groomers: recurring patterns in their language use. *J. Pragmat.* 155, 15–27.
- Lorenzo-Dus, N., Evans, C., Mullineux-Morgan, R., 2023. *Online Child Sexual Grooming Discourse*. Cambridge University Press.
- Lorenzo-Dus, N., 2022. *Digital grooming: Discourses of Manipulation and Cyber-Crime*. Oxford University Press.
- Ningrum, P., Kusbaryanto, K., 2021. Factors influence sexual behavior in adolescent: a literature review. *Indones. J. Nurs. Midwifery* 9 (2), 119.
- Nini, A. (2020). Corpus analysis in forensic linguistics. In *The concise encyclopedia of applied linguistics* (pp. 313–320).
- NSPCC (2024). Online grooming crimes against children increase by 89% in six years. <https://www.nspcc.org.uk/about-us/news-opinion/2024/online-grooming-crimes-increase>.
- Pérez-Sabater, C., García-Montes, A., Lorenzo-Dus, N., 2024. Representación discursiva del ciberacoso sexual a menores: del chat a la sentencia judicial. *J. Lang. Law* 82, 183–203. <https://doi.org/10.58992/rld.i82.2024.4231>.
- Papacharissi, Z., 2013. A networked self: identity performance and sociability on social network sites. In: Lee, F.L.F., Leung, L., Qiu, J.L., Chu, D.S.C. (Eds.), *Frontiers in New Media Research*. Routledge, pp. 207–221.
- Partington, A., 2010. Modern Diachronic Corpus-Assisted Discourse Studies (MD-CADS) on UK newspapers: an overview of the project. *Corpora (Online)* 5 (2), 83–108.
- Pennebaker, J.W., Boyd, R.L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. <https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015LanguageManual.pdf>.
- Quayle, E., Jonsson, L., Löf, L., 2012. Online Behaviour Related to Child Sexual abuse. Interviews with Affected Young people. ROBERT, Risktaking online behaviour, Empowerment Through Research and Training. European Union & Council of the Baltic Sea States.
- Sardá, T., Natale, S., Sotirakopoulos, N., Monaghan, M., 2019. Understanding online anonymity. *Media Cult. Soc.* 41 (4), 557–564.
- Schneevogt, D., Chiang, E., Grant, T., 2018. Do perverted justice chat logs contain examples of overt persuasion and sexual extortion? A research note responding to Chiang and Grant (2017, 2018). *Lang. Law* 5 (1), 97–102.
- Seymour-Smith, S., Kloess, J.A., 2021. A discursive analysis of compliance, resistance and escalation to threats in sexually exploitative interactions between offenders and male children. *Br. J. Soc. Psychol.* 60 (3), 988–1011.
- Taylor, C., Marchi, A., 2018. *Corpus Approaches to Discourse*. Routledge.
- Thomas, K., Hamilton-Giachritsis, C., Branigan, P., Hanson, E., 2023. Offenders’ approaches to overcoming victim resistance in technology-assisted child sexual abuse. *Child Abuse Negl.* 141, 106–143.
- van Gijn-Grosvenor, E.L., Lamb, M.E., 2016. Behavioural differences between online sexual groomers approaching boys and girls. *J. Child Sex. Abus.* 25 (5), 577–596.
- Vartapetian, A., Gillam, L., 2014. Our little secret’: pinpointing potential predators. *Secur. Inf.* 3 (1), 1–19.
- Whittle, H., Hamilton-Giachritsis, C., Beech, A., Collings, G., 2013. A review of online grooming: characteristics and concerns. *Aggress. Violent. Behav.* 18 (1), 62–70.
- Wright, D., 2025. *Corpus Approaches to Discourse in Forensic and Legal Contexts*. Taylor & Francis.