This is the peer reviewd version of the followng article:

The communicative modus operandi of online child sexual groomers: Recurring patterns in their language use / Lorenzo-Dus, N., Kinzel, A., Di Cristofaro, M.. - In: JOURNAL OF PRAGMATICS. - ISSN 0378-2166. - 155:(2020), pp. 15-27. [10.1016/j.pragma.2019.09.010]

Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

15/11/2026 02:15

(Article begins on next page)

The Communicative *Modus Operandi* of Online Child Sexual Groomers: Recurring Patterns in their Language Use

Nuria Lorenzo-Dus, Anina Kinzel and Matteo Di Cristofaro

To appear in Journal of Pragmatics 155 (2020), pp 15-27

1. Introduction

Grooming of children online is a legally punishable form of child sexual abuse. In the UK, for instance, the Serious Crime Act has since 2015 made it a criminal offence for an adult to send a sexual message to a child.[1] Nevertheless, this offence is on the increase. The UK-based charity NSPCC (National Society for the Prevention of Cruelty to Children), for example, reports a 16.8% growth in police-recorded cases of online child sexual grooming in England and Wales from 2015/16 to 2016/2017 (Bentley et al., 2018, p. 19). Moreover, these figures underestimate the real scale of the issue, which is known to be significantly under-reported by their victims because of a variety of reasons. These include the inability in some cases to realise for some time that they are being abused online, believing instead that they have a friendship-based and/or romantic relationship with their online groomers (e.g., Davidson and Gottschalk, 2011; NSPCC, 2018[2]).

The seriousness of the offence and its increasing prevalence over time may account for a growing body of academic scholarship investigating online child sexual grooming over the past decade or so. This scholarship has been primarily conducted within the disciplines of Criminology, Psychology and, to a lesser extent, Computational (primarily, Machine Learning) Text Analysis (See Section 2). In contrast, Linguistics scholarship into online child sexual grooming is still "in its infancy" (Chiang and Grant, 2018, p. 2). This is sadly ironic given that online child sexual grooming is an internet-enabled *communicative*

1 http://www.legislation.gov.uk/ukpga/2015/9/section/67/enacted
2 https://learning.nspcc.org.uk/media/1596/courage-talk-childline-annual-review-2017-18.pdf

process of entrapment in which an adult uses language and other semiotic modes (e.g. images) to lure a minor into taking part in sexual activities online and, at times, also offline.

Against this backdrop, the present study is the first to identify the linguistic patterns used regularly by a large number of convicted child sexual offenders when grooming their targets online. The paper is structured as follows. In Section 2 we review the relevant literature into online child sexual grooming, with a focus on Linguistics / language-based studies. In Section 3 we describe our corpus and the procedures and methodological framework we applied to analysing it. Section 4 presents key results, including identification of a set of three-word collocations regularly used by the groomers in the corpus under examination in order to fulfil their sexually abusive goals online. Section 5 reflects upon the implications and significance of these results for both detecting and contributing to prevent online child sexual grooming.

2. Online Child Sexual Grooming – A Progressive Focus on Language Use

There is a substantial body of research into online child sexual grooming, most of which has been conducted within Criminology, Psychology and, to a lesser extent, Computational Text Analysis. Alongside victim vulnerability factors and offender typologies (see, e.g., Gottshalk, 2011; Whittle, Hamilton-Giachristis, Beech and Collings, 2013; Babchishin, Hanson and VanZuylen, 2015;), this scholarship has focused on the process of online child sexual grooming. This has resulted in the development of several models of online child sexual grooming, the components of which generally include – under different labels (e.g. themes, tactics) – the following groomer goals: developing children's trust; separating them from their support networks; determining how likely they are to go along with the behaviour being proposed to them; introducing sexual content; and, in the case of so-called "contact-driven groomers",[3] arranging offline contact for sexual purposes (see, e.g., O'Connell, 2003; Williams, Elliott and Beech, 2013; Van Gijn-Grosvenor and Lamb, 2016; Quayle and Newman, 2017; Winters, Kaylor and Jeglie, 2017; Kloess, Hamilton-Giachritsis and Beech, 2017).

[3] The boundaries between contact-driven online groomers and those who do not seek to meet their victims off-line (so-called "fantasy-driven" online groomers) are, nevertheless, blurred (see Broome, Izura and Lorenzo-Dus, 2018).

Online child sexual groomers seek to fulfil the above goals through language (e.g. paying compliments to develop the child's trust in them) and other semiotic modes (e.g. sharing indecent images to desensitise the child to sexual content). However, systematic examination of the linguistic and/or visual means via which these goals are realised is still largely absent in the literature, which to date has primarily relied upon Thematic and/or Content Analysis methods. The comparative dearth of Linguistics studies into online child sexual grooming constitutes an important gap in knowledge. For example, attempts within the field of Machine Learning to develop online child sexual grooming detection software can be significantly enhanced through micro-level, contextualised description of the linguistic means via which groomers seek to fulfil their abusive goals online (see, e.g. , Kontostathis, Edwards and Leatherman, 2009; Cano Basave, Fernandez and Alani, 2014; Bogdanova, Rosso and Solorio, 2014; Vartapetiance and Gillam, 2014).

This gap in knowledge has only recently begun to be filled within, arguably, the most relevant Linguistics approaches, given their focus on examining language in its context of use, namely: Discourse Analysis and Pragmatics. Lorenzo-Dus, Izura and Pérez-Tattam (2016) used Computer-Mediated Discourse Analysis (Herring, 2013), specifically Pragmatic Act Theory (Mey, 2001) and Relational Work (Locher and Watts, 2005), to examine online child sexual grooming chat logs. Each chat log was the record of a full online grooming interaction, as published on the website Perverted Justice. This website is a project of Perverted Justice Foundation Inc. (www.pjfi.org), a non-profit foundation based in the United States that specialises in fighting online child sexual grooming. Perverted Justice Foundation Inc. has a number of adult volunteers who pose as children online (decoys). These decoys build profiles on social networking sites and/or enter chat rooms on a regular basis. As per the Perverted Justice Foundation rules of engagement, when contacted by an adult, decoys engage in an online conversation with him/her. If the conversation turns sexual, they continue the interaction and, in collaboration with law enforcement, try to secure the arrest of the adult when he/she arranges an offline meeting with the decoy (whom the adult presumes to be a child). If the adult is subsequently convicted of child sexual abuse, Perverted Justice Foundation Inc. makes available the relevant chat log on its website, along with the screen name, real name, age, photograph (if available), email address, and conviction notes of the adult.

The analysis conducted by Lorenzo-Dus et al (2016) resulted in the first empirical model of online grooming *discourse* (OGDM, henceforth). The OGDM comprises five, partly overlapping, discourse processes, as schematically represented in Figure 1.

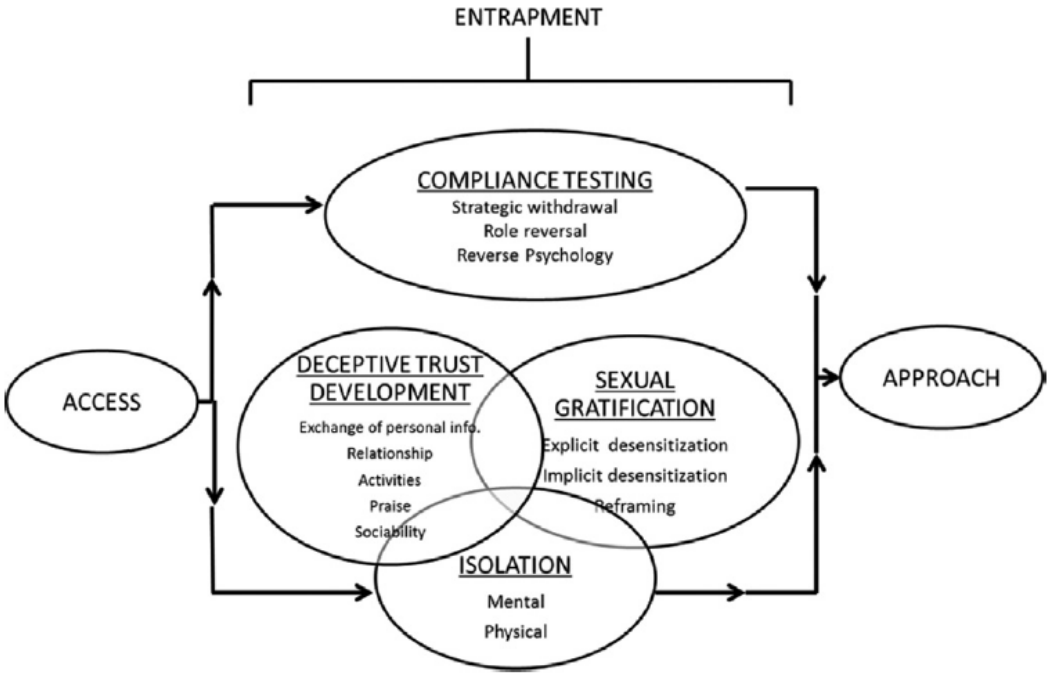

Fig. 1 Online Grooming Discourse Model (Lorenzo-Dus et al., 2016, p. 44)

Access signals groomers entering the online communicative environment and making contact with their target. Approach consists of groomers' attempts to secure a physical meeting for sexual purposes. In between Access and Approach there are four processes, which collectively comprise what the authors term the "online grooming communicative entrapment network". They are: Deceptive Trust Development, Sexual Gratification, Compliance Testing and Isolation.

Within the Deceptive Trust Development process, groomers hide their goal of wanting to engage their target in sexual activities by building a trust-based relationship with him/her in which friendship and/or romantic love is emphasised. This is realised via five strategies, namely: eliciting and sharing information, discussing activities and hobbies, making small talk, talking about relationships and praising the target.

Sexual Gratification covers groomers' attempts to involve their target in sexual activities online and, in the case of "contact-driven" groomers, preparing him/her for sexual interaction offline. This is realised by introducing sexual content, either explicitly discussing sexual topics using vulgar language and/or graphic images (explicit desensitisation) or using vague language to connote sexual content (implicit desensitisation). Sexual Gratification also includes presenting sexual activity between groomer and target as beneficial to the latter, that is, discursive reframing.

As part of the Compliance Testing process online groomers check whether the target is underage as well as the likelihood of his/her participating in whatever sexual activity they may propose to him/her. This process is discursively realised via three strategies, whereby groomers challenge the target (reverse psychology), seemingly put him/her in control of the grooming relationship (strategic withdrawal) and assume an unexpected low-risk attitude from him/her (role reversal).

Finally, the process of Isolation covers discursive work whereby the groomers distance the target from other meaningful people in his/her life while concomitantly emphasising the specialness and secrecy of the grooming relationship. This is done by arranging to talk to the target privately and alone, making sure that he/she is not supervised and instructing him/her to delete any trace of (online) interaction with the groomer (physical isolation). It is also done by eliciting self-disclosing talk from the target, attempting to become his/her confidant, while severing his/her emotional ties with others by means of "impoliteness talk" (Culpeper, 2011), such as criticism, targeted at these other meaningful others (mental isolation).

In a further study, Lorenzo-Dus and Izura (2017) focused on a particular strategy within the Deceptive Trust Development process, namely praise. Noting that praise was primarily realised via the "pragmatic act" (Mey, 2001) of paying compliments, the authors manually identified and extracted all the complimenting sequences (n = 1268) in sixty-eight Perverted Justice chat logs. In order to ascertain whether grooming duration affects frequency and type of compliment used within these sequences, they divided the chat logs under examination into three duration-based groups: "fast", "average" and "slow". They analysed the topic, syntactic patterns and discourse functions of each complimenting sequence within each of the three groups. The analysis revealed (i) a prevalence of sexually-oriented compliment topics, principally physical appearance, and that this prevalence increased as grooming

duration decreased; (ii) a high syntactic formulaicity for compliments, regardless of grooming duration; and (iii) a multifunctional use of compliments, also regardless of grooming duration.

For their part, Chiang and Grant (2017, 2018) used Discourse Analysis – specifically Genre Analysis – to examine, respectively, twenty online grooming chat transcripts from seven Perverted Justice chat logs and one offender interacting with several children from UK police data. Their analysis yielded several "rhetorical moves" that were typically used by the online groomers whose discourse they analysed. In Chiang and Grant (2018) sixteen such moves were identified,[4] which displayed "substantial overlap" (Chiang and Grant, 2018, p. 7) with the fourteen moves identified in Chiang and Grant (2017). The two new moves, labelled "overt persuasion" and "extortion", designated groomers' attempts to exert power over their victims by commanding and threatening them directly and indirectly, blaming them and "stating 'contractual terms' linked to their online relationship" (Chiang and Grant, 2018, p. 11).

Four other studies have focused on the language of online child sexual grooming, relying on significantly larger data sets that may hence enable the identification of recurring language patterns, namely: Black, Wollis, Woodworth and Hancock (2015); Drouin, Boyd, Hancock and James (2017); Schneevogt, Chiang and Grant (2018); and Lorenzo-Dus and Kinzel (2019). The first two studies used the Linguistic Inquiry and Word Count (LIWC)[5] software for their analyses. The software contains pre-defined dictionaries that are aligned to "variables" that concern psychological metrics (for example, different types of emotions: anger, sadness, happiness, and so forth), semantic fields (for instance, sex, family, home) and grammatical categories (for example, verbal tense and word class). LIWC works by counting words within a given variable and comparing them to the overall word count of the data set being examined, assigning a percentage to each variable (Tausczik and Pennebaker, 2010; Pennebaker, Boyd, Jordan and Blackburn, 2015a). LIWC thus performs analyses of textual data sets at decontextualized, word level.

[4] Greeting, maintaining conversation, meeting planning, reprimanding, sign-off, rapport, assessing likelihood and extent of engagement, assessing criteria fulfilment, assessing and managing risk, assessing role, sexual rapport, initiating sexual topics, maintaining/escalating sexual content, immediate sexual gratification, overt persuasion, extortion.

[5] http://liwc.wpengine.com/.

In Black et al.'s (2015) study each of the forty-four Perverted Justice chat logs examined was divided into five consecutive segments. The authors then mapped the results of the LIWC analysis of these segments onto five online grooming processes, as identified in O'Connell (2003), namely: friendship forming, relationship forming, risk assessment, exclusivity and sexual content. Their analysis revealed that some of the processes tended to occur within certain segments, for example, risk assessment occurred mostly in the first and second segments of the chat logs, while exclusivity was primarily used in the fourth and fifth segment. The sexual process was introduced as early as the first segment of the chat logs. Words contributing to the friendship and relationship forming processes occurred throughout the chat logs.

For their part, Drouin et al.'s (2017) study divided the 590 Perverted Justice chat logs they considered according to supposed decoy gender. The analysis entailed comparing groomer and decoy language. They found that (i) groomers used more sexual words than decoys, even more so when grooming female decoys, (ii) groomers used more words overall than decoys, and (iii) groomers had higher "clout" scores than decoys. Clout is one of the variables in LIWC. It measures "expertise" (Pennebaker et al., 2015b, p. 22), understood in terms of a person's place in a social hierarchy.[6]

The last two studies – Schneevogt et al. (2018) and Lorenzo-Dus and Kinzel (2019) – used Corpus Linguistics analytic tools to examine Perverted Justice chat logs. Lorenzo-Dus and Kinzel (2019) sought to determine whether – and if so, to what extent – LIWC and Corpus Linguistics methods were compatible. To do so, and focusing on the online child sexual grooming processes most closely examined by Black et al. (2015) and Drouin et al. (2017), that is, seeking sexual gratification and isolating the target, Lorenzo-Dus and Kinzel (2019) ran their corpus of >600 chat logs through both LIWC and Corpus Linguistics (CQPweb[7]) software. Through the latter Lorenzo-Dus and Kinzel (2019) identified a number of words that were used frequently with sexual connotations by many of the groomers in the corpus but that did not feature in the sexual content variable in LIWC, for example "wet" and "suck". The analysis also revealed important variation as regards im-/explicitness and (in)directness levels in the use of words within the LIWC sexual content variable. Moreover, it showed that words

[6] Although the underlying algorithm for this variable has – to our knowledge – not been disclosed in the LIWC literature or manuals, Pennebaker et al. (2015b) connect it in part to pronoun use.
[7] http://cwb.sourceforge.net/cqpweb.php

thematically mapped in Drouin et al. (2017) to the goal of isolating targets, namely family terms, were actually used to advance simultaneously various processes within the OGDM, specifically: Isolation, Sexual Gratification and Deceptive Trust Development. Whilst acknowledging that LIWC should be credited with placing language on the radar of online child sexual grooming research, Lorenzo-Dus and Kinzel (2019, p. 30) concluded that, if used, the results of LIWC analyses should only constitute "a starting point for identifying some lexico-grammatical properties of online groomers' language." The absence of general language comparison scores for those resulting from the LIWC variables for given data sets, the fact that the algorithm that supports some of the LIWC variables (e.g. clout) is not open-source and that the software performs decontextualized, word level analyses all mean that other Linguistics methodologies, operating at a Discourse Analysis / Pragmatics level, should be deployed for the analysis of online groomers' communicative *modus operandi*.

In a research note responding to Chiang and Grant (2017, 2018), Schneevogt et al. (2018, p. 98) sought to identify "examples of overt persuasion or sexual extortion" – that is, the two moves identified as new in Chiang and Grant (2018) – within 622 Perverted Justice chat logs. The methodological procedure entailed eliciting a list of statistically significant "key terms (words, bigrams and trigrams)" linked to overt persuasion or sexual extortion in the corpus by using both a reference corpus (the Brown corpus[8]) and findings from Chiang and Grant (2018). The corpus software used for the analysis – AntConc[9] – means that the statistical measure used for deriving key terms did not take into account the different size of the corpora being compared (Perverted Justice and Brown). The research note did not specify which spelling standardisation method was applied to the Perverted Justice data set prior to its analysis, beyond noting that "some manual editing" was used to "clean up the files" (Schneevogt et al., 2018, p. 99). A systematic approach to spelling standardisation is important given extensive spelling variation in the Perverted Justice data (see Section 3.2) and the impact it can have on statistical results.[10] Schneevogt et al. (2018) identified no examples of extortion and a considerably smaller number of examples of overt persuasion than those in Chiang and Grant (2018). However, the authors noted that "where clear examples [of overt persuasion] were found in the PJ [Perverted Justice] data, these did

8 http://clu.uni.no/icame/manuals/BROWN/INDEX.HTM

9 http://www.laurenceanthony.net/software/antconc/

10 The research note does not include any statistical results to help to ascertain the specific significance of the key terms reported.

seem to be parallel to examples found in Chiang and Grant (2018, pp. 10-11)" (Schneevogt et al., 2018, p. 101). The study concluded that, although data featuring adult decoys "is not truly representative" of interactions between child victims and their online groomers (2018, p. 97), it "may still be useful for asking some important questions" (2018, p. 101). about online child sexual grooming.

The above is a conclusion that we endorse. Having, on the one hand, carefully considered the empirical evidence regarding differences between groomer-decoy and groomer-child data and, on the other hand, the need for micro-level, Discourse / Pragmatics analysis of large data sets to inform detection software development, our study uses Perverted Justice chat logs to identify recurring patterns in online groomer language use. This is on the clear understanding that the groomers whose language we examine were interacting with adults whom they believed to be children.

3. Methodology

3.1 The data

The corpus for this study was built from all the chat logs (n=622) published on the Perverted Justice website at our data collection point (February 2018). Each chat log contained all the individual chat sessions (total n=6,968) between a groomer and one or more decoy(s) over a period of interaction that ranged from 17 to 10,597 minutes. Our analysis focused on the language produced by the groomers within these chat logs, which amounted to 3,297,475 words and 656,746 individual messages (writing turns). The majority of the chat logs came from Instant Messenger platforms. They were all uploaded to the website between 2004 and 2016.

3.2 Framework

Our study adopts a Corpus-AssistedDiscourse Studies (CADS, henceforth) methodology, which entails "the investigation and comparison of features of particular discourse types, integrating into the analysis, where appropriate, techniques and tools developed within corpus linguistics" (Partington, 2010, p. 88; see also Taylor and Marchi, 2018). CADS follows a "serendipitous" journey of discovery (Partington, 2006, p. 12) such that software-assisted linguistics searches of corpora are treated as an initial "'map' […] pinpointing areas of interest for a subsequent close analysis" (Baker, Gabrielatos, KhosraviNik,

Krzyzanowski and McEnery, 2008, p. 284). Its underlying premise, therefore, is that software-enabled, quantitative and manual, qualitative analyses of corpora must be synergised.

The analytic process in this methodology tends to move from generating distributional information about a corpus (e.g. general quantifications, frequency-ranked wordlists) to deriving particular details thereof. A typical analytic step in this process entails calculating collocations of a given search word (or words) – that is, the lexical units with which that word statistically tends to co-occur. Collocations are important because the meaning of a word is derived from the relationships it establishes with other words "which tend to occur in its environment" (Leech, 1974, p. 20).

The analytic process also requires extensive, close reading of the texts that comprise a corpus, which is facilitated by concordancing software tools that compute statistical information. A concordance is the collection of all the examples containing a search word in the corpus. In our study close reading of extended concordance lines was informed by extant knowledge of the communicative process of online child sexual grooming, as we next detail.

3.3 Procedure

3.3.1 Data extraction and formatting

The chat logs were (i) scraped from the Perverted Justice website using a custom Python script, (ii) transferred to a data frame, and (iii) saved to different csv files (one for each chat session within a given chat log). A metadata spreadsheet was also created, where further details for each chat log were manually annotated, namely the total amount of time each groomer spent chatting with the decoy(s), the groomer and decoy(s) usernames, and the chat platform through which they interacted. A custom Python script was then used to convert the extracted csv files into XML, which is the format supported by the CQPweb software.

The script was also used to (iv) insert each message into a tag to separate them into single writing turns; (v) identify any emojis, replacing them with their respective labels (e.g. :-), :-(, :-*) → happy-smiley, sad-smiley, kiss-smiley); and, for each writing turn, (vi) add the timestamp, username and a custom tag (user-type), indicating whether the user was a groomer (g) or decoy (d). We subsequently, and for each

chat session, indicated its total duration in minutes – as manually annotated in the metadata spreadsheet.

Prior to uploading the formatted corpus to CQPweb, and on account of the high presence of non-standard spellings and variations in the data (acronyms, abbreviations and typographical errors), we standardised its spelling. To do so, we extracted a set of 187,500 random messages (approximately 1/4 of the entire corpus) and imported them into the spelling standardisation software VARD 2.5.4.[11] The variations therein were manually identified and standardised, before being used as training data to standardise the spelling in the whole corpus automatically. This was then uploaded onto CQPweb alongside the non-standardised corpus. All the analyses were conducted using the standardised corpus. This minimised skewed statistical results, which would have been derived from treating – and statistically computing – different spellings of the same word (e.g. thanks and thx) as different words.

3.3.2 Data analysis

3.3.2.1 Lexical Frequency and Dispersion

The plain text files (one for each chat log) containing the groomer corpus were loaded into the Corpus Linguistics software #LancsBox (Brezina, McEnery and Wattam, 2015), where lexical dispersion values (DP_{Norm} scores) were calculated. A DP_{Norm} score indicates the normalised dispersion of a word in a corpus, based on its frequency combined with the number of elements (texts) in which the word appears (Gries, 2008). The score ranges from 0 to 1. Values around zero mean that "the relative frequencies of occurrence of the linguistic expression in the corpus are directly proportional to the sizes of the corpus parts whereas large values close to 1 mean the linguistic expression is distributed very unevenly among the corpus parts." (Gries, 2010, p. 274-75). From the resulting list we selected words with a DP_{Norm} score <0.6 so as to ensure that subsequent analyses were based on sufficiently dispersed terms.

3.3.2.2: Collocation and Concordance Analyses

[11] http://ucrel.lancs.ac.uk/vard/about/

We calculated the collocates of all the content words with a <0.6 DP_{Norm} value using a Dice coefficient metric – a measure whose results reflect a combination of significance (amount of evidence) and effect size (strength of connection) (Baker and Levon, 2015; Gabrielatos, 2018). On their own, Dice coefficient scores can provide a partial picture of the data, as a given collocation may be extremely salient but only appear in a limited number of texts. We therefore visually rendered all the Dice coefficient scores for all the collocations alongside the relevant observed frequencies and DP_{Norm} scores in a dynamic plot (see Figure 2). We then made the decision to focus on the collocations in the corpus with a Dice coefficient ≥0.12, which reflected collocational strength, an observed frequency >400 and a DP_{Norm} score <0.6..

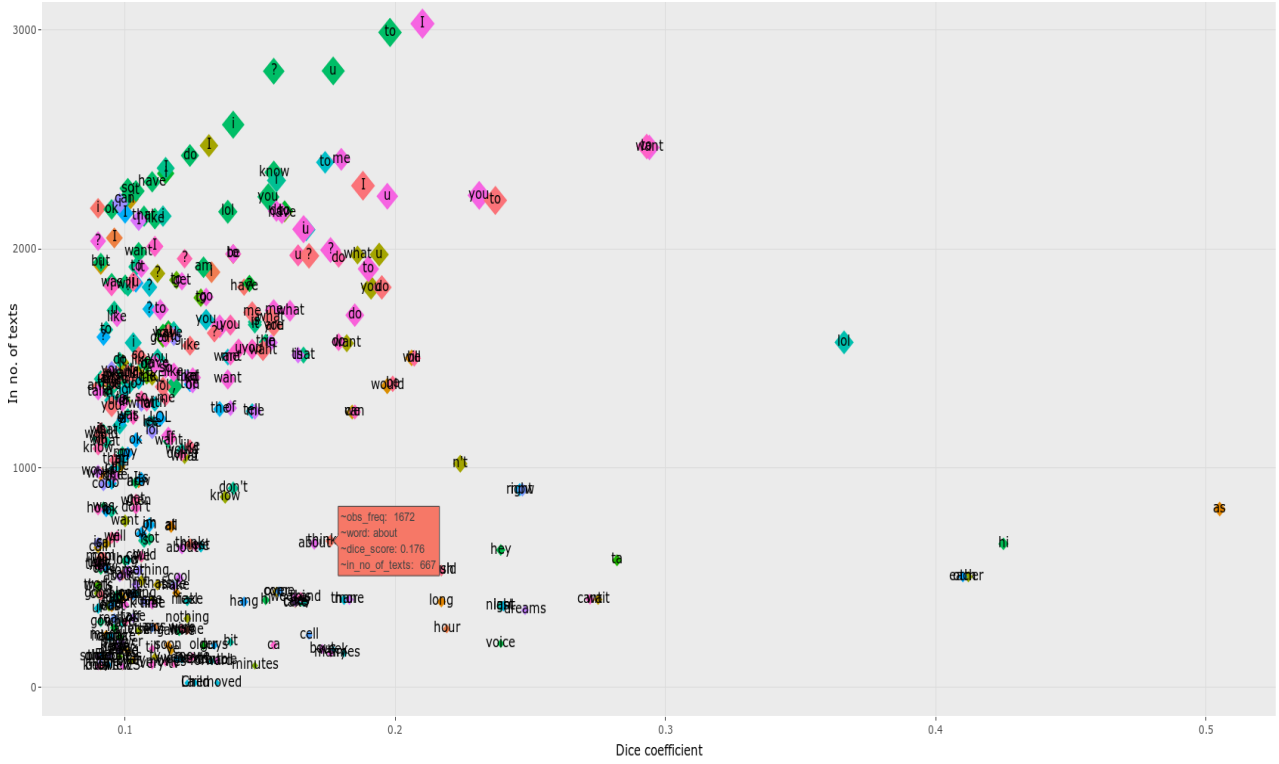

Fig. 2 Screen capture of the dynamic visualisation plot. Details of a two-word collocation (thinking + about)

In Figure 2 the horizontal axis represents the Dice coefficient scores and the vertical axis the number of chat logs in which the collocation appeared. Each of the diamonds represents one of the words within the selected DP_{Norm} cut-off point. By hovering next to each of the diamonds, an information box (the rectangle in Figure 2) appears that shows how many times the word was used with the relative collocate

in the corpus. Figure 2 illustrates this for the word 'about', as it collocated with 'thinking'. The statistical significance of this collocation ('dice_score') was 0.176, the frequency of use in the corpus ('in_no_of_texts') was 667.

Analysis of the plot visualisation of the corpus revealed a number of two-word collocations – like 'thinking + about' in Figure 2 – that displayed both significant salience (as measured by Dice coefficient scores) and high dispersion (as measured by DP_{Norm} scores and observed frequency). A random sample of fifty extended concordances was selected for each of these two-word collocations and examined manually to determine whether their pragmatic meaning (i.e., their meaning in context) supported the fulfilment of online child sexual grooming goals. We then derived a list of three-word collocations by calculating the collocates of all the two-word collocations, once again proceeding to identify those with significant frequency/dispersion and collocational strength values.

Finally, we randomly selected fifty extended concordance lines of each of the three-word collocations, and, once again, manually determined their meaning in context, noting whether they were used to realise one or more of the discourse processes identified in the OGDM (see Section 2; Figure 1) and, if so, which one/s. While pioneer in its use of Discourse Analysis / Pragmatics to understand the discourse of online child sexual grooming, the predictive power of this model is somewhat limited by the relatively small size of the data used to derive it (twenty-four chat logs from a pool of >600 chat logs available from the Perverted Justice website) and the semi-random chat log selection method used (chat logs that included a minimum of three separate groomer-decoy interactions, and in which the decoys had successfully been involved in at least ten groomer convictions). In addition to assisting our main study objective to provide a contextualised description of the linguistic patterns typically used by online groomers, therefore, this final analytic step also enabled us to determine whether the components of the OGDM featured regularly in a much larger, non-randomised corpus.

It is worth noting that in some cases the concordances we examined corresponded to pragmatic acts, including canonical speech acts, such as requests to meet in person within the Approach process (e.g. 'can I come over this weekend?').[12] In many other cases, however, the concordances that we manually

[12] Mey (2001, p. 221) uses the term pragmatic act, or pract, to designate each instantiation, in context, of a "general situational prototype, capable of being executed in the situation" – which he in turn labels a "pragmeme" - such as

assigned to one or more online child sexual grooming discourse processes did not correspond to any particular subtype of pragmatic act. For instance, the Sexual Gratification process entailed multiple linguistic realisations by means of which groomers introduced implicit and/or explicit sexual vocabulary to their target. Some of these described groomers' sexual experiences or desires (e.g. 'thinking about u makes me wet lol'); others conjured up hypothetical scenarios regarding their future sexual encounters (e.g. 'it would be romantic for us to both be in a bed naked kissing and feeling and fondlin each other'), and so forth. For this reason, rather than attempting to derive a taxonomy of pragmatic acts for online child sexual grooming from the concordance lines we selected, we analysed them to determine whether groomers used them to realise one or more of the processes in the OGDM. This opened up the possibility of both identifying new processes and/or discarding existing ones.

Overall, the combination of the above analytic steps enabled us to achieve our main objective of identifying and describing recurrent linguistic patterns (in our case, statistically frequent and highly dispersed three-word collocations) used by online child sexual groomers in order to fulfil their abusive goals. Additionally, we were also able to determine the validity (or otherwise) of the components of the OGDM and, more broadly, of the model itself.

4. Results

We identified a list of seventy three-word collocations that characterise the communicative *modus operandi* of online child sexual groomers in our corpus (Table 1). In this section we first consider the results of the lexical dispersion and collocation analysis (Section 4.1). We then examine the online child sexual grooming goals that these collocations, when used in context, typically fulfilled (Section 4.2).

4.1 Lexical Dispersion and Collocation Analysis

The lexical dispersion analysis generated one hundred and twenty-seven content words with a DP_{Norm} score <0.6. The meaning of a small number of these words (n=31) appeared to support the fulfilment

denying, bribing, complimenting, warning, inciting and so forth (2001, p. 221). He further notes that traditional or canonical speech acts (e.g. paying compliments) are "amongst the tools that we have at our disposal to control our environment and adapt to it in various ways" (2001, p. 215). This means that "while speech acts, when uttered in context, are pragmatic acts, pragmatic acts need not be speech acts (not even indirect ones)." (2001, p. 216).

of online child sexual grooming goals, which was confirmed through manual examination of a random sample of extended concordances containing them. These included (underlined in the examples below): two explicitly sexual words that were used in concordances that supported the Sexual Gratification process (e.g. 'do u want sex movie'; 'i wish u had a really sexy pic to show mesexy'); eight words linked to location or time, which related to the (Physical) Isolation process (e.g. 'are you sure no one will come home when we're together'; 'don't u have school today?'; 'do u want me to park my behind ur house?'; 'if i ever had a chance to come see u would u want me to come see u in pewrson one day?; 'what time did your mom get back?'); and twenty-one words that supported the Deceptive Trust Development (Relations) or Sexual Gratification (Desensitisation) processes (e.g., 'if u wanna do stuff like that [sexual act]'; 'you never play with it and get real wet?'; do you want to kiss me?'; 'yeah you will love it [sexual act]').

Plot visualisation of the collocations of the words with high dispersion (DP_{Norm} <0.6) and frequency (>400 observed frequency) led to identifying a total of forty-five two-word collocations with a Dice coefficient ≥0.12. Manual analysis of a random sample of fifty extended concordances per collocation showed that eleven of these were used to realise one or more processes in the OGDM: 'each + other'; 'want + to', 'can't + wait', 'wish + could', 'thinking + about', 'come + over', 'older+ guys', 'home + alone', 'tell + me', 'we + meet', 'this + weekend'.

Sexually explicit terms were absent in this list of two-word collocations. In some cases, the concordances in which the collocations were used contained sexual explicitness, for example, 'if we meet will let me give u sexual pleasure'. In other cases, the collocation connoted sexual intent, for instance, 'that is...if you want to see how it feels sometime'. In contrast, there were several collocations that pointed towards inter-personal relationships of a non-sexually explicit kind, such as 'tell me something fun to do', 'and we can belive in each other', and 'ive been thinking about us alot this week!'). How exactly these collocations were used by the groomers in our corpus is something that we were able to ascertain via the next iteration of the collocation and concordance analysis, at the three-word level.

Calculation of the collocates for each of the eleven two-word collocations listed above and analysis of fifty randomly selected extended concordances for each[13] generated seventy three-word collocations that were used to realise one of more processes in the OGDM (see Table 1).

Two-word collocation	Three-word collocation
each + other	+ *holding;* + *kissing*; + *touching*; + see; + *hold*; + seeing; + *feeling*; + loving (n=8)
want + to	+ do: + me; + see; + come; + meet (n=5)
cant + wait	+ either; + till; + see; + *hold*; + *taste;* + meet (n=6)
wish + could	+ help; + see; + hear; + marry; + come; + spend; + *kiss* (n=7)
thinking + about	+ stop; + *doing;* + *getting;* + *hard;* + *excited;* + *horny;* + *wet* (n=7)
come + over	+ able; + could; + can; + tomorrow; + want; + sometime; + maybe (n=7)
older + guys	+ like; + *into;* + young; + girls; + talking (n=5)
home + alone	+ leaves; + weekend; + stay; + leaving; + gonna; + leave; + left; + tonight (n= 8)
tell + me	+ what; + about; + please; + something; + how; + where; + can; + when; + more (n=9)
we + meet	+ when; + after; + can; + where; + first; + could (n=6)
this + weekend	+ plans; + doing (n=2)

Table 1 Three-word collocations used by groomers in the corpus

As can be seen in Table 1, the eleven two-word collocations generated a different number of three-word collocations. For instance, there were five three-word collocations for 'older + guys' ('like/into/young/girls/talking'); two for 'this + weekend' ('plans/doing'), and so forth. Some of the three-

[13] This was the case when there were more than fifty occurrences of the three word collocations in the corpus. When there were less than fifty occurrences, we analysed them all.

word collocations (n=15; *italicised* in Table 1) denoted sexual content and could thus be tied to the Sexual Gratification process within the OGDM.

Table 1 also contains a number of three-word collocations (n=14; underlined) from which information could be gleaned regarding the type of either sexually implicit or ostensibly non-sexual relationship that groomers deceptively constructed with their targets. This ranged from marriage ('wish + could + marry') and love (e.g. 'each + other + loving') to help (e.g. 'wish + could + help') and friendship (e.g., 'can't + wait + see'; 'older + guys + talking'). Although the lexical dispersion and two-word collocation results had already revealed the presence of relationship/friendship language, their frequency, distribution and variety became more evident through the three-word collocation analysis. In other words, analysis of one-word and two-word linguistic structures had under-estimated the salience of two key OCSG processes: Sexual Gratification and Deceptive Trust Development.

4.2. *Discursive Processes in Online Child Sexual Grooming*

As noted earlier, the collocation analysis also entailed manually examining a significant number of concordance lines (c. 5,000 overall), identifying the particular discourse process within the OGDM that each realised. In this section we report the results of this qualitative analysis. It is important to note that no new components of the OGDM were identified through this analysis. Moreover, we found frequent instances of all the existing components, even though the task of aligning each collocation/concordance to one single component was not always straightforward, as we next discuss.

Table 2 shows the three-word collocations the concordances of which were always linked to *one* OGDM process, specifying what that process is.

OGDM process	Three-word collocation
Sexual Gratification	each_other_holding; each_other_touching; cant_wait_hold; cant_wait_taste; wish_could_marry; thinking_about_getting; thinking_about_hard; thinking_about_excited; thinking_about_horny; thinking_about_wet
Isolation	home_alone_leaving; home_alone_left; home_alone_gonna
Approach	come_over_able; come_over_maybe

Compliance Testing	older_guys_girls

Table 2 Three-word collocations mapped to *one* process in the OGDM

As Table 2 shows, sixteen of the seventy three-word collocations always realised only one process in the OGDM. The process with the highest number of these collocations was Sexual Gratification (n=10), followed by Isolation (n=3), Approach (n=2) and Compliance Testing (n=1).

In ten of the remaining fifty-four three-word collocations, their respective concordances realised more than one process in the OGDM – the percentage distribution across these processes being broadly similar. In eight collocations, the concordances were linked to two ('we + meet + when', 'home + alone + tonight', 'older + guys + into', Figure 3) or three ('older + guys + young', 'wish + could + spend', 'cant + wait + see', 'cant + wait + till', 'want + to + see', Figure 4) processes.

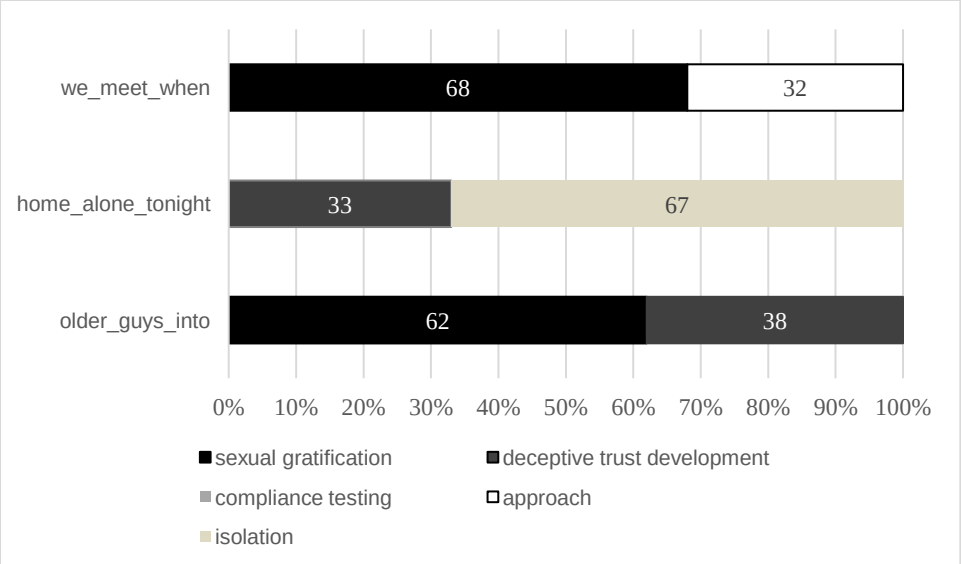

Fig. 3 Three-word collocations linked to two OGDM processes in %

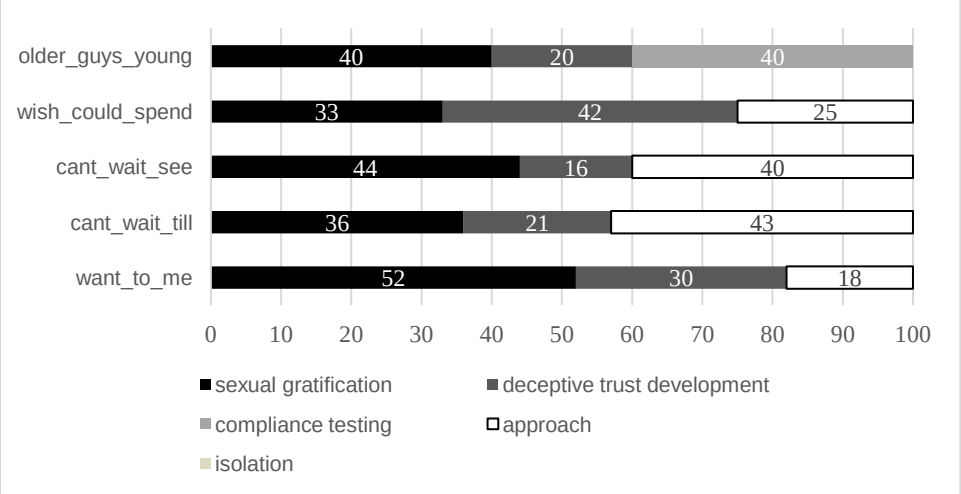

Fig. 4 Three-word collocations linked to three OGDM processes in %

The collocations in Figures 3 and 4 displayed a weaker association, in terms of frequency of use, with a given process than those in Table 2. For example, 33% of the concordances containing the collocation 'wish + could + spend' (Figure 4) were used to advance the process of seeking Sexual Gratification, 42% the Deceptive Trust Development process, and the remaining 25% the Approach process. As such, and while being used to advance some groomer goals, these collocations could not be considered as being typical realisations of a particular process in the OGDM.

The concordances containing the remaining two of these ten collocations ('thinking + about + do' and 'cant + wait + either') were used to realise three OGDM processes, two of which displayed very similar – and higher - percentage frequencies of use (Sexual Gratification and Deceptive Trust Development) than the third one (Approach): 'thinking + about + do' (46% Sexual Gratification; 50% Deceptive Trust Development; 4% Approach) and 'cant + wait +either' (46% Sexual Gratification; 44% Deceptive Trust Development; 10% Approach).

As for the concordances corresponding to the remaining forty-four three-word collocations, they realised more than one OGDM process, too. However, as shown in Table 3 below, one process was much more frequent than the rest in all of them, suggesting that the particular collocations were predominantly associated with that OGDM process. For example, in forty-six (92%) of the fifty concordances containing the collocation 'wish + could + kiss' the groomers sought Sexual Gratification (Examples: 'i wish i could kiss u'; 'ty for tha kiss i wish i could give ya one in person'); in the remaining four (8%) concordances

they used the same collocation to advance another (only one more) process, such as ‘wish i could kiss it and make the pain go away’ to advance Deceptive Trust Development. Similarly, in thirty-two (64%) of the concordances of the collocation ‘each + other + see’ that we analysed the groomers arranged to meet the target offline for sexual purposes, thus using this collocation to fulfil their Approach goal (Example: ‘and if all goes well we will see each other in person shortly after that’); in the remaining eighteen concordances sampled (36%) the same collocation was used to realise two other OGDM processes, namely, Sexual Gratification (Example: ‘Will you let me kiss you when we see each other?’) and Deceptive Trust Development (Example: ‘i wish we lived close so we cud see each other all the time’).

2 OGDM processes	3 OGDM processes	4 OGDM processes	5 OGDM processes
each_other_hold (97%)	each_other_see (64%)	want_to_do (50%)	tell_me_about (72%)
each_other_feeling (87%)	each_other_seeing (50%)	want_to_see (54%)	tell_me_please (56%)
each_other_loving (86%)	want_to_come (70%)	wish_could_come (66%)	we_meet_can (60%)
wish_could_help (76%)	want_to_meet (80%)	come_over_want (80%)	older_guys_like (64%)
wish_could_kiss (92%)	cant_wait_meet (67%)	home_alone_stay (66%)	
thinking_about_stop (81%)	wish_could_hear (89%)	tell_me_what (62%)	
come_over_sometime (83%)	wish_could_see (62%)	tell_me_something (60%)	
home_alone_leaves (87%)	come_over_could (76%)	tell_me_how (56%)	
home_alone_weekend (89%)	come_over_can (84%)	tell_me_where (48%)	
home_alone_leave (92%)	come_over_tomorrow (88%)	tell_me_can (42%)	

this_weekend_plans (88%)	we_meet_could (84%)	tell_me_when (48%)	
older_guys_talking (80%)	thinking_about_doing (50%)	tell_me_more (62%)	
	cant_wait_either (46%)	we_meet_after (40%)	
		we_meet_where (54%)	
		we_meet_first (66%) this_weekend_doing (78%)	

Table 3 *- Three-word collocations linked to predominant OGDM processes (%)*

Table 3 Key

Sexual Gratification	Isolation	Deceptive Trust Development	Compliance Testing	Approach

As shown in Table 3, three of the OGDM processes had a considerable number of three-word collocations that were predominantly associated with them: fifteen collocations for Approach, twelve collocations for Sexual Gratification, and twelve collocations for Deceptive Trust Development. The remaining two processes had a lower number of collocations, the concordances of which were predominantly associated with them: four for Isolation and one for Compliance Testing.

The analysis of the concordances of the seventy three-word collocations provides insights into the varying strength of association that these recurring linguistic structures had with a given OGDM process. In a minority of cases (10/70) there was no clear strength of association to a particular process. The remaining sixty collocations, however, were used to realise one OGDM process exclusively (Table 2) or predominantly (Table 3). Of the sixteen collocations that were used *exclusively* to realise one OGDM process, eleven were tied to Sexual Gratification. Across the remaining forty-four collocations

that were predominantly linked to one OGDM process, the percentage of concordances in which these collocations were thus used varied. As it decreased for a given collocation, its strength of association to the relevant process also decreased.

5. Conclusion

This study has unveiled recurring patterns in online groomers' language use. Specifically, our study has identified a list of linguistic structures (three-word collocations) used regularly by many of the groomers in the 622 chat logs examined in order to fulfil their sexually abusive goals online. The identification of no less than sixteen three-word collocations through which online groomers always perform the discursive process of Sexual Gratification in a corpus comprising c. 3.7 million words is a case in point. Importantly, this list has been derived from a combination of, on the one hand, replicable statistical metrics and tools (DP_{Norm} for lexical dispersion; Dice coefficient for collocational strength) and, on the other, manual examination of a large number of concordance lines across the various analytic steps taken. The analyses progressively performed at the single word (lexical dispersion and frequency), two-word collocation, three-word collocation and concordance levels thus bring us closer to understanding these individuals' communicative *modus operandi*.

The results also contribute to furthering our understanding of online child sexual grooming as a communicative entrapment network. All the language structures identified in this study were tied to one or more of the processes in the OGDM and no new processes emerged from the analysis we conducted. As such, our results provide broad support of the discourse of online child sexual grooming model developed by Lorenzo-Dus et al (2016), which was inductively derived from a smaller data set than that used in the current study. Additionally, our results advance knowledge of how groomers fulfil discursively specific abusive goals, also pointing towards future avenues of exploration.

The Sexual Gratification process, for instance, was linked to over a third of the three-word collocations in the corpus (38%; 23/60), all the concordances that we examined being either exclusively or strongly associated with that process. This confirms its salience within online child sexual grooming. Yet the collocations (and extended concordance lines) through which the Sexual Gratification process was realised in the corpus were not always as sexually explicit as might have been expected. This may

partly explain why some victims of online child sexual grooming do not recognise the sexual abuse they experience as such. Failure to do so may stem also in part from online groomers' use of language that connotes – rather than explicitly denotes – sexuality, as evidenced by the relatively high proportion of collocations and concordances used to realise the Sexual Gratification process that contained vague and / or sexually implicit language, as well as by the also relatively high proportion of collocations (20%; 12/60) that were linked to the Deceptive Trust Development process. This latter group of collocations, our qualitative analysis revealed, were predominantly – rather than exclusively - associated with the Deceptive Trust Development process. This is perhaps unsurprising given that the Deceptive Trust Development process is particularly broad, containing a combination of canonical speech acts (e.g. compliments on sexually and non-sexually oriented topics, requests for information) and other pragmatic acts that are not easily typified, not even as indirect speech acts (e.g. engaging in small talk and talking about relationships). These findings both support the salience of the Deceptive Trust Development process within the communicative process of online child sexual grooming and also call for closer, Discourse / Pragmatics analysis of the different strategies by which groomers seek to gain their target's trust in them, including their strategic use of inter-personal indirectness and sexual implicitness (see Lorenzo-Dus and Kinzel 2019).

As for Approach, this process was linked to a considerable number of collocations (28%, 17/60 of the total) in the corpus, too. It is worth reminding ourselves that all the individuals whose language we examined were so-called contact-driven online groomers, that is, they all arranged a physical meeting for the purpose of child sexual abuse. Additionally, these groomers were interacting with decoys who had received training in how to secure child sexual abuse convictions, a key evidence of which is arranging to meet a child offline for sexual purposes. Although we did not examine the decoys' language, the training received likely meant that, during their online interactions, they were willing to comply with the groomers' requests to meet offline. These data-related factors may explain the salience of the Approach process in our corpus.

Finally, the discourse processes of Isolation and Compliance Testing were tied to comparatively fewer collocations within the corpus, seven (10%) and one (1.4%), respectively. This may indicate significant overlap between these and other processes within the OGDM. In the case of Isolation, all the three-word collocations were derived from the two-word collocation 'home + alone'. Therefore, they were all

linked to the groomer goal of securing exclusive access to the victim's home environment (Physical Isolation). We did not identify recurring collocations within concordances linked to the goal of isolating the target emotionally. This suggests that the Emotional Isolation component within the OGDM may considerably overlap with other components, possibly that of discussing relations within Deceptive Trust Development. A similar explanation may apply to Compliance Testing, where the only collocation predominantly linked to it referenced some aspect of age-inappropriate relations ('older + guys + girls' and 'older + guys + talking'). Two other linguistic structures were found to be used in relation to this process ('older + guys + young' and 'home + alone + tonight'), albeit that their association strength was considerably weaker (see Figures 3 and 4).

Despite different degrees of association with a given online grooming discourse process, the groomers used all seventy collocations identified in our study regularly and frequently in order to fulfil their abusive goals. The fact that most of the collocations supported more than one online grooming process is an important finding, which speaks to the interconnectedness of online groomers' tools of entrapment. The particular combinations of discourse processes aligned to specific collocations, including the strength of association with each process, that we have – to our knowledge – for the first time identified in this study provide the kind of detailed linguistic knowledge that must, in response to calls from Computational Text Analysis (Machine Learning), inform development of sophisticated algorithms underpinning online child sexual grooming detection software (see Morgan et al., under review). The findings of this study may also be used as part of prevention programmes aimed at professionals working to protect children and young people from online child sexual abuse. These may rely on examples of how groomers typically use language to achieve their abusive goals. The sexual implicitness in some of these examples may, for instance, help to raise awareness of how communicatively sophisticated – albeit insidious – the discourse process of online child sexual grooming is. This may in turn contribute to resilience development interventions underpinned by evidence-based research into online groomers' communicative *modus operandi*.

References

Babchishin, Kelly M., Hanson, R. Karl., & VanZuylen, Heather. (2015). Online child pornography offenders are different: A meta-analysis of the characteristics of online and offline sex offenders against children. *Archives of Sexual Behavior*, *44*(1), 45-66.

Baker, Paul, Gabrielatos, Costas, KhosraviNik Majid, Krzyzanowski, Michal, and McEnery, Tony (2008) A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press, *Discourse & Society* 19(3), 273–306.

Baker, Paul, and Levon, Erez, (2015). Picking the right cherries? A comparison of corpus-based and qualitative analyses of news articles about masculinity. *Discourse & Communication*, 9(2), 221-236. Doi: 10.1177/1750481314568542

Bentley, Holly, Burrows, Andy, Clarke, Laura, Gillgan, Abbie, Glen, Jazmin, Hafizi, Maria, Letendrie, Fiona, Miller, Pam, O'Hagan, Orla, Patel, Priya, Peppiate, Jessica, Stanley, Kate. Starr, Emily, Vasco, Nikki and Walker, Janaya, (2018). *How safe are our children? The most comprehensive overview of child protection in the UK.* London: NSPCC. https://learning.nspcc.org.uk/media/1067/how-safe-are-our-children-2018.pdf

Black, Pamela, Wollis, Melissa, Woodworth, Michael and Hancock, Jeffrey T., (2015). A linguistic analysis of grooming strategies of online child sex offenders: Implications for our understanding of predatory sexual behavior in an increasingly computer-mediated world. *Child Abuse & Neglec*t, 44, 140-149. Doi: 10.1016/j.chiabu.2014.12.004

Bogdanova, Dasha, Rosso, Paolo and Solorio, Thamar, (2014). Exploring high-level features for detecting cyberpedophilia. *Comput. Speech Lang*, 28(1), 108–120.

Brezina, Vaclav, McEnery, Tony and Wattam, Stephen, (2015). Collocations in context: A new perspective on collocation networks*. International Journal of Corpus Linguistics*, 20 (2), 139-173.

Broome, Laura, Izura, Cristina and Lorenzo-Dus, Nuria, (2018). A systematic review of fantasy vs. contact driven internet initiated sexual offences: Discrete or overlapping typologies?. *Child Abuse and Neglect,* 79, 434-444.

Cano Basave, Amparo, Fernandez, Miriam and Alani, Harith, (2014). Detecting child grooming behaviour patterns on social media. In: *SociInfo 2014: The 6th International Conference on Social Informatics*, http://oro.open.ac.uk/41394/1/predatorBehaviour.pdf

Chiang, Emily and Grant, Tim, (2017). Online grooming: moves and strategies. *Language and Law/ Linguagem e Direito*, 4(1), 103-141.

Chiang, Emily and Grant, Tim, (2018). Deceptive identity performance: Offender moves and multiple identities in online child abuse conversations. *Applied Linguistics*, 1-25. Doi: 10.1093/applin/amy007

Davidson, Julia and Gottschalk, Petter, (2011). Characteristics of the Internet for criminal child sexual abuse by online groomers. *Criminal Justice Studies*, 24(1), 23-36.

Drouin, Michelle, Boyd, Ryan, Hancock, Jeffrey T. and James, Audrey, (2017). Linguistic analysis of chat transcripts from child predator undercover sex stings. *The Journal of Forensic Psychiatry & Psychology*, 28(4), 437-457. Doi: 10.1080/14789949.2017.129170.

Gabrielatos, Costas, (2018). Keyness analysis: nature, metrics and techniques. In Taylor, C. and Marchi, A. (Eds.). *Corpus Approaches to Discourse: A Critical Review*, 225-258.

Gottschalk, Petter, (2011). A dark side of computing and information sciences: Characteristics of online groomers. *Journal of Emerging Trends in Computing and Information Sciences, 9,* 447-455

Gries, Stefan T., (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, 13(4), 403-437. Doi: 10.1075/ijcl.13.4.02gri

Gries, Stefan T., (2010). Useful statistics for corpus linguistics. *A mosaic of corpus linguistics: Selected approaches*, 66, 269-291.

Hardie, Andrew, (2012). CQPweb - combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17(3), 380–409. Doi: 10.1075/ijcl.17.3.04har

Herring, Susan, (2013). Discourse in Web 2.0: Familiar, reconfigured, and emergent. In Tannen, D. and Trester, A.M. (Eds). *Georgetown University round table on languages and linguistics 2011:*

Discourse 2.0: Language and new media. Washington, DC: Georgetown University Press, 1-25.

Kloess, Juliane A., Hamilton-Giachritsis, Catherine E. and Beech, Anthony R., (2017). Offense processes of online sexual grooming and abuse of children via internet communication platforms. *Sexual Abuse,* 1-24. Doi: 10.1177/1079063217720927

Kontostathis, April, Edwards, Lynne and Leatherman, Amanda, (2009). ChatCoder: Toward the tracking and categorization of internet predators. *Workshop held in conjunction with the Ninth SIAM International Conference on Data Mining (2009*).

Locher, Miriam and Watts, Richard, (2005). Politeness theory and relational work. Journal of Politeness Research, 1 (1): 9-33.

Lorenzo-Dus, Nuria, Izura, Cristina and Pérez-Tattam, Rocío (2016). Understanding grooming discourse in computer-mediated environments. *Discourse, Context and Media*, 12, 40-50. Doi: 10.1016/j.dcm.2016.02.004

Lorenzo-Dus, Nuria and Izura, Cristina (2017). ‘‘cause ur special’’: Understanding trust and complimenting behaviour in online grooming discourse. *Journal of Pragmatics*, 112, 68-82. Doi: 10.1016/j.pragma.2017.01.004

Lorenzo-Dus, Nuria and Kinzel, Anina (2019). ‘So is your mom as cute as you?’: Examining patterns of language use by online sexual groomers. *Journal of Corpora and Discourse Studies,* 2, 14-39.

Mey, Jacob, (2001; 2nd edition). *Pragmatics – An Introduction*. London: Wiley Blackwell.

Morgan, Jay, Paiement, Adeline, Lorenzo-Dus, Nuria, Kinzel, Anina and Di Cristofaro, Matteo (under review). Hybrid Deep Neural Networks with Prior Corpus Linguistics Knowledge: An Application for Detecting Online Grooming.

O'Connell, Rachel, (2003). *A typology of child cybersexploitation and online grooming practices*. Preston, England: University of Central Lancashire. Retrieved from: http://image.guardian.co.uk/sys-files/Society/documents/2003/07/17/Groomingreport.pdf

Quayle, Ethel and Newman, Emily, (2017). An exploratory study of public reports to investigate patterns and themes of requests for sexual images of minors online. *Crime Science*, 5(2), 1-12. Doi: 10.1186/s40163-016-0050-0.

Partington, Alan, (2006). *The Linguistics of Laughter A Corpus-assisted Study of Laughter-talk*. London: Routledge.

Partington, Alan, (2010). Evaluating evaluation and some concluding reflections on CADS. In John Morley and Paul Bayley (eds). *Corpus-Assisted Discourse Studies on the Iraq Conflict,* London: Routledge, 261-304.

Pennebaker, James W., Boyd, Ryan L., Jordan, Kayla, and Blackburn, Kate, (2015a). *The development and psychometric properties of LIWC2015*. Austin: University of Texas at Austin, TX. Retrieved from: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf

Pennebaker, James W., Booth, Roger J., Boyd, Ryan L. and Francis, Martha E., (2015b). *Linguistic Inquiry and Word Count: LIWC2015: Operator's Manual*. Austin: University of Texas at Austin, TX. Retrieved from: https://s3-us-west2.amazonaws.com/downloads.liwc.net/LIWC2015_OperatorManual.pdf

Schneevogt, Daniela, Chiang, Emily and Grant, Tim, (2018). Do Perverted Justice chat logs contain examples of overt persuasion and sexual extortion? A research note responding to Chiang and Grant (2017, 2018). *Language and Law = Linguagem e Direito*, 5(1), 97-102.

Tausczik, Yla R. and Pennebaker, James W., (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psycholog.* 29(1), 24-54. Doi: 10.1177/0261927X09351676.

Taylor, Charlotte and Anna Marchi (Eds.) *Corpus Approaches to Discourse. A Critical Review*. London: Routledge.

van Gijn-Grosvenor, Evianne L. and Lamb, Michael E., (2015). Behavioural differences between online sexual groomers approaching boys and girls. *Journal of Child Sexual Abuse*, 25(5), 577–596. Doi: 10.1080/10538712.2016.1189473.

Vartapetiance, Anna and Gillam, Lee, (2014). ‘Our little secret’: Pinpointing potential predators. *Security Informatics,* 3 (1): 1-19.

Williams, Rebecca, Elliott, Ian A. and Beech, Anthony, (2013). Identifying sexual grooming themes used by internet sex offenders. *Deviant Behavior*, 34(2), 135-152. Doi: 10.1080/01639625.2012.707550.

Winters, Georgia M., Kaylor, Leah E. and Jeglic, Elizabeth L., (2017). Sexual offenders contacting children online: an examination of transcripts of sexual grooming. *Journal of Sexual Aggression*, 23(1), 62-76. Doi: 10.1080/13552600.2016.1271146.