

The Language of Online Child Sexual Groomers - A Corpus Assisted Discourse Study of Intentions, Requests and Grooming Duration

Anina Kinzel, BA (Hons), MA

Submitted to Swansea University in fulfilment of the requirements for the Degree of Doctor of Philosophy in Applied Linguistics

Swansea University, 2021

Abstract

Online grooming has become a wide-spread and worryingly fast increasing issue in society. This thesis analyses a corpus of online grooming communication, made available by the Perverted Justice (PJ) archive, a non-profit organisation that from 2004 until 2019 employed volunteers, who pretended to be children and entered chat rooms to catch and convict groomers, collaborating with law enforcement. The archive consists of 622 grooming chat logs and approx. 3.7 million words of groomer language. A corpus of this database was built, and a Corpus-Assisted Discourse Studies (CADS) approach used to analyse the language therein. Specifically, the language was compared to a reference corpus of general chat language data (PAN2012) and duration of online grooming and manipulative requesting behaviour were also investigated. The following research questions were answered:

- 1) What are the features of a corpus of online groomer language compared to that of a general digital chat language reference corpus? Is online groomer language distinct? How are online grooming intentions realised linguistically by online groomers?
- 2) Does duration of grooming influence the grooming process/intentions? Is usage of specific words/specific grooming intentions associated with different duration of grooming? Can different duration profiles be established and, if so, what are the cut-off points for these duration profiles?
- 3) How are requests realised in online grooming and how does duration influence this? How do groomers make requests and what support move functions do they use? Does duration influence how requests are made, and the type of support move function that are used?

The thesis newly identifies nuanced linguistic realisations of groomers' intentions and strategies, proposing a new working terminology for discourse-based models of online grooming. This is based on a review of the literature followed by an empirical analysis refining this terminology, which has not been done before. It finds evidence for two distinct duration-based grooming approaches and yields a fine-grained qualitative analysis of groomer requests, also influenced by grooming duration. There have only been very few studies using a CADS analysis of such a large dataset of groomer language and this thesis will lead to new insights, implications and significance for the successful analysis, detection and prevention of online grooming.

Declaration and Statements

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed  (candidate)

Date: 30/06/2021

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed  (candidate)

Date: 30/06/2021

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed  (candidate)

Date 30/06/2021

STATEMENT 3

I hereby confirm that Swansea University's ethical procedures have been followed and ethical approval for this thesis was granted in December 2017.

Signed  (candidate)

Date 30/06/2021

Content warning

This thesis analyses a potentially distressing issue (online child sexual grooming) and includes examples from online grooming chat logs, some of which include explicit language use. These examples have been carefully chosen but might still be upsetting to the reader.

Contents

Abstract	1
Declaration and Statements	2
Content warning	3
Contents	4
Acknowledgments	8
List of Figures	9
List of Tables	9
Abbreviations	11
1. Chapter 1: Online Grooming - A Pressing Social Challenge.....	12
1.1 A Language-Based Approach to a Social Challenge (Online Grooming) .	14
1.2 Purpose/Aim of Study	18
1.3 Significance of Study	20
1.4 Organisation of Thesis	21
1.5 Conclusions on Online Grooming as a Societal Issue	26
2. Chapter 2: The Literature on Online Grooming and Digital Discourse	27
2.1 Towards a Definition of Online Grooming	28
2.2 The Online Grooming Process	30
2.2.1 Online Grooming Models	30
2.2.1.1 Sequential Models of Online Grooming	30
2.2.1.2 Non-Sequential Models of Online Grooming	39
2.2.2 The Duration of the Online Grooming Process.....	49
2.3 Participants in the Online Grooming Process	50
2.3.1 Online Groomers	50
2.3.1.1 Groomer Typologies – Psychological Approaches.....	51
2.3.1.2 Groomer Language Profiling	53
2.3.2 Decoys.....	56
2.3.3 Children.....	59
2.4 Digital Discourse Research	62
2.4.1 The Three Waves of Digital Discourse Research	62
2.4.2 Participation Frameworks in Digital Spaces	63
2.4.3 Online Dating Language	64
2.5 Summary: Taking Stock and Moving Forward.....	66
2.5.1 Overview of Online Grooming Communication	66
2.5.2 Overview of Participant Features.....	67
2.5.3 Overview of Online Grooming Communicative Intentions.....	68
2.5.3.1 Seeking Sexual Gratification	72
2.5.3.2 Developing Deceptive Trust	73
2.5.3.3 Arranging Further Contact	74
2.5.3.4 Assessing Risks and Isolating	75
2.5.3.5 Negotiating Power Dynamics	76
2.5.4 Conclusions on Existing Online Grooming Literature	77
3. Chapter 3: Choosing the Right Methodology	78

3.1	Data	79
3.1.1	Approaches to the Study of Online Grooming Communication	79
3.1.2	Two Quantitative Approaches to Online Grooming Language	82
3.1.3	Data Sources	86
3.1.3.1	Groomer-Decoy Data (Perverted Justice Archive)	87
3.1.3.2	Groomer-Child Data	88
3.1.4	Ethical Considerations	90
3.1.4.1	Ethical Considerations Using Groomer-Child Data.....	90
3.1.4.2	Ethical Considerations Using Groomer-Decoy Data	91
3.2	Framework: Corpus-Assisted Discourse Studies (CADS).....	92
3.2.1	What is CADS?	92
3.2.2	CADS Toolkit	95
3.2.2.1	Corpus Linguistics: Keywords and Keyness Measures	95
3.2.2.2	Corpus Linguistics: Collocations	98
3.2.2.3	Discourse Analysis: Speech Acts and Im/Politeness	100
3.2.3	Choosing the Right Corpus Linguistics Software: CQPWeb	105
3.3	Procedure	106
3.3.1	Data Extraction	106
3.3.2	Data Normalisation Process	109
3.3.3	Second Stage of Data Normalisation	112
3.3.4	Creating Sub-Corpora	114
3.3.5	Data Analysis	115
3.3.5.1	Keyword Analyses	115
3.3.5.2	Requests	119
3.4	Conclusions on Methodologies	121
4.	Chapter 4: The Features of Online Groomer Language.....	122
4.1	Chapter Methodology	124
4.2	Comparing Online Grooming to Digital Language (Keywords)	126
4.3	The Online Grooming Discourse Model (Grooming intentions).....	128
4.3.1	Developing Deceptive Trust	131
4.3.1.1	Relationship terms.....	132
4.3.1.2	Small talk/Conversation Filler	135
4.3.1.3	Bonding.....	139
4.3.1.4	Exact Location	141
4.3.2	Seeking Sexual Gratification	143
4.3.2.1	Sex Acts	144
4.3.2.2	Appearance/Clothes	145
4.3.2.3	Physical Contact.....	146
4.3.2.4	Contraception	148
4.3.2.5	Other.....	149
4.3.3	Arranging Further Contact	151
4.3.3.1	Offline	152
4.3.3.2	Phone.....	153
4.3.3.3	Day/Time	154
4.3.3.4	Online.....	155
4.3.4	Assessing Risks and Isolating	156
4.3.4.1	Secrecy	156
4.3.4.2	Supervision.....	157
4.3.4.3	Replace Parental Figure	157
4.3.4.4	Criticise Parental Figure.....	158

4.3.5	Negotiating Power Dynamics	159
4.3.5.1	Legality	160
4.4	Conclusions on the Features of Online Grooming.....	161
5.	Chapter 5: Duration of Online Grooming and its Impact on Discourse ..	165
5.1	Duration of Online Grooming – an Overview of the Corpus	166
5.2	One Corpus with Two Grooming Approaches?.....	171
5.2.1	Developing Deceptive Trust	175
5.2.1.1	Exact Location	176
5.2.1.2	Relationship Strengthening	178
5.2.1.3	Activities	180
5.2.1.4	Visual Information	180
5.2.1.5	Sociability	181
5.2.2	Arranging Further Contact	182
5.2.2.1	Offline	183
5.2.2.2	Phone.....	184
5.2.2.3	Time	184
5.2.3	Seeking Sexual Gratification	185
5.2.4	Assessing Risks and Isolating	186
5.3	Three or Two Grooming Duration approaches?	188
5.3.1	Developing Deceptive Trust	190
5.3.2	Arranging Further Contact	192
5.3.3	Negotiating Power Dynamics	193
5.4	Long or short Grooming? Investigating Long Duration Grooming	196
5.4.1	Developing Deceptive Trust	199
5.4.2	Seeking Sexual Gratification	203
5.4.3	Assessing Risks and Isolating	204
5.4.4	Arranging Further Contact	205
5.5	Towards a First Mapping of Two Grooming Duration Approaches.....	206
6.	Chapter 6: Power and Manipulation in Online Groomer Requests.....	212
6.1	Approach to Examining Requests in Online Grooming	213
6.2	Results	218
6.2.1	Overview	218
6.2.1.1	Requesting Behaviour	218
6.2.1.2	Support Move Functions	225
6.2.2	Requesting Behaviour in Online Grooming.....	228
6.2.2.1	Suggestory Formulae – Developing Deceptive Trust Focus	229
6.2.2.2	Obligation Statements – Developing Deceptive Trust Focus	239
6.2.2.3	Want Statements – Assessing Risks and Isolating Focus	257
6.3	Conclusions on Manipulative Requesting Behaviour and Duration	264
7.	Chapter 7: Conclusions on Online Groomers’ Discourse	268
7.1	Introduction	268
7.2	RQ1: Are the Features of Online Groomer Language Distinct?.....	268
7.2.1	Online Groomer Language Features are Distinct.....	269
7.2.2	Implications for Online Grooming Research: A Revised Terminology 276	
7.2.3	New Resources for Online Grooming Detection Algorithms	279
7.3	RQ2: Does Duration Affect the Online Grooming Process?	280
7.3.1	Duration Affects the Grooming Process: Two Different Approaches	281

7.3.2	Implications of Two Online Grooming Duration-Based Approaches	.283
7.3.3	Tailoring Law Enforcement Efforts to Tackle Online Grooming284
7.4	RQ3: Does Duration Influence Groomers' Requesting Behaviour?285
7.4.1	Duration Influences Online Groomers' Requesting Behaviour286
7.4.2	Towards a Description of Two Groomer Duration Profiles288
7.5	Drawing Conclusions on a Study of Online Groomer Language291
7.5.1	Limitations294
7.5.2	Future Direction in Online Grooming Language Research296
7.6	Conclusion297
8.	References298
9.	Appendix316
10.	Glossary336

Acknowledgments

First, I would like to thank my supervisor Prof Nuria Lorenzo-Dus for her unwavering belief in me, for providing guidance and feedback when I needed it and for helping me shape this project into a thesis. My thanks also go to my first second supervisor Professor Emeritus James Milton, who retired before I finished my PhD (perhaps I took too long ☺). Thank you for your invaluable insights and advice, which helped to shape the corpus research design of my thesis. Thanks, Dr Alexia Bowler, my second supervisor, who agreed to take over second supervisor duties and who gives excellent feedback and advice. Thanks also to the Department of Applied Linguistics at Swansea University for providing such a warm and welcoming research environment. I have thoroughly enjoyed my time with you during my three undergraduate years and (almost) four years as a PhD student. I am also very grateful to my supportive research team. Thank you Keighley, Ruth and Rosie for everything you do. I would also like to thank Mr Jay Morgan and Dr Di Cristofaro for help with data extraction and preparation.

I would not have been able to get this far without my parents, Holger and Andrea Kinzel, who are some of the best people I know – but I am biased. Thank you for always encouraging me, providing invaluable support, keeping me motivated and most importantly, letting me videocall my cat! I would not be where I am without you, thank you both. I would also like to thank my sister Jana simply for being my favourite sister. Thanks also go to my partner Francis for always being there for me and for believing in me. I am also deeply grateful to my friends and support network here in Swansea and further afield. You are amazing, and I am so happy to have found my people. A special shoutout to those of you asking when I will finally be done with my PhD. Can confirm, it will be very soon! I am aware this acknowledgments section could never do any of you justice but from the bottom of my heart – thank you.

List of Figures

Figure 1: Online grooming model introduced by Lorenzo-Dus et al (2016: 70).....	41
Figure 2: The online grooming techniques and effect by Whittle et al. (2014b: 419)	60
Figure 3: Online groomer communicative intentions & studies reporting them.....	69
Figure 4: Overview of “seeking sexual gratification” strategies.....	72
Figure 5: Overview of "developing deceptive trust" strategies.....	74
Figure 6: Overview of "arranging further contact" strategies	74
Figure 7: Overview of “assessing risks and isolating” strategies.....	75
Figure 8: Overview of “negotiating power dynamics” strategies.....	76
Figure 9: PJ archive example chat log.....	87
Figure 10: Screenshot of a KWIC using the keyword "test" in GN corpus on CQPWeb	98
Figure 11: CQPWeb screenshot of the easy-to-use front-end interface	106
Figure 12: VARD screenshot of acn suggestions.....	110
Figure 13: VARD screenshot of example word that is added as a variant.....	111
Figure 14: VARD screenshot in the second stage of normalisation (before normalisation).....	113
Figure 15: VARD screenshot in the second stage of normalisation (after normalisation).....	113
Figure 16: Screenshot of CQPWeb visualisation of PJ normalised corpus.....	114
Figure 17: Step One - Examine instances of keyword	116
Figure 18: Step Two - calculate collocational database	117
Figure 19: Step Three - Examine instances of the strongest collocates from step 2	118
Figure 20: Example of more extensive analysis of 50 instances of the keyword.....	118
Figure 21: Duration overview - Number of chat logs per number in 100-minute steps.....	169
Figure 22: General patterns in first analysis	207
Figure 23: The two grooming approaches.....	208
Figure 24: Short grooming duration similarities	208
Figure 25: Comparing G5 and G6	209
Figure 26: Comparison of short and long grooming analyses.....	210
Figure 27: Distribution of online grooming intentions in request concordance lines	221
Figure 28: Request support in request concordance lines (n=567) in G3 in percentage	222
Figure 29: Request support in request concordance lines (n=1004) in G4 in percentage	222
Figure 30: Support move functions in three head act types in G3 (in percentage)	228
Figure 31: Support move functions in three head act types in G4 (in percentage)	228
Figure 32: Revised terminology for online grooming communicative intentions.....	278
Figure 33: The two duration-based grooming approaches	288

List of Tables

Table 1: Number and percentage of strategies to carry out intentions	72
Table 2: Groomer normalised (GN) corpus details	79
Table 3: Online grooming language/communication studies, chronologically	80
Table 4: Six sub-corpora created for chapter five (sub-sections 5.2, 5.3 and 5.4)	115
Table 5: Search queries sorted into Blum-Kulka et al.’s (2011) head act framework.....	120
Table 6: Word count for GN and PAN2012.....	126
Table 7: Top five keywords in GN compared with PAN2012	127
Table 8: Top keywords in GN compared with PAN2012	128
Table 9: Keywords sorted into grooming intentions	129
Table 10: Keywords that did not easily align with one or more grooming intention.....	131
Table 11: Five strategies and their keywords in developing deceptive trust.....	132
Table 12: Five strategies and their keywords in the seeking sexual gratification	143
Table 13: Four strategies and their keywords in arranging further contact.....	152
Table 14: Four strategies and keywords in assessing risks and isolating	156

Table 15: Strategy and keywords in negotiating power dynamics.....	159
Table 16: Overview of duration sub-corpora and analyses	171
Table 17: Details of duration sub-corpora G3 and G4	171
Table 18: Top 50 keywords of G3 compared with G4	172
Table 19: Keywords of G3 and G4 sorted into grooming intentions	174
Table 20: Four strategies in G3 within developing deceptive trust.....	175
Table 21: Three strategies in G4 within developing deceptive trust	176
Table 22: Three strategies in G3 within arranging further contact.....	183
Table 23: Three strategies in seeking sexual gratification in G3.....	185
Table 24: Details of G1 & G2.....	188
Table 25: Keywords of first two sub-corpora.....	189
Table 26: Keywords of G1 and G2 sorted into grooming intentions	190
Table 27: Details of fifth and sixth sub-corpora.....	196
Table 28: Top 50 keywords of G5 compared with G6	197
Table 29: Keywords of fifth sub-corpus sorted into grooming intentions	198
Table 30: Request strategy types taken from Blum-Kulka et al. (2011: 349-350).....	214
Table 31: The formulaic request head act types and corpus search queries.....	216
Table 32: Breakdown of number of request concordance lines per head type act	217
Table 33: Online grooming intentions in request head act types in G3.....	218
Table 34: Online grooming intentions in request head act types in G4.....	219
Table 35: Support (in %) in G3 and G4 per head act type	223
Table 36: Selection of head acts and number of concordance lines analysed	224
Table 37: Support move functions in online grooming request concordance lines.....	225
Table 38: Overall number of support move functions.....	227
Table 39: Support move function in suggestory formulae concordance lines in G3.....	229
Table 40: Support move function in suggestory formulae concordance lines in G4.....	229
Table 41: Support move function in obligation statement concordance lines in G3	239
Table 42: Support move function in obligation statement concordance lines in G4	239
Table 43: Support move function in want statement concordance lines in G3	257
Table 44: Support move function in want statement concordance lines in G4	257
Table 45: Extract of table of 116 keywords identified within grooming intentions.....	279
Table 46: Extract of table of 52 keywords with overlapping grooming intentions	280
Table 47: Extract of keywords associated with the two different duration approaches	285
Table 48: Positive and negative politeness strategies with examples from GN corpus	316
Table 49: Impoliteness strategies with examples from GN corpus	317
Table 50: Keywords of GN compared with PAN2012.....	318
Table 51: Keywords of G3 and G4.....	322
Table 52: All keywords of G5 and G6.....	326
Table 53: Full list of newly identified keywords in the first empirical chapter	327
Table 54: Keywords belonging to more than one online grooming intention	330
Table 55: Unique keywords associated with the two grooming duration approaches	332
Table 56: Keywords with more than one grooming intention.....	332

Abbreviations

<i>Abbreviation</i>	<i>Meaning</i>	<i>Page</i>
AFC	Arranging Further Contact	69
AIDA	Action-Implicative Discourse Analysis	56
AOL	America Online	65
ARISO	Assessing Risks and Isolating	69
CADS	Corpus Assisted Discourse Studies	17
CDA	Critical Discourse Analysis	95
CL	Corpus Linguistics	95
CLEF	Conference and Labs of the Evaluation Forum	126
CMDA	Computer-Mediated Discourse Analysis	41
CSA	Child Sexual Abuse	90
CSEM	Child Sexual Exploitation Material	51
CSV	Comma-Separated Values	110
DDT	Developing Deceptive Trust	69
DHA	Discourse Historical Approach	95
EFL	English as a Foreign Language	103
FIRE	Forum for Information Retrieval	126
GN	Groomer Normalised	80
HTML	HyperText Markup Language	108
IM	Instant Messenger	57
IRC	Internet Relay Chat	126
KWIC	Keyword in Context	83
LIWC	Linguistic Inquiry and Word Count	22
LL	Log Likelihood	96
LR	Log Ratio	97
NF	Normalised Frequency	97
NPD	Negotiating Power Dynamics	70
OG	Online Grooming	42
OGDM	Online Grooming Discourse Model	21
OP	Online Predator	56
PJ	Perverted Justice Foundation	16
SAT	Speech Act Theory	102
SMS	Short Message Service	110
SSG	Seeking Sexual Gratification	69
URL	Uniform Resource Locator	108
VARD	Variant Detector	21
XML	Extensible Markup Language	109

1. Chapter 1: Online Grooming - A Pressing Social Challenge

This introduction chapter will give a brief overview of online grooming as a societal issue and will explain why it is worth investigating, specifically looking at its effects, rise and speed. It will then introduce relevant debates on online grooming in the academic literature that are guiding this work. These debates relate to whether online grooming is a sequential or non-sequential process and the key linguistic literature on online grooming will be presented in 1.1. It will also give an overview of the Perverted Justice Archive as a data source in the same sub-section. The aims and three research questions will be presented in sub-section 1.2 and the significance of the study will be outlined in sub-section 1.3. Then, the organisation of this study will be laid out in sub-section 1.4 and brief conclusions will be drawn in sub-section 1.5.

Online child sexual grooming is not easy to define, but it is often described as an adult using the internet to befriend an underage person with the intent of abusing them sexually, either online, “offline” or both (see chapter two, section 2.1 for a discussion about definitions and terminology). It is a widespread problem that is on the rise. Children and young people who have been groomed are affected by feelings of self-blame, intrusive thoughts, nightmares, social withdrawal, depression, and low self-esteem. Some experience suicidal ideation, post-traumatic stress disorder and severe anxiety, which may lead to long-term difficulties and psychological repercussions later in life. The impact of online grooming is as harmful as “offline” grooming, yet young people sometimes feel like they receive less support if they were groomed online (Hamilton-Giachritsis et al., 2017). Many children and young people may not realise they have been groomed, believing themselves to be in a trusting relationship with the groomer, rather than being manipulated for the purpose of sexual abuse/exploitation (Lorenzo-Dus & Kinzel, 2021). It can be difficult for children who have been groomed to trust others again, as their trust has been broken (Marchenko, 2017).

In the UK, Online grooming has been steadily increasing and seen a sharp rise in 2013, likely owing to several high-profile grooming cases in Rotherham, Rochdale and Oxford covered in mainstream media, which encouraged people to disclose their own abuse (BBC, 2015). In December 2015, the UK’s National Society for the Prevention

of Cruelty to Children (NSPCC) reported receiving almost 50% more inquiries into online grooming than in 2014. This number further rose in 2017, specifically in the North East England, where online grooming offences rose by 325% compared to previous years (ITV, 2017). Part of the reason for this exponential rise was an addition to the Serious Crime Act¹ making it a criminal offence for an adult to send a sexual message to a child in the UK, which also came into effect that same year. This was previously not illegal in Wales and England. In Scotland, the Sexual Offences Act had been introduced in 2009 and came into effect in 2010. Part 4, section 24 made it a criminal offence to communicate indecently with an underage person, which includes written messages². A total number of 10,391 of sexual cases involving an online element were noted in 2019/2020, an increase of 16 % compared to the previous year, across all police forces in the UK (NSPCC, 2020a). Most recently, during the first 2020 covid-19 lockdown, online grooming instances increased as well. The NSPCC reported a 16 % rise in counselling sessions with children about grooming increased during lockdown (NSPCC, 2020a). Another report from late 2020 stated that 1,220 offences were carried out online during the first three months of lockdown (NSPCC, 2020b). In the US, the FBI warned that every day there are more than 500,000 online groomers active using multiple accounts and offering game cheat codes to lure in their targets (Nelson, 2021). They also cautioned that online grooming could lead to sextortion, a victim producing sexually explicit material of themselves and sending it to the groomer. This can in turn result in blackmail and suicide, as was the tragic case of Evan McDaniel, a Fourteen-year-old boy who was groomed and blackmailed online in early 2021 (Nelson, 2021). The Australian Centre to Counter Child Exploitation (ACCCE) also reported a recent increase in online child sexual exploitation. They stated to have received 21,000 reports of child sexual exploitation and intercepted more than 250,000 child abuse material files in just 12 months in 2020 (ACCCE, 2020).

These numbers and statistics are based on the online grooming cases that have been reported, but there are many cases that go unreported either because the children and young people are ashamed, blame themselves or do not realise they were groomed.

¹ <https://www.legislation.gov.uk/ukpga/2015/9/section/67/enacted>

² <https://www.legislation.gov.uk/asp/2009/9/part/4>

While there has been a steady increase in online grooming in recent years, there is also a concern of how fast groomers groom children. There have been reports that grooming can happen in very short amount of time, but duration of grooming is nonetheless an area that has hardly ever been researched. As a 2012 UNICEF technical report put it: ‘These processes of online grooming may take minutes, hours, days or months, depending on the goals and needs of the abuser and the reactions of the young person.’ (UNICEF, 2012: 15). The government cited grooming could happen in under 45 minutes (Home Office, 2018) while the NSPCC quoted a timeframe of under 20 minutes from initial contact (NSPCC, 2018). The groomer convinces their target to meet face-to-face for sexual purposes in that short amount of time (Davis, 2016).

1.1 A Language-Based Approach to a Social Challenge (Online Grooming)

What has the academic community done about the problem of online grooming? Most research on online grooming has been done primarily in Psychology, Criminology and Computer Science, more specifically Machine Learning, and little attention has been paid to the *language* of online groomers, which will be reviewed in more detail in chapter two. The few studies that have analysed the language of online grooming have sought to model the online grooming process and the kinds of behaviours identified in it. They looked at the elements of the online grooming process and labelled them, while also trying to figure out how grooming happened and whether there was an order to these elements. These studies can be divided into two camps: sequential and non-sequential proposed models, which encompass these identified elements or processes of online grooming. Four sequential models have been introduced by Egan et al. (2011), Gupta, et al. (2012), Black et al. (2015) and Kloess et al. (2017a). The former two models were developed from a Machine Learning perspective while the latter two came from the field of Psychology. Even though they examined the language of online grooming, they had no linguistic input or analysis. These models are partly based on the first model of online grooming put forward by O’Connell (2003) which proposed five sequential stages of online grooming. The other camp, the non-sequential models, were developed by Williams et al. (2013), Lorenzo-Dus et al. (2016), van Gijn-Grosvenor and Lamb (2016), Quayle and Newman (2017) and Chiang and Grant

(2017, 2018). Although more recent work into online grooming regards the online grooming process as non-sequential, the sequential models have not been disproved (see chapter two, section 2.2.1 for a discussion about sequential and non-sequential models). Non-sequential grooming is also the stance adopted in this thesis, which will move beyond this debate of sequential or non-sequential and will examine a more nuanced view of online grooming language. Research on the language of online grooming, especially from a linguistic viewpoint, has been quite underdeveloped in comparison to other areas such as groomer characteristics and motivations and victim characteristics and risk factors.

Within the few studies focusing on the language of online grooming, there have only been a handful of studies conducted within Linguistics, even though online grooming is primarily a language-based phenomenon. Lorenzo-Dus et al. (2016) used Discourse Analysis informed by Pragmatic Act Theory and Relational Work to develop the first online grooming *discourse* model, which encompasses a complex network of five overlapping online grooming processes. In a follow up study, Lorenzo-Dus and Izura (2017) analysed compliments in online grooming chat logs based on speed of grooming. Chiang and Grant (2017, 2018) used Genre Analysis and developed a list of rhetorical moves offenders typically used in order to achieve their goals of grooming targets. Three new moves were identified in Chiang and Grant (2018). Schneevogt et al. (2018), Lorenzo-Dus and Kinzel (2019, 2021), and Lorenzo-Dus et al. (2020) used Corpus Linguistics tools to examine online grooming language (see chapter three, section 3.1.1). The thesis will contribute new insights to this recent Linguistics scholarship. Specifically, it will analyse one aspect of online grooming that has not been looked at systematically: Speed or duration of grooming, which as mentioned in section 1, is a significant concern. There is no known average length of time groomers spend with their targets. A few studies mention different durations such as between ten minutes and five hours of interaction over a timeframe of a day to six months. One study looking at compliments in online grooming divided their data into three speed groups which ranged from under four hours to over 11 hours (Lorenzo-Dus & Izura, 2017). Another study found evidence that just under half of the groomers interacted with their target for less than a day before arranging to meet “offline” (Briggs et al., 2011). There is also evidence in the literature that sexual content has been introduced into the conversation very early on (Black et al., 2015; Kloess et al.,

2017a) with one study citing 33 minutes into the conversation (Winters et al., 2017). This suggests that there is a wide range of duration in online grooming with different grooming processes and strategies introduced at different points in the process. This makes sense, since a person trying to convince someone of something will use different language if they only spend 20 minutes on the task, compared with hundreds of hours. It is important to delve deeper into this duration aspect, since groomers will likely take different approaches depending on the time they spend interacting with their targets. Specifically, the thesis examines the impact duration has on the online grooming process and different groomer approaches. Capturing these different approaches and creating different duration profiles for groomers could help with the detection of online groomer language and might be used by law enforcement to prevent online grooming from happening. It could be used to inform algorithms that detect online grooming on social media platforms. Knowledge gained from an analysis into grooming duration could also help inform training for professional groups with safeguarding responsibilities (e.g., children charities, teachers, therapists), which would help detect and prevent online grooming from happening. It might also be used to educate and train children about what to watch out for when interacting in an online space.

While the academic community has certainly done research on online grooming, most of the above research has relied on the same data source: The Perverted Justice Foundation archive. The Perverted Justice (PJ) Foundation was a non-profit organisation founded in 2003 and based in California and Oregon who had a number of adult volunteers or decoys that entered chat rooms online in search of groomers. In 2006, the number of trained decoys was 65 (Salkin, 2006). These decoys waited for groomers to approach them and if the conversation turned sexual, they collaborated with law enforcement to have the online groomers arrested and convicted. The resulting chat log, which is a record of the entire interaction between groomer and decoy from first contact to the agreed face-to-face meeting, was then uploaded to the website after a successful conviction (see chapter three, section 3.1.3). The archive consists of 623 convicted groomers, 622 that were based on chat log evidence and one research conviction. In 2004 PJ collaborated with the TV channel NBC to air 11 undercover sting operations in a tv series called *To Catch a Predator*, which ran from 2004 until 2007 on *Dateline NBC* and was presented by Chris Hansen. In it, an online

groomer or predator was given an address of a seemingly underage person they had been chatting to online and whom they were hoping to sexually abuse. The address was actually a house in which an undercover sting operation had been set up, and the groomer would be greeted by Chris Hansen and a camera team when they turned up. In later episodes, police would also wait at the house to make an arrest (Salkin, 2006). In 2019, PJ ceased all operation, and the website was turned into an archive (Perverted Justice, 2019). The founder, Xavier von Erck, cites difficulty in keeping up with technological advances and groomers using social media sites, rather than chatrooms, making it more difficult to catch them, as the reason for this decision (Kozłowska, 2019).

Though studies into online grooming have used the PJ archive as a data source (see chapter three, section 3.1.3 for a more detailed discussion about data sources in online grooming), there are only very few studies which have analysed the entirety of the dataset using quantitative methods. One quantitative study looked at 590 chat logs, which was the entirety of the archive in 2017 at the time the study was published (Drouin et al., 2017). The three studies mentioned above using Corpus Linguistics have used the whole archive since the foundation's termination in 2019³ to analyse communicative patterns in online grooming language (Schneevogt et al., 2018; Lorenzo-Dus & Kinzel, 2019, 2021; Lorenzo-Dus et al., 2020). Lorenzo-Dus and Kinzel (2019, 2021), and Lorenzo-Dus et al. (2020) have been the first to apply a Corpus-Assisted Discourse Analysis (CADS) approach (see chapter three, section 3.2 for an overview of CADS) to online grooming, combining quantitative analyses of the communicative patterns of online groomers with more fine-grained qualitative analysis, analysing different aspects of online grooming. These studies show that a CADS approach lends itself well to studying online grooming. This thesis will add value to this limited body of research, adopting the same approach to analyse duration-based patterns in online grooming language across the entirety of the PJ archive for the first time.

³ No new chat logs were added to the PJ archive after 2016. The total number remains 622 chat log convictions and one research conviction.

As the above has shown, online grooming is an under-researched area, especially in Linguistics, even though online grooming is broadly understood to be a linguistic phenomenon. Other cybercrime contexts, such as terrorism and online extremism have benefitted from more linguistic scrutiny. In recent years, terrorism propaganda has been analysed using Corpus Linguistics and CADS methods (see Kinzel, 2016; Baker & Vessey, 2018; Lorenzo-Dus et al., 2017; Lorenzo-Dus et al., 2018). In the case of terrorism, integrating ‘linguistic tools of enquiry can advance current knowledge of jihadist ideology groups’ (Lorenzo-Dus et al., 2018: 1). Lorenzo-Dus et al. (2017) argue that a Discourse Analytical view of analysing terrorism ‘can add significant value to the predominant language-based Content Analysis conducted in the field of Terrorism Studies’ (Lorenzo-Dus et al., 2017: 167). It does so by combining qualitative and quantitative research methods and concepts from other disciplines. This is very similar to online grooming research, which has mostly relied on language-based Content Analysis and would benefit from more linguistic inquiry and study. This thesis will build on the recent work that has focused on integrating qualitative and quantitative linguistic methods and will add new insights through focus on duration of online grooming, which will be outlined next.

1.2 Purpose/Aim of Study

The aim of this study is to analyse the language of online groomers and establish its features using the PJ archive of chat logs between convicted groomers and their targets (adult volunteers pretending to be children). Throughout the thesis, decoys will be referred to as targets. It is assumed that the online groomers involved in the PJ archive expected these decoys to be underage and talked to them like they would other young people. Pronouns will also be consistently used, groomers will be referred to by male pronouns, because the entirety of the PJ archive is made up of male groomers. Targets will be referred to by gender neutral pronouns, as there were PJ volunteers pretending to be both male and female underage children.

When conducting analyses in which features of language are established, it is useful to have a comparator within the same broad register or genre. In this case, this comparator will be general digital chat language use, which is similar enough to the data for this thesis, online grooming chat logs. Although some prior knowledge about

the features of online groomer language exists, as introduced in sub-section 1.1 above and in the literature review chapter (see section 2.2 & 2.3.1.2), most of this research conducted Content Analysis on the language and used qualitative approaches and had no linguistic input. This thesis will build on recent work providing linguistic knowledge and will integrate quantitative and qualitative methods to examine the features of online groomer language and comparing it to general digital chat language. It will also analyse whether the language of online groomers is distinct from digital chat language. As such, the first research question in the first empirical chapter (chapter four, section 4.1) is as follows:

1. What are the features of a corpus of online groomer language compared to that of a general digital chat language reference corpus?
 - a. Is online groomer language distinct?
 - b. How are online grooming intentions realised linguistically by online groomers?

The study will also examine what role duration plays in the online grooming process and how this influences the grooming processes and strategies groomers use in their discourse. As mentioned above, this has been an overlooked aspect of the grooming process. This will be done by comparing six different grooming duration groups based on two duration cut-off points to each other. The research question in the second empirical chapter (chapter five, section 5) is:

2. Does duration of grooming influence the grooming process/intentions?
 - a. Is usage of specific words/specific grooming intentions associated with different duration of grooming?
 - b. Can different duration-based grooming profiles be established, and if so, what are the duration cut-off points?

Thirdly, the study will look at how groomers manipulate their targets and negotiate power dynamics. Broadly speaking, online grooming is a form of manipulation, which is created out of an inherent power imbalance in the groomer's favour since the groomer is older and more experienced than their supposed target. Manipulation is achieved when the interlocutor does not notice or resist the manipulation. This can be

because of lack of relevant information, their fundamental, strong emotions, or their social position (van Dijk, 2006). In the case of online grooming, the target's social situation, strong emotions and lack of relevant information would make them susceptible to manipulation, because the groomer's ulterior motive is to sexually abuse their targets, which they keep hidden from them. The groomers will attempt to build relationships with their targets, which is sure to produce strong emotions. Groomers will also attempt to convince their targets to take part in sexual activities online (such as sharing explicit photos or videos, discussing sex acts, and instructing targets to masturbate). Even just eliciting private information from targets, such as their phone number or address are acts of manipulation. These specific attempts can be described as requests, which can be manipulative, especially in online grooming contexts. Therefore, this study will examine how groomers perform requests in interactions with their targets and how duration may influence this. The third research question, which will be answered in the last empirical chapter (chapter six) is as follows:

3. How are requests realised in online grooming and how does duration influence this?
 - a. How do groomers make requests and what support move functions do they use?
 - b. Does duration influence how requests are made, and the type of support move function that are used?

This last empirical chapter is guided by findings from the other two empirical chapters and focuses on a more qualitative, Corpus Pragmatics approach.

1.3 Significance of Study

Guided by the aims outlined above, the study makes several important contributions. Firstly, as mentioned above, very few studies have systematically examined the entirety of the Perverted Justice (PJ) archive, which this thesis does. This archive has potential to answer questions about online grooming that have not been asked yet, like the research questions above.

Additionally, the study will advance knowledge of the Online Grooming Discourse Model (OGDM) developed by Lorenzo-Dus et al. (2016) using a bigger dataset and a novel methodology, and therefore contributing new insights to academic knowledge of online grooming. It will provide more nuanced aspects of how the online grooming intentions introduced in the OGDM are actually realised linguistically by groomers, which could be used to inform Machine Learning algorithms to develop online grooming detection software. A list of specific words or communicative patterns associated with short and long duration grooming based on statistic measures as well as detailed language-based strategies might be used to train Artificial Intelligence to spot these specific words in private social media conversations, text messages or other communication online, which would significantly help law enforcement flag risky conversations and catch online groomers.

Another specific contribution to the online grooming scholarship is the analysis of duration of grooming, an overlooked aspect in the literature with the potential to make online grooming detection more specific by providing detailed analysis of different groomer approaches based on how long groomers interact with their targets. By identifying specific words and online grooming processes that are tied to different durations of grooming and developing duration profiles, law enforcement's efforts to detect, disrupt and prevent online grooming could become more specialised and tailored to the groomers' duration approaches. Different counter-grooming profiles could be established and specific training about the differences between shorter duration grooming and longer duration grooming and what to look out for could be developed and given to individuals with safeguarding responsibilities, parents, and children. Moreover, the study will test tools used in Corpus Linguistics, such as Variant Detector (VARD) used for normalising corpora, which was not developed for modern English or digital language.

1.4 Organisation of Thesis

The thesis is organised as follows: In the next chapter, the literature will be reviewed. This will be done by firstly defining the concept of online grooming and clarifying terminology by discussing existing definitions of online grooming in the literature and then outlining the definition used in this study in sub-section 2.1. The terms online

grooming and online child sexual grooming will be used throughout this study. The chapter will then move on to examining the grooming process by reviewing studies focusing on online grooming language. Proposed sequential and non-sequential grooming models are reviewed. The sequential models rely on small data sizes, self-reporting and trusting Linguistic Inquire Word Count (LIWC) scores without validation and two of them find strategies overlapping, thus there is little evidence the online grooming process is sequential. The non-sequential models are able to identify more complex models and strategies, as they do not rely on O'Connell's (2003) five stages. These studies still rely on mostly small data sizes which is one of the gaps in the literature. Duration is the next area reviewed, showing that it is an overlooked aspect of online grooming but there is evidence that it makes a difference to the language of online groomers. The third section of the literature review is concerned with participants in the online grooming process examining online groomers. More specifically, existing online groomers typologies and studies on groomer language profiles will be covered. Then, the focus will be on decoys and children as participants involved in the online grooming process. Digital Discourse research will be covered, looking at how the field emerged and changed over the last 30 years, examining participation frameworks in the context of online grooming and online flirting, and dating language research. The next section of the chapter will summarise and draw some conclusions by providing an overview of online grooming communication and participant features. The literature review will conclude by proposing a new terminology comprising five online grooming communicative intentions that make up the online grooming process and which have been referred to by different terms in the literature. A short conclusion paragraph follows.

Chapter three is the methodology chapter, which features three broad sub-sections. In the first one the data size and source for this study will be presented and then different data sizes and approaches to online grooming are compared. To do so, 17 papers on online grooming with different methods ranging from purely quantitative to purely qualitative are reviewed. They also differ in their data sizes showing that only recently a new line of enquiry using large datasets has emerged which this study will contribute to by using the entirety of the PJ archive and an approach combining quantitative and qualitative methods. In the next sub-section, different data sources and ethics will be reviewed. This section will point out the difficulties in securing groomer-child data

rather than groomer-decoy data which most studies have used. Ethics around online child sexual grooming research will also be considered in the next sub-section and the decisions for this study concerning anonymisation, use of decoy data, and researcher safety and wellbeing will be outlined. The framework, Corpus-Assisted Discourse Studies (CADS) used in this study is explained next by describing what CADS is and the toolkit it consists of, focusing on the Corpus Linguistics contribution to it first and then moving on to the Discourse Analysis concepts used in the study, namely Speech Act Theory and Im/Politeness. The procedure will be outlined in the third broad sub-section by first explaining the data extraction process using Python scripts, then describing the data normalisation process using the Variant Detector (VARD) software and lastly stating how the data analysis in the three empirical chapters (chapters four-six) was carried out.

The study will then move on to data analysis in the three empirical chapters. Chapter four is the first empirical chapter, which consists of a keyword analysis of the online grooming corpus compared with a general reference corpus (PAN2012). First, the data is described and both study corpus and reference corpus are defined. Next, the research question is introduced which concerns what the features of online groomer language are and how they compare to a more general digital language corpus. The keywords are shown and briefly analysed, showing many high frequency keywords that also feature some online grooming processes. A keyword analysis comparing the groomer corpus to PAN2012 is carried out and the keywords are sorted into existing online grooming processes taken from the OGMD by Lorenzo-Dus et al. (2016). The chapter finds that all grooming intentions outlined in the literature chapter and described in Lorenzo-Dus et al.'s (2016) OGDM are identified in the keyword analysis. The grooming intentions are each analysed separately, identifying specific keywords used for strategies within the grooming intentions. The chapter shows that online grooming is distinct from other digital language, represented by PAN2012.

Based on the finding that online grooming language is indeed distinct, the second empirical chapter delves into the online grooming process further by focusing on the duration of grooming and whether it influences the online grooming process and intentions. It also asks whether different duration profiles can be established and what the cut-off points are. The data for the chapter is introduced first. The chapter features

three keyword analyses with three duration cut-off points. These cut-off points are defined by looking at a duration overview and data on how much teenagers chatted online which is used as a benchmark. Then, the six sub-corpora are introduced. The three keyword analyses find that duration does make a difference and influences the specific online grooming intentions and strategies groomers use. The more time groomers spend interacting with the decoys, the fewer intentions are identified at the keyword level with short duration groomers using four intentions and long duration groomers using two. These are the two distinct duration approaches identified in the chapter. Short duration groomers make use of a variety of grooming intentions focusing on elicitation of personal information, location sharing, arranging further contact with the decoy, seeking out visual information and discussing explicitly sexual themes. Long duration groomers use terms of endearment, express feelings, and emotions, use emoticons and sociability to give the decoy the impression they are in a loving and caring relationship. The second keyword analysis within this chapter suggests that the two short duration groomer corpora are fairly homogenous with few differences. The chapter concludes that there are two distinct duration profiles, but the cut-off point cannot be established and only estimated. Negotiating power dynamics only appears on the keyword level in one sub-corpus keyword analysis and seems like a more elusive grooming intention than previously thought.

The final empirical chapter is chapter six, which is guided by the previous two empirical chapters, specifically the finding that duration influences which grooming processes are used. It will examine how groomers manipulate their targets. Manipulation underpins the online grooming process as a whole. This is what the third empirical chapter focuses on by zooming in on how groomers use request structures in their manipulative discourse, and how they use politeness and impoliteness strategies to support their manipulative requests. It will compare the requesting behaviour of the two main duration groups from the second empirical chapter. This chapter is a qualitative analysis of the language of online grooming using Corpus Pragmatics. It focuses on a sub-set of the groomer corpus, which are request sequences in context. First, a short overview of requests in the literature is provided and the chapter introduces request head act types. Then, it will outline how the request sequences and expanded concordance lines were extracted from the corpus using search queries that reflect six of the nine head act types. The research question of this

chapter concerns how requests are realised in online grooming and how duration influences the requests. It also asks about the support move functions groomers use. The head act types will be compared generally before going into more detail about four specific head act types. The chapter finds that short and long duration groomers use requests and support move functions differently and strategically in these four head act types in combination with politeness and impoliteness strategies. While short duration groomers minimise the imposition, long duration groomers use vague language to create a coded language with their targets, creating trust. Intensifying is used by short duration groomers to harass, and by long duration groomers to clarify information. Terms of endearments are used by both short and long duration groomers to evoke trust. Minimising is also used by short duration groomers to make requests less direct while long duration groomers intensify requests. Trust becomes a currency to long duration groomers which can be earned or withdrawn. Long duration groomers use negative politeness strategies such as hindering their targets linguistically, threatening them and invoking guilt in request sequences. Both long and short duration groomers establish a student-teacher relationship to educate the target about sex. Short duration groomers use impoliteness, specific and open-ended threats, and inciting violence in specific requesting sequences. This is in contrast to long duration groomers who use a more subtle approach of positive politeness strategies and increasing the target's dependence on them.

Chapter seven is the discussion and conclusion chapter. It features an introduction section, takes stock of what the empirical chapters have done and then outlines the structure of the discussion and conclusion chapter. The first part of the chapter is divided into three parts corresponding to the three empirical chapters, first summarising the main research findings, and answering the research questions, then describing the significance of the findings and applications to real world contexts and then discussing the findings in relation to online grooming language literature. An updated terminology for the online grooming communicative processes proposed in the literature review chapter (section 2.5.3) in light of the findings will be presented and two duration-based groomer language approaches will be sketched out. The second part of the chapter will begin by revisiting the aims of the thesis set out in the first chapter and describing how the thesis has met these aims. It will then reflectively

summarise the thesis, outline limitations and further research avenues. A short conclusion paragraph summarising the chapter will also be provided.

1.5 Conclusions on Online Grooming as a Societal Issue

This chapter served as an introduction to the thesis. It first introduced the problem of online grooming by providing some worrying statistics about the effects and rise of online grooming cases from 2013 until 2020 and discussing duration of online grooming. The chapter then covered the existing literature on online grooming to answer the question what the academic community can do and has done about the problem this far. The literature shows that there has been a debate about whether the online grooming process is sequential or non-sequential. Most studies focusing on the language of online grooming have not been done from a linguistic perspective. One aspect that has not been studied systematically is duration of online grooming which this study will examine. The chapter has outlined the thesis structure and chapter contents, the next chapter will review the existing literature on online grooming, including proposed models on the online grooming process and will look at what the academic community has done to tackle the pressing issue of online grooming.

2. Chapter 2: The Literature on Online Grooming and Digital Discourse

Having set out the aims of this thesis along with the research questions and the significance of the research and providing an overview of online grooming as a complex social issue, this chapter will examine the key existing literature around online grooming research. The first section will review online grooming language or communication literature. This will be done by first taking different terminologies used to refer to the process and different definitions by scholars and unpacking them in pursuit of a definition of online grooming that will be used across the thesis in sub-section 2.1. The second section will look closer at the online grooming process as a whole by providing an overview of the main linguistic models of online grooming. This section is divided into two parts, the first reviewing sequential models, such as work by O'Connell (2003) the first linguistic model of online grooming and Black et al. (2015) who empirically test O'Connell's (2003) model in sub-section 2.2.1.1. The next subsection, 2.2.2.2 will review non-sequential models of online grooming by Williams et al. (2013), Lorenzo-Dus et al. (2016), Quayle and Newman (2017) and Chiang and Grant (2017). These models will be reviewed concerning the processes and strategies within them, which will be used to propose a new terminology to use in the thesis in sub-section 2.5.3. Section 2.2. will investigate literature on the duration of the online grooming process.

Section 2.3 will examine the participants of the online grooming process by first looking at online groomers in sub-section 2.3.1, which is split into two parts. The first part, 2.3.1.1, reviews groomer typologies, mainly derived from psychological approaches, while the second part, 2.3.1.2 outlines groomer language profiling, mostly from a Computational Linguistics point of view, which includes work by Pendar (2007), Bogdanova et al. (2014) and Pranoto et al. (2015). Decoys as participants of the online grooming process will be reviewed in sub-section 2.3.2 and children in sub-section 2.3.3. Digital Discourse research will be reviewed in section 2.4, with an overview of the three waves in sub-section 2.4.1, a closer look at the concept of participatory frameworks in the context of online grooming (sub-section 2.4.2) and aspects of online dating/flirting language in sub-section 2.4.3. The chapter will be summarised in section 2.5, which provides an overview of participant features in sub-section 2.5.1 and an overview of identified communicative intentions of online

groomers. These intentions will each be defined and reviewed in their own sub-sections, starting with 2.5.3.1 Seeking Sexual Gratification, 2.5.3.2 Developing Deceptive Trust, Arranging Further Contact, 2.5.3.3., Assessing Risks and Isolating, 2.5.3.4 and finishing with sub-section 2.5.3.5 Negotiating Power Dynamics. A short conclusion paragraph follows in sub-section 2.5.4.

2.1 Towards a Definition of Online Grooming

Online grooming is difficult to define. This section will look at existing definitions from both “offline” and online contexts to establish a definition and define the terms that will be used throughout this thesis.

Grooming in “offline” contexts has been defined as a preparatory seductive stage prior to sexual abuse that some offenders engage in (Bennett & O'Donohue, 2014). There is some debate about what this concept entails. Some scholars note the seductive aspect of it (Howitt, 1995; Bennett & O'Donohue, 2014) with Salter (1995) stating that ‘The grooming process itself often seems similar from offender to offender, largely because it takes little to discover that emotional seduction is the most effective way to manipulate children’ (Salter, 1995: 74). This statement shows tentative assumptions that the grooming process may follow certain patterns that are universal across groomers. The seductiveness mentioned by these scholars tries to put into words the combination of affection and trust that is built in the grooming process and the sexual abuse that takes place, comparing it with adult relationships.

Other definitions concentrate on the deceptive aspect of grooming, which entails gaining the target’s compliance, isolating, and desensitising them, lowering their inhibitions, maintaining secrecy, and gaining the target’s trust and friendship, blurring boundaries and roles, and making them complicit in sexual interactions (SgROI, 1982; Leberg, 1997; Berson, 2003; Craven et al., 2006; Knoll, 2010; Brackenridge, 2001). Gillespie (2004) emphasises that grooming is a precursor to gaining access to a child, which is echoed by Craven et al. (2006).

There are different terms to refer to the process of online grooming, too. One of these terms is *sexual solicitation*, which is primarily used in the United States and another

one is *online grooming*, primarily used in the United Kingdom (Kloess et al., 2017a). It has also been referred to as sexual online exploitation of children/adolescents (Ly et al., 2016) and cyberpedophilia (Bogdanova et al., 2014). With this variation of terminology comes an equally important inconsistency in defining what the process of grooming entails. The term chosen that will be used in this dissertation is online grooming, as it seems to encompass the entire process and the complexity of the issue, rather than just summarising it as the outcome the groomer pursues (sexual online exploitation) or reducing the main purpose for online grooming to just one (sexual solicitation).

A first very tentative definition of online grooming by O’Connell (2003) is the following:

A subset of cybersexexploitation is grooming, which may or may not involve explicit conversations of a sexual nature, or indeed online enactment of fantasies but still falls under the umbrella of cybersexexploitation because the intention is to sexually abuse a child in the real world, but one of the points of contact occurs in cyberspace. (O’Connell, 2003: 6)

Surprisingly, the most frequently cited definition of sexual grooming in online grooming literature is that by Craven, Brown, and Gilchrist (2006):

a process by which a person prepares a child, significant adults and the environment for the abuse of this child. Specific goals include gaining access to the child, gaining the child’s compliance, and maintaining the child’s secrecy to avoid disclosure. (Craven et al., 2006: 297)

This definition, despite being based on research into “offline” grooming, has been chosen as a starting point by many scholars in trying to define the online grooming process (Black et al., 2015; Chiang & Grant, 2017; Drouin et al., 2017; Lorenzo-Dus et al., 2016; Quayle & Newman, 2017; van Gijn-Grosvenor & Lamb, 2016; Williams et al., 2013; Winters et al., 2017). Most of the grooming models that will be reviewed in the following sections use the term *online grooming* consistently to describe this process. Kloess et al. (2017a) point out inconsistencies in defining the term and what it involves and does not involve. They notably define the process as being completed once sexual abuse occurs, summarising it as follows:

Based on the definition by Craven et al. (2006), sexual grooming characterizes the process of preparation for the abuse of a child, and it is therefore suggested that interactions which move beyond this process more accurately constitute sexual exploitation and abuse rather than grooming. (Kloess et al., 2017a: 3)

Chiang and Grant (2017) appear to take a similar stance, as they point out that Craven et al.'s (2006) definition seems to describe all sexualised conversation between an adult and a child. (Chiang & Grant, 2017). They state that grooming is a 'preparatory act' (Chiang & Grant, 2017: 105) and that most grooming conversations include sexual acts, which should be separated from the grooming itself:

We do freely use the term "grooming" and we recognise that engagement in less severe sexualised activity can groom a child for more severe actions but where possible we restrict our use of the term grooming to its preparatory meaning.

However, many other scholars see both sexual grooming and exploitation online to be interlinked as will be discussed in reviewing different models of online grooming in the next section.

2.2 The Online Grooming Process

2.2.1 Online Grooming Models

2.2.1.1 Sequential Models of Online Grooming

Within the literature on online grooming communication, the focus is mostly on the online groomer, rather than the target. This might be because of data issues, such as groomer-child data, not groomer-decoy data, which will be discussed further in the methodology chapter (chapter three, sections 3.1.3.1 and 3.1.3.2). It is also widely believed that the online grooming process is groomer-led, further justifying this focus on groomer language. This section will review the online grooming process models that have been introduced, many of which are based on groomer behaviour. These models tend to differ in terms of number and type of processes or stages of online grooming, introducing several different terminologies. They also perceive the online grooming process differently, with some studies seeing it as a sequential process with distinct stages the groomers goes through, while others regard it as a non-sequential model with different strategies the groomer may use, with no order. For this reason, the following section will be structured according to this distinction: first the

sequential models will be reviewed and in sub-section 2.2.1.2, the non-sequential models will be reviewed.

The first study examining the language within the online grooming process emerged in 2003 (O'Connell, 2003), a few subsequent studies were published between 2003 and the early 2010s (e.g., Pendar, 2009; Egan et al., 2011; Gupta et al., 2012; Williams et al., 2013) and further studies were done in the second half of the 2010s (e.g., Black et al., 2015; Lorenzo-Dus et al., 2016; Chiang & Grant, 2017). Some studies that aim to explore and analyse the language/communication of online grooming are not grounded in linguistic theories or concepts.

The first study to employ Linguistics to analyse online grooming as a sequential process was O'Connell (2003) which sought to develop an online grooming typology by conducting over 50 hours of 'research in chat rooms conducted intermittently over five years' (O'Connell, 2003: 5). The author entered chat rooms pretending to be an eight, 10 and 12-year-old girl and waited for groomers to approach her. Her persona's backstory represented typical signs of social isolation (moving to a new city, parents that are constantly fighting, no friends). O'Connell used Sociolinguistic analytical techniques, which were not further specified or outlined, and proposed a five stage online grooming model that consists of: friendship forming stage, relationship forming, risk assessment, exclusivity, and sexual stage. These stages include fantasy re-enactment, child erotica/pornography creation/distribution, damage limitation and a 'hit and run' (O'Connell, 2003: 13) tactic. Importantly, the study also highlighted the significance of accessibility and opportunity.

O'Connell (2003) described different behaviours of target selection, proposing that some groomers send a generic description to the whole chat room, pretending to be a child similar in age to their potential targets. They then start a private chat with their selected targets. Other groomers lurk in chat rooms and do not send messages until they identified a target, whom they then contact privately. O'Connell (2003) stated that some groomers do not lie about their age or pretend to be teenagers. Although O'Connell's model is sequential, the author acknowledged that different behavioural patterns were closely linked to motivations and different groomers spent more time on some stages than others (O'Connell, 2003:8):

The order and number of stages will vary and these variations provide clues of the user with ill-intent's motivations. Furthermore, whereas some stages, for example the risk assessment stage, have specific and identifiable goals, the goals of other stages are psychological and relate closely to both the aims of the adult and his perceptions of, for example, how malleable a child is in terms of meeting his requirements.

The following are the five proposed stages in more detail: Friendship forming is an initial stage in which the groomer gets to know the child. Questions include whether the child has a picture, which the groomer will then demand to see, which can be used to identify the child in real life and confirm the age of the child to the groomer. This stage is followed by relationship forming, which according to O'Connell (2003) is an extension of the previous stage and will involve small talk, such as talking about life at home and school. The main goal of this stage is to become the child's close or best friend, although not all groomers engage in this stage (O'Connell, 2003). The risk assessment stage sees the groomer enquire about where the computer they are using is located and how many other people use it. This is done to assess the risk of getting caught by family members of the child. The exclusivity stage relies heavily on a feeling of mutuality. The groomer conveys to the child that they are understood and can talk about anything. Trust is also an important aspect of this stage. This leads to the next stage, the sexual stage, in which the groomer asks intimate questions, which is based on the intense trust they have previously established with the child. O'Connell (2003) identified different motivations, such as maintaining a relationship with the child, which intensified the relationship forming stage. The groomer wants to be regarded as a mentor or future lover. If the child expresses being uncomfortable, the groomer will express regret, which will lead to forgiveness and stronger ties between groomer and child. The sexual stage may range from gentle suggestions to more explicit descriptions of sex acts, such as oral sex. According to O'Connell (2003) 'this pattern of conversation is characteristic of an online relationship that may progress to a request for a face-to-face meeting and arguably most closely resembles the course of conduct the 'anti-grooming' legislation is designed to combat' (O'Connell, 2003: 10).

While being ground-breaking research, O'Connell's research relied on self-reporting and the data collection and data analysis procedures were not outlined in the paper,

calling the replicability of the study into question. Additionally, it is unclear how the article arrived at the proposed model and no data was provided alongside the research.

In 2012, Gupta et al. (2012) used LIWC (Linguistic Inquiry Word Count)⁴ to annotate 75 chat conversations of paedophiles taken from the PJ website, analysing six stages of online grooming proposed by O'Connell (2003): friendship forming, relationship forming, risk assessment, exclusivity, sexual and conclusion. The study found that relationship forming was the most dominant stage, limiting software detecting sexually explicit words to detect paedophiles, as these were not the most prominent feature of groomer communication. This is a significant finding, as it validates the study into the relationship building, or deceptive trust development phase of online grooming and confirms that this is the central part of the process.

Black, Wollis, Woodworth and Hancock (2015) viewed online grooming as sequential but did not necessarily agree with O'Connell's stages. The study pointed out that O'Connell's (2003) model was not empirically tested and conducted a study of 44 chat logs taken from the PJ archive. In the article, Black et al. (2015) sought to examine the similarities and differences between face-to-face grooming and online grooming using O'Connell's (2003) model of online grooming.

The 44 chat logs were examined using Linguistic Inquiry Word Count (LIWC), a computer tool that analyses frequencies of words of a given text and compares them to pre-determined dictionaries, resulting in percentages of use related to overall word count (Tausczik & Pennebaker, 2010). To do so, Black et al. (2015) divided the chat logs into five segments corresponding to O'Connell's five stages of online grooming: friendship forming stage, relationship forming stage, risk assessment stage, exclusivity stage and sexual stage (O'Connell, 2003). Next, to determine whether stage-relevant terms occurred in their corresponding stages, a mixed model statistical analysis was conducted. The stages were defined as the dependent variable with the 'grooming process stage' as the independent variable. A Content Analysis of the chat

⁴ LIWC (n.d.) is a computer software that reads a text and compares it to a dictionary created by the user to calculate how many words correspond to certain categories, such as emotions, parts of speech and social concerns. It is primarily based on lexis and syntax.

logs was carried out after initial results showed a mismatch between the five stages and words related to these stages.

The Content Analysis was used to determine the times at which strategies belonging to these five stages were introduced and used in the chat logs. The strategies were: flattery/compliments, sexuality in the context of relationships, parents' work schedule, online sting operations, inappropriateness of behaviour, dangers of Internet/Assurance of safety, trusting relationship and travel plans. Black et al. (2015) found that friendship and relationship terms were not used more frequently in their respective stages than in other stages. Risk assessment was used more frequently in stages one and two, than in stage three. Exclusivity was used more in stages three and four and sexual terms were not statistically significant. This stage was further divided into 'sexual terms and 'meeting terms', which led to the finding that stage five was not made up of more sexual terms than the other stages. Instead, stage three had more sexual terms than stage one. The Content Analysis revealed that all offenders used at least one of the eight strategies. The strategies relating to risk assessment were first used in the first stage by half of the groomers. An important claim Black et al. (2015) made is that the online grooming process is not linear, as previously proposed by O'Connell (2003). Risk assessment takes place in earlier stages than in face-to-face grooming, which might be because groomers can easily move on to another target if targets are too high risk. Black et al. (2015) concluded that even though online grooming is similar to face-to-face grooming in terms of strategies used, their occurrence and progression differs. Risk assessment and assessing potential for victimisation is evaluated much earlier, as is sexual content that is introduced.

One shortcoming of Black et al.'s (2015) research was that the statistical findings, although presented and discussed, were not clarified by means of examples taken from the data. Moreover, it was not apparent how much textual context was considered during the analysis, which is a self-criticism of LIWC (Tausczik & Pennebaker, 2010). LIWC takes a word and compares it to several pre-defined dictionaries, not considering the context, which could change the meaning of the word, such as whether *produce* is a noun or a verb. Examples of a coding framework for the Content Analysis that was carried out by Black et al. (2015) was also not provided, which makes the study difficult to replicate. However, a general summary of the strategies is outlined

in the methods section. Additionally, no error rate was reported in the article. The sample size of 44 chat logs seems comparatively small and might not be representative of the existing database on the PJ website or indeed online groomers' language, which makes the findings difficult to generalise. Despite these issues, Black et al.'s (2015) article made a significant contribution to online grooming communication research. It was one of the first studies to use linguistic methods to analyse grooming communication and one of the first to empirically test existing hypotheses. Black et al. (2015) claimed that it was also one of the first studies to analyse 'the psychological underpinnings and strategies that can be derived from the language used in the transcripts' (Black et al., 2015: 142).

In 2011, the third sequential model of online grooming was developed by Egan, Hoskinson and Shewan, which conducted a Content Analysis of the PJ website to analyse the language of online groomers, independent of proposed stages by O'Connell (2003). The study used 'relational content analysis' (Egan et al., 2011: 4) to analyse 20 chat logs and found eight 'recurrent themes' (Egan et al., 2011: 1), suggesting support for a sequential online grooming model. These themes were implicit/explicit content, online solicitation, which was divided into two sub-themes, initiation and transference, fixated discourse, use of colloquialism, conscience, which was subdivided into 'empathic and unempathic' (Egan et al., 2011: 13), acknowledging illegal/immoral behaviour, minimizing risk, and preparing to meet offline. The implicit/explicit content referred to conversations related to sex, either unbeknownst to the target or made explicit. Implicitly, this happened by e.g., the groomer giving compliments that were more and more sexual in nature or framing sex as a game. Egan et al. (2011) claimed that the explicit content 'clearly undermines any subtle grooming strategy and renders redundant any implicit technique.' (Egan et al., 2011: 7). No further comments on the possible interplay of implicit and explicit strategies to gain the target's trust were made.

The initiation phase of online solicitation was any conversation that led the target to perform a sexual act, while online. Transference referred to the process by which the groomer shifted the responsibility of performing a sex act onto the target. This might be disguised as asking the target for consent. The fixated discourse theme had not been picked up as its own theme by other studies reviewed below until 2014 when

Bodganova et al. (2014) further analysed it. This process referred to the groomer dismissing or ignoring the target to advance their own interests and agenda. Egan et al. (2011) found this theme was strongly related to the explicit content and online solicitation themes. The groomer insisted on continuing talking about e.g., a specific fantasy or instructions on how to masturbate, even if the target did not respond favourably. This was concluded to ‘perhaps substantiating the claim that offenders see the Internet grooming process as a convenient opportunity for immediate gratification, becoming a substitute to committing a contact offence.’ (Egan et al., 2011: 12). Use of colloquialism referred to groomers using idioms or netspeak to appear attractive to the target. This was not a thematic category, which points to a set of flawed codes.

The empathetic and unempathic themes ‘links closely’ (Egan et al., 2011: 10) to acknowledging that the target may get upset about the inappropriate conversations or suggestions. Egan et al. (2011) only found one instance of an empathic theme, while unempathic seemed to be more prevalent. This was related to the fixated discourse theme. The groomer may acknowledge the inappropriateness of the behaviour, which is the sixth theme, however, he usually continued his advances, especially if he was likely to gain sexual gratification. Notably, Egan et al. (2011) found that minimising risk was not very prevalent in their dataset. Groomers took risks, which might be facilitated by the internet’s anonymity. While groomers took on a different persona, they frequently stated their correct age. The last theme, preparing to meet offline, referred to the arranging of a meeting with the target offline to ‘finalise their grooming procedure’, as Egan et al. (2011: 11) put it.

Egan et al.’s, (2011) model does not seem very robust, and the article did not report on how it measured relatedness of different themes, which makes the model non-replicable. The analysis was thematic with no quantitative analysis to back up claims and the qualitative approach was flawed with a non-consistent coding set, which includes non-thematic codes. Furthermore, the study analysed 20 chat logs and made generalisations about the online grooming process based on Thematic Analysis, which did not feature micro level analyses of language, but was rather loosely based on analysing language without a robust framework or methodology.

Another sequential model, in the form of developing an offense process diagram of online sexual grooming and abuse, was proposed by Kloess, Hamilton-Giachritsis and Beech (2017a), who examined five online grooming-child cases by qualitatively analysing 29 chat logs and police reports collected from three UK police forces ‘to provide an inclusive and realistic’ (Kloess et al., 2017a: 1) overview of the online grooming process. The time frame covered the end of 2009 until the beginning of 2012. The offenders were between 27 and 52 years old with targets between 11 and 15 years old. The offenders had different demographic characteristics, but notably two offenders worked closely with children. Kloess et al. (2017a) used Thematic Analysis to inductively let themes and patterns emerge. Stages of online grooming, which described the ‘development and progression of offending behavior in a temporal sequence from initial contact to end of contact’ (Kloess et al., 2017a: 6) were initial approach, grooming, cycle of online sexual activity, closure, and degree of contact (Kloess et al., 2017a: 6). These stages had different strategies, for example sexual activity was mostly done by the offender exposing themselves to the child and masturbation. Furthermore, the offender encouraged the child to engage in several sexual activities, ranging from disrobing to performing sexually humiliating acts (Kloess et al., 2017a).

Offense process diagrams of each case were created and turned into a general diagram, serving as their proposed model. The initial approach of the children differed across the small sample Kloess et al. (2017a) analysed. The data of first contact was not available in three cases, as the offender and child met and interacted on public or private channels for which chat logs were not collected (e.g., public chat room, adult dating website, a video chat website). The remaining two offenders contacted the children using social networking sites. Kloess et al. (2017a) found that there were two approaches, one was gentle while the other one was more direct and ‘highly sexualized’ (Kloess et al., 2017a: 9).

As mentioned above, Kloess et al. (2017a) took Craven et al.’s (2006) definition very literally and regarded the grooming process to end once any sexual activity (also cybersex) was involved. As such, the article found that only two of the five offenders engaged in online grooming, based on this definition. Kloess et al. (2017a) claimed that the other two offenders utilising a more direct approach, did not make use of any

grooming strategies to prepare the target. The two offenders who, according to the study, did use grooming strategies made use of a variety of strategies. These were gaining a child's compliance, overcoming a child's resistance, and maintaining secrecy through giving compliments and flattery, queries about the targets' relationship status and sexual hints. Another grooming strategy identified by the study was introducing a meeting. According to Kloess et al. (2017a), in most cases this strategy was introduced after a relationship had been built and the offender knew the child would be compliant. However, one offender asked for a meeting almost immediately. Kloess et al. (2017a) also suggested that once sexual activities had been introduced, the conversation was more likely to include sexual talk, which was apparent in one case.

The last aspect of the grooming process Kloess et al. (2017a) focused on was closure; The way the interaction ended. Kloess et al. (2017a) found various endings, such as attempts to reinitiate contact, which were ignored, natural conversation endings due to e.g., the time of day, abrupt endings and one ending due to police involvement. Although the article is informative, the small sample size makes the findings difficult to generalise. The study examined 29 chat logs from five offenders, which is relatively small. However, the study managed to collect real world data, which is quite rare in online grooming research. The definition of the scope of online grooming is somewhat problematic, as it categorises sexual behaviours as abuse, without looking at them in detail to understand how they are part of the escalating process of online grooming.

As this review of sequential model has shown, there is little evidence that the online grooming process occurs sequentially. The studies regarding it as such suffer from important caveats: O'Connell's (2003) model is based on self-reports and cannot be replicated or generalised. Black et al. (2015) and Gupta et al. (2012) used LIWC, a software which is not validated to be used for online grooming language (Broome et al., 2018). Black et al. (2015) relied on LIWC's output, which is quite non-transparent and additionally investigate the online grooming process thematically. Egan et al. (2011) and Kloess et al. (2017a) also used Thematic Analysis, which did not show much evidence of sequential grooming. Moreover, these studies relied on a small sample size, ranging from 20 to 44 chat logs. The next section will examine the alternative: non-sequential models to describe the online grooming process.

2.2.1.2 Non-Sequential Models of Online Grooming

In this section several studies that see the process of online grooming as non-sequential are reviewed. These are: Williams et al. (2013), Lorenzo-Dus et al. (2016), van Gijn-Grosvenor and Lamb (2016), Quayle and Newman (2017) and Chiang and Grant (2017).

Williams et al. (2013) who analysed the initial hour of eight PJ transcripts supported a move away from a sequential model of online grooming. The article noted that while its methodology did not entail examining theme sequencing, the identified themes did not appear in any particular order and themes overlapped and co-occurred. Williams et al. (2013) used Thematic Analysis to analyse themes related to online grooming and identified three main ones: rapport-building, sexual content, and assessment. These themes are similar to O'Connell's (2003) friendship and relationship building phases, the sexual phase and risk assessment phase. These three themes were further arranged into sub-themes. Williams et al. (2013) found that the groomers attempted to make the target feel comfortable and created a 'special bond' (Williams et al., 2013: 140) with them. This was done through coordination, mutuality, negative and positive rapport-building. Positive rapport building consisted of the groomer finding common interests, adopting the same colloquialisms, and presenting themselves in a positive light, while negative rapport-building was made up of impatience and passive aggressiveness. This occurred in the form of repeatedly asking questions. Williams et al. (2013) decided to focus on the positive rapport, which consisted of the following sub-strategies: Coordination, which was related to the groomer's behaviour and how they tried to synchronise behaviours with their target. This was originally labelled 'roles'. This behaviour ranged from pretending to be closer to the child in age, adopting netspeak and treating the child as an adult to raise them to the groomer's level. Mutuality originally labelled 'common interests' (Williams et al., 2013: 140) was concerned with gathering information about the child's hobbies, interests, attitudes and circumstances. Groomers then aligned their own interests and attitudes with that of the target, tried to find common ground or similarities, which then extended to giving advice and guidance. Positive related to how the groomer was trying to get the target to see them.

The groomer attempted to portray themselves in ‘an unrealistically positive light’ (Williams et al., 2013: 142). Self-attributes included being trustworthy, friendly, harmless, polite, responsible, sensible, caring and generally having good intentions. This prevented the target from sensing danger and lulled them into a false sense of security. This ‘positive environment’ (Williams et al., 2013: 143) was turned negative by the groomer using passive-aggressiveness and impatience to get what they wanted. This was done by repeating a question if the child did not answer and leaving conversations, which is reminiscent of the fixated discourse theme identified by Egan et al. (2011).

The second theme, sexual content, was introduced and maintained or escalated. This was done in different ways. The groomer introduced it as a game, as advice or as a mutual fantasy based on common interests. Maintenance and escalation were done by repetition or force. The third theme, assessment, concerned the target and their environment. The child’s trust, vulnerability and receptiveness were assessed to see whether the target was receptive, desensitised and whether there was a risk of detection. The environment of the child was also assessed, specifically potential obstacles that could ‘interfere with the grooming process’ (Williams et al., 2013: 148) were evaluated, such as, the target’s physical location, home situation, the location of the computer in the house and their parents’ schedules. Additionally, most of O’Connell’s stages were present in the first hour of chat logs that were analysed. Williams et al. (2013) thus concluded that ‘(...) the grooming process may not follow a sequential order; the order in which each theme becomes prevalent will depend instead on the offender as an individual.’ (Williams et al., 2013: 150). Further, it stated that a bigger sample size might increase the validity of results, which is an important point, given the range of sample sizes within online grooming communication research mentioned above.

Lorenzo-Dus et al.’s (2016) online grooming model was based on the comprehensive offline grooming model by Olson, Ellevold and Rogers (2007) who developed a theory of ‘luring communication’ (Olson et al., 2007: 234). Most notably this theory features a cycle of entrapment at the core of luring, with deceptive trust development as the main component. Lorenzo-Dus et al. (2016) modified this theory based on a review of existing online grooming literature and a Computer Mediated Discourse Analysis

(CMDA) of 24 chat logs taken from the Perverted Justice (PJ) website. The procedure involved a language-focused Content Analysis to develop a set of coding categories, focussing on speech acts and relational-work. This initial taxonomy was then tested using a fraction of the data and revised as necessary. Figure 1 shows Lorenzo-Dus et al.'s (2016) Online Grooming Discourse Model.

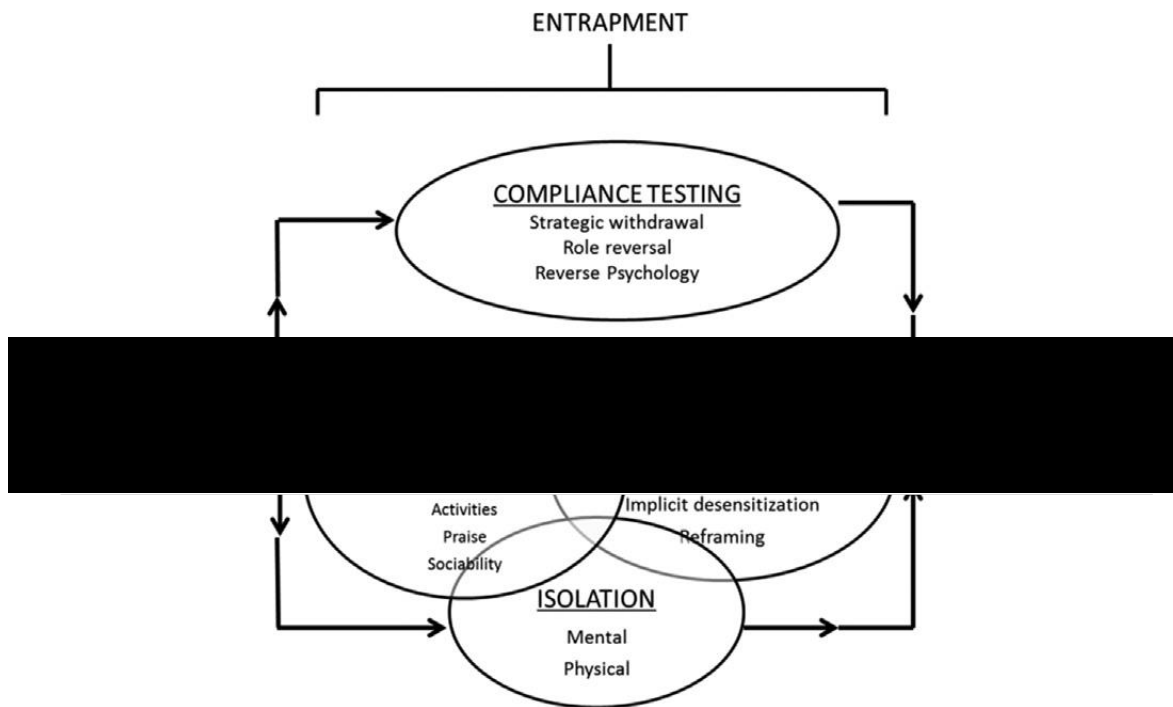


Figure 1: Online grooming model introduced by Lorenzo-Dus et al (2016: 70)

As can be seen, the model is made up of three phases: access, entrapment, and approach. The entrapment phase is further divided into four main components, which are deceptive trust development, compliance testing, sexual gratification, and isolation. These components each have their own set of online grooming strategies. The access phase, in which the groomer initially contacted the minor and the approach phase, in which the groomer arranged a face-to-face meeting with the minor, did not appear in O'Connell's (2003) grooming model. They are new additions, a variation of which can also be found in Olson et al.'s (2007) theory of luring communication. According to Lorenzo-Dus et al. (2016) the entrapment phase, which is at the core of their online grooming communication model, is the most complex of the three phases. The aim of this phase is to 'lure targets into different forms of sexual behaviour, including soliciting and/or sharing indecent images of children and/or adults.' (Lorenzo-Dus et al., 2016: 44). Lorenzo-Dus et al. (2016) reported two new strategies

within deceptive trust development, praising and sociability, that had not been reported by research into “offline” grooming. Exchange of personal information, activities and relationships are part of previous online grooming models. Compliance testing is also a newly described strategy with three sub-strategies: strategic withdrawal, role reversal and reverse psychology. Lorenzo-Dus et al. (2016) conducted statistical tests to find out strategy frequency and inter-relations. The article found that deceptive trust development was more frequent than the other strategies and that sexual gratification was more frequent than compliance testing, isolation, and approach. Compliance testing correlated with sexual gratification and isolation and deceptive trust development correlated with the other three processes. This suggested that the more the compliance was tested, the more sexual gratification and isolation strategies were used. It also implied that establishing trust encouraged groomers to make use of the other strategies. This led Lorenzo-Dus et al. (2016) to conclude that deceptive trust development ‘may be paramount to OG⁵, over and above any other process.’ (Lorenzo-Dus et al., 2016: 45).

Unlike O’Connell’s (2003) model, Lorenzo-Dus et al.’s (2016) combined qualitative close examination of the text with statistical measures. The article also reported on its coding framework and provided plenty of examples to illustrate the complex nature of online grooming strategies that are used. Lorenzo-Dus et al. (2016) also stressed that online grooming should not be regarded as a cyclical or sequential process, but rather an interdependent network of entrapment, in which many of the strategies are more interdependent as previously believed (Lorenzo-Dus et al., 2016). One of the conclusions, based on the analysis of face-work in the data, suggested that researchers should move away from categorising online groomers as suffering from psychiatric and personality disorders and substance abuse that make them unable to use social skills:

Whatever disorders a number of them may suffer from, they do not seem to be conditions that significantly affect their sociopragmatic competence. It is therefore important, especially for detection purposes, that we understand the “accomplished” nature of OG discourse. Adding speech act realisation and relational work analyses to the lexical analysis tools that currently inform online grooming prevention software may lead to improved detection levels. (Lorenzo-Dus et al., 2016: 48)

⁵ Online Grooming

The study called for more research adopting an approach that combines qualitative analyses, such as Speech Act Theory and relational work with lexical analyses.

Van Gijn-Grosvenor and Lamb (2016) analysed 101 chat transcripts taken from the PJ archive to find out how online groomers communicated with male (49 chat logs) and female (52 chat logs) decoys. The study supported a non-sequential approach to online grooming and proposed categories that could be summarised in an online grooming model. They focused on contact-driven offenders, a distinction introduced by Briggs et al. (2011), which stated that offenders could be classified as either contact-driven, which entailed receiving sexual gratification from arranging an offline meeting with the target, or fantasy-driven, meaning only interacting with the target online was enough to gain sexual gratification (Briggs et al., 2011). Van Gijn-Grosvenor and Lamb (2016) believed 'contact-driven' groomers to be a greater danger to children. The study used a mixed method approach and coded the chat logs according to the following categories: offense characteristics, rapport building, sexual matters, ways to conceal contact.

Van Gijn-Grosvenor and Lamb (2016) found that the offenders age ranged from 19 to 69 (average 38.11 years), offenders grooming female targets were significantly younger than offenders grooming male targets. Female targets were asked to describe their physical appearance, whereas male targets were asked to describe their genitals, offenders targeting males were more explicit and more focused on sexual gratification while offenders targeting females wanted to make them feel special by telling them they loved them and giving compliments.

The study also found that the online contact was longer between groomers and female targets, than between groomers and male targets. Van Gijn-Grosvenor and Lamb (2016) only used statistics to describe the online grooming process, not validating findings by examining the text of the chat logs. The study did not specifically attempt to find out whether the process is sequential. However, judging from the lack of segmenting the chat logs and using the target gender as the dependent variable, it viewed the online grooming process as non-sequential. Although its data size was bigger than that of other work, the findings were not very informative or surprising.

Gijn-Grosvenor and Lamb's (2016) rationale that contact-driven groomers posed more risk than fantasy-driven groomers is a dangerous one to make, especially taking into consideration that there is no evidence that supports this distinction (Broome et al., 2018), which will be addressed in more detail in section 2.3.1.1.

Quayle and Newman (2017) conducted a study analysing reports by children that have been groomed online submitted to Cybertip.ca to find out more about the online grooming process, groomers, and children/young people. Like Lorenzo-Dus et al. (2016) and Williams et al. (2013), this study did not support a sequential process model of online grooming and found evidence of some of the strategies proposed by Black et al. (2015). It concluded that Briggs et al. (2011) were correct in suggesting a division into fantasy-driven and contact-driven online groomers who behave differently.

Quayle and Newman (2017) examined reports made on Cybertip.ca, a Canadian website that receives reports from the general public about potential online sexual abuse in the form of child pornography, online luring, child exploitation through prostitution, child trafficking, among others. Quayle and Newman (2017) wanted to investigate the behaviours that led the members of the public to make a report, information about the child/young person and the suspect and how they interact. The anonymised data, which comprised 264 reports that were suspected to involve online grooming between 2007 and 2011, was analysed using SPSS, counting frequencies and cross-tabulations of their 'numerical data' (Quayle & Newman, 2017: 4). Some reports included additional content, which was analysed using a Content Analysis identifying themes and a coding framework consisting of 13 items. The results concerned demographic information and Content Analysis coding. In terms of demographics, most reports (86%) involved a female child/young person and a male suspect (90%), and the children and young people were between nine and 17 years old. The mean age of female child/young person was more concentrated around the age of 13, whereas the age of male young people was more evenly spread. The suspect age in relation to gender revealed that suspects grooming females were slightly older (mean age 26) than suspects grooming male young people (mean age 24).

The additional text was provided in the forms of transcripts of Instant Messengers and other means of online communication (games, social networking sites, online voice and video calls, online dating sites, email software etc.), this was available in 60% of the reports. The following themes emerged from the Content Analysis: request for sexual images, voyeurism, exhibitionism, contact request, resistance, threat, self-taken, cell phone, vigilantism, peer sex, deception, money, and vulnerability. Quayle and Newman (2017) found similarities with previous research, supporting Briggs et al.'s (2011) division into contact-driven and fantasy-driven offenders.

Additionally, the article did not find evidence of a 'clear process model' (Quayle & Newman, 2017: 10), as mentioned above. Furthermore, it did not find evidence for some of the strategies identified in previous online grooming research, such as enquiring about the parents' schedules, deception about the offender's age, being targeted based on sexual identity. Quayle and Newman (2017) stated more research into male targets of online grooming was needed, as this group seemed to be largely ignored, possibly due to data constraints. This article did not suffer from the caveat of authenticity, which will be discussed further in chapter three (Section 3.1.1), as it was one of the few studies using data other than the PJ archive. However, the reliability of this data cannot be guaranteed. Especially because not all suspects were convicted as a result of the reports and false reports cannot be detected easily. The same suspects could have been involved in more than one report as well.

Chiang and Grant (2017) took a different approach to online grooming, by investigating it using Swales (1981) move analysis, which is typically used to describe structures of a given genre. The model is a non-sequential model, as they found evidence of 'the interaction being multi-tracked with moves being developed together rather than sequentially one after another' (Chiang & Grant, 2017: 117). The study analysed seven online grooming transcripts collected from the PJ archive to find out whether this framework is useful in contributing to understand the grooming process better. The transcripts were selected based on three criteria. One: Some preparatory grooming features were present, and the groomer did not immediately begin sexual communication. Two: All chat logs took place between 2006 and 2009. Three: They were full interactions, i.e., most of the interaction was recorded in the form of chat

logs and those interactions that consisted of unrecorded interactions (such as emails, texts, phone calls) were excluded.

Chiang and Grant (2017) were interested in the pragmatic uses of Swales' theory, wanting to find out what the conversational goals of online groomers were and analysing whether online grooming moves could be identified as being sequential. Moves convey the communicative function of parts of texts, which in turn inform the communicative functions of a whole genre (Chiang & Grant, 2017). There is some freedom and variety in moves and strategies authors can implement, which is why Chiang and Grant (2017) accounted for more than one linguistic strategy for every move. The analysis was more complex than expected and the authors had to code every groomer utterance for its primary rhetorical purpose and then grouped them with other moves into functional and semantic themes. A pilot study was conducted colour coding each move and then all moves were colour coded. The study observed additional themes to the ones previously identified by scholars (O'Connell, 2003; Williams et al., 2013; Black et al., 2015). The move *assessing and managing risk* was selected for a closer linguistic analysis.

Chiang and Grant (2017) identified 14 rhetorical moves and 87 strategies. Greetings were words or expressions initiating a conversation. Building and maintaining a friendship with the target was done through building rapport, which was carried out by giving compliments, asking about the target's age, sex and location and current activities, using, and eliciting statements of trust and reassurance and sympathy, positive evaluations, requesting email addresses, photographs, and webcam calls. Sexual rapport, which was the establishment and maintenance of a positive sexual connection was realised through sexual compliments, describing sex as enjoyable and beneficial and sexual guidance, giving the target topic control, lessening anxieties, retracting sexual comments and questions, trying to lessen the severity and intensity of sexual content. Maintaining current interaction was the attempt to continue the conversation and was done by backchanneling, checking whether the target was present in the conversation and explaining technical difficulties.

Assessing likelihood and extent of engagement was the willingness of the target to engage in sexual activities and offline meetings, which the groomer gauged. This

move was realised by the following strategies: eliciting responses to hypothetical sexual scenarios, probing ideas for sexual activities, asking about previous sexual partners and activities, among others. *Assessing accessibility* was the groomers' efforts to find out where the target was physically located and assessing other obstacles, such as the parents' schedule, the target's immediate surroundings (family and friends) and relationship status. *Assessing and managing risk* concerned managing non-detection and risks associated with detection. The groomer asked about the target's home situation, talked about the age gap between them and the inappropriateness of their relationship and elicited private or secret statements from the target. Assessing personal criteria fulfilment referred to the groomer trying to establish whether the target fulfilled their personal and physical preferences. Strategies of this move included asking about age, virginity and sexual orientation and requesting photos, videos and phone calls and general descriptions of the target.

Assessing own role was defined as assessing the target's compliance as well as what was needed to get full compliance. Strategies included questions about sexual experience and self-disclosure about the groomer's sexual experience levels. *Introducing sexual content* was any dialogue that introduces sexual topics. *Immediate sexual gratification* referred to attempts to achieve immediate sexual arousal, which was done by sharing or requesting sexual photos, videos or phone and video calls, as well as giving and eliciting sexual descriptions and hypothetical scenarios.

Maintaining and escalating sexual content were attempts to make sure sexual content continued to be part of the conversation, which was done by introducing fantasy planning, expressing sexual desires, and normalising behaviour. *Planning and arranging contact* was defined as making plans to physically meet, this included offering and eliciting details, such as time and location and contact information. *Sign off* was the messages groomers sent before they left the conversation. Chiang and Grant (2017) found that only two strategies, *Planning and arranging contact* and *sign off* were not present in all chat logs, and considered them optional. All other strategies were present in the seven transcripts and suggested some consistency, making up the core of the model. Chiang and Grant (2017) identified *building rapport* as the most frequently used move in the data, followed by *maintaining and escalating sexual content*. This was surprising to them, as they attempted to exclude chat logs that

contained sexual acts and communication. Chiang and Grant (2017) concluded that even in these interactions, sexual abuse was frequently occurring. Another tentative finding was that online grooming did not follow the conventions of a genre:

Finally, it should be noted that while this move analysis identified several shared communicative goals, the range of strategies used to achieve them is fairly broad. This suggests that chatroom groomers exercise a considerable amount of linguistic freedom in pursuing their aims, and that chatroom grooming interactions are not bound by rigid conventions as traditional genre types are. (Chiang & Grant, 2017: 115)

However, Chiang and Grant (2017) emphasised that this is true for most internet genres. The variety of strategies and moves might indicate that groomers tend to adopt individual grooming styles (Chiang & Grant, 2017). Chiang and Grant (2017) concluded that move analysis could help but was not a perfect framework. The study also raised the question whether online grooming was a genre and whether groomers were a discourse community. This paper took a novel approach to online grooming analysis by investigating it using Swales (1981) move analysis and attempting to establish whether this is useful in future research. The small sample size is a shortcoming of this paper and makes it difficult to generalise findings. However, with a qualitative analysis which involves colour coding applying it to a bigger sample size might not be realistic. Some strategies were similar to previously proposed components and strategies by Lorenzo-Dus et al. (2016) and O'Connell (2003).

The non-sequential studies reviewed above showed a more complex picture of online grooming, as they did not limit themselves to O'Connell's originally introduced five stages and instead attempted to describe the process inductively. This gave a much more nuanced view, such as the introduction of compliance testing by Lorenzo-Dus et al.'s (2016). These models used different methodologies, with the focus being on either using a mixed methods approach incorporating quantitative and qualitative analyses or using a Thematic Analysis. This PhD will regard the online grooming as non-sequential, as there is little evidence that it is sequential and more evidence supporting the claim that it is not. While these models provide a foundation of online grooming communication, they still only provide a language-based Content Analysis of online grooming lacking the micro-level analysis and detail a Linguistics approach can add, which this thesis does.

2.2.2 The Duration of the Online Grooming Process

This section will look at the duration of the online grooming process, which has not yet been established. Some studies gave a general indication of the fact that some groomers spent more time grooming children than others (O'Connell, 2003). The degree of contact, as Kloess et al. (2017a: 16) called it, has not been the focus of much research. Kloess et al. (2017a) reported on the average duration of the main conversations between a groomer and one or more targets and the overall time frame they covered. They noted that most interactions were relatively short, lasting between one day and six months, while conversations lasted between 10 minutes and five hours. The average of these 'main' (p.16) conversations was two hours and 25 minutes; However, this was not reported in their paper. It should also be noted that the study looked at a very small sample size of 29 transcripts taken from five groomer-child cases, this average is therefore not reliable for generalisations.

Lorenzo-Dus and Izura (2017) divided their data into three grooming speed categories (fast, average, and slow) depending on the duration of the interaction and the number of times the groomer logged into the chat room. These speeds corresponded to the following groups: fast (under four hours), average (between five and 11 hours) and slow (over 11 hours). Lorenzo-Dus and Izura (2017) isolated all the compliments from 68 chat logs, taken from Perverted-Justice.com, to find out how online groomers used the praise strategy within the Deceptive Trust Development process of the OGD. More specifically, they sought to answer whether compliment topics differed across groomers and what syntactic structures were used most and least often. The findings showed that slow groomers used more compliments than fast groomers with a prevalence of compliments concerning the physical appearance of the targets (Lorenzo-Dus & Izura, 2017). Groomers' compliments were also found to support other online grooming processes, such as isolation and sexual gratification in the form of desensitization. Importantly, Lorenzo Dus and Izura (2017) found a link between groomer speed and the number and type of compliments they used, indicating that the language of groomers who spent more time grooming differed from the language of groomers who spent less time. This has not been analysed further, showing an

important gap in the literature, which might show different vocabulary use and strategy use or point to different models for different groomers. This study also highlights the importance of a Discourse Analysis-based approach to online grooming communication research that can further develop the above outlined framework.

Some other references concerning the time span of grooming interactions were made by Briggs et al. (2011), who computed that most groomers communicated with their targets for less than a week, with about 40% waiting less than a day (24 hours) to arrange further offline contact (Briggs et al., 2011). Fantasy-driven groomers were reported to have interacted with their targets for an average of 32.9 days. This is in direct contrast with Wolak et al.'s (2014) reported time span of more than a month of interaction between groomer and target. Other studies concentrated on which stages or strategies were introduced at which point in the grooming interaction. Black et al. (2015) found that groomers introduced sexual topics 'within the first 20% of contact time' (Black et al., 2015: 146). This was confirmed by Winters et al. (2017), which claimed that it only took an average of 33 minutes for the groomer to present sexual content. It should be noted, however, that Black et al. (2015) and Winters et al. (2017) did not work with the entirety of the PJ archive, but rather with 44 and 100 chat logs, respectively. Other studies were vaguer about the duration of grooming and did not include specific times. For instance, Kloess et al. (2017a) said that sexual topics were introduced 'relatively promptly' (Kloess et al., 2017a: 10), but failed to define what this meant explicitly.

2.3 Participants in the Online Grooming Process

2.3.1 Online Groomers

Another growing body of research on online grooming has been focused on identifying and profiling online groomers. This was divided into two paths. First: Groomer internal states and their intention to move across online-offline realm in online grooming, mainly developed in Psychology. Second: Groomers' language compared to other people (decoys, children, teenagers, adults in sexual conversation), which has been developing in the areas of Computational Linguistics, Machine Learning and Natural Language Processing parallel to the linguistic studies outlined above. These

studies focused on the automatic detection of online grooming, while still considering different areas of language, such as individual words, word classes or combinations of words.

2.3.1.1 Groomer Typologies – Psychological Approaches

Developing online groomer typologies has been an area of academic interest, especially in Psychology, to better understand the groomers and their motivations. These typologies are often based on typologies of sex offenders in general. This includes the broad distinction between situational offenders and preferential offenders, made by Lanning (2010). The use of child pornography has also been noted as a behaviour of some online offenders (Krone 2004; Elliott & Beech, 2009). A distinction can be made between offenders who use pornography and others who do not. Another branch of research coming out of Psychology investigated mental processes of those offenders who use child pornography, also called CPOs by Merdian et al. (2013). One of these studies was done by Kettleborough and Merdian (2017) and examined permission-giving thought patterns. The study used Thematic Analysis and found four overarching themes: perceived nature of children, non-sexual engagement with child sexual exploitation material (CSEM), denial of harm and expression of general sexual preference.

Another distinction was made by Briggs et al. (2011) who analysed clinical and behavioural data from 51 convicted online grooming offenders and concluded that there are two distinct groups of online-groomers: fantasy-driven offenders and contact-driven offenders. Notably, none of the participants were diagnosed with paedophilia. Fantasy-driven offenders engaged in cybersex and more sexual behaviour than contact-driven offenders, who sought sexual activities to take place in offline meetings. Fantasy-driven offenders were more often diagnosed with paraphilia and narcissistic personality disorder. Briggs et al. (2011) concluded that offenders have different motivations for grooming teenagers online, but more research was needed. This distinction is widely accepted, and some evidence has been found that different motivation is connected to different behaviour.

DeHart et al. (2017) analysed 200 offender chat logs, email communication and social media posts to generate an offender typology. The study used a mixed methods approach to both qualitatively and quantitatively analyse and code the data. Cases were classified into three groups: those engaging in real-time masturbation and who did not attempt to schedule (cybersex), those who attempted to schedule but did not engage in cybersex (schedulers) and those who did both (cybersex/schedulers). A fourth group emerged from further analysis: buyers. Those who chatted to a third party about sex trafficking. DeHart et al. (2017) found that interactions lasted between 10 minutes and four years. The article supported Briggs (2011) distinction of fantasy-driven and contact-driven offenders.

Merdian et al. (2013) proposed three dimensions of online child pornography offending, which they developed by reviewing the existing literature and comparing models with each other. The three dimensions were: fantasy-driven vs. contact-driven offending, motivation, and the social component. Merdian et al. (2013) stressed the importance of the distinction made by Briggs et al. (2011) for child pornography offending. The article suggested that most CPOs are fantasy-driven with only a quarter of them escalating to contact-driven offending, as reported in the literature (Merdian et al., 2013). The model allowed for offenders to move to other subcategories over time. Merdian et al. (2018) attempted to empirically test their previous distinction between fantasy-driven and contact-driven child sexual exploitation material (CSEM) users (Merdian et al., 2013) by analysing self-reported survey answers of 68 male sex offenders and prison inmates. Because of the size of the dataset, they used cluster analysis and principal component analysis to find out whether there are any observable differences in motivation of fantasy-driven and contact-driven CSEM users. Four item groupings were found to be different between offender types: social exclusion and escape, justification, children as sexual agents and power and entitlement. Of these items, only the first two were significant in predicting offender subtype. This was a first attempt to empirically test whether contact-driven, and fantasy-driven offenders can be classified based on their responses and behaviours.

Some offender typologies and models relied on self-reporting, which might not be the most reliable method. However, the literature suggests that a clear distinction can be made between fantasy and contact-driven online offenders. Broome et al. (2018)

conducted a thematic literature review of contact- and fantasy-driven groomers. They reviewed 22 studies that looked at the proposed distinction of fantasy- and contact-driven offenders and found that both groups engage in both activities, making the distinction vague. Broome et al. (2018) proposed adopting the European Online Grooming Project's typology (Webster et al., 2012), which focused on the intensity of the interaction. They classified offenders into intimacy seeking, adaptable and hypersexualised groups (Webster et al., 2012).

2.3.1.2 Groomer Language Profiling

This section will look at research that has compared groomer language to that of decoys, children, or teenager language to for example develop Machine Learning algorithms that are able to detect online grooming, or to analyse the differences in language.

Guice (2016) sought to explore the language of online groomers, adult decoys and teens and compiled three corpora, one consisting of predator language isolated from 60 chat logs taken from the PJ archive, another one consisting of the decoy language of these chat logs and a third corpus containing data from teen chat rooms. The three corpora sizes were quite different with the predator corpus being the largest (302,235 words) and the teen corpus being the smallest (20,085). The decoy corpus contained 215,598 words. The study had the following hypotheses. One: The predator corpus will contain the most sexual terms. Two: The decoy corpus will show the most frequent use of textisms. Three: The teen corpus will use the most emoticons. These hypotheses were confirmed by Guice's (2016) analysis. Limitations of the research are a gap in time between the predator and decoy corpora and the teen corpora, the latter was collected in 2016, while the others were randomly collected from the PJ archive. The analysis suffered from these time inconsistencies in the two datasets. In fact, Guice (2016) acknowledged that most of the chat logs were from the early to mid 2000s. This thesis did show evidence that a corpus-based approach is an acceptable methodology to use for analysing chat log data of online grooming as Corpus Linguistics lends itself well to this kind of comparative research.

Pendar (2007) attempted to build an automatic recognition system of online groomers. The study reported on a pilot study, with two datasets, which were both taken from the PJ archive, to test modules that might be able to detect the difference between decoy and groomer language. Pendar (2007) was working on the assumption that a decoy child would use very different language than an online groomer, even though the chat lines were broadly on the same topic. The article reportedly collected 701 chat logs from the perverted justice website, which were defined as “conversations”, but did not correspond to different groomers, as there are currently 622 chat logs available in the archive, which shut down operation in 2019. Pendar (2007) concluded that it is possible to automatically distinguish groomer language from decoy language, with ‘very high accuracy’ (Pendar, 2007: 240), without specifying numbers. This paved the way for detecting online grooming amongst other types of conversations, such as between two consenting adults or decoy and children.

Kontostathis et al. (2009) developed a prototype software ChatCoder, which had a similar goal to Pendar’s (2007) research. The developers sought to distinguish between groomer and decoy and groomer and normal chat. Kontostathis et al. (2009) used 288 chat logs from the PJ archive as well as general chat room data for comparisons. Kontostathis et al. (2009) reported studying 20 PJ chat logs carefully to develop a codebook and dictionary. The study carried out an experiment to test if the software could correctly identify whether a conversation was grooming or not, reaching a 60% accuracy. The second experiment used the general chat data and PJ data, with the software being able to correctly identify 93% of the data. They also used a clustering technique to identify which phrases were used by predators and concluded that there were four distinct groups using different language patterns. These patterns are not disclosed; However, it is an interesting finding, which shows that further research into the language patterns of online groomers is needed. Kontostathis et al. (2009) pointed out that some groomers spend more time grooming their targets and might use different strategies to other groomers, which has implications for the algorithm.

Bogdanova, Rosso and Solorio (2014) analysed 60 groomer-decoy chat logs, also taken from the PJ archive, a general chat dataset (NPS) and a dataset of non-

paedophiles⁶ using explicitly sexual language (cybersex logs) to find out how to distinguish between these two types of communications and predict whether a given text is written by a paedophile (Bogdanova et al., 2014), which is one step further than Pendar's (2007) research. The study was done on the assumption that paedophiles have low self-esteem and show emotional instability (Hall & Hall, 2007). Their analysis was on the word level, focussing on positively charged words, negative words, and other emotions, such as anger, fear, and sadness (Bogdanova et al., 2014). The study also borrowed concepts from McGhee et al. (2011), which were turned into features, such as information words, desensitization words, approach words relationship nouns, communicative and family words. Additionally, the article acknowledged Egan et al.'s (2011) concept of fixated discourse and attempted to model this in the dataset using lexical chains. Bogdanova et al. (2014) ran experiments to test their algorithm and found varying degrees of accuracy. For example, the emotional features and features borrowed from McGhee et al. (2011) reached a high degree of accuracy in the PJ and cybersex data but did not perform well with the NPS data: 'The best accuracy of 97% on the cybersex data is achieved by combining emotional, fixated discourse features and those from McGhee et al. (2011)' (Bogdanova et al., 2014: 118). A bigger dataset was required to validate their findings (Bogdanova et al., 2014).

Pranoto et al. (2015) analysed 111 transcripts taken from the PJ archive and compared them to 48 scripts taken from literotica.com to find out whether a logistic model could be established that is able to detect online grooming language. The study found that the most frequent OG strategies were: Using word in biology, body, and sexual category, Introduced sexual stage, Using word in feeling category, Arrange further contact and meeting, Telling the sexual preference or desire and sexual experience, and Calling intimate part using popular name or using slang word. The model could detect grooming conversations with 95% accuracy including 96% true positive and 93% true negative, four percent false positive and seven percent false negative. However, this did not consider any textual context and was purely based on quantitative statistical tests and Mathematics.

⁶ This is the term for online groomers used in this study

2.3.2 Decoys

The language of online groomers has also been compared to that of PJ decoys. This was the topic of a PhD thesis by Buchanan (2016). More specifically, Buchanan (2016) sought to answer how decoys communicatively addressed the dilemma of actively encouraging online groomers, without appearing suspicious and how groomers communicatively handled seducing teenagers, or decoys in this case, without scaring them off and being found out. Forty chat transcripts were analysed using an action-implicative Discourse Analysis (AIDA) framework. Buchanan (2016) identified the following communicative strategies decoys use: Target Presentation, OP (Online Predator) Safety, Sexual/Relational Contribution Management and Bust Facilitation. These all had sub-categories. Target presentation was concerned with how the decoy presented themselves to encourage the groomer, while also seeming realistic and not directly engaging in sexual communication. OP safety referred to making sure that the online predator feels safe in their luring activities. The sexual/relation contribution system was the third strategy and bust facilitation the last one. Buchanan (2016) concluded that decoys successfully handled the dilemma of seducing predators while making sure they did not entrap them:

Generally, these strategies served to help PJMs set up an ideal luring situation for OPs, while setting up an ideal conviction situation for themselves. These ideal luring situations allowed OPs to feel comfortable in their luring while PJMs gathered incriminating information. By utilizing the [sic] aforementioned strategies, PJMs were able to accomplish their goals of conviction while avoiding the risks of scaring OPs away or having their cases dismissed on the grounds of entrapment. (Buchanan, 2016: 205)

This sheds light on how the decoys use language in order to influence groomer and do their jobs, but it does not provide an overview of their training.

McLeod and Grant (2017) reported on identity disguise as part of the Pilgrim program, a training programme for specialist online undercover officers (OCOs). As part of this training program, the officers received linguistic training sessions, three to four hours of input on vocabulary, orthography, Pragmatics, and topic development concerning identity disguise. The goal was to be able to successfully take over chats of children being groomed in order to convict the online groomer. McLeod and Grant (2017) called this ‘authorship synthesis’ (McLeod & Grant, 2017: 2). In the article McLeod

and Grant (2017) reviewed this linguistic training prospective OCOs receive. In doing so, they collected two sets of IM (Instant Messenger) Yahoo chat logs between one or more trainees and one instructor created during role-playing exercises. One set was collected prior to training and one set collected after training. These two sets were compared focussing on grammar, lexis, punctuation, and model of production. Misspellings, emoticons, and omissions among others were identified in the spelling and vocabulary choices. McLeod and Grant (2017) identified different spellings of yes, no, what and g-clippings as illustrative examples.

The willingness to engage in sexual topics was also reviewed by McLeod and Grant (2017) and the article found that some trainees were reluctant in initiating sexual talk or acts and declined to participate in them when the offender, in this case the instructor, initiated them. McLeod and Grant (2017) stated that this needed to be overcome, as it was a vital part of the online grooming process. After linguistic training, trainees ‘particularly appreciated how the language analysis can protect against accusations of acting as an *agent provocateur*.’ (McLeod & Grant, 2017: 18, emphasis in original). McLeod and Grant (2017) also concluded that the easiest discrepancies to spot were structural ones, such as changes in spelling and punctuation. These errors were fewer post linguistic input, evidencing an improvement in the linguistic identity disguise of the trainees.

McLeod and Grant (2017) was one of the first articles to present some training undercover online agents receive, which has been criticised before. This can be extremely useful for other agents who are entering chat logs to catch groomers and even more important for law enforcement and agencies in charge of these agents, in terms of developing their own linguistic training sessions and reflecting on linguistic choices as well as other aspects, such as speed of typing and reluctance to use webcams, which McLeod and Grant (2017) point out cannot be improved by providing linguistic input.

Another study that compared groomer language and decoy language was done by Drouin, Boyd, Hancock, and James (2017) and analysed 590 chat logs from the Perverted Justice website, which made up the entire database at the time the article was written, using Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al.,

2015a). The study was interested in overall word count, sexual content, and clout and whether that affects sentencing outcome. Another goal was to create an online grooming database that can be compared to other databases. The first analysis comparing female decoys to male decoys revealed that groomers that groomed female decoys were younger, used more words overall and had higher clout scores, while the decoy used fewer sexual words. Based on these differences, the authors decided to carry out two separate analyses.

Overall, groomers used more sexual words than decoys. This was only inverted in nine percent of the cases, where the decoy used more sexual words than the groomer. Groomers who groomed female decoys used 3.5 times more sexual words than groomers communicating with a male decoy. The groomers also used more words overall. The clout scores of groomers ranged from 21.93 to 99 with two thirds having a score of more than 80. Decoys had a greater score of 80 in 12% of the individual cases. The last analysis, which related word count, sexual words, and clout score to jail time, showed a correlation between overall words and length of jail time, while no correlation between sexual words and jail time and clout and jail time was found. The groomers who used fewer words than their decoys received on average 12 months less jail time, than those using the same number of words or more than their decoy. Regarding sexual words, groomers who used fewer sexual words than their decoys received nine months less jail time and having lower clout scores than their decoys resulted in four months less jail time compared with groomers who used more sexual words and had higher clout scores than their respective decoys.

Drouin et al. (2017) also found that sexual words played an important role in the online grooming process and suggested most groomers who were caught following a sex sting operation engaged in 'sexual grooming of targets' (Drouin et al., 2017: 451). The high variety of sexual word usage was attributed to hypersexualised groomers and those who spend time building intimacy. Drouin et al., (2017) suggested that jurors should pay attention to this, as sexual words might reflect the grooming style of the individual. The finding related to overall word count was explained by the fact that groomers spend different amounts of time grooming their victims. Drouin et al. (2017) concluded that clout might be the most useful measure to use as evidence in courtrooms 'as it may help them [the jurors] determine whether the offender was

leading the conversation with their underage target or whether the agent was implicitly conveying control and status' (Drouin et al., 2017: 452).

2.3.3 Children

The grooming process has been attempted to be described from the children's perspectives by Whittle et al. (2014a), who conducted a series of studies based on interviews of perpetrators and children who had been groomed and analyse these using Thematic Analysis. In 2014, Whittle et al. (2014a) examined risk and protection factors in children that have been groomed online. They interviewed eight children who had been groomed online. The study identified family as a very important factor, as most children had temporary or long-term problems in this area. Internet safety was also not covered enough and should be introduced at primary school level. Reasons for engaging with the online groomer included loss of family protection and risky behaviours online.

Additionally, Whittle et al. (2014) identified three different child vulnerability scenarios based on the eight children: multiple long-term factors, trigger events and online behavioural risk. The first group, multiple long-term factors, displayed increasing risk factors in their life while having few protective factors and were considered vulnerable offline. The second group had protection in place until a trigger event occurred, which made them vulnerable both online and offline. The third group engaged in risky behaviour online despite having protective factors in place and few risk factors. These findings provide insights into why particular teenagers were targeted for online grooming. However, due to the small sample size, this cannot be generalised easily.

In another study Whittle et al. (2014b), using the same dataset as in their previous study, looked at the experiences of children who had been groomed online, finding similarities and differences. Interestingly, the female children all considered the groomer to be their boyfriend, supporting the understanding of relationship building and intimacy as part of the online grooming process, as evident in Lorenzo-Dus et al.'s (2016) online grooming discourse model. The grooming process also 'varied

dramatically in duration' (p. 409) and more than half of the groomers were grooming multiple children.

Whittle et al. (2014b) identified consistent grooming themes within this sample, which included manipulation, deception, regular/intense contact, secrecy, sexualisation, kindness and flattery, erratic temperament and nastiness, and simultaneous grooming of those close to the child. Whittle et al. (2014b) found that these grooming strategies did not occur in a linear order, they instead proposed a cyclical model, which can be seen in Figure 2.

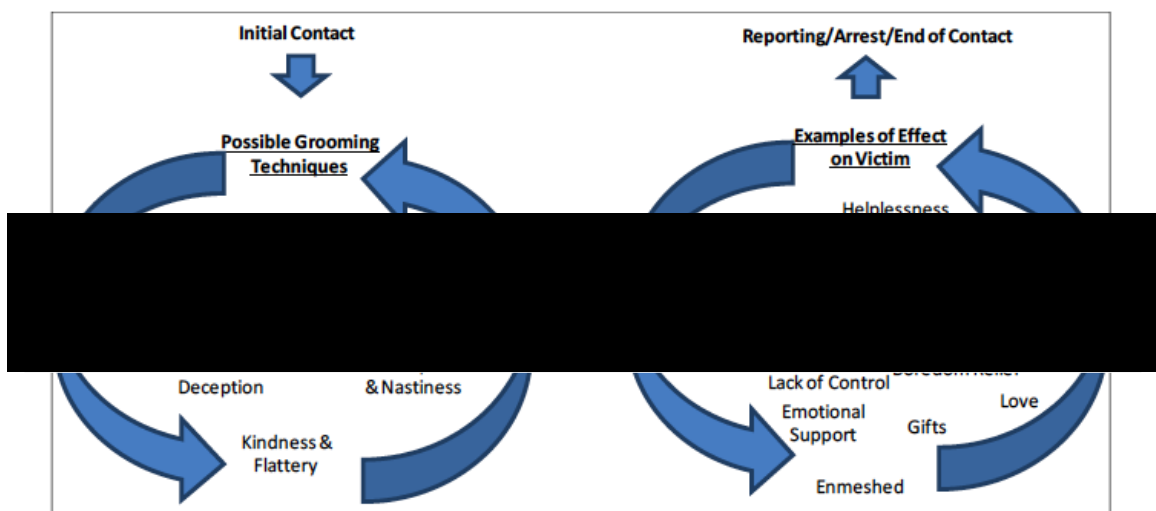


Figure 2: The online grooming techniques and effect by Whittle et al. (2014b: 419)

Whittle et al. (2014b) concluded that any phase of online grooming could occur at any stage of the grooming process and that not all groomers were contact-driven, as not all of them arranged offline meetings with children.

In 2015, Whittle et al. (2015) interviewed three different children who had been groomed and their groomers, separately, to find out more about how children and groomers perceived different stages of the online grooming process. The study noted whether the children and groomers agreed or disagreed on different things, such as who initiated sexual contact, whether sexual photos were sent, whether they felt they were in a relationship and whether there was a future to this. Most disagreement was recorded concerning the phases of sexual contact, with children putting the responsibility on the groomers and vice-versa. Whittle et al. (2015) concluded that

although overarching themes of grooming could be identified, the online grooming process was experienced uniquely by all the children who had been groomed and the process may vary according to the individuals involved (Whittle et al., 2015). This study is distinctive in that it interviewed children who had been groomed and groomers and relied on their account of the process, rather than evidence, such as chat logs.

Kloess et al. (2017b) also focused on the perspective of children who had been groomed online by analysing 22 groomer-child chat logs of cases and analysing how children behaved and reacted to online grooming. The article pointed out that no research had used 'real-world data' (Kloess, 2017b: 621) to identify children's behaviours and responses to online grooming, which was what they aimed to achieve. Kloess et al. (2017b) also used Thematic Analysis to identify overarching themes and found five key themes: getting to know each other, seeking assurance regarding relationship status, levels of engagement, secrecy of contact and child vulnerabilities. Within the first theme, the study found that groomers either used a direct or indirect approach with most children seeming to regard the groomer as a friend. The levels of engagement of the children varied, with some using excuses or reasons not to engage with the groomer via webcam, using face-saving strategies to get out of a situation without a confrontation with the groomer. Kloess et al. (2017b) identified risk factors and behaviours which might be contributing towards riskier and sexual behaviour and willingness. The small sample size was acknowledged by Kloess et al. (2017b) who stated that this type of data is difficult to access. The study admitted that caution should be taken when interpreting findings and making generalisations.

Another study investigating the children's perspective was done by Katz (2013), which interviewed 20 children who were suspected of being groomed online that led to offline sexual abuse to understand the narrative of children who had been groomed online. The study identified five overarching themes: the suspect's grooming process, the offline meeting, "he can do anything", the secret and the dynamic between the children and the interviewer. Katz (2013) employed Thematic Analysis to let themes emerge from the narratives, rather than coding them into predefined categories. Notably, eight of the 20 children refused to make allegations against their alleged abusers, despite evidence that abuse had taken place. Katz (2013) acknowledged the small sample size and admitted that the findings could not be generalised.

Nevertheless, the study provides insights into children's narratives and the grooming process of contact-driven offenders.

These studies have all been analysed using qualitative Thematic Analysis, an inductive 'method for identifying, analysing, and reporting patterns (themes) within data' (Braun & Clarke, 2006: 79). This approach may have been used because of the small sample sizes of these studies and the ethnographic approach they take. As Kloess et al. (2017b) pointed out it is also difficult to gain access to data of groomer-child interactions.

2.4 Digital Discourse Research

Online grooming happens in an online space, which needs to be conceptualised. As such, research into the three waves in Digital Discourse research, participatory frameworks and studies on dating/flirting language will be reviewed next.

2.4.1 The Three Waves of Digital Discourse Research

Digital Discourse research has been framed in terms of three waves with the first one starting in the 1990s as a consequence of the invention and wide-spread use of the internet (Herring, 2003). Researchers realised the potential of analysing computer-facilitated communication, that is, 'communication produced when human beings interact with one another by transmitting messages via networked or mobile computers' (Herring & Androutsopoulos, 2015: 127). Language scholars published research on digital discourse as early as 1985, but the single piece of research that made linguists aware of its vast potential as a research field, *Interactive written discourse as an emergent genre* was not published until 1991 by Ferrara, Brunner and Whittmore (Herring, 2003). This first wave of digital discourse research resulted in text-based descriptive linguistic approaches attempting to label a single variety of internet language (Herring, 2003; Garcés-Conejos Blitvich & Bou-Franch, 2019). In the second wave, this characterisation of Computer-Mediated Discourse as one genre and the tendency to generalise findings about how people behave online was criticised (Herring, 2004). As a result, the second wave focused more on social practices in

digital discourse, such as ‘linguistic variability, social diversity, issues of identity and community formation and maintenance’ (Garcés-Conejos Blitvich & Bou-Franch, 2019: 5).

The most recent third wave of Digital Discourse Analysis has moved towards analysing multimodality and how people ‘transcend different media’ (Garcés-Conejos Blitvich & Bou-Franch, 2019: 5). This was partly because of an emerging idea from Media Studies that internet users behaved differently online when there was an unknown number of audience members to address, called ‘context collapse’. The idea was that ‘[s]ocial media technologies collapse multiple audiences into single contexts, making it difficult for people to use the same techniques online that they do to handle multiplicity in face-to-face conversation.’ (Marwick & boyd, 2011). This concept gained traction but ‘has been criticized by several discourse analysts (Georgakopoulou 2017; Tagg, Seargeant & Brown 2017; Szabla & Blommaert 2018)’ (Blommaert et al., 2018). Instead, a focus on multimodality and embracing the dichotomy of online and offline selves emerged: ‘Digital practices always transverse boundaries between the physical and the virtual, and between technological systems and social systems.’ (Jones et al., 2015: 3).

Online grooming takes place in a digital environment. The PJ groomers analysed in this thesis all produce Digital Discourse data in the form of chat logs with their targets. They also all arrange meetings with their targets, which means they move from a digital space to a real-world space and their actions online have consequences for the offline world. Their online and offline selves cannot be separated. The initial approach of groomers sometimes happens in public chat rooms with multiple users, too and groomers sometimes target several targets at once, which means different participation frameworks are at play, which will be briefly reviewed next.

2.4.2 Participation Frameworks in Digital Spaces

As mentioned, some groomers use a ‘scatter-gun approach’ (Broome et al., 2018: 435) to target many targets at once. This means, they likely send the same messages to many people and their discourse could be regarded as one-to-many communication.

Marwick and boyd (2011) argued that '[s]ocial media thus combines elements of broadcast media and face-to-face communication.' (Marwick & boyd, 2011: 123), which is relevant in this context. Groomers talking to multiple targets using Internet Relay Chats is similar to YouTube comment sections, where participants produce one-to-many, many-to-many or one-to-one communication within the context of inter-group communication (without an "over-hearing" audience), which Dynel (2014) argues 'can be classified as multiparty interaction' (Dynel, 2014: 38). Grooming communication happens on an inter-personal communication level. Both grooming chat log interactions and YouTube comment sections can also be asynchronous and synchronous Computer-Mediated Communication, which Dynel (2014) says 'manifests various participatory configurations' (p. 38). Both also adhere to turn-taking and coherence, which can be difficult to signal in YouTube comment sections made up of many participants, but evidence has been found that turn-management signals and cohesion devices are prevalent (Bou-Franch et al., 2012). Overall, there are different participation contexts at play in online grooming communication. In the present context, these interactions will be one-to-one asynchronous interactions.

2.4.3 Online Dating Language

Flirting in online dating spaces has been the subject of previous research, albeit not a significant amount since 'romantic interpersonal communication has been termed a 'black box' in language and interaction research' (Mortensen, 2017: 582) because people are understandably reluctant to give researchers access to their spontaneous, private and intimate conversations. Mortensen (2017) analysed flirting in digital language (Instant Messages and emails) taken from two Danish online dating sites to find out how people negotiate romantic connections in an online space. The author focused on implicitness as an overlooked but 'key aspect of flirtation' (Mortensen, 2017: 583). Findings include participants creating an 'imagine togetherness' (Mortensen, 2017: 581) at some point in the future, which can take the form of an 'offline meeting [which] becomes an important potential future event for the participants to point to and negotiate (Mortensen, 2017: 584). This future togetherness is intentionally kept vague and is also 'not necessarily connected to specific spaces, but can in some cases be centered on an activity rather than a space' (Mortensen, 2017:

587). Closure is specifically avoided to keep the imagery of the togetherness implicit and create potential for fantasising:

[T]he implicit workings of flirting might also serve another function above that of politeness, namely that of nourishing (...) the excitement of possibility. (...) the participants put a lot of work into postponing the fulfilment of the pleasurable imagined togetherness that they jointly construct. By suspending concrete arrangements for offline dates, I argue that the participants create a tantalizing and playful space for fantasizing. (Mortensen, 2017: 583).

This is called ‘postponement of pleasure’ by Mortensen (2017: 594). In the context of online grooming, flirting is part of the key communicative intention of deceptive trust development, which groomers use to develop relationships with their targets. Face-to-face meetings also play a central role in the PJ data used in this thesis. These concepts of a shared togetherness and postponing pleasure could therefore be relevant to the analysis of online groomers’ discourse.

Two further studies by del-Teso-Craviotto (2006, 2008) sought to analyse ‘how people flirt, express attraction, or use erotic talk in dating Internet chat rooms’ (del-Teso-Craviotto, 2006: 461) and the issue of authentication in these online spaces before they can engage in ‘the intricacies of online desire and eroticism’ (del-Teso-Craviotto, 2008: 251). Del-Teso-Craviotto (2006) investigated five 30-minute conversations from five English-speaking AOL (America Online) chat rooms and four Spanish-speaking mIRC chat rooms oriented towards specific sexual orientations, religious affiliations, or ethnic groups. Specifically, flirting language was examined to see how people discursively construct sexual desire and attraction. Findings included participants constructing the online space as a game, which meant

participants seem to establish a safe distance between their real (critical) selves and their virtual (enjoying) selves (...) Framing interactions as a game also creates a safety net for the participants, since breaches of social norms and expectations regarding the appropriateness of sexuality in public situations can be excused because all is done in jest (del-Teso-Craviotto, 2006: 466).

Additionally, del-Teso-Craviotto (2006) found that this framing helped build a sense of solidarity, friendship, and intimacy, while also being a face-saving strategy. Participants were ‘able to derive their enjoyment not from physical contact with other

people, but from the unrealized or unrealizable promise of it' del-Teso-Craviotto, 2006: 475). Flirting strategies were also identified and included laughter symbolised by onomatopoeia, emoticons, and acronyms and humour. These strategies were sometimes used to mitigate sexual content and protect participants' vulnerability.

In del-Teso-Craviotto (2008), the author analysed the same data as in the previous study. The expectation was for participants in these chat rooms to be authentic and orient themselves to the sexual identity, gender and age group indicated by the channel name, e.g., 'Lesbian 30s' (del-Teso-Craviotto: 2008: 253). This was done by using an opening turn of 'a/s/l', revealing the age, sex, and location of the participant. The location information was included 'in case the relationship wants to be transferred from the chat room to the offline world' (del-Teso-Craviotto: 2008: 256). This opening turn and other descriptors of the participants identity were expected to align with the chosen chat room. Del-Teso-Craviotto (2008) concluded that '[t]his does not necessarily mean that everyone that participates in a chat believes that everyone else is being truthful about the identity they display, but that belief is a communicative precondition for flirtation and eroticism to take place in the room' (Del-Teso-Craviotto, 2008: 266). This is similar to online grooming interactions, in which the opening turn of 'a/s/l' also appears. Arguably, the groomer and target both have to authenticate themselves before they can interact with each other meaningfully. Furthermore, the groomer will also check whether the target is underage, as they claim to be, so the authentication process is ongoing (see section 2.5.3.4).

2.5 Summary: Taking Stock and Moving Forward

2.5.1 Overview of Online Grooming Communication

This literature review chapter has first looked at the online grooming process, which has different definitions and terminologies attached to it without any consensus. The term used in this thesis is online grooming, as it seems to encompass the entire process and the complexity of the issue. The chapter then reviewed proposed online grooming process models, divided into sequential models and non-sequential models. The sequential models by O'Connell (2003), Black et al. (2015), Egan et al. (2011) and Kloess et al. (2017a) suffered from some limitations, such as self-reporting, relying on

qualitative analyses and O'Connell's (2003) originally proposed five stages, small data sizes and trusting LIWC scores without validating claims based on close readings of the text. Furthermore, there is little evidence that the online grooming process is sequential, as both Black et al. (2015) and Gupta et al. (2012) found evidence for certain strategies occurring outside their dedicated stages.

The non-sequential models developed by Williams et al. (2013), Lorenzo-Dus et al. (2016), van Gijn-Grosvenor & Lamb (2016), Quayle and Newman (2017) and Chiang and Grant (2017) did not suffer from the constraints of O'Connell's (2003) five stages and are therefore able to introduce more complex models and strategies. Their data size was still quite small, and their data mostly relied on mixed method approaches, Thematic Analysis, or quantitative statistics only, without considering the context. In this thesis, the online grooming process will be regarded as non-sequential, rather than sequential.

2.5.2 Overview of Participant Features

The participants of the online grooming process were reviewed next, which was divided up into three parts: groomers, decoys, and children. The groomer section reviewed groomer typologies often informed by sex offender typologies. Typologies that were proposed included situational offenders and preferential offenders (Lanning, 2010), using child pornography, and not using child pornography (Krone, 2004; Elliot & Beech, 2009), fantasy-driven offenders and contact-driven offenders (Briggs et al., 2011) and engaging in real-time masturbation, cybersex, schedulers, (cybersex/schedulers) and buyers (DeHart et al., 2017). Broome et al. (2018) suggested that there is little evidence for Briggs et al.'s (2011) fantasy-driven and contact-driven distinction and proposed to adopt the European Online Grooming Project's typology, which classifies groomer into intimacy seeking, adaptable and hypersexualised groups.

Attempts to detect or profile online groomer language, mostly using Machine Learning, were reviewed next. Guice (2016) compared a corpus of PJ groomer language to that of teenagers and PJ decoys while Pendar (2007), Kontostathis et al. (2009), Bogdanova et al. (2014) and Pranoto et al. (2015) all attempted to detect or

predict groomer language compared to general language or decoy language. These studies, while analysing language, were not informed by tools or frameworks from Linguistics and concentrate on comparing groomer and decoy language, which this thesis does not do.

The literature review moved on to discussing three studies that (partially) analysed PJ decoy language. Buchanan (2016) examined how decoys address the communicative dilemma of actively encouraging and contributing to grooming conversations without attracting suspicion. McLeod and Grant (2017) reported on a study of linguistic training for online undercover officers (OCOs) who take over communication with a groomer and have to assimilate the victim's language, also called authorship synthesis. Lastly, Drouin et al. (2017) compared PJ groomer language to that of PJ decoys using three specific LIWC categories: sexual word count, overall word count and clout.

The third participation group were children who have been groomed. The studies reviewed included Whittle et al. (2014a, 2014b, 2015), Kloess et al. (2017b) and Katz (2013) that identified several risk factors, children's' behaviours and effects of online grooming on children. Data sizes were small, likely because of the difficulty in obtaining groomer-child data (see chapter three, section 3.1.3.2). The literature review will now move on to proposing a new working terminology of online groomer language informed by the key literature on online grooming language.

2.5.3 Overview of Online Grooming Communicative Intentions

This section provides an overview of the communicative grooming intentions and their strategies (at various levels) collectively identified in the literature, which will help with chapter four that seeks to establish the features of online groomer language compared to general digital chat language. It will also subsequently help in chapters five and six to consistently refer to the same terminology. The term communicative intentions has been chosen in this thesis, because it analyses the discourse of online groomers whose overall goal is to groom their targets and they do so by using digital language. These communicative intentions are made up of strategies, which are the linguistic realisations of these overall intentions.

Figure 3 shows the communicative intentions of online groomers that the literature introduced, which will be explained in more detail below. These intentions are Seeking Sexual Gratification (SSG), Developing Deceptive Trust (DDT), Arranging Further Contact (AFC), Assessing Risks and Isolating (ARISO) and Negotiating Power Dynamics (NPD). These terms will be used throughout the rest of the thesis, unless the term in the original study is specifically highlighted or part of differences in terminology being discussed.

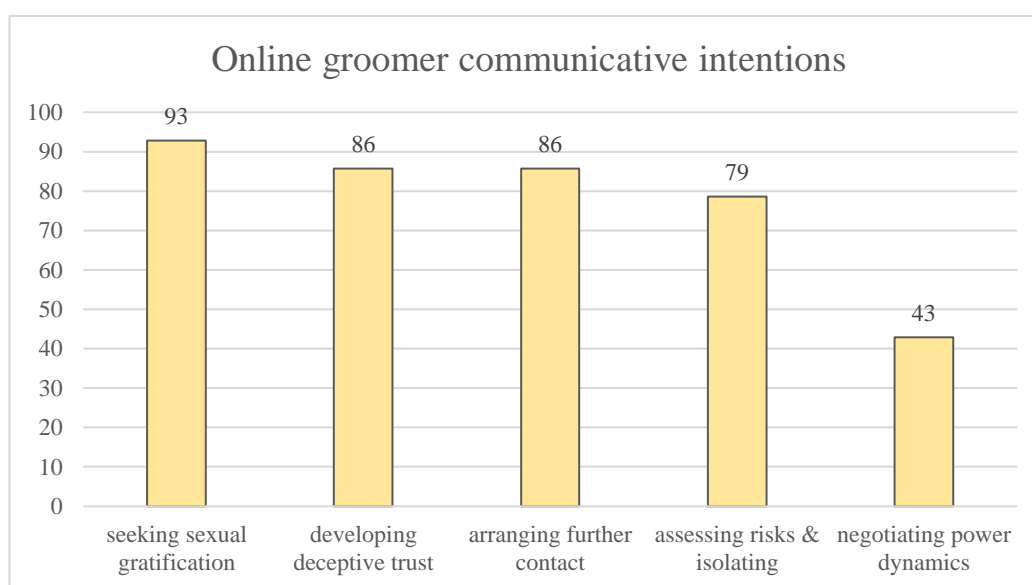


Figure 3: Online groomer communicative intentions & studies reporting them

These terms, which will be defined in the next few paragraphs, have been derived from reviewing the literature and summarising the communicative intentions. Each term is realised through a set of strategies that will be outlined in more detail below.

Seeking sexual gratification refers to instances where the groomer receives instant sexual gratification from the conversation, as suggested by Lorenzo-Dus et al. (2016), by either sharing sexual content with the target, such as pornographic photos, discussing the target’s sexuality or disclosing sexual information about themselves, among others. Developing deceptive trust comprises of strategies, such as discussing the groomer – decoy relationship, offering a compliment or gift and eliciting personal information, which leads to a development of deceptive trust and may seem like a trusting relationship. However, the ulterior motive of the online groomer is to

manipulate the target into sexual activity. These are instances where the groomer seemingly strengthens the relationship to the target.

Arranging further contact refers to instances in which the groomer proposes an offline meeting, and the logistics of this future meeting are discussed. This communicative intention has been adapted to include any arrangement of further contact, e.g., by phone, email, or other platforms. Assessing risks and isolating refers to those bits of conversation, in which the groomer makes sure that the target will not report him, and they will not be detected or arrested. This is done by eliciting information about the target's parents' schedule, the location of the computer in the home and gauging the target's reaction to certain pieces of information, such as the groomer's real age or sexual comments, which may be retracted based on this reaction. Negotiating power dynamics is defined as instances in which the groomer influences or manipulates the target. This can entail adopting low-risk behaviour, challenging the target, blackmailing them, or putting pressure on them.

The most frequently reported component in 93% of the core literature is seeking sexual gratification, which is unsurprising, as online grooming is the process of preparing a target for sexual abuse, which does not end when sexual acts are introduced into the conversation. Other terms to describe this component are sexual stage (O'Connell, 2003; Black et al., 2015), sexual content (Williams et al., 2013; Winter et al., 2017), sexual gratification (Lorenzo-Dus et al., 2016), sexual rapport (Chiang & Grant, 2017), desensitization (Bogdanova et al., 2014) and implicit/explicit grooming (Egan et al., 2011). The only study that does not explicitly refer to "sexual content" is Quayle and Newman (2017), as it reviews reports from the public that may or may not include sexual content or abuse. Their focus is different from other online grooming studies, which seek to describe the process and how the groomer uses language to convince a target to e.g., carry out sexual acts. The next most frequently reported component with 86% is developing deceptive trust, also called friendship forming stage (O'Connell, 2003; Black et al., 2015; Gupta et al., 2012), deceptive trust development (Lorenzo-Dus et al., 2016), rapport building (Williams et al., 2013; Chiang and Grant, 2017), establishing a relationship (Kloess et al., 2017a) and befriending the target (Egan et al., 2011), which Quayle and Newman (2017) as well as Gauz (2016) do not explicitly

mention. Gauz (2016) focuses on salient themes in the first 20 lines of chat logs, which may explain why relationship forming is not included in this study.

Arranging further contact is also reported by most of the core literature. It is also called approach (Lorenzo-Dus et al., 2016), talk about meeting at offender's house (van Gijn Grosvenor & Lamb, 2016), arranged in-person meetings (Winters et al., 2017) and arranging further contact and meeting (Pranoto et al., 2015). It does not occur in O'Connell (2003), Williams et al. (2013) and Bogdanova et al., (2014). Assessing risks and isolating is another intention of online groomers which is supported by most studies. This is also called risk assessment (O'Connell, 2003; Black et al., 2015; Gupta et al., 2012; Winters et al., 2017), assessment/managing risks (Chiang & Grant, 2017), assessment (Williams et al., 2013), questions to assess the risk of the conversation (Pranoto et al., 2015) and minimising risk (Egan et al., 2011). It is not mentioned by Gauz (2016), Quayle and Newman (2017) and Bogdanova et al. (2014), who conduct a Machine Learning study on how to detect online grooming. Their features do not include risk assessment, as they focus on the emotional instability and immaturity of the groomers. They do not seek to explain the online grooming process; However, their findings and approach are still relevant to this discussion.

The last intention negotiating power dynamics is reported by 43% of the studies. The studies that mention it explicitly are O'Connell (2003), Williams et al. (2013), Pranoto et al., (2015), Kloess et al. (2017a), Egan et al. (2011), Lorenzo-Dus et al. (2016) and Bogdanova et al. (2014). This intention is also referred to as fixated discourse (Egan et al., 2011; Bogdanova et al., 2014), coercion (Kloess et al., 2014), blackmail (Williams et al., 2013) and compliance testing (Lorenzo-Dus et al., 2016). As the whole process of online grooming is subject to underlying coercion and deceit, the remaining studies chose not to explicitly refer to this intention in their models.

The different intentions are realised through strategies, which can be seen in Table 1. Overall, 22 strategies are identified. Of these 22 strategies, a fourth (23%) make up the seeking sexual gratification intention, the negotiating power dynamics intention and the developing deceptive trust intention, four (18%) are used to realise assessing the risks and isolating intentions and three strategies (13%) are used to carry out arranging further contact.

Table 1: Number and percentage of strategies to carry out intentions

Intention	Number	%
Seeking sexual gratification	5	23
Negotiating power dynamics	5	23
Developing deceptive trust	5	23
Arranging further contact	3	13
Assessing risks & isolating	4	18
Overall	22	100

This does not suggest that seeking sexual gratification, negotiating power dynamics, and developing deceptive trust are the most important intentions, only that more strategies have been identified to implement them. The next sections provide a breakdown of the different strategies that are used to carry out the communicative intentions and how many studies identify or mention these strategies.

2.5.3.1 Seeking Sexual Gratification

Figure 4 shows the strategies used to advance the intention of seeking sexual gratification.

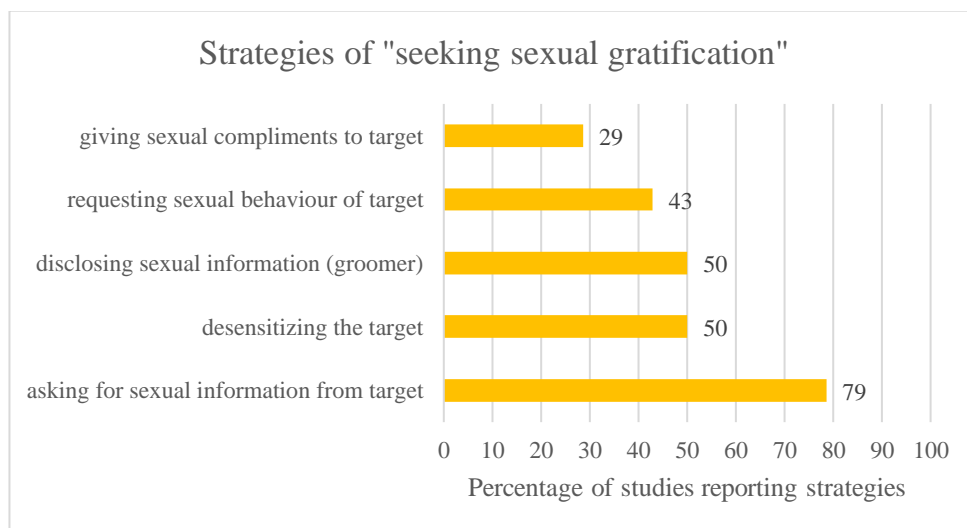


Figure 4: Overview of "seeking sexual gratification" strategies

The most frequently reported strategy of seeking sexual gratification is asking for sexual information from the target (79%), which is done by asking for a sexual

description/photo, asking intimate questions, such as enquiring about the sexual experience of the target or of those close to the target (friends, ex-partners) and asking about the target's sexuality and sexual desires.

Half of the studies report desensitizing the target as a strategy of seeking sexual gratification. This strategy includes reframing sexual activity as beneficial or a game, using 'child language' (Pranoto et al., 2015: 360) to describe sexual organs and diminishing sexual anxieties. Disclosing sexual information is also reported in half the studies (50%), which includes disclosing sexual fantasies of the groomer and sharing sexual information about the groomer's ex-partner. Requesting sexual behaviours of the target, the fourth most widely reported strategy (43%) is done by asking the target to masturbate and explaining/discussing masturbation to the target. The last strategy, giving sexual compliments to the target, is reported by just under a third of the studies (29%).

2.5.3.2 Developing Deceptive Trust

The strategies used to carry out the communicative intention of developing deceptive trust are displayed in Figure 5. The literature reports that eliciting personal information and strengthening the target – groomer relationship are the most frequently identified strategy to advance to intention of developing deceptive trust. They both appear in 79% of the examined studies and are realised by asking for pictures, contact information (email address, address phone number), asking for a non-sexual description, enquiring about ASL (age, sex, location), talking about hobbies and activities of the target, and asking about the target's relationship status, for the former. The latter is done by giving the target compliments, discussing the trusting relationship between groomer and target, asking for the target's opinion on the groomer's appearance and giving sympathy and presents to the target.

Disclosing personal information is the third most frequently reported strategy (36%), which is realised through disclosing the groomer's hobbies, sexual experiences and attempting to impress the target by disclosing positive self attributes.

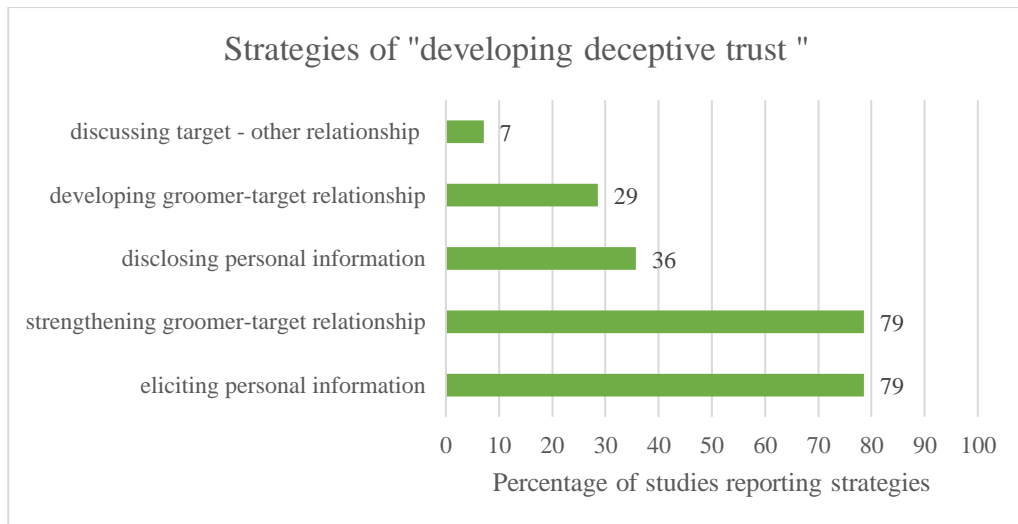


Figure 5: Overview of "developing deceptive trust" strategies

The fourth most frequently identified strategy is developing the target - groomer relationship, which is reported by 29% of the studies. Small talk is the sub-strategy to advance this strategy. The last strategy is discussing target – other relationships, which includes talking about the target’s current or ex-partners.

2.5.3.3 Arranging Further Contact

Arranging further contact has three strategies associated with it (See Figure 6). The literature reports that arranging further offline contact, which involves suggesting an offline meeting and discussing logistics, is the most prominent of the strategies, as it appears in 79% of the reviewed studies.

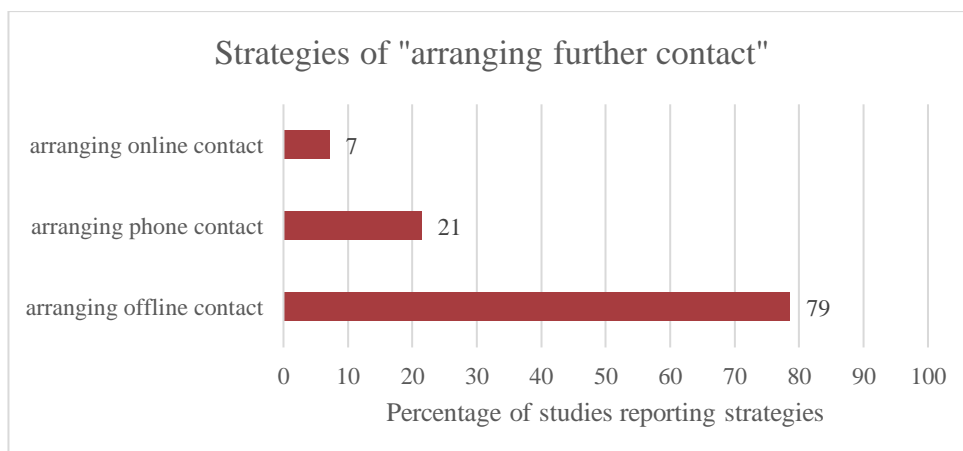


Figure 6: Overview of "arranging further contact" strategies

Arranging further contact by phone appears in 21% of the reviewed studies and arranging further contact online is only discussed in seven percent of the studies. This is likely due to the focus on contact-driven versus fantasy-driven online groomers.

2.5.3.4 Assessing Risks and Isolating

Figure 7 shows the assessing risks and isolating strategies, which can be summarised into four strategies, pictured below. All studies that were reviewed report checking whether it's safe to proceed, which includes asking about the parents' schedule, the location of the computer in the home and the number of people using it, confirming the age of the target, asking about the home life of the target, checking the immediate surroundings of the target, and gathering information about how many people have access to the target's online accounts.

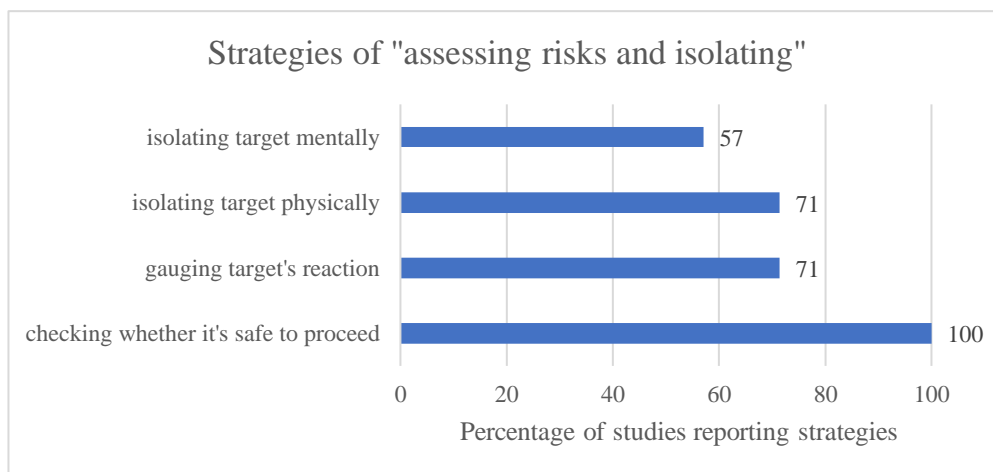


Figure 7: Overview of “assessing risks and isolating” strategies

The second strategy, gauging the target's reaction, sees the groomer mentioning the inappropriateness of the situation, online sting operations and the possibility of the target being an undercover police agent, pointing out the age gap between the groomer and the target and retracting sexual comments based on the reaction of the target. Isolating the target physically, which was reported in 71% of the reviewed studies, is done by emphasising the secretive nature of the conversation and thus making sure the target deletes chat logs, texts, photos, and previous conversations. This also involves arranging to spend more time with the child. Mental isolation, which occurred in 57%

of the literature, sees the exclusivity of the groomer – target relationship put into focus, which includes increasing the target’s dependency on the groomer.

2.5.3.5 Negotiating Power Dynamics

The strategies advancing the last communicative intention negotiating power dynamics can be seen in Figure 8. Just over a third of the studies (36%) report giving the target control, coercing the target, and challenging the target as strategies. Giving the target control involves outwardly checking the target’s consent, offering perceived control over sexual topics, and using strategic withdrawal, which means the groomer lets the target make decisions. Coercing the target is done by emotionally blackmailing the target, using aggression to get what the groomer wants or using other strategies to coerce the target.

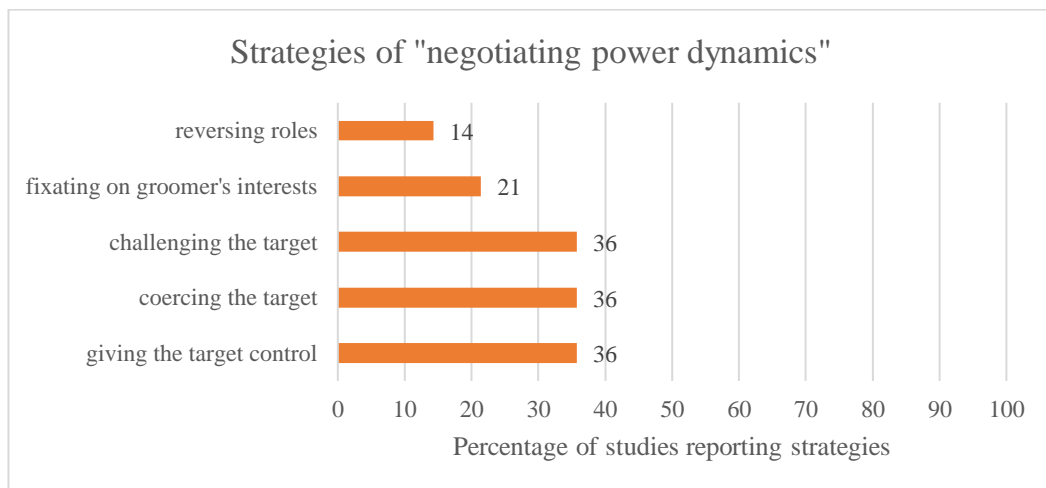


Figure 8: Overview of “negotiating power dynamics” strategies

Challenging the target includes pressing boundaries or otherwise challenging the target. The next most frequently reported strategy, fixating on groomer’s interest, is only reported by 21% of the literature. It involves a repetition of sexual themes or other conversational topics to advance the groomer’s interest and snub the target. The last strategy, reversing roles, is reported by 14% of the literature. This involves the groomer adapting low risk behaviour that the target would otherwise display, such as suggesting a meeting in a public place.

2.5.4 Conclusions on Existing Online Grooming Literature

This chapter has reviewed the existing literature on online grooming with a particular focus on online grooming language. It first examined different definitions of online grooming to clarify the terms used in this study which are online grooming or online child sexual grooming. It then moved on to looking at the online grooming process as a whole and taking a closer look at sequential and non-sequential models of online grooming language. The processes and strategies within these models were used to derive a new online grooming terminology in the last part of the chapter. The few studies that mention duration of online grooming were also reviewed. The next main section of the chapter looked at participants in the online grooming process: online groomers and their typologies and Computational Linguistics language profiling, decoys, and children. Then the chapter provided an overview of Digital Discourse research to situate the study and provide background information for the first empirical chapter. The chapter was summarised before moving on to the newly developed new terminology of online groomer communicative intentions, which will be used throughout the thesis. The study will now move on to outlining the methodology, specifically the data size and source of this study, ethical considerations, the framework, and procedure.

3. Chapter 3: Choosing the Right Methodology

Having reviewed the relevant literature in the previous chapter and identified areas that could be developed into research questions, the methodology of this thesis will be described next, outlining how these research questions will be studied and addressed. The chapter will first define the data size and source of this study in section 3.1.1. It will situate the study by reviewing different approaches to the analysis of online grooming data in extant literature, with a focus on data sizes and methodological approach. One particular approach, using Linguistic Inquiry and Word Count (LIWC) as the only other quantitative approach to the study of online grooming language, will be examined in sub-section 3.1.2. Different data sources (groomer-child data and groomer-decoy data) previously used in online grooming research will be discussed concerning their advantages and disadvantages in sub-sections 3.1.3.1 and 3.1.3.2 respectively. The chapter will then discuss ethical considerations (section 3.1.4) of using groomer-child data (section 3.1.4.1) and groomer-decoy data (section 3.1.4.2). The chapter will also discuss researcher wellbeing, in the context of working with distressing data and describing resilience techniques to overcome this. Ethics will also be covered in the same sub-section, describing the ethics review process for this study. The framework of the thesis will be explained next by giving an overview of what a Corpus-Assisted-Discourse-Studies (CADS) approach is (sub-section 3.2.1) and describing its “standard” analytic toolkit (sub-section 3.2.2). This toolkit is divided into Corpus Linguistics (sub-section 3.2.2.1), specifically keyness measures, collocation calculation (sub-section 3.2.2.2), and a Discourse Analysis framework (sub-section 3.2.2.3), with a focus on Speech Act Theory and Politeness Theory. The analytic procedure this PhD followed will be described in the third section (3.3) with data extraction, data normalisation and corpus and sub-corpus creation and data analysis outlined in respective sub-sections. The data analysis steps for the quantitative keyword analyses will be described in sub-section 3.3.5.1 and the data analysis procedure for the qualitative chapter on requests will be described in the next sub-section, 3.3.5.2. Conclusions will be drawn in the last sub-section, 3.3.6.

3.1 Data

3.1.1 Approaches to the Study of Online Grooming Communication

The data in this study, the Groomer Normalised (GN) corpus, is made up of groomer chat logs from 644 groomer usernames scraped from the Perverted Justice (PJ) archive. This amounts to a total of 3,604,926 words and 636,781 writing turns, as shown in Table 2. The number of groomer usernames is 644 instead of 622⁷ which is the number of chat logs in the complete PJ archive, as some groomers use various usernames.

Table 2: Groomer normalised (GN) corpus details

Words	Writing turns	groomers
3,604,926	636,781	644

Including all groomer usernames was done so that all of the chat logs in the PJ archive would be captured and analysed. Most of the literature into online grooming that used PJ as a data source has only used part of the PJ archive (see chapter two, section 2.2.1).

There have been several different methodological approaches to the study of online grooming chatlog data. As outlined in the previous chapter (chapter two, section 2.2.1), many studies into online grooming interactions based on chat log data relied on a qualitative approach, using Thematic Analysis (Braun & Clark, 2006) or Content Analysis of grooming language in chat logs to let categories emerge, sometimes with very small sample sizes, which do not support generalising findings. The sample size, especially concerning language data also has a broad range. Table 3 shows an overview of the data size, type and methodology used in the studies reviewed to develop a working terminology of online grooming language (see chapter two, section 2.5.3).

⁷ There are 623 convictions in the PJ archive (<http://www.perverted-justice.com/index.php?archive=full>) based on 622 chat log convictions and one research conviction that did not include a chatlog.

Table 3: Online grooming language/communication studies, chronologically

Study	Data	Method
O'Connell (2003)	50 hours in chat rooms (groomer-decoy)	Sociolinguistics analysis
Egan et al. (2011)	20 groomer-decoy chat logs	Content analysis
Gupta et al. (2012)	75 groomer-decoy chat logs	LIWC
Williams et al. (2013)	eight groomer-decoy chat logs	Thematic analysis
Bogdanova et al. (2014)	20 groomer-decoy chat logs	Machine Learning models
Black et al. (2015)	44 groomer-decoy chat logs	LIWC, Content Analysis
Pranoto et al. (2015)	111 groomer-decoy chat logs	Logistic mathematical models to classify online grooming conversation
Lorenzo-Dus et al. (2016)	24 groomer-decoy chat logs	Inductive language-based Content Analysis, supported by relational work
Gauz (2016)	100 groomer-decoy chat logs	Grounded theory
van Gijn Grosvenor & Lamb (2016)	101 groomer-decoy chat logs	Mixed methods (Qualitative coding & descriptive statistics)
Chiang & Grant (2017)	seven groomer-decoy chat logs	Rhetorical move analysis
Chiang & Grant (2018)	20 groomer-child chat logs	Rhetorical move analysis
Winters et al. (2017)	100 groomer-decoy chat logs	Content analysis based on information on PJ website & review of literature
Kloess et al. (2017a)	29 groomer-child OG chat logs	Thematic analysis
Quayle & Newman (2017)	267 Cybertip.ca reports	Mixed methods (Content Analysis & numerical data analysis)
Schneevogt et al. (2018)	622 groomer-decoy chat logs	Corpus Linguistics
Lorenzo-Dus & Kinzel (2019)	>600 ⁸ groomer-decoy chat logs	LIWC & CADS
Lorenzo-Dus et al. (2020)	622 groomer-decoy chat logs	CADS
Lorenzo-Dus & Kinzel (2021)	>600 ² groomer-decoy chat logs	CADS

⁸ The data in these studies was comprised of the groomer contribution of the whole PJ database

As can be seen, data sizes ranged from seven chat logs (Chiang & Grant, 2017) to over 600 chat logs (Schneevogt et al., 2018; Lorenzo-Dus & Kinzel, 2019; Lorenzo-Dus et al., 2020). The studies also employed different methodologies, ranging from content/Thematic Analysis (Egan et al., 2011; Williams et al., 2013; Gauz, 2016; Winters et al., 2017; Kloess et al., 2017a), which are primarily qualitative, to logistical models (Pranoto et al., 2015) and LIWC (Gupta et al., 2012; Black et al., 2015), which are primarily quantitative. Some studies also used a mixed methods approach (Lorenzo-Dus et al., 2016; van Gijn Grosvenor & Lamb, 2016; Black et al., 2015; Quayle & Newman, 2017; Lorenzo-Dus & Kinzel, 2019; Lorenzo-Dus et al., 2020; Lorenzo-Dus & Kinzel, 2021).

Drouin, Boyd, Jeffrey, Hancock, and James (2017), which used LIWC to compare groomer language to that of the decoys, was the first study to use the entirety of the existing Perverted Justice database (590 chat logs). It used 590 chat logs, which was the entire database during their data collection. The number rose to 622 in 2016. Quantitative studies using LIWC lend themselves well to large datasets. Drouin et al. (2017) focused their analysis on sexual words, word count and clout, which were examined without context and might not be an accurate representation of the PJ archive language they analysed. Lorenzo-Dus and Kinzel (2019) used a larger dataset than that (>600 PJ chat logs) when they examined sexual words using both LIWC and Corpus Linguistics software. Chiang and Grant (2017) only analysed seven transcripts in detail, making use of qualitative methods, rather than quantitative ones. Winters et al. (2017) selected 100 transcripts from the PJ website to analyse offender characteristics, victim characteristics and conversation characteristics. The only studies that have used the entirety of the PJ archive before it shut down operation in 2019 were Schneevogt, Chiang and Grant (2018) and Lorenzo-Dus, Di Cristofaro and Kinzel (2020). Schneevogt et al. (2018) used Corpus Linguistics methods to find out whether the PJ archive included instances of overt persuasion or sexual extortion, two moves initially identified in Chiang and Grant (2017, 2018). Lorenzo-Dus et al. (2020) sought to identify regular communicative patterns used by online groomers aligned to the grooming intentions identified in a previous study based on a sub-set of the PJ archive (Lorenzo-Dus et al 2016). The models that have tried to characterise online grooming language have thus used very different approaches and data sizes, which also resulted in differences in what they contain. There is no consistency in these

approaches and studies. However, these studies had different aims and research questions justifying their data sizes and methods.

Another point to discuss is the methodological approach these studies have taken. Work undertaken using Machine Learning usually takes underlying linguistic models or concepts, such as O'Connell's (2003) online grooming model or McGhee et al.'s (2011) concepts and builds their algorithmic models on them. This can be problematic, as the literature review chapter highlighted, models about online grooming communication have not been very methodical and thorough in their approach to analyse and characterise groomer language.

A small sample size might be beneficial for a close examination of the text and online grooming strategies, while bigger datasets could point out general trends and patterns that all online groomer communication has in common. A combination of the two would be the ideal way to analyse online groomer language by empirically analysing online groomer behaviour. This suggests there is a gap in the literature of studies that are able to analyse large datasets of online grooming from which communicative patterns can be derived, without losing the ability to examine the language qualitatively. In other words, using a combination of qualitative and quantitative analyses to analyse the phenomenon of online groomer language further. A Corpus-Assisted Discourse Studies (CADS) approach combines these two methods by providing both a quantitative overview and close readings of the text via KWIC (Key word in context) analyses and concordance lines. This will be discussed further in subsection 3.2. The only study to date to have done so was Lorenzo-Dus et al. (2020), which analysed communicative patterns online groomers typically use. The corpus in this study fulfils these requirements for a bigger dataset and adds specific knowledge about the duration of online grooming and manipulative requesting behaviour.

3.1.2 Two Quantitative Approaches to Online Grooming Language

Having examined different data sizes in previous online grooming research, the next section will examine an alternative methodology to CADS, the only other quantitative methodology in the study of online grooming using large datasets (Gupta et al., 2012; Black et al. 2015; Drouin et al., 2017; Lorenzo-Dus & Kinzel, 2019) similar to the one

in this study, which warrants a closer examination. The tool is Linguistic Inquiry and Word Count (LIWC), and its potential and drawbacks will be reviewed by example of one article that used LIWC, Drouin et al. (2017), which analysed 590 PJ chat logs according to sexual words, word count and clout, three of LIWC's categories.

LIWC is a computerised text analysis tool that was developed by Psychologists Pennebaker, Tausczik, Francis, Booth and Boyd (Pennebaker et al. 2015a, 2015b), who wanted to create an automated tool that 'simply looked for and counted words in psychology-relevant categories across multiple text files' (Tausczik & Pennebaker, 2010: 27). LIWC was first developed in 1994 and was updated in 1997, 2007 and 2015. At LIWC's core there is a dictionary that was developed and defined by Pennebaker et al. (2015a) and cannot be changed by the user. This dictionary is made up of almost 6,400 words, word stems and emotions in the current version of LIWC (LIWC2015). Categories include summary language variables (analytical thinking, clout, authentic, emotional tone, words/sentences, words > 6 letters, dictionary words), linguistic dimensions (function words, pronouns, articles, prepositions, conjunctions and others), word classes (e.g. verb, pronoun, adjective), which are called 'other grammar' and psychological processes, which are divided into 10 sub-processes (affective, social, cognitive, perceptual, biological, drives, time orientations, relativity, personal concerns, informal language) (Pennebaker et al., 2015a: 3-4). LIWC works by counting the overall words and comparing each word to its dictionary. It then computes the percentage of the specific category certain words belong to compared to the overall word count, e.g., five percent out of the words that were analysed belong to the sexual word dictionary. A word can be categorised into multiple categories, e.g., *crying* is sorted into verbs, past focus, negative emotion, sadness, and overall affect (Tausczik & Pennebaker, 2010).

First, the above dictionary generation procedure was justified by explaining that words 'do not adhere to traditional psychometric laws that we see in questionnaires' (Tausczik & Pennebaker, 2010: 28) and can thus not be validated using reliability statistics or validation tests. Tausczik and Pennebaker (2010) also stated that style word, or function words, were much more interesting to a psychologist than content words, which merely 'convey what they [people] are saying' (Tausczik & Pennebaker, 2010: 29). Style words, on the other hand, 'are much more closely linked to measures

of people's social and psychological worlds' (Tausczik & Pennebaker, 2010: 29). LIWC was developed from a psychological point of view with little to no knowledge of Linguistics beyond basic grammatical word classes. This has important implications when it comes to using LIWC as a linguistic measure or for analyses that are interested in more than just counting words. A researcher should have a basic understanding of the psychological studies behind LIWC before using it for linguistic analyses and interpreting of their findings. Otherwise, misinterpretation is likely to occur. This is not an easy task, as LIWC is quite non-transparent and refers to literature in the field of Psychology, but rarely explains what its measures are based on exactly.

Another criticism of LIWC is that it considers a word in isolation and matches it to one or more of its dictionaries. This was noted by Ventura (2012), which analysed deception and faking in personality assessments and concluded that LIWC may not be the best tool to use, as it is based on 'pre-defined word sets' (Ventura, 2012: 172). Franklin (2015), which discussed theoretical considerations associated with text analysis software, agreed:

It only considers single words. In doing so, LIWC assumes that words have meaning in isolation (A3). There is also an implicit assumption (A4) that inaccuracies due to negation, word order, particles (e.g. in phrasal verbs), ambiguity of word senses, type of discourse and other context-dependent factors are negligible or unimportant. (Franklin, 2015: 9).

This assumption is a risky one to make, especially when it comes to non-literal language use, such as irony, idioms, and sarcasm. This lack of context is something the creators of LIWC acknowledge, stating that the word *mad* is currently categorised as an anger word, when it could occur in constructions such as *mad about someone* and not be an expression of anger (Tausczik & Pennebaker, 2010). They caution anyone who tries to rely on word use to know 'people's true selves' (Tausczik & Pennebaker, 2010: 30). This decontextualisation has implications for Drouin et al.'s (2017) sexual word category. It is entirely possible that words belonging to that category were coded incorrectly as not belonging to it, because their meaning is contextual and non-obvious. Simultaneously, words that were coded as belonging to it, might not belong to this category based on the context of the chat log, although this is more unlikely, given the online grooming context of the study.

In fact, one study tested this and found a number of words regularly used by online groomers that had sexual connotations but were not part of the sexual word category in LIWC (Lorenzo-Dus & Kinzel, 2019). This limitation found in Lorenzo-Dus and Kinzel (2019) is related to another criticism, which is the content of the sexual word category. It is only briefly described as containing ‘sexual words such as *horny*, *love* and *sex*, as well as explicit and non-explicit words pertaining to sexual organs and sexual acts’ (Drouin et al., 2017: 441). However, no full list of items belonging to this group is given. The fact that *love* is part of the sexual word category is a problem. *Love* is not inherently sexual. It can become sexual in contexts, such as *I would love to have sex with you*, however even in this example the word *love* only amplifies the sexual content. Love would be better categorised into a *relationship* category or a *positive emotion* one⁹, even if the underlying meaning in this context might be deceptive, but that would require some examination of words in their context.

Taking the above limitations and criticisms into account, LIWC might not be the best tool to use to profile online grooming language, especially in isolation. Studies may not make claims beyond the percentage and LIWC scores, such as the linguistic features of online groomer language without analysing it in detail. Drouin et al. (2017) claim even without Content Analysis their analysis gives ‘vital insights into the linguistic features of chat transcripts of child predator sex stings’ (Drouin et al., 2017:453), which is inaccurate, as they do not analyse the LIWC output beyond percentages and do not describe the linguistic features of online grooming in detail.

LIWC might not be the best program to use in isolation. Gonçalves et al. (2014) tested eight sentiment measuring methods (Emoticons, LIWC, SentiStrength, SentiWordNet, SenticNet, SASA, Happiness Index and PANAS-t) using a large database of tweets and messages coded by people and focussing on six major topics to find out which method was the best at detecting polarity. The study compared these methods and tested the coverage of the tweets the methods accurately/inaccurately captured and agreement between the individual methods (Gonçalves et al., 2014). It found that LIWC has a coverage of between 45% and 72% and did not perform better

⁹ In Pennebaker et al. (2015a) *love* is an example of both the positive emotion category and the sexual category. The sexual category further includes *incest* and has 131 words/word stems in general, which are not outlined further.

than other methods. When combining different methods, a coverage of 95% was reached. This suggests that LIWC should not be used without other methods, as it is not very reliable when it comes to sentiment analysis. However, Gonçalves et al. (2014) used LIWC2007, so these findings might not apply to the latest version of LIWC, both used by Drouin et al. (2017) and Lorenzo-Dus and Kinzel (2019). Additionally, Wang et al. (2014) pointed out that LIWC was not created to analyse Computer-Mediated Communication, such as Facebook status updates, or chat log interactions:

An additional consideration is that the LIWC 2007 dictionary was designed for traditional forms of written language, which may be unsuitable for the language used in the on-line environment. For example, it does not contain smileys (e.g. =)), abbreviations (e.g. LOL), or fashionably misspelled words (such as sx‘BOOOORED!!!!!!!!!!!!’ or ‘‘H-A-P-P-Y’’) (Wang, 2014: 489).

However, Wang et al. (2014) also used LIWC2007 to analyse sentiments in Facebook status updates. LIWC2015 includes some of these features, as outlined in the LIWC2015 manual (Pennebaker et al., 2015b). Netspeak is a new language dimension that was added, which includes words that are used frequently in social media and basic punctuation-based emoticons. The manual does not explain how these frequent words were derived (Pennebaker et al., 2015b). Drouin et al. (2017) mention utilising a spelling standardisation procedure to correct ‘netspeak’ (Drouin et al., 2017: 443) and misspellings. However, the study does not report how emoticons were handled. How emoticons were handled in this thesis is covered in the data extraction subsection 3.3.1. As this discussion has shown, LIWC is not the best tool to use in isolation and to analyse online grooming language, as it was not developed for such purposes.

3.1.3 Data Sources

One aspect that almost all linguistic research into online grooming (except O’Connell, 2003, Quayle & Newman, 2017; Kloess et al., 2017a, 2017b and Chiang & Grant, 2018) to date has in common is that the data is taken from the PJ archive (see Table 3, section 3.1.1), but there are two other sources of online grooming language, namely reports and chat logs taken from Cybertip.ca (Quayle & Newman, 2017) and groomer-child chat logs and police reports obtained from police forces in the UK (Kloess et al.,

2017a, 2017b; Chiang & Grant, 2018). The PJ data is labelled groomer-decoy data, and the other data is labelled groomer-child data. The three data sources will be described in more detail below.

3.1.3.1 Groomer-Decoy Data (Perverted Justice Archive)


As outlined in chapter one (section 1.1), the Perverted Justice Foundation was a US based Foundation made up of volunteers who pretended to be children, entering chat rooms and other online spaces (e.g., Yahoo Instant Messenger, AOL Instant Messenger, gay.com, meetme.com) and waiting for online groomers to start a conversation with them. If the conversation turned sexual, they collaborated with law enforcement to convict the groomers. The resulting record of the whole conversation between a groomer and a PJ volunteer (chat log) was only uploaded to the archive if the groomer was successfully convicted. PJ volunteers also added comments to the chat logs. The Foundation stopped operating in 2019 (see chapter one, section 1.1). An example of a PJ archive chat log can be seen in Figure 9.

Perverted-Justice.com Archives

Cody had to schedule his meet to molest a child until after his meeting with his probation officer.
Bust by [Retired Contributor](#) @ 10/1/2013 12:55 AM PST

[Conviction]

Perverted Justice mark: Cody ██████████ 20
Yahoo IM: perfect_buddy_ga [[perfect_buddy_ga@yahoo.com](#)]
Location: Alpharetta, Georgia
[\[send perfect_buddy_ga a message \]](#)



This wannabe pedo tried to solicit maddieguri92, a 13 year old girl
... or so they thought!

Currently reading with the contributor's comments (in blue).
[Turn commentary off](#)

maddieguri92 (07/19/06 10:28:39 PM): do i know u?
perfect_buddy_ga (07/19/06 10:28:52 PM): I dont know
perfect_buddy_ga (07/19/06 10:28:58 PM): your name sounds real familiar
perfect_buddy_ga (07/19/06 10:29:02 PM): how old are you?
maddieguri92 (07/19/06 10:29:22 PM): 13
maddieguri92 (07/19/06 10:29:32 PM): how old r u?
perfect_buddy_ga (07/19/06 10:29:39 PM): Oh then probably not
perfect_buddy_ga (07/19/06 10:29:46 PM): 20
perfect_buddy_ga (07/19/06 10:29:53 PM): too old right?
maddieguri92 (07/19/06 10:30:17 PM): cool ur cute

Conviction Counter
Number of predators convicted due to Perverted-Justice.com since June 2004:
623
Chatlog Convictions: 622
Research Convictions: 1

Real Stories Project
For the next six months, we're documenting actual cases of internet abductions and molestations. One case a day for at least six months. Beware, these stories can be disturbing.

Updates from around the Foundation

Site News
08/22/09: Michigan October 2008 sting - 21 arrested


PeeJ Most Wanted


Figure 9: PJ archive example chat log

The groomer's name in this example was Cody and his age (20), email address or Yahoo IM (perfect_buddy_ga)¹⁰ and location (Alpharetta, Georgia) were displayed next to his photo. Other metadata included the name of the PJ volunteer (a retired contributor in this case, usually there was a hyperlink with all the other convictions from data by the same contributor embedded in their name), the date and time the chat log was added to the archive (1 October 2013 at 12:55 AM PST) followed by the actual conversation between the groomer (perfect_buddy_ga) and the target (maddigurl92). The extract of the chat log took place between the 19 July 2006 at 10:28PM and ended at 10:30AM on the same day. The PJ volunteer commentary can also be turned off so only the chat log remains.

3.1.3.2 Groomer-Child Data

The other, much rarer data source is groomer-child data shared by the police. This can be difficult to obtain, store, anonymise and analyse. There are a number of privacy and sensitivity data issues, which will be discussed in more detail among other ethical considerations when working with groomer-child and groomer-decoy data in subsections 3.1.4.1 and 3.1.4.2.

Quayle and Newman (2016) conducted a study on reports of online grooming submitted to Cybertip.ca and they found resistance to be among the strategies children used frequently to stop the groomer:

About one third of of [sic] young people (32.4 %) showed resistance to the contacts by suspects and in spite of a lot of sexually explicit and persistent behaviour managed to terminate the contact. Sometimes this was by disclosing what had happened, but also through simple strategies such as blocking or deleting the suspect. (Quayle & Newman, 2016: 10)

This would suggest that decoy language might not be an accurate representation of the victim's part of the online grooming process. Chiang and Grant (2017) also pointed out that there is no existing literature on decoy language and the Perverted Justice did not provide any information about how decoys were trained (Chiang & Grant, 2017).

¹⁰ This chapter will be the only time groomers' usernames are referred to in this thesis, for illustration purposes and to show what kind of metadata the PJ archive made publicly available.

However, the study noted that it is in the decoy's interest to continue the conversation, which might lead to different behavioural strategies than real grooming victims would use. Winters et al. (2017) suggested that the decoys might even steer the conversation into specific directions to gather more evidence to successfully charge an offender (Winters et al., 2017). Winters et al. (2017) called for more research into the decoys' language. This suggests that there are some differences between decoy language and real online grooming survivors' language. Some researchers also called for research to use groomer-child data instead of groomer-decoy data, as talking to a decoy pretending to be underage might create 'different interactional patterns with these online offenders' (Chiang & Grant, 2018: 19), which might not translate to the groomer having to overcome resistance, distrust or rejection (Chiang & Grant, 2018). Chiang and Grant (2018) identified two new moves, overt persuasion, and extortion, using groomer-child data, which other studies did not identify using PJ data. They concluded their findings 'perhaps undermine[d] the continued use of Perverted Justice data as good proxy data for research into genuine online CSA¹¹ conversations' (Chiang & Grant, 2018: 19). In a research note responding to this article, Schneevogt et al. (2018) did find similar patterns in the overt persuasion move in the PJ archive, albeit fewer examples. The study acknowledged that 'PJ data may still be useful for asking some important questions.' (p. 101). This was echoed by Lorenzo-Dus et al. (2020), which stated:

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. (Lorenzo-Dus et al., 2020: 19)

This is also the approach taken in this study. The research aims do not involve analysing the target language and the under-researched areas that will be analysed (duration of online grooming, groomers' requesting behaviour) have not been examined before.

¹¹ Child Sexual Abuse

3.1.4 Ethical Considerations

3.1.4.1 Ethical Considerations Using Groomer-Child Data

As mentioned, obtaining groomer-child data comes with some ethical considerations, such as data anonymisation, storage and using the data for publication. Grant and McLeod (2020) described their data collection of groomer-child data to train undercover police officers as follows:

The genuine chat logs have been provided to us through a UK police force and are subject to a data sharing agreement, the provisions of which focus mostly on the security of the data and the protection of policing tactics used in this domain. (Grant & McLeod, 2020: 52).

Chiang and Grant (2018) also provided some details about data sharing agreements, stating that ‘This agreement provides for the secure storage of data using encrypted devices, anonymization of all transcripts, and for the psychological support of researchers.’ (Chiang & Grant, 2018: 8). The other study using groomer-child data, Kloess et al. (2017b) mentioned that all material had to be anonymised by police officers and added a further security step saying, ‘the principal researcher received vetting clearance to undertake research activities as part of the Child Exploitation Investigation Team at a U.K. police force.’ (Kloess et al., 2017b: 623).

Potential harm to victims of online grooming were also considered in Grant and McLeod’s (2020) study and the authors decided not to approach the survivors of online grooming to ask for consent to use the data they received from a UK police force, as ‘any such approach contained its own ethical risks in terms of potential further harm to those victims’ (Grant & McLeod, 2020: 53). Instead, the data was anonymised to the point where not just place names and other identifiers were redacted but quotations were also checked for identifying information. The study also granted suspects who had not been convicted of crimes anonymity and only named convicted individuals if there was a prominent media presence of their crimes. Undercover police officers were asked for their consent to use their data and it was also anonymised.

In terms of potential researcher harm, Grant and McLeod (2020) reported they provided psychological support via police resources. The study also discussed whether helping undercover police officers was ethical, saying that ‘[u]ndercover policing is

necessarily deceptive and deception might be considered to create harm both against the deceived individual and against a wider idea of a civil society through a general erosion of trust.’ (Grant & McLeod, 2020: 55). The study argued that because this type of deception prevents serious harms like sexual abuse of children, it should be permissible to help undercover police officers and provide linguistic training to them to improve their performance.

3.1.4.2 Ethical Considerations Using Groomer-Decoy Data

In terms of groomer-decoy data, because the PJ data is in the public domain and online groomers have been convicted, most previous studies did not mention anonymising data (see Black et al., 2015; Egan et al., 2011; Williams et al., 2013; Winters et al., 2017). In fact, Williams et al. (2013) used the groomer usernames in its examples. Chiang and Grant (2017) justified their decisions to not include usernames as follows:

All transcripts are publicly available on the P[]J website. Where phone numbers or addresses appear in transcripts, these are removed by P[]J and replaced with “*edit number/address*”. Thus the acquisition or use of this data might generally be considered acceptable; even so, all transcripts were further anonymised in that all participants’ formerly displayed usernames were removed and replaced with O1, O2, etc. for offenders, and D1, D2, etc. for decoys. (Chiang & Grant, 2017: 109)

This trend could also be observed in Lorenzo-Dus et al. (2016) and Lorenzo-Dus et al. (2017), which did not specifically mention anonymisation but included examples without usernames, replacing them with OG for online groomer and T for target. This is also the approach taken in this study, using G for groomers and T for targets. The language of these decoys, or targets as they will be referred to throughout this thesis with the understanding that groomers were under the impression they were talking to an underage person, was not analysed in detail. It was not removed to be able to preserve the conversational structures and turn-taking, but a sub-corpus for the groomer language was created for some of the analysis, which will be explained in more detail in sub-section 3.3.4.

In terms of researcher safety/wellbeing, an ethics review for this thesis was sought and potential effects on mental health due to the sensitive and disturbing nature of the data was pointed out. This was further addressed by explaining that, although the research

could potentially cause psychological stress and anxiety, the researcher was made aware of the wellbeing services at the university and knew how to contact them to alleviate the possible stress and anxiety. The researcher also had previous experience with sensitive data, which did not cause psychological stress and was not deemed a primary concern. Ethical approval was secured in December 2017. Throughout the study, the researcher received group training along with other researchers working with distressing data in March 2020 by an accredited psychotherapist whose specialities include trauma related issues, especially connected to child sex abuse/sex abuse, trafficking, and exploitation. The researcher also attended an online workshop ‘Understanding Vicarious Trauma and its Path to Resilience’ delivered by The Chicago School of Professional Psychology in April 2020 to develop resilience and counteract a potential effect on their mental health and wellbeing. Furthermore, the researcher also attended stress control sessions in August 2018, June 2019 and January 2021 provided by the university’s wellbeing services to deal with general stress related to doing a postgraduate research degree, dealing with a heavy workload and stress related to conferences and presentations and completing a postgraduate research degree during a pandemic.

3.2 Framework: Corpus-Assisted Discourse Studies (CADS)

After outlining the data size and source of this study and previous research into online grooming language research and exploring ethical considerations, the next section will describe the framework used, CADS.

3.2.1 What is CADS?

A Corpus-Assisted-Discourse Studies (CADS) approach is ‘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: 88). It makes use of quantitative Corpus Linguistics tools to narrow a large corpus down to a subset of data which is worth analysing, e.g., based on a keyword or collocation analysis (see sub-section 3.2.2.1 and 3.2.2.2) using frameworks from Discourse Analysis (Baker et al., 2008). These frameworks are

qualitative in nature, which gives the researcher the opportunity to examine concordance lines closely to observe the phenomenon (Baker 2006, 2015). The aim of CADS is ‘the uncovering, in the discourse type under study, of what we might call non-obvious meaning, that is, meaning which might not be readily available to naked-eye perusal.’ (Partington, 2008: 3). CADS does so by letting the researcher familiarise themselves with the discourse type under investigation and interacting with it, rather than looking at the corpus as a whole (Partington, 2008). As Partington (2008), one of the pioneers of CADS, put it:

‘This search is performed by combining the corpus techniques of statistical overview – frequency and key-item listing, along with the rapid search facility generally known as concordancing – with the traditional virtues of discourse analysis, that is, close-reading backed up with introspection.’ (Partington, 2008: 208).

The approach has also been praised for preventing “cherry-picking” data to confirm or reject a given hypothesis (Baker & Levon, 2015) and instead letting the data drive the analysis and using an inductive approach (Gabrielatos & Baker, 2008), which Hardt-Mautner (1995) described as follows: ‘concordancing effectively heralds a breaking down of the quantitative/qualitative distinction, providing as it does the basis for quantitative analysis without ‘deverbalising’ the data, that is, without transferring it, through human intervention, to the numerical mode.’ (Hardt-Mautner, 1995:24). A CADS approach can also be used deductively to test out theories using a reference corpus (Gabrielatos & Baker, 2008). Corpus Linguistics makes diachronic studies and synchronic ones possible. Diachronic studies use several corpora spanning years or decades that are compared to search for evidence of language change, whereas synchronic studies show a snapshot of language at a certain point in time (Partington et al., 2013). Other studies use a reference corpus, a large collection of a particular language type, such as general English language use to compare a study corpus to (Baker, 2009).

Although a CADS approach reduces researcher bias, such as “cherry-picking”, it does not eliminate it. Baker (2006) stressed that the researcher needs to reflect on how involved he or she is in their research. Baker et al. (2008) called CADS approaches a

‘useful methodological synergy’ (Baker et al., 2008: 273) and concluded that ‘CL¹², in general, and concordance analysis, in particular, can be positively influenced by exposure and familiarity with CDA¹³ analytical techniques, and the theoretical notions and categories of DHA¹⁴ can inform the quantitative CL analysis’ (Baker et al., 2008: 297). Partington (2004) highlighted that ‘detailed statistical analysis’ is one of the advantages Corpus Linguistics offers to Discourse Analysis via a CADS approach (Partington, 2004: 10). More recently, Mautner (2019) argued that CADS has come a long way since its inception 20 years ago and has now entered the mainstream, ‘for a particular kind of project — typically located at the language and society interface and using large datasets — it is now arguably the ‘go-to’ approach’ (Mautner, 2019: 2). However, the study pointed out five key challenges CADS researchers face while designing their research project (Mautner, 2019: 5). These were:

1. Finding the right fit between the research question(s), data, method, and the underlying theory.
2. Preventing analytical tools from dictating the research process.
3. Maintaining the distinction between quantitative and qualitative tools.
4. Ensuring that empirical claims are commensurate with the representativeness of the data.
5. Avoiding the misinterpretation of findings.

All of these have to work together and be matched properly, otherwise a project fails. The study also highlighted that a CADS piece of research has to consider how the two areas are combined, ‘[o]therwise, two sets of apparently unrelated results are simply placed side by side, with links between them asserted rather than demonstrated.’ (Mautner, 2019: 8). This is something that will be further discussed in chapter seven (section 7.5). Mautner (2019) also pointed out that CADS researchers are ‘at the mercy of the tools that the original software developers have devised, including a raft of background computational operations that most discourse analysts are likely to find unfathomable’ (Mautner, 2019: 7). Concordance software will be discussed in further detail in sub-section 3.2.3. Overall, CADS is a useful methodology that needs to be carefully thought about and designed to work effectively. Next, the CADS toolkit

¹² Corpus Linguistics

¹³ Critical Discourse Analysis

¹⁴ Discourse Historical Approach

will be examined and statistical measures for keyword and collocation analyses in this study will be explored.

3.2.2 CADS Toolkit

The following section will outline the “standard” CADS toolkit, which is also used in this study. It will first give an overview of keywords as a way into the data and will then compare different measures to calculate keyness, in sub-section 3.2.2.1. Sub-section 3.2.2.2 will look at collocate calculation. The next sub-section 3.2.2.3 will cover the specific theoretical and analytic concepts taken from Discourse Analysis, namely Speech Act Theory and Politeness Theory.

3.2.2.1 Corpus Linguistics: Keywords and Keyness Measures

A CADS approach usually starts with a keyword or collocation analysis (see section 3.2.2.2) that helps to narrow the corpus down to some words of interest. Keywords are defined as words in a given corpus that, compared to another corpus, occur ‘with unusual frequency’ (Scott, 1997: 236), indicate what the text is about. Keywords can be calculated in different ways. The most traditionally used keyword measure is Log Likelihood (LL) (Dunning, 1993), which entails calculating the statistical significance of word frequencies compared to word frequencies in a reference corpus, measuring whether the difference is based on chance or a statistical difference is present. This measure was regarded as ‘the standard measure’ (Lorenzo-Dus & Di Cristofaro, 2016: 44) up until 2015 and has been used across many studies using Corpus Linguistics (Lorenzo-Dus & Di Cristofaro, 2016; Rayson, 2015; Durán-Muños, 2019).

However, more recently some scholars argue that Log Likelihood is not appropriate to calculate keywords, or at least not in isolation, as it measures statistical significance, in other words, whether a difference between corpus A and corpus B exists, but not how big that difference is (Hardie, 2014; Gabrielatos, 2018). To measure this difference, corpus linguists have to consider effect size, which measures the magnitude of the difference, alongside the significance of the difference (Log Likelihood). Proposed effect size measures include ratio (Kilgarriff, 2009), odds ratio (Everitt, 2002), Log Ratio (LR) (Hardie, 2014), %DIFF (Gabrielatos & Marchi, 2011)

and difference coefficient (Hofland & Johansson 1982). Gabrielatos (2018) summarised these metrics efficiently.

Ratio examines the normalised frequencies of the items in the corpora under investigation. A 'value of '1' indicates that the item has equal normalised frequency in the two corpora, which higher/lower values indicating higher/lower NF¹⁵ in C1' (Gabrielatos, 2018: 235). Odds ratio works on raw frequencies and very similarly to ratio (Gabrielatos, 2018). Log ratio is 'the binary logarithm of the ratio of normalised frequencies' (Gabrielatos, 2018: 236). A value of 0 indicates a comparable frequency in both corpora, while a value of 1 indicates that the item is twice as frequent in one corpus. A Log Ratio value can also be negative, indicating the item is more frequent in the reference corpus than the study corpus (Gabrielatos, 2018). %DIFF also works on normalised frequencies (NF) with a value of 0 indicating equal frequencies in the corpora. '[E]very increase of '100' adds one to the difference – for example a value of '500' indicates six times higher frequency' (Gabrielatos, 2018: 236). The difference coefficient also works with normalised frequencies with scores ranging from 1 to -1. A value of '1' indicates that the item only exists in C1 (...) '0' indicates that the item has the same normalised frequency in the two corpora (...) '1' indicates that the item only exists in C2' (Gabrielatos, 2018: 236).

There is no consensus among corpus linguists regarding which of these measures is the most appropriate for keyword analyses. As Brezina (2018) pointed out '[c]urrently, the question of which statistic best suits the identification of keywords is an open one.' For this study, Log Ratio with a Log Likelihood filter was chosen to make sure both statistical significance (Log likelihood filter) and effect size (Log Ratio) are considered. Log Ratio is a term coined by Hardie (2014) who developed the measure. It is calculated by dividing the relative frequency of a word in the study corpus by the relative frequency of a word in the reference corpus. The resulting score is then turned into a binary logarithm. Hardie (2014) explained that a Log Ratio score of 0 means that the word has a comparable frequency in both corpora. A score of close to or 1 means the word is twice as common in the study corpus compared with the reference corpus. A score of 2 means it is four times as frequent and so on (Hardie, 2014).

¹⁵ Normalised Frequency

Negative Log Ratio values indicate the same, in the reference corpus, rather than the study corpus. The Log Likelihood filter excludes words that do not have a high enough statistical significance.

While particularly high or low Log Ratio values are mentioned in studies, a Log Ratio cut-off point has not been universally agreed on. Lorenzo-Dus and Di Cristofaro (2016) highlighted a Log Ratio value of over 3, meaning a word being eight or more times as common, as being particularly interesting due to the high frequency compared to other semantic domains analysed, while Demjén (2016) specified a minimum value of 0.58. Sometimes, results are presented in Log Ratio increment steps of 1, for example, words with Log Ratio = 0, words with Log Ratio = 1, words with Log Ratio = 2 etc. (Alipour & Nooreddinmoosa, 2018; Durán-Muñoz, 2019). Bednarek (2018) chose a wider range of 0 to 1.99, 2 to 2.99 and 3 or more Log Ratio values to effectively filter out keywords to analyse more thoroughly: 'Just like range, frequency, or statistical significance, Log Ratios can be used for down-sampling, helping the analyst to choose which key words to focus on.' (Bednarek, 2018: 152). These key words were also called 'candidate key items' (p.238) by Gabrielatos (2018). These studies generally had one study corpus that was compared against a reference corpus, unlike this study, which has different analyses that feature two study sub-corpora at a time. The Log Ratio cut-off point was decided in the context of these comparative analyses and was thus different for each analysis (see chapter five, sections 5.2, 5.3 and 5.4).

Once keyword calculation has happened and a quantitative analysis informed by Corpus Linguistics has taken place, CADS adds another layer, which comes in the form of expanded concordance lines and close readings of the text to make sure the text is not decontextualized (Berger et al., 2017). A concordance line or a Keyword in Context (KWIC) is a single instance of the word or string of words that is being investigated in its context, normally 10 to 20 words on either side of the word or string of words under investigation. This is displayed in a Corpus Linguistics software, such as Antconc (Anthony, 2020), CQPWeb (Hardie, 2012) or WordSmith (Scott, 2020). These will be further discussed in section 3.2.3. The concordance line shows the word in context and the window of context words can be modified so more context is shown. The researcher can also select a specific concordance line and see more of the

interaction in context, which is called an extended or expanded concordance line. An illustration of this in CQPWeb can be seen in Figure 10. The keyword, in this example *test* is displayed centrally with the context to its right and left. The number of hits (n=136) can be seen at the top contained in the corpus, which is 3.6 million words large. The frequency per million words (37.73) is also displayed.

No	Filename	Solution 1 to 50	Page 1 / 3
1	v1sEIFIwqj43895412	so i get out early yay !! N:53 U:1 a:1 UT:1g go luck on your	N:54 U:1 b:1 UT:1d its all week its so boring N:55 U:1 a:1 UT:1g are you gon na tell your
2	DmLRa6QzGa62411850	did the scarlet letter N:79 U:1 a:1 UT:1g boring N:80 U:1 a:1 UT:1g so i would turn in my own	with the F orig: already already on it N:81 U:1 b:1 UT:1dorig: omg oh my god lol r
3	ZJLSh39z0u26702759	go to temple to the va hospital N:10 U:1 a:1 UT:1g good news all of my	were ok N:11 U:1 a:1 UT:1g u know how brighten sor N:12 U:1 b:1 UT:1d thats good
4	ZJLSh39z0u26702759	but y did u go r u sick ? N:13 U:1 a:1 UT:1g they did some	to make sure i am ok N:14 U:1 a:1 UT:1d _SAD-SMILEY_ \$ N:15 U:1 a:1 UT:1g y the _SAD-SMILEY_ \$
5	ZJLSh39z0u26702759	they can treat it sooner N:23 U:1 b:1 UT:1d k orig: gues guess so N:24 U:1 a:1 UT:1g trust me u want	done N:25 U:1 a:1 UT:1g u heard about the new gu
6	HbKbl_5fM1x22073450	my the end of class on tues of next week have a	tomorrow he said we would get out early too N:26 U:1 b:1 UT:1d did u study
7	44o1oKHLvr75067885	N:8 U:1 b:1 UT:1dorig: nothin nothing much how was school last night ? N:9 U:1 a:1 UT:1g good pass me written	he did not get the grads out yet N:10 U:1 b:1 UT:1d cool _HAPPY-SMILEY_ \$ N:11 U:1 a:1 UT:1g yep N:12 U:1 a:1 UT:1g
8	jx8oquaC5d42607880	N:13 U:1 a:1 UT:1g r on the hand on stuff N:14 U:1 b:1 UT:1d _HAPPY-SMILEY_ \$> N:15 U:1 a:1 UT:1g still have a final	book test N:16 U:1 a:1 UT:1g that is next thursday N:17 U:1 b:1 UT:1dorig: ull you'll do great or N:18 U:1 a:1 UT:1g it

Figure 10: Screenshot of a KWIC using the keyword "test" in GN corpus on CQPWeb

Duguid (2010) emphasised this need to examine concordance lines and says expanded concordance lines need to be examined to confirm or reject hypotheses. This also makes the quantitative/qualitative dichotomy somewhat ambiguous, as ‘qualitative’ findings can be quantified, and that ‘quantitative’ findings need to be interpreted in the light of existing theories, and lead to their adaptation, or the formulation of new ones.’ (Baker et al., 2008: 297). However, this is an advantage CADS has over an approach using Corpus Linguistics in isolation, as it can not only examine the context of the word in question, but consider wider social, political, and cultural contexts.

3.2.2.2 Corpus Linguistics: Collocations

Collocation calculation is another “standard” tool in Corpus Linguistics. This measures how words often appear in proximity to other words. As Rayson (2015) described it:

In practical terms, collocation as a method refers to the counting of the co-occurrence of two words in a corpus depending on their relative proximity to one another, and usually includes the calculation of a statistic or metric to assign significance values to the amount or type of co-occurrence relationships. (Rayson, 2015: 41-42)

There are different ways to calculate this statistic or metric and it can also be affected by the window span between node word (word that is being studied) and collocate (word that the node word co-occurs with) (Rayson, 2015). Other considerations include the word boundary and whether to include minimum frequencies for both node word and collocate. Rayson (2015) mentioned frequency, Mutual Information, Log-Likelihood, Z-score, T-score, MI2 or MI3 as options, which can also be combined. Mutual information is computed by ‘dividing the observed frequency of the co-occurring word in the defined span for the node word by the expected frequency of the co-occurring word in that span and then taking the logarithm of the result; (Xiao, 2015: 106).

T-scores are calculated ‘on the basis of the difference between the observed and expected means, scaled by the variance, to determine the probability of a particular sample of that mean and variance with the assumption of the normal distribution of the dataset’ (Xiao, 2015: 109). Z-scores is a metric that ‘adjusts for the general frequencies of the words involved in a potential collocation and shows how much more frequent the collocation of a word with the node word is than one would expect from their general frequencies’ (Xiao, 2015: 110). These tests and measures all have advantages and disadvantages. According to Jaworswa and Nada (2018), Mutual Information leans towards emphasising ‘low-frequency words’ (p. 387), while t-scores is the opposite and tends to favour ‘relatively high frequency’ (p. 387) words like function words.

LogDice is another collocation measure first used in SketchEngine, but which has since been made available to use in other Corpus Linguistics software (see section 3.2.3). LogDice ‘expresses collocation typicality based on the frequency of node (terror*)-collocate and the frequency of the collocation in the whole corpus (Kilgarriff et al., 2014).’ (Almaged, 2021: 3). McEnery, McGlashan and Love (2015) highlighted that LogDice ‘has been shown to work well with corpora of different sizes’ (McEnery et al., 2015: 241), which is an advantage. Baker and Levon (2015) described it as ‘an

effect size statistic which is a measure of strength of association between two words' (Baker & Levon, 2015: 226). Jaworska and Nada (2018), which also chose to use it, described it as follows:

LogDice, which is based on the Dice coefficient, can be positioned in the middle, as it combines the relative frequency of the relation X (headword) + Y (collocate) with frequencies of X in the same syntactic position and with any collocate, and Y in any syntactic position (Rychlý 2008) (Jaworska & Nada, 2018: 387)

The study went on to say that '[s]ome researchers see LogDice as the best method of determining collocations (cf. Baker 2014)' (Jaworska & Nada, 2018: 387). This is also the viewpoint adopted in this study and dice coefficient was used to calculate collocates (see section 3.3.5.1 for further detail).

3.2.2.3 Discourse Analysis: Speech Acts and Im/Politeness

Once the Corpus Linguistics tools above are used and the corpus is narrowed down from millions of words to a list of keywords with collocates, these keywords can be examined in context using concordance lines (see section 3.2.2.1). These concordance lines need to be examined manually to add a more qualitatively analysis to them and understand them from a discursive perspective. Discourse Analysis is informed by many different academic disciplines like Anthropology and Linguistics, Communication, Psychology, Philosophy, Literary Criticism, and Artificial Intelligence (Schiffrin et al., 2015). It can be defined as the 'study of language in use' (Schiffrin et al., 2015: 1) and has a broad toolkit. As Schiffrin et al. (2015) put it:

Our own experiences in the field have led us to the conviction that the vastness and diversity of discourse analysis is a strength rather than a weakness. (...) we find the theoretical and methodological diversity of discourse analysis to be an asset (Schiffrin et al., 2015: 5).

In the context of online groomer language and the research questions outlined in section 1.2 and specifically analysing interactions and communicative intentions, it made sense to work with the notion of speech acts, more specifically the speech act of requesting. Groomers regularly request actions or information from their target, corresponding to their communicative intentions (e.g., engage in sexual discussions,

meet face-to face, send a photograph). The ulterior motive is to sexually abuse the target, which takes manipulation. (Further discussed in chapter six, section 6.).

Speech Act Theory (SAT) originated in the field of Philosophy. In his book *How to do thing with words* Austin (1975) stated that by saying specific sentences the interlocutor performs an action. Austin (1975) also stated speech acts take one of three forms: locutionary act, illocutionary act or perlocutionary act (Austin, 1975). Searle (1969) built on this work and introduced more conditions to what constitutes a speech act. In the wider field of Pragmatics, this has been further developed and reached popularity in the 1980s and 1990s. Speech acts were utterance based and sentences were examined in isolation. A more recent interpretation of speech acts sees them as context-dependent and not bound by sentences (Taylor, 2015; Aijmer, 2015, 2019; Webman Shafran, 2019; Van Olmen, 2019; Frediani, 2019). Speech acts are also not necessarily classified into Austin's three types but include speech acts such as requesting, apologising, complimenting, and thanking. They are still used in recent research in Pragmatics and Discourse Analysis but not necessarily by referring to Speech Act Theory (see for example Webman Shafran, 2019; Aijmer, 2019; Van Olmen, 2019), but also referring to Blum-Kulka's (1984, 2011) request act type taxonomy and contextualising their research in Politeness and Impoliteness Theory terms, analysing politeness markers, for example.

Speech acts in this study are not considered to be utterance based but are context dependent. The analysis takes a Discourse Analysis approach, in which the whole request sequence between groomer and target, which includes up to 23 interactional turns is analysed, as we shall see in chapter six (see example request sequence 3.1 below). The example request sequence 3.1 involves 14 turns between a groomer (g) and their target (t) and the request can be found in turn eight (*you don't have to come*). This was discarded, for the actual analysis, as the request appears in the target turn, not the groomer turn. In this case, the target responses and reactions to the groomer's discourse are used to understand the context (a face-to-face meeting between the groomer and their target and potential repercussions) but are not analysed beyond this.

Example 3.1

Turn	Speaker	Message
1	g	never have done this before
2	t	lol me either
3	g	i 'm scared shitless ...
4	g	do you understand ?
5	t	yes
6	g	if this is a setup i i'll never forgive you
7	t	aww its not a setup
8	t	you don't have to come
9	g	i just want you to know how serious it could be
10	g	if anything goes wrong
11	t	ok
12	g	i 'm a nice person ...
13	t	I believe you
14	t	§_HAPPY-SMILEY_§

A substantial amount of research on speech acts has been carried out in the areas of English as a Foreign Language (EFL), child language acquisition and cross-cultural Pragmatics. Different patterns of requests have been identified in different cultures (Webman Shafran, 2019). One influential project by Blum-Kulka and Ohlstein (1984), which was further developed by Blum-Kulka, House and Kasper (2011), used Discourse Completion Tests to investigate the realisation patterns in the speech acts of requesting and apologising by speakers from different cultural backgrounds. To do so, they developed a taxonomy to classify requests into head act types and support types (see section 3.3.5.2 and chapter six, section 6.1), paying attention to how directness and indirectness influence requests and apologies.

Combining this requesting framework and integrating it with quantitative Corpus Linguistic tools and analysis, which this study will do, is a methodological challenge. This is concerned with the connection between language form and function and the fact that only speech acts that have regular patterns or formulae can be searched for in corpora and even then, not all formulae and instances will likely be found in the corpus. As Aijmer (2019) put it:

It is not self-evident how we can use corpora and corpus-linguistic methods to study pragmatic phenomena such as speech acts. Speech acts depend on context for their interpretation and they do not have a specific form. They are

closely associated with spoken language and the use of language for interaction. Corpus linguistics, on the other hand, presupposes that we can identify a relationship between form and function. (Aijmer, 2019: 259)

Jucker et al. (2008) stated that two approaches can be taken to the analysis of speech acts using corpora, in their case compliments. The corpus can be searched for the speech act verb, or it can be searched for ‘syntactic patterns or lexical elements that are typical for actual compliments’ (Jucker et al., 2008: 1616). Since only a very limited number of corpus-based research has been done on directives (which requests are part of) and this research usually involved pragmatically tagged corpora (see Flöck & Geluykens, 2015; John, Brooks, & Schrieffer, 2019; Culpeper & Tantucci, 2021), the second approach was taken in this study. Pragmatics and Discourse Analysis studies were reviewed, and search queries that expressed head acts in requests were extracted and mapped onto Blum-Kulka et al.’s (2011) head act type taxonomy, which will be further outlined in section 3.3.5.2 and chapter six (sub-section 6.1).

Requests can be face-threatening acts because they ‘can threaten the addressee’s freedom of action and may also threaten the speaker’s face if the addressee is reluctant to comply with the request’ (Webman Shafran, 2019), but this is context dependent. Brown and Levinson (1987) regarded some acts as inherently face-threatening. These were defined by Brown and Levinson (1987) as ‘certain kinds of acts intrinsically threaten face, namely those acts that by their nature run contrary to the face wants of the addressee and/or of the speaker’ (Brown & Levinson, 1987: 313). A few impoliteness and politeness frameworks have been developed based on Brown and Levinson’s theory (1987) of politeness. In this theory, Brown and Levinson distinguished between negative and positive face, based on the original concept of face by Goffman (1967). They defined negative face as ‘the want of every ‘competent adult member’ that his actions be unimpeded by others’ (p. 312) and positive face as ‘the want of every member that his wants be desirable to at least some others’ (p.312). Face threatening acts were then further divided into positive face threatening acts: ‘those acts that primarily threaten the addressee’s (H’s) negative-face want, by indicating (potentially) that the speaker (S) does not intend to avoid impeding H’s freedom of action’ (Brown & Levinson, 1987: 313), which included disapproval, criticism, contempt, ridicule, complaints, reprimands, accusations, insults, expression of violent

emotions, mention of taboo topics, including those that are inappropriate in the context, among others.

Negative face threatening acts were defined as ‘those acts that threaten the positive-face want, by indicating (potentially) that the speakers does not care about the addressee’s feelings, wants, etc. – that in some important respect he doesn’t want H’s wants’ (Brown & Levinson, 1987: 314). These included orders, requests, suggestions, advice, reminders, threats, warnings, and dares, among others. Brown and Levinson (1987) also developed notions of positive and negative politeness, saying:

Positive politeness is oriented toward the positive face of H, the positive self-image that he claims for himself. Positive politeness is approach based; it ‘anoints’ the face of the addressee by indicating that in some respects, S wants H’s wants (...). Negative politeness, on the other hand, is oriented mainly toward partially satisfying (redressing) H’s negative face, his basic want to maintain claims of territory and self-determination. Negative politeness, thus, is essentially avoidance based, and realizations of negative-politeness strategies consists in assurances that the speaker recognizes and respects the addressee’s negative face wants and will not (or will only minimally) interfere with the addressee’s freedom of action (Brown & Levinson, 1987: 317).

As seen above, Brown and Levinson (1987) saw requests as an inherent face-threat to the hearer’s negative face.

The analysis of this thesis focuses on face-related aspects of requesting behaviour by groomers in the PJ data collected in this study¹⁶. The context of this data concerns illegal and immoral behaviour by groomers since the ulterior motive of online groomers is to abuse children sexually. As such, groomers will bring up illegal topics, among other things. However, depending on what online grooming intention the requests are aligned to (e.g., request to meet offline for sex, request to provide personal information, request to use explicit sexual language etc.) and other contextual factors, the groomer may believe them to be more or less face threatening. Groomers will use different forms of face redress (politeness), or issue explicit, unmitigated face-threatening requests.

¹⁶ This methodology section cannot go into detail about literature debates about politeness and impoliteness theories, how they relate to the notion of face or critiques of them. (See for example Wilson, Kim & Meischke, 1991; Mao 1993; Eelen 2002; Watts, 2010; Haugh & Watanabe, 2017; Haugh & Culpeper, 2018).

Therefore, request concordance lines will be analysed using a politeness and impoliteness framework. This Politeness and impoliteness framework has also been used across empirical chapters (four, five and six) to understand how groomers use politeness and impoliteness in their language to contribute to their grooming intentions. The specific impoliteness taxonomy used in this study is the one developed by Culpeper (2005), based on Brown and Levinson's politeness theory. This theory was revised in 2011 and the resulting taxonomy of impoliteness strategies (Culpeper 2011) does not directly mirror that of Brown and Levinson's positive and negative politeness strategies anymore. It is instead based on Spencer-Oatey's (2005, 2007, 2008) notion of face, which includes different facets, namely: quality face, identity face, relational face, equity rights, association rights. Culpeper (2011) added two categories to these existing ones, which were Taboo and Physical self. Culpeper's (2005) and (1996) taxonomies are still used in recent research (for example Dobs & Garcés-Conejos Blitvich, 2013; Perelmutter, 2018; Garrido Ardila, 2019, Khazraie, & Talebzadeh, 2020). A full overview of the politeness and impoliteness taxonomies including strategies referred to in the empirical chapters can be found in the Appendix (section 9).

3.2.3 Choosing the Right Corpus Linguistics Software: CQPWeb

Different Corpus Linguistics software has been developed and used extensively. Antconc developed by Anthony (2020) is freely available across platforms. Wmatrix was developed by Rayson (2008) and provides key word and key semantic tagging of corpora (Rason, 2008). It has a free one-month trial and licences cost £50 per year. Wordsmith was developed by Scott (Scott, 2020). It is not a freeware software and licences also cost £50. SketchEngine was developed by Kilgarriff et al. (2014). It is also not freeware and licences cost between 4.83€ and 8.33€ per month. The last Corpus Linguistics tool, CQPWeb, was developed by Hardie (2012), which is free to use but has a very complicated back-end development.

These different tools also have different capabilities. It was decided to use CQPWeb, for its ability to work efficiently with large corpora and the variety of up-to-date measures to calculate keyness and collocates that it affords. At the time of corpus creation, it was one of the only Corpus Linguistics software that had an option to

choose Log Ratio with a Log-Likelihood filter as the measure for calculating keywords and dice-coefficient as the collocate measure. The front end of it is also very user-friendly and intuitive. A screenshot of the user interface can be seen in Figure 11.

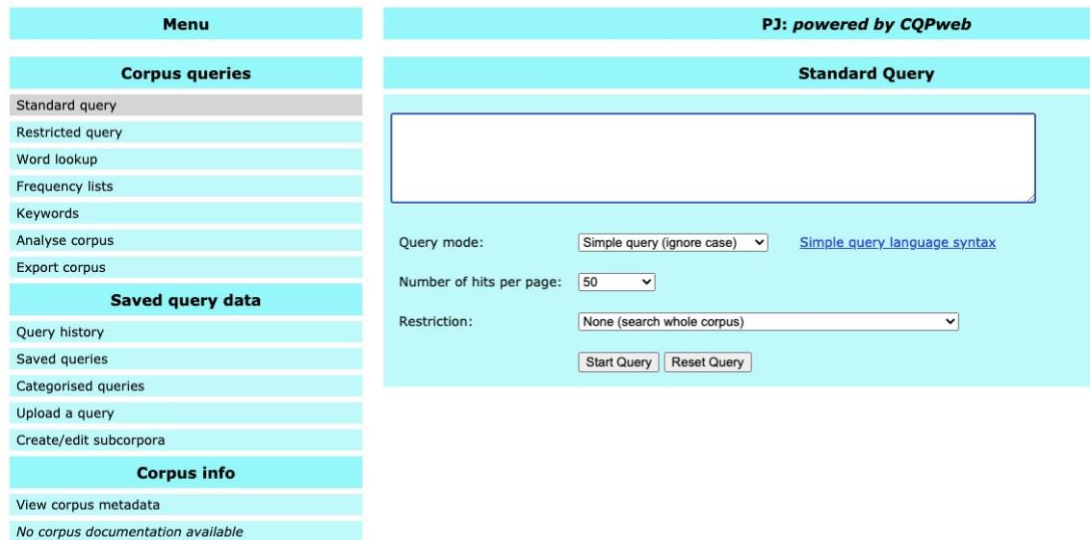


Figure 11: CQPWeb screenshot of the easy-to-use front-end interface

3.3 Procedure

The previous section has explained the framework (CADS) used in this study by outlining the components and “standard” toolkit. The following sub-sections will outline the steps involved in data extraction and pre-processing, leading to corpus creation. how the corpus was built from extracting the data (sub-section 3.3.1), normalising it (sub-sections 3.3.2 and 3.3.3) to creating the corpus (sub-section 3.3.4). It also briefly outlines how sub-corpora were created once the corpus was uploaded to CQPWeb. Corpus analysis in the empirical chapters will be outlined in section 3.3.5.

3.3.1 Data Extraction

The chat logs were hosted on the Perverted Justice website archive¹⁷ and were extracted using Python 3.4.3. Each chat log had its own web page with its own unique URL (Uniform Resource Locator). There was a stable part of this URL

¹⁷ <http://www.perverted-justice.com/?con=full>

“<http://www.perverted-justice.com/?archive=>” and a varying part, which is the groomer name, e.g., the URL of “the_donjuan02” was:

http://www.perverted-justice.com/?archive=the_donjuan02

The HTML (HyperText Markup Language) tag *div class post* captured the part of the web page that was relevant, i.e., only the chat log and additional information, such as metadata information, was captured in this tag and no other elements were included. A python function was created to extract this tag *div class post* with the help of the stable URL and usernames. A list of groomer usernames (*groomers_list*) was created. This function was the main part of an initial script that used BeautifulSoup to download the archived pages and saved them as HTML files with the extension *.post* to differentiate them from the future extracted data files. The script read the file *groomers_list* and extracted the content of the *div class post* tag.

Another script read the *.post* files created in step one and checked whether a chat log contained more than one conversation (text) in the HTML structure, which was indicated by more than one *codechat* tag. If it found more than one of these tags it saved each new conversation (text) in a new file. This *codechat* tag was not used consistently in the PJ archive, so sometimes all chat log sessions appeared under one *codechat* tag and could not be separated by the script. They were all saved into one file. The next step was to remove the HTML tags, which was done by writing a third script which used BeautifulSoup to remove all HTML tags in the *codechat* files created in the previous step. It is generally advised against using regular expressions to work on data. However, due to the inconsistent and chaotic nature of the database, it was decided to do that. The third script was also modified to delete the comments made by the PJ decoys, which were marked by the tag *code.c*. The tags as well as the content of the tags were deleted. Once this process was completed, the file was saved to a file with a different extension.

The chat logs followed very different formats, especially across platforms (e.g., Yahoo Instant Messenger, AOL Instant Messenger, gay.com, meetme.com and text messages). For example, in the chat logs using meetme.com there were either no time stamps or they appeared at the end of the line, rather than at the beginning, which was

the case for messages sent via Yahoo IM and AOL IM. AOL IM used square brackets and only showed the time in hours and minutes, while Yahoo IM used normal brackets and usually included hours, minutes, and seconds and sometimes date information in the time stamp. Initially, it was decided to focus on extracting the chat logs using Yahoo IM, as these were the most consistent chat logs across the database. They also made up most of the database (534 out of 621). Although the structure was quite consistent, there were still some inconsistencies, e.g., the time stamps either followed this format:

atonomous2000 (10:29:19 PM)

or the time stamps included the date:

donni1957_male (4/21/2006 6:03:36 PM):

It was decided to double check the groomer usernames using Yahoo IM, which resulted in a further reduction to 486 chat logs, after excluding problematic chat logs. As this was felt to be too restrictive, a fourth python script was written that captured the entire database, regardless of the Instant Messenger platform used. This resulted in 622 of the 623 web pages, as one of these was a research conviction and did not include a chat log. Modifications to the script were made to save the individual chat logs into separate folders and the regular expression to capture different formats of the time stamps was amended to include further variations.

The user type, differentiating between groomer (g), decoy (d) or na (not applicable due to unknown usernames), was determined by another Python script and based on a metadata table that was created and that included congregated information about all 622 chat logs (including groomer and decoy usernames, gender, start and end dates, total time chatting and platform). The total time chatted was also added to the chat log and the chat log was turned into XML (Extensible Markup Language) format which is readable by the software to be used for analysis: CQPWeb. Timestamps were not added to the final chat logs, as they proved too difficult to extract by the Python script. Three types of emoticons were transliterated into decryptions using the Python script: Happy-smiley, sad-smiley, kiss-smiley. Other emoticons were not considered based on too many variants existing. They were left in the data and were not transliterated. Additionally, the language of targets of online grooming, in this case decoys, was not

analysed. The focus is the language of the online groomers. Decoy language is only used to provide context for fine-grained qualitative analyses of interactions.

3.3.2 Data Normalisation Process

This section outlines the data normalisation process of the scraped PJ data described in the previous sub-section. It will first outline how the training data was created, how it was used to train the normalisation software and the initial stages of normalisation. Sub-section 3.3.3 will outline the second stage of data normalisation.

Corpus Linguistic software is based on statistics and processing and counting words, which means a corpus, especially one consisting of digital data, needs to be cleaned or normalised. This is the case for specific types of language that are more prone to variants, as Baron and Rayson (2009) pointed out:

Spelling variation is a feature in many corpora, particularly in historical corpora such as from the Early Modern English period but also in modern language varieties such as web-based texts. Whilst this orthographical variation is often of linguistic importance, abnormal spellings can have a detrimental effect on the accuracy of automatic corpus linguistic techniques. (Baron & Rayson, 2009: 21)

The normalisation process of the online grooming database was carried out using VARD 2.5.4, (Variant Detector) a software that was originally developed for normalising variants in Early Modern English spelling (Baron & Rayson, 2009). The software has been modified since its initial development (Rayson et al., 2005) to give the researcher more options to adapt it to their research and type of language. More recently, it has been used to normalise digital language, such as SMS (Tagg et al., 2012) and the Teenage Health Freak Corpus (Smith et al., 2014).

The initial 486 chat logs were extracted into an excel CSV (Comma-Separated Values) Excel file, as it was easier to read and edit. This Excel table had the following columns: turn number, username, date (m), date (d), date (y), time stamp, am/pm, text message. The script also checked whether the time stamp included the data, so whether it followed the prominent Yahoo IM date structure:

donni1957_male (4/21/2006 6:03:36 PM):

If that was not the case, the script checked if the chat log followed the other format:

atonomous2000 (10:29:19 PM)

The script then automatically filled in NA (not applicable) in the date column of these scripts, to be amended manually at a later stage. If the structure did not follow either of these structures, it was discarded. The last column of the CSV files, column [7] (in Python) was read by a python script and collected, so that only the text messages stored in different files were considered. This amounted to 860 CSV files from 486 groomers who used Yahoo IM. The extracted messages amounted to 750,485 messages. It was decided to take a third of these messages to train VARD 2.5.4, so 250,000 random messages were extracted and manually processed. Given that the PJ archive data was collected in the USA and after examining VARD's default dictionary and noticing it did not include American English spellings, e.g., *colour* was not marked as a variant, but *color* was, an American dictionary (Hunspell American Dictionary) that is recommended by VARD's developer was located and downloaded, so that the base dictionary included British English and American English spellings, and these were not marked as variants and highlighted. Twitter normalisation lists¹⁸ were also included in the training information in VARD, so that VARD suggested alternative to variants, such as *kool*, which was changed to *cool*.

VARD highlighted words it did not recognise and gave the user the option to tag them as a variant of a list of words it suggests, e.g., for the misspelled word *acn* it suggested *acne*, *can*, *ago*, *aachen* and *accent* (see Figure 12).

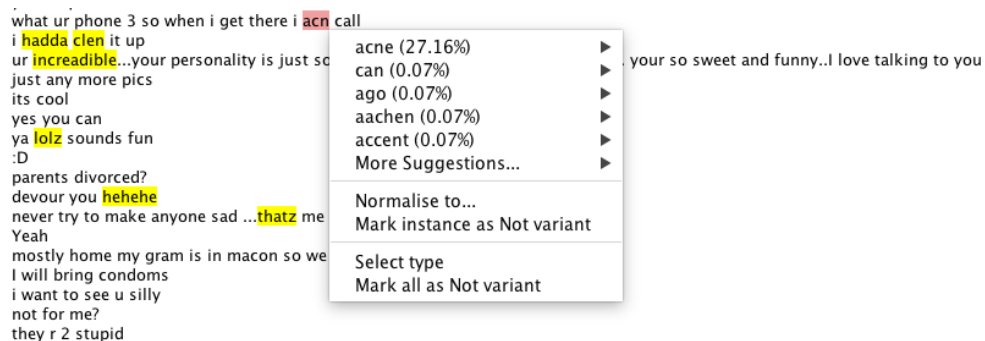


Figure 12: VARD screenshot of *acn* suggestions

¹⁸ <https://noisy-text.github.io/norm-shared-task.html>

As this word had not been encountered before, the most likely suggestion, *acne*, was incorrect. If the user selected *can*, this selection increased the percentage of *can* and the next time this word was encountered, *can* had a higher percentage.

The user could also introduce their own variants, by selecting “normalise to” and typing in their correction. If this correction was not already contained in its dictionary, VARD asked whether it should be added to the dictionary, further influencing the displayed percentages. The user could also normalise all instances based on their selection, which changed each instance. This could also be undone if a mistake had been made. Words that were not highlighted could also be added as variants and were then highlighted and suggestions provided. This was frequently the case with *u* (you), *b* (be), *r* (are/or) and *y* (why), which were not flagged up by VARD as variants of a word but appeared very frequently in the training data. In Figure 13, the *r* was highlighted and the option “mark instance as variant” was selected. The number 2 also appeared in the training data, which either meant *to* or *too* depending on the context.

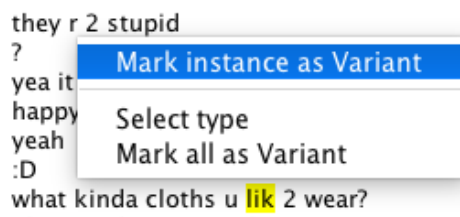


Figure 13: VARD screenshot of example word that is added as a variant

The training files were used to further train VARD, which then got better at recognising which words needed normalising. In this first step of training VARD, three types of words were encountered: very frequent ones with over 20 occurrences, infrequent words with less than 20 occurrences but more than 10 occurrences and very infrequent words with one to five occurrences. The very frequent words were normalised, which consisted mostly of acronyms and paralinguistic features, such as laughter. The infrequent words were most likely to be caught by VARD but needed some manual annotation. The very infrequent words were left unaltered, as they amounted to around 40,000 tokens in total, which were negligible in the corpus that is

around 3.6 million words large. Decisions were made to normalise acronyms and the following are some examples:

idk	→ I don't know
omg	→ oh my god
brb	→ be right back
gf	→ girlfriend
sry	→ sorry
ttyl	→ talk to you later
yw	→ you're welcome
ppl	→ people
b4	→ before
2day	→ today
2morrow	→ tomorrow
2night	→ tonight
gpa	→ grandpa
gma	→ grandma
jk	→ just kidding
thx	→ thanks
hw	→ homework

Other variants that were normalised were missing apostrophes (thats → that's) misspellings of words, e.g., taht, thta → that or were subject to g-clippings (doin → doing, goin → going, nothin → nothing). These variants were displayed in the standardised form on CQPWeb but there was an indication the word had been normalised and the original spelling was also retained in case of mistakes.

3.3.3 Second Stage of Data Normalisation

VARD also has an “auto normalise” option, which normalises all variants with a specific threshold, for example 70%. This option was used to test the training files in a second step of the training. A sample (250,000 messages) from the training data extracted in step one was normalised. The number of words (not variants), variants and normalised words at the beginning of this step can be seen in this screenshot (Figure 14):

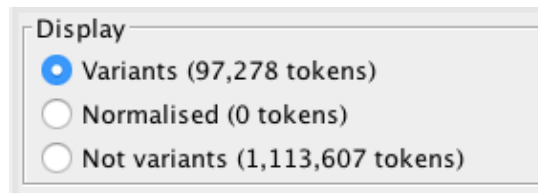


Figure 14: VARD screenshot in the second stage of normalisation (before normalisation)

This step could not be performed using the auto-normalise function, as an error message occurred. Instead, all variants which had a percentage of 70% or more were normalised manually and then variants up to 15 occurrences followed, normalising 71,076 tokens in total:



Figure 15: VARD screenshot in the second stage of normalisation (after normalisation)

The reason why there were more not variants than before the process was that some words were incorrectly highlighted as variants, such as phone numbers, email addresses or words in another languages, such as Spanish words. Words that were not highlighted as variants were largely ignored, as VARD did not highlight single letters mentioned in the previous paragraph, such as b (be), r (are, or), y (why) etc.

During this second stage of normalisation training, a few further decisions were made about how to treat certain variants:

- Treat lolz as variant of lol → changed to “lol”
- Treat yay as word → added to dictionary
- Treat realz as variant of real → “4 realz” changed to “for real”
- Muah (to denote kiss) → replaced with “kiss” similar to laughter replaced with “haha”
- Whatcha → left as whatcha, and other spellings corrected to “whatcha” (e.g., watcha)

The training data was then used to normalise the rest of the data automatically after the error message was resolved.

3.3.4 Creating Sub-Corpora

This section will outline how the corpus was uploaded to CQPWeb and how sub-corpora were created. The normalised corpus was uploaded to CQPWeb in the next step. The visualisations for the chat log were adjusted to show the turn number (N), username (U), user type (UT) and message, which can be seen in Figure 16. The example lasts five turns (1135-1139), two users are talking to each other, ACAR556 is the groomer (g) and JetPackSideKick is the target (d). Some spelling has been normalised, which is indicated by the *orig:* tag, e.g., *doin* has been normalised to *doing*:

```
N:1135 U:ACAR556 UT:g yes
N:1136 U:JetPackSideKick UT:d i rode up 2 get a dew
N:1137 U:ACAR556 UT:g it is orig: absolutely absolutely freezing
N:1138 U:JetPackSideKick UT:d i lived
N:1139 U:JetPackSideKick UT:d what u orig: doin doing now
```

Figure 16: Screenshot of CQPWeb visualisation of PJ normalised corpus

The overall corpus included the decoy turns of the conversation and consisted of 6,850,073 words and 6,968 texts. A text in this instance was a conversation between target and groomer (as outlined in section 3.3.1). One whole chat log was made up of different conversations, different times the groomer and target interacted with each other. The end of a conversation was sometimes signalled by the “code chat” tags (see section 3.3.1). Since only the groomer part of the conversation needed to be included in Corpus Linguistics statistical analyses, a sub-corpus containing only the groomer turns had to be created. In order to analyse RQ2 (duration), this groomer sub-corpus had to be split up into a further six sub-corpora (see chapter five). The sub-corpora were created in CQPWeb with a regular expression using all the groomer usernames in the metadata table:

```
[_u_username="(username01|username02|username03)"]
```

CQPWeb was queried with the option “CQPWeb syntax” rather than the standard simple query.

The specific sub-corpora for the analysis in the second empirical chapter (chapter five, sections 5.2, 5.3 and 5.4) were based on timings, also taken from the metadata table. The groomer usernames in the above regular expression were replaced by groomer usernames that spent, e.g., less than 300 minutes chatting with the decoy and those that spent more e.g., more than 300 minutes chatting with the decoy. The following sub-corpora listed in Table 4 were created (see chapter five):

Table 4: Six sub-corpora created for chapter five (sub-sections 5.2, 5.3 and 5.4)

Sub-corpus	Time span
G1	0-99 minutes
G2	100-199 minutes
G3	0-299 minutes
G4	300-10,600 minutes
G5	300-799 minutes
G6	800-10,600 minutes

3.3.5 Data Analysis

The following sub-sections will outline the analytical steps taken to compare sub-corpora (sub-section 3.2.5.1 and 3.2.5.2) in order to address RQ1 and RQ2 (see chapter four and five). The analytic procedure adopted to address RQ3, which entailed a qualitative analysis, is outlined in sub-section 3.2.5.2.

3.3.5.1 Keyword Analyses

For empirical chapters one and two (chapters four and five), a keyword analysis was chosen as the starting point of the analysis. Keywords were calculated using Log Ratio (LR) with a Log Likelihood (LL) filter (see section 3.2.2.1). The Log Ratio cut-off point was determined based on the range in the scores in both keyword analyses, as there is no agreed-on cut-off point. The Log Ratio scores ranged from 5 to 0.15 in chapter four with roughly 70% of the keywords having a Log Ratio score below 3.5. Thus, only the top 30% of keywords were considered. In chapter five, in the keyword

The four instances seem to indicate that *quest* is used most frequently with *map* to indicate the groomer referring to *mapquest*. Next, the collocates were calculated to see the strongest collocations (paying attention to number of hits of collocates and dice-coefficient scores). The collocational database in Figure 18 shows that *map* is indeed the strongest collocate with a dice-coefficient score of 0.334. The other collocates are much weaker with dice-coefficient scores between 0 and 0.002.

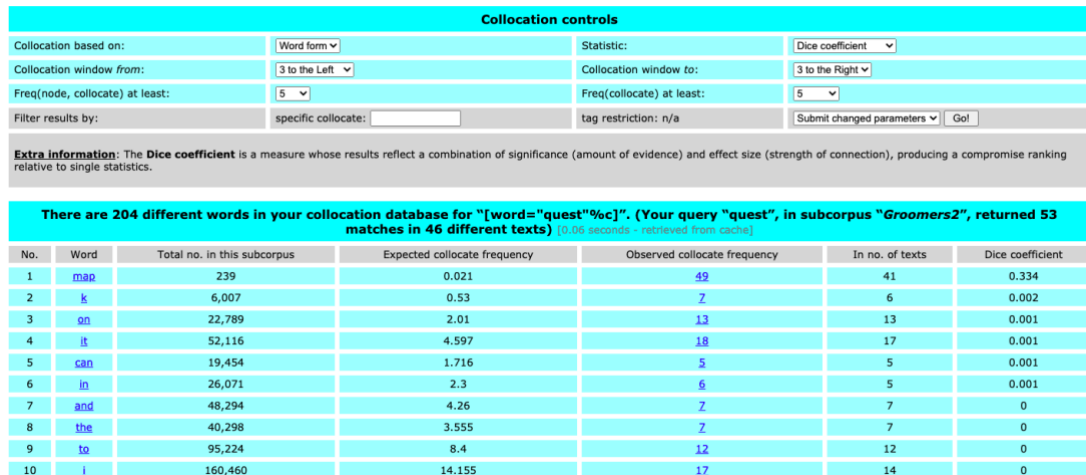


Figure 18: Step Two - calculate collocational database

In this case, the assumption that *mapquest* is the strongest collocate is a safe one to make. The last steps involved looking at hits of the strongest collocates, typically 10, to see how groomers use collocation with the keyword under observation (Figure 19) and examining the instances for patterns. In the case of *mapquest*, the instances suggest that this is used by the groomer to contribute to arranging further contact (examples one, two, three, five). Example four is not conclusive and would warrant more analysis.

Your query "quest", in subcorpus "Groomers2", returned 53 matches in 46 different texts (in 3,604,926 words [636,781 units]; frequency: 14.70 instances per million words), collocating with map (49 hits) [0.052 seconds]

< << >> >| Show Page: 1 Line View Show in random order New query Go!

No	Filename	Solution 1 to 49	Page 1 / 1
1	xPQMtmKZ1K1239230	what do i need to bring ? N:15 U: UT:g u how do i map quest	N:16 U: UT:d you go to mapquest.com once i give u the addy N:17 U: UT:g give me
2	163AY8x1E14590755	u gon na come sun ? N:26 U: UT:g ok let me find out map quest	N:27 U: UT:d cool N:28 U: UT:d thats the best orig: probs probably N:29 U: UT:g hold on N:30 U: UT:g ok N:31 U: UT:g i got it orig: sweetie sweetie
3	IYvB(AA)PN7216920	for it N:350 U: UT:g let me see if I can find this on map quest	N:351 U: UT:d k N:352 U: UT:gorig: Well! Well orig: itll it'll be a dark mini van . For now N:353 U: UT:d ok
4	QRISCMnERq79957481	oo : P N:1290 U: UT:g yup quest	N:1292 U: UT:g k N:1293 U: UT:d k N:1294 U: UT:g its farther than fayetteville N:1295 U: UT:g lol N:1296 U: UT:d lol i said its by
5	xHCYxGCbge62131208	drive is it 4 u ? N:856 U: UT:g 7am N:857 U: UT:d yah N:858 U: UT:g wow N:859 U: UT:g go to map quest	N:860 U: UT:g and put in your home address and then the city and town

Figure 19: Step Three - Examine instances of the strongest collocates from step 2

In chapter four, this was in the context of a reference corpus, whereas in chapter five, sub-corpora from the groomer corpus were compared to each other. Extended concordance lines were also examined in this step, taking a random sample of 50 per keyword, using CQPWeb's random reproducible option, if there were more than 100 or more concordance lines.

In total, roughly 8,900 concordance lines were examined in chapter four and approximately 10,750 concordance lines were analysed in chapter five. After sorting the keywords into their respective grooming intentions, they were analysed further using extended concordance lines to find out how they were used by groomers. This was to confirm or adjust initial patterns/hypotheses noted in step three but involved more extensive analysis of the keyword in its context (see Figure 20).

Displaying extended context for query match in text xPQMtmKZ1K1239230

File info for text xPQMtmKZ1K1239230 Go! Show tags

u really want to do this
 N:5 U: UT:d what all we orig: doin doing i got ta know what to prepare for
 N:6 U: UT:g well partying i said i hate to drink olne
 N:7 U: UT:d lol i havent gotten drunk before
 N:8 U: UT:g and if u say u haveto be home at a orig: certian certain time u will be orig: ther there
 N:9 U: UT:d not really i mean my dads gone until monday
 N:10 U: UT:g well orig: dont don't get drunk for me
 N:11 U: UT:d lol id like to orig: itd it'd be fun
 N:12 U: UT:g well how do i reach u
 N:13 U: UT:g if i leave now to guide me
 N:14 U: UT:d can u mapquest ?

what do i need to bring ?

N:15 U: UT:g u how do i map quest
 N:16 U: UT:d you go to mapquest.com once i give u the addy
 N:17 U: UT:g give me a minute
 N:18 U: UT:d oh okay what do i need to bring ?

N:19 U: UT:g nothing but your friendship
 N:20 U: UT:d lol u want me to walk out naked ?

Figure 20: Example of more extensive analysis of 50 instances of the keyword

In this step, politeness and impoliteness was also noted. Keywords that occurred less than 10 times and ones that were only used by fewer than four groomers were not analysed further, as they were not deemed representative of the 3.7-million-word corpus. There is no agreed-on cut-off point for this either, but four out of 622 groomers (0.64%) did not seem representative.

Coding the keywords into communicative grooming intentions from the OGDM was done over a period of time and the codes were revisited throughout the study. In terms of intra-coder reliability or ‘consistency in how the same person codes data at multiple time points.’ (O’Conner & Joffe, 2020: 2), coding was iterative. The first coding that was done using the keyword and its collocates was followed up by closer reading of the keyword in context. When the intentions were each analysed and strategies within these intentions were identified, this again involved close-readings of the text and researcher reflexivity, which led to adjustment of some coding. This was repeated to make sure coding and grooming intentions were consistent between chapters.

In terms of inter-coder reliability, the coding process and categories above were regularly discussed with the supervisory team, who interrogated a number of assumptions to fine-tune the codes. This was for example done by reviewing chapter drafts, which led to follow-up discussions. This is also called intercoder consistency:

Many qualitative research teams include an element of comparison between individual team members’ impressions of the data, but may refrain from quantifying the degree of consensus. (...) Thomas and Harden’s (2008) approach of thematic synthesis suggests that independent researcher identification of themes could be followed by group discussion of overlaps and divergences. (O’Conner & Joffe, 2020: 2).

3.3.5.2 Requests

Chapter six had a different focus which was on requests, as mentioned in sub-section 3.2.2.3 above. It was also a qualitative analysis focused on a subset of the corpus. This subset had to be determined first. It is difficult and there is no agreed methodology in the recent field of Corpus Pragmatics to query a non-pragmatically annotated corpus using a pragmatic category (requests) that is not just lexically realised and thus consists of more than one word (see Flöck & Geluykens, 2015; John, Brooks, &

Schriever, 2019; Culpeper & Tantucci, 2021, for examples of pragmatically-annotated corpora and Garcia McAllister, 2014 for an example of manually identified pragmatic categories). To do this, a list of 24 corpus search queries was created from literature into requests and sorted into Blum-Kulka House and Kasper's (2011) head act type categorisation (see chapter six, section 6.1 for more detail).

Table 5: Search queries sorted into Blum-Kulka et al.'s (2011) head act framework

Request strategy type	formulae
performatives	asking you, asking u, askin u, askin you
hedged performatives	would like you, would like u
obligation statements	you will have to, you'll have to, u will have to, u'll have to
want statements	wish you would, wish u would, want you to, want u to
suggestory formulae	how about, what about
query preparatory	would you mind, would u mind, can you, could you, can u, could u, can I, could I

These search queries had to cover different spellings of *you* due to the difficulty in normalising single letter variants, such as *u* (you), *b* (be) and *y* (why) mentioned in sub-section 3.3.2. The extended concordance lines of the head act types were then coded for one or more grooming intentions they were pragmatically aligned to, and whether any support move types were also used such as minimising, intensifying, giving grounds and enhancing interpersonal relations. The analysis was further narrowed down to three of the nine head acts that showed frequency differences between the two duration groups. Only the instances that featured support moves were extracted to see how groomers used e.g., minimising/intensifying to make their requests more or less direct and more or less polite. The support moves were examined and a taxonomy for their functions was developed bottom-up. The requests were analysed in their communicative context and up to 23 conversation turns were considered per requesting sequence. These interactions were analysed according to their support move functions and politeness/impoliteness strategies and compared in the two duration-based sub-corpora that had previously shown the most differences in groomer behaviour (see chapter five).

3.4 Conclusions on Methodologies

This chapter has described the methodology of this study by first contextualising the data source (PJ archive) and size (622 groomer-decoy chat logs) of the study by examining data sizes and approaches in existing studies on online grooming language. Most of the extant literature has used small data sizes and favoured Thematic/content Analysis of the PJ archive. The only other data sources used came from a handful of studies (Kloess et al., 2017a; Quayle & Newman, 2017; Chiang & Grant, 2018; Grant & McLeod, 2020). Ethical considerations, such as data anonymisation, storage and handling using groomer-child and groomer-decoy data were also discussed. The specific choices considering not anonymising the data but avoiding usernames in this study were outlined and steps to ensure researcher wellbeing and resilience to the disturbing nature of the data were also outlined. The chapter then moved on to discussing the other quantitative approach used in online grooming literature: LIWC and why it is not a good tool to use in isolation. The next main section described the framework of the thesis: CADS was first explained using key literature and the “standard” toolkit drawing from Corpus Linguistics and Discourse Analysis were described. The specific tools used in this thesis are keywords, collocations, KWIC analysis, requests and politeness and impoliteness strategies. The third main section described the procedure of this study. Data extraction, normalisation, corpus, and sub-corpus were all outlined and then the data analysis for the empirical chapters summarised. The study will now move on to the first empirical analysis on establishing the features of online groomer language using a general chat language reference corpus.

4. Chapter 4: The Features of Online Groomer Language

This chapter will analyse online grooming language by comparing it with a general digital chat language reference corpus, with a focus on the language used by online groomers. As introduced in chapter one, there is a small but growing body of recent work within Linguistics using qualitative and quantitative methods analysing online grooming, an under-researched area (Schneevogt et al., 2018; Lorenzo-Dus & Kinzel, 2019; Lorenzo-Dus et al., 2020). Online grooming language has been analysed predominantly qualitatively using mostly small datasets. There is a body of research into online grooming language use that aimed at describing the grooming process and groomer behaviour but that does not include linguistic analysis (Egan et al., 2011; Williams et al., 2013; Black et al., 2015; van Gijn-Grosvenor and Lamb, 2016; Kloess et al., 2017a; Quayle & Newman, 2017). Another strand of research in Machine Learning focused on the automatic detection of online grooming (Kontostathis, 2009; Inches & Crestani, 2012; Cheong et al., 2013; Cardei & Rebedea, 2017; Kim et al., 2020), which involved comparing online grooming language to other language in online settings. However, these Machine Learning studies have not used Discourse Analysis or CADS approaches to inform their analysis, which are highly contextualised approaches. Quantitative analysis paired with fine-grained qualitative analysis like CADS can enable the identification of communicative patterns as well as more nuanced analysis of language features. The thesis and this chapter specifically will contribute to this recent body of work in Linguistics.

When establishing the features of a specific type of language, it is useful and common practice in CADS to have a comparator or reference corpus that is from the same genre or using the same register, in this case, digital language. This reference corpus helps by being a representation of more general digital language use and providing a benchmark. In the case of online grooming, the ideal reference corpus would be of adults engaging in chat logs about consensual discussions of sex and dating. However due to data privacy reasons, no such reference corpus is readily available for analysis. The research into Machine Learning has therefore tended to use a dataset called PAN2012, a digital chat language corpus that does feature such discussions (see subsection 4.1 for details).

Research into digital language use has evolved over the last 30 years (see chapter two, section 2.4.1). The most recent new directions point researchers towards examining how people make sense of the online and offline world and ‘transcend different media’ (Garcés-Conejos Blitvich & Bou-Franch, 2019: 5). Users see the online world as an ‘extension of offline interactions’ (Bolander & Locher, 2020: 2) and drawing a distinguishing line between the offline and online worlds has recently been questioned by researchers (Thurlow, 2018; Garcés-Conejos Blitvich & Bou-Franch, 2019; Bolander & Locher, 2020). Thus, considering multimodal and multi-semiotic language rather than just textual data is essential. Semiotic language includes language conventions that have arisen as part of the emergence of technology (Jones et al., 2015), for example the use of abbreviations, acronyms, and emoticons to represent non-verbal elements of language often seen in online chats is used to assimilate face-to-face language (del-Teso-Craviotto, 2008). While the use of technology has been described as limiting by some, these new media also ‘allow users to establish new types of communication and language use that are sensitive to the temporal, spatial, and channel characteristics of the chat rooms (...)’ (del-Teso-Craviotto, 2008: 253).

Specific features of digital language use that are relevant to the context of online grooming may be found in research on online dating and flirting in online spaces. We know, for example, that groomers also use flirting strategies to build a deceptive bond with their targets (Black et al., 2015; Lorenzo-Dus et al., 2016; Chiang & Grant, 2017; 2018). Linguistic strategies used by adults engaged in online dating are an imagined togetherness, constructing a shared future, avoiding closure, and postponing pleasure, among others, which are used to advance online romantic relationships that might not transcend this online space (Mortensen, 2017). Other strategies include laughter, represented by onomatopoeia, emoticons, and acronyms. They have been identified as a face-saving strategy in online flirty conversations, which is used to mimic the gaze in offline interactions (del-Teso-Craviotto, 2008). While not systematically investigated in the context of online grooming, these strategies are particularly relevant to the process of deceptive trust development in online grooming, for example groomers’ attempts to engage their victims in discussion about romantic relationships (Lorenzo-Dus et al., 2016).

This chapter will first describe the methodology used in the next sub-section, 4.1, by stating the aim and research questions of this chapter. The reference corpus (PAN2012) will be discussed in more detail in the same sub-section and both the study and reference corpora will be described. The results will be examined next. Sub-section 4.2 will include a brief keyword overview to analyse the keywords that signal what the GN corpus is about and provide frequency scores for both GN and PAN2012. The top keywords and their frequencies will be discussed in this sub-section. Individual grooming intentions informed by the OGDM will be analysed in more detail in sub-sections 4.3.1 to 4.3.5, drawing on KWIC analyses and concordance lines and using the OGDM as a basis. Conclusions will be drawn in section 4.4.

4.1 Chapter Methodology

The aim of this chapter is to explore the features of online groomer language in more detail and to find out if groomer language is distinct from other digital language represented by a general digital language reference corpus (PAN2012). The analysis examines whether keywords that are distinctive to the groomer normalised (GN) corpus, when compared to PAN2012, can be mapped onto known online grooming intentions derived from the literature into online grooming communication, specifically the Online Grooming Discourse Model. As part of this, the keywords will be analysed and sorted into the online grooming intentions from the model and then examined in more detail.

The research question as such is:

1. What are the features of a corpus of online groomer language compared to that of a general digital chat language reference corpus?
 - a. Is online groomer language distinct?
 - b. How are online grooming intentions realised linguistically by online groomers?

Whether online grooming language is distinct will be measured by conducting a keyword analysis using a reference corpus. If keywords that are very frequent are also aligned with the OGDM intentions, this is a sign that groomer language is distinct.

The analysis will also examine how grooming intentions are realised linguistically, which may mean specific strategies within the intentions such as *relationship* within developing deceptive trust will emerge from specific frequent keywords. Analysing these keywords will lead to phrases or sentences that are aligned with the grooming intention strategies. The previous analysis that led to the development of the OGDM provided a taxonomy with examples, but that list is not exhaustive. This chapter will contribute to this list of examples of online grooming intentions and strategies by using a different methodology and bigger dataset.

The keyword analysis in this chapter uses a reference corpus (PAN2012) to compare to a normalised groomer language corpus (GN), using a CADS approach. PAN2012 is a dataset created for a competition held in 2012 by the Conference and Labs of the Evaluation Forum (CLEF)¹⁹. PAN competitions pre 2012 were focused on authorship attribution plagiarism and vandalism detection and competitions since 2012 have been focused on authorship analysis, authorship profiling and verification in different digital contexts (e.g., deception, online grooming, and fake news). Since 2011 PAN has also branched out into PAN at FIRE (Forum for Information Retrieval) in South Asia, and tackles tasks such as detecting plagiarism, legal information access and mixed script information retrieval²⁰. The goal of the 2012 PAN competition was to detect online grooming language using Machine Learning algorithms. The dataset is a digital language corpus made up of chat language taken from Internet Relay Chat (IRC) about generic topics, online groomer interactions from the PJ archive and chats about consensual sex amongst adults - taken from Omegle (Cheong et al., 2013). The dataset can be divided into four categories: (1) PJ chat, (2) Omegle chat about sex/cybersex, (3) IRC general chat and (4) IRC technical chat. The chats are sixty percent one to one conversations with the rest being multi-party conversations (Cardei & Rebedea, 2017). The Omegle chat part of the corpus has been described by the competition organisers as ‘perfect to augment the level of false positives in our collection’ (Inches & Crestani, 2012: 3). This suggests that the Omegle chats are similar in style and content to those of the PJ archive. The PJ archive chats were

¹⁹ <http://www.clef-initiative.eu/>

²⁰ <https://pan.webis.de/events.html>

removed from the PAN2012 dataset, and the remaining data was normalised before it was used as a reference corpus (see chapter three section 3.3 for more details).

Table 6: Word count for GN and PAN2012

Corpus	No. of Words
GN	3,712,063
PAN2012	21,491,341

The word counts of the two corpora can be seen in Table 6. The normalised groomer language corpus (GN) is the study corpus and contains roughly 3.7 million words; PAN2012 is 21 million words large. Keywords were calculated using Log Ratio with a Log Likelihood filter (see chapter three, section 3.2.2.1). The second part of the analysis involved sorting the keywords into the previously established grooming intentions taken from the OGDM based on collocates that were calculated and examining extended concordance lines (see chapter three, section 3.3.5.1).

4.2 Comparing Online Grooming to Digital Language (Keywords)

The analysis will now move on to an overview of the results by presenting and briefly discussing the top keywords of GN compared with PAN2012. The full list of considered keywords can be seen in the appendix (section 9). Overall, there were 893 keywords. The Log Ratio (LR) cut-off point was defined as 3.5, which left 178 keywords to be analysed further. Most of the 178 keywords identified (n=104), have an LR between 3.5 and 5. This means they are roughly eight to 32 times more frequent in GN compared with PAN2012. A fourth of the keywords have an LR between 5 and 6, which means they are 32 to 64 times more frequent in the GN corpus. Ten keywords are more than 64 times as frequent in GN compared to PAN2012 (LR between 6-7). Five percent of the keywords (n=9) have an LR between 7 and 8, meaning they are 128 to 256 times more frequent in GN. The minority of the keywords (n=5) have an LR higher than 8, which means they are more than 256 times more frequent in GN compared with PAN2012. The high proportion of keywords that are frequent (LR higher than 3.5) in GN compared with PAN2012 and the small number of keywords that are highly frequent (Log Ratio above 5) suggest that the language of online groomers is distinct. The top keywords will be looked at in more detail next.

The top five keywords that are more than 256 times more frequent in GN can be seen in Table 7. The top keyword are two dots that are a misspelling of ellipses or used to signal a longer pause in GN (*yes i wont promised*). They are only used three times in PAN2012. *Gon*, the second top keyword, is part of the contraction *gonna* (spelled as *gon na*), meaning *going to* in GN (*i think i 'm gon na play it*). There are only 29 instances of *gon* in PAN2012 and they are used to mean *going to* or *gone* without the extra syllable (*haters always gon hate*).

Table 7: Top five keywords in GN compared with PAN2012

#	Word	Freq 1 (per mill)	Freq 2 (per mill)	Log ratio	LL
1	..	268.71	0.19	10.5	756.34
2	gon	1136.4	1.35	9.72	3128.41
3	=P~	170.74	0.37	8.84	452.6
4	KC	103.56	0.23	8.8	273.9
5	Mapquest	47.58	0.19	8	119.54

The third top keyword is frequently used as an emoticon in the GN corpus and only appears nine times in PAN2012, where is it also used as an emoticon. *KC* is the name of a PJ decoy involved in several chat logs with groomers, which is why this keyword is salient in GN. In PAN2012, it only appears five times and refers to Kansas City (*Last weekend on friday I went to KC to watch the Jays pound the Royals for 3 games*). The last top five keyword, *mapquest*, is used in GN to ask about the target’s exact location (*hey whats ur address so i can use mapquest*), which is evidence of the developing deceptive trust and arranging further contact grooming intentions. This is the first keyword that is aligned with online grooming intentions from the OGDM.

The next most frequent keywords can be seen in Table 8. They have a Log Ratio of seven or above. More keywords emerge that seem to be aligned to the OGDM. The sixth most frequent keyword, *wan*, is another example of a word being split into two. In this case it is *wanna*, meaning *want to*. This is similar to *gon*. *Wan* is frequently used in GN to engage the target in exchange of personal information (*anything u wan na know about me ?*). *Mwah*, which is an onomatopoeic textual representation of a kiss was used by less than four groomers in GN. It seems to be aligned with the

developing deceptive trust grooming intention. The same is true for the kiss smiley. *B/F*, an abbreviation of boyfriend, is also related to the grooming intention of developing deceptive trust and specifically the relationship strategy.

Table 8: Top keywords in GN compared with PAN2012

#	Word	Freq 1 (per mill)	Freq 2 (per mill)	Log ratio	LL
6	wan	1713	8.61	7.64	4178.26
7	<i>Mwah</i>	97.97	0.6	7.34	232.3
8	b/f	134.35	0.88	7.25	315.62
9	§_KISS-SMILEY_§	2001.3	13.21	7.24	4700.21
10	robby	27.99	0.19	7.23	65.65
11	Catcha	33.59	0.23	7.17	78.29
12	<i>privates</i>	19.59	0.14	7.13	45.47
13	<i>313</i>	25.19	0.19	7.08	58.13

Robby is only used by two groomers, as it is the name of a decoy and one other groomer talks about his friend, Robby. *Catcha* is used by a small number of groomers to express catch you, as in *catch you later*. *Privates* and *313* are also used less than ten times or by fewer than four groomers. They do echo online grooming intentions such as seeking sexual gratification and arranging further contact.

The above brief analysis of the top keywords shows that the language of online groomers in the top keywords seems to be distinct from that of a digital chat language reference corpus and online grooming intentions begin to emerge as part of the top keywords.

4.3 The Online Grooming Discourse Model (Grooming intentions)

The chapter will now move on to analysing the specific online grooming intentions found in the GN corpus. The 178 keywords were sorted into grooming intentions from the OGD, which can be seen in Table 9, by calculating their collocates using dice-coefficient and examining extended concordance lines. When there were more than 100 occurrences, a random sample of 50 of them were analysed. Overall, around 5,800 concordance lines were examined in this analysis. A full list of keywords can be seen

in the appendix (section 9). Keywords that did not easily align pragmatically with the grooming intentions are shown in Table 10 and were not analysed further. Keywords that had less than 10 occurrences (shaded in grey in Table 9) or were used by fewer than four groomers (italicised in Table 9) were not analysed in depth either.

Due to the complex and interconnected nature of the online grooming process, and by extension the OGD, one keyword did not necessarily always correspond to only one grooming intention. In most cases, one keyword contributed to at least two grooming intentions; However, there was usually one primary intention that could be identified by frequency through analysing extended concordance lines. The following are illustrative examples of the keyword *fell*:

- (A) G: thats not true i **fell** asleep after i got out of the shower im sorry
- (B) G: i hope i can make you **fell** like you can tell me anything or ask anything
- (C) G: show yo how it would **fell** and teach you about your self and me with you

Fell is used for small talk in 82% of its concordance lines, an example of which is A. This was the primary grooming intention. In a subset of the concordance lines (16%) groomers used *fell*, a misspelling of *feel*, for bonding and seeking sexual gratification purposes. This was a secondary grooming intention for this particular keyword, which B and C are examples of. This is similar to the study by Lorenzo-Dus et al. (2020) that analysed three-word collocates of the same data and found clear primary grooming intentions for 60 out of 70 three-word collocations (Lorenzo-Dus et al., 2020). In Table 9, the primary grooming intention for the GN keywords was chosen.

Table 9: Keywords sorted into grooming intentions

Grooming intention	Keywords
Developing deceptive trust	<i>Mwah</i> , b/f, <i>cuttie</i> , nosy, <i>amor</i> , sweetie, <i>papi</i> , kindness, comfy, g/f, blush, hun, <i>Thxs</i> , smiling, nite, gifts, snack, dreams, shower, laundry, Detroit, TA, roads, Disney, nada, pee, Princess, dreaming, youngest, cutie, smile, MAD, sweetheart, MISS, bills, precious, addy, baby, girlfriend, UPSTAIRS, adventure, chattin, busy, comfortable, Aww, k, bout, fell, PROMISE, TY, beach, house, moms, grounded, dork, dressed, tonight, Motorcycle, Chevy, truck, Ford

Seeking sexual gratification	<i>Privates, tingly, doggie, pyjamas, condoms, thongs, gentle, bras, foreplay, cloths, oral, blanket, massage, excited, pill, shave, panties, Thong, bikini, swallow, sexually, orgasm, cotton, DAM, Snuggle, kisser, hugging, tickles, kisses, cuddle, cuddling, tickle, kissing, kiss, Wear, tummy, hug, kissed</i>
Arranging further contact	Mapquest, 313, 714, Daytona, tues, cedar, motel, mikes, yahoo.com, mall, thurs, bowling, thur, 9pm, Sunday, Saturday, cal, Park, Theater, gas, tonight, Sat, noon, Anytime, voice, nervous
Assessing risks and isolating	Sneak, neighbors, grandma, aunt, tuck
Negotiating power dynamics	jail, cops, cop

Developing deceptive trust is very prevalent in the corpus with 61 keywords, making up just under half (46%) of the considered keywords. This is similar to Lorenzo-Dus et al.'s (2016) finding that developing deceptive trust was more frequent than the other processes. The grooming intention of seeking sexual gratification is also present with 38 keywords, as is arranging further contact with 26 keywords. The approach intention in the original OGDM only referred to verbal lead ins for an offline face-to-face meeting with the target, rather than any arrangement of further contact. The former was also quite prevalent in Lorenzo-Dus et al.'s (2016) dataset, while the latter was not considered as such. In Lorenzo-Dus et al. (2020), three-word collocations contributing to one online grooming intentions were most frequently aligned with seeking sexual gratification (16 three-word collocations out of 70).

There are several keywords (n=5) that belong to the assessing risks and isolating grooming intention. Finally, three keywords contribute towards negotiating power dynamics. Lorenzo-Dus et al. (2020) found three three-word collocations only aligned with isolation and approach and one three-word collocation aligned with compliance testing. Forty-four of the 70 three-word collocations contributed to more than three online grooming intentions, with one being the predominant intention. Out of these, 15 three-word collocations were predominantly used for approach, 12 were used for sexual gratification and deceptive trust development. Four were predominantly used for isolation and one for compliance testing. Importantly, only ten out of the 70 three-word collocations did not have a clear main online grooming intention associated with it (Lorenzo-Dus et al., 2020).

Table 10: Keywords that did not easily align with one or more grooming intention

Other	Keywords
Miscellaneous -	..., ..., <i>Catcha</i> , <i>whatelse</i> , <i>icic</i> , <i>name*</i> , <i>tub</i> , ICKY reals, crystal, <i>wolf</i> , <i>lil</i> , Comp, booted, UPSTAIRS
Names	KC, robby, Marshmallow, allie, brooke, jaz, willow, Annie, Luke, kelsey, maya, Susie, mercy, stephen
Misspelling	Gon, wan, na, haveing, tommorrow, felling, havin,
Textual representations of paralinguistic features	=P~ §_KISS-SMILEY_§ -/ b- -w lol .lol <i>hehhee</i> OHHHHHHHHH ohhhhhhhh hh

In terms of the “miscellaneous” keywords, some of them are PJ decoy names (e.g., *KC*, *Allie*, *Mercy*), while others are misspellings of words (*haveing*, *felling*) that were not standardised during the normalisation process using Variant Detector (VARD) before analysis, as explained in chapter three (sections 3.3.2 and 3.3.3). Other words are part of the PJ decoy’s effort to redact identifiable information, such as names (*name**). There are also emoticons/part of emoticons (=P~) and other textual representations of paralinguistic features and non-verbal behaviour, such as onomatopoeic laughter (*hehhee*, *lol*, *.lol*) that are present in the groomer corpus and seem to be less present in the reference corpus. Features like these have been found to be strategies in online flirting and dating to mimic face-to-face interactions (del-Teso-Craviotto, 2008). However, this thesis will not focus on them. Abbreviations of words, such as *comp* for *computer* and *lil* for *little* can also be found in the Table 10. Some words are only used by a few groomers (*whatelse*, *hehhee*) and seem to be a specific feature of their language, rather than representing grooming language in general. Tables 9 and 10 show that online grooming language emerges strongly in the keywords and all grooming intentions in the original OGDМ are found in the keywords. The analysis will now move on to the individual intentions and analysing how they are realised through language in this groomer corpus.

4.3.1 Developing Deceptive Trust

The individual grooming intentions will be considered next. Within developing deceptive trust there are four strategies (see Table 11): relationship

building/strengthening (17 keywords), small talk/conversation filler (24 keywords), bonding (14 keywords) and exact location (four keywords), which is connected to the arranging further contact intention. These strategies broadly correspond to the developing deceptive trust strategies in the OGDM.

Table 11: Five strategies and their keywords in developing deceptive trust

Strategy	Keywords
Relationship terms	<i>Mwah</i> , b/f, <i>cuttie</i> , <i>amor</i> , <i>sweetie</i> , <i>papi</i> , g/f, <i>hun</i> , <i>Princess</i> , <i>cutie</i> , <i>sweetheart</i> , <i>precious</i> , <i>baby</i> , <i>girlfriend</i> , <i>adventure</i> , <i>bout</i> , <i>dork</i>
Small talk/conversation filler	<i>Snack</i> , <i>shower</i> , <i>nosy</i> , <i>laundry</i> , <i>TA</i> , <i>roads</i> , <i>nada</i> , <i>pee</i> , <i>youngest</i> , <i>bills</i> , <i>k</i> , <i>fell</i> , <i>Chattin</i> , <i>busy</i> , <i>Thxs</i> , <i>Disney</i> , <i>dressed</i> , <i>tonight</i> , <i>nite</i> , <i>dreams</i> , <i>motorcycle</i> , <i>chevy</i> , <i>truck</i> , <i>ford</i>
Bonding	<i>Kindness</i> , <i>comfy</i> , <i>blush</i> , <i>smiling</i> , <i>gifts</i> , <i>smile</i> , <i>dreaming</i> , <i>mad</i> , <i>miss</i> , <i>comfortable</i> , <i>Aww</i> , <i>PROMISE</i> , <i>TY</i> , <i>grounded</i>
Exact location	<i>Detroit</i> , <i>addy</i> , <i>beach</i> , <i>house</i>

4.3.1.1 Relationship terms

The first strategy within the developing deceptive trust intention is connected to building/strengthening and discussing the groomer – target relationship. *Sweetie*, *hun*, *princess*, *cutie*, *sweetheart*, and *baby* are all used as terms of endearment, which strengthens the deceptive relationship. The following are illustrative examples, in which G indicates the groomer talking²¹:

- (1) G: So are **b/f** and **g/f**
- (2) G: not today **sweetie** just staying in
- (3) G: TALK TO YOU LATER **hun** , bye
- (4) G: yes you can be my **girlfriend**...lol

B/f (boyfriend) and *g/f* (girlfriend) are used to negotiate and define the relationship, as example one shows. This negotiation also contributes towards assessing risks and isolating and negotiating power dynamics (*can we be secret b/f n g/f*). As example four

²¹ This will be the case for all subsequent examples. In some extended examples below T indicates the target talking.

shows, *girlfriend* seems to be used in similar ways, both to negotiate relationship terms (example four) and to refer to the target (*Hey girlfriend*). However, the latter is used by one specific groomer and is not representative of online groomers in general. Similarly, *mwah*, *cuttie*, *amor*, and *papi* are used by fewer than four groomers and will not be analysed further. This relationship negotiation seems to be a specific linguistic realisation of the relationship strategy within the developing deceptive trust intention in the OGDM. Examples two and three show how groomers intersperse these terms of endearments with other developing deceptive trust strategies, such as small talk (example three) and maintaining the conversation (example four). Groomers are strategically combining different strategies within the same intention to maximise their success at building trust with their target. This connection emerges as part of examining keywords using a CADS approach and adds nuance to the OGDM concerning the relationship between individual strategies within a grooming intention.

The remaining keywords did not adhere to the pattern above, but still contribute towards the building, defining, and dismantling of the groomer-target relationship. They will be analysed in more detail next. *Dork* is used to strengthen and negotiate the groomer-target relationship, however slightly differently than the terms of endearments above. It is used by the groomer to reassure the target that they are not a dork (*seen the real you no dork I can tell*). Being a dork is portrayed as unfavourable here. It is also used by the groomer to mitigate awkwardness around discussing feelings and relationships (*aww I feel like a dork in situations like that lol*). The groomer takes the brunt of the awkwardness and normalises feeling uneasy in a way to encourage and comfort the target while downplaying the illegal and unethical relationship-dynamics that are at play. *Precious* is also used to strengthen the groomer-target relationship. The groomer uses it to describe the target and to emphasise how important they are, thus creating deceptive trust:

(5) G: Take care of the most **precious** thing in the world... . you!!

(6) G: u would be shareing sometihngs so **precious** with me

The groomer pays the target a compliment in example five. Lorenzo-Dus et al. (2016) found that praise had a positive correlation with implicit desensitisation and mental isolation. Compliments were not only used as a ‘vehicle to enhancing trust’ (Lorenzo-

Dus et al., 2016: 46), but also for sexual gratification purposes and isolating the target (see also Lorenzo-Dus & Izura, 2017). Compliments were not a prominent strategy in the keyword analysis above, likely owing to the different methodology used since it is difficult to identify compliments using a keyword analysis. Compliments often occur via specific multi-word strings, rather than just being one-word units. This is confirmed by Jucker et al. (2008) who found nine syntactic patterns for compliments by searching a general British English language corpus (Jucker et al., 2008).

Example six shows that *precious* is also used as a euphemism for sex reinforcing the interconnectedness of developing deceptive trust and seeking sexual gratification. *Precious* is also used in connection with the concept of virginity (*Are you sure you want to give up your most precious gift to me ...*), thus also contributing towards the seeking sexual gratification grooming intention. *Adventure* is also used in different ways to build a stronger bond with the target. This is achieved by making small talk (example seven) isolating the target from their support network (example eight) and using *adventure* as a euphemism for sex and the grooming relationship, thereby mitigating the perceived negative face threat to the target (example nine):

- (7) G: i like action **adventure**, obviousle science fiction
- (8) G: i want us to keep this our little **adventure**.
- (9) G: that should be a wonderful **adventure**

This shows that grooming intentions, such as assessing risks and isolating, developing deceptive trust, and seeking sexual gratification are intertwined, and can all be used by the groomer simultaneously, confirming the link between developing deceptive trust and the remaining intentions, also found in the original OGD. Both *previous* and *adventure* seem to be used by groomers to try to avoid making explicit reference to relationship terms, which then becomes a more implicit form of relationship negotiation than using terms of endearment.

Bout (misspelled about) is used as part of the key collocation ‘*think(ing) + about*’ (dice co-efficient: 0.26) to build trust by showing affection towards the target. It is used by the groomer to tell the target they are thinking about them and showing affection (*ok i b thinking bout u*). *Thinking + about* is also used to introduce praise (*i will be thinking*

bout your pretty eyes lolz), which reinforces the deceptive trust. Another pattern pertains to a physical meeting between the target and groomer (*was here just thinking bout if we ever gon na see each other*). *Thinking + about* is also used for seeking sexual gratification by describing a sexual scenario between the target and the groomer or discussing sex in general (*who u been thinking bout havin sex with ?*). This is an example of a collocational pattern that serves different grooming intentions, in this case, developing deceptive trust, arranging further contact and seeking sexual gratification with deceptive trust being the most prominent of these. In an analysis using the same corpus, Lorenzo-Dus, Kinzel and Di Cristofaro (2020) found that the two-word collocational pattern *thinking + about* was used primarily to contribute to the grooming intention of seeking sexual gratification (Lorenzo-Dus et al., 2020).

4.3.1.2 Small talk/Conversation Filler

The second strategy within developing deceptive trust is connected to maintaining the conversation. The following keywords (*laundry, nada, bills, chattin*) are used to make small talk, which simultaneously keeps the conversation going and contributes towards developing deceptive trust between the groomer and their target:

- (10) G: sorry had some **laundry** to finish done now all your
- (11) G: **Nada** it rained all day so am drinking
- (12) G: lol no fun nope but it pays the **bills**
- (13) G: yea thats good u **chattin** with ur friends ?
- (14) G: nice r u **busy**?
- (15) G: yep **roads** driving this morning were HORRIBLE !!

In example 10, the groomer provides a reason for not responding to the target's message and picks the conversation back up, promising undivided attention going forward (*now all your*). Examples 11, 13 and 14 show exchanges of activities the groomer and/or their target are engaged in. Example 12 is related to the groomer's job, which he does not do because he enjoys it, but because it *pays the bills*. Example 15 illustrates that *roads* is mostly used for small talk related to driving conditions. In a subset of the concordance lines (17%) this is used to discuss meeting details between the groomer and their target e.g., the target's road, roads near the target, roads the

groomer can take to get to the target. Activities, which is a prominent strategy within the OGD, does not emerge as prominent in these keywords. However, the original strategy is a very broad category, including general likes, dislikes, planned and previous activities and discussions about the online interaction. CADS makes it difficult to identify salient words that express such activities with too many possibly occurring to detect statistically significant keywords. The reference corpus also plays a part, as some of these activities will likely appear in the general chat interactions and instances of strangers chatting to each other, taken from Omegle, which would mean they do not emerge as keywords in this analysis.

Discussions of vehicles are also used for small talk purposes using the keywords *motorcycle, chevy, truck, ford*. They are often used to answer the question “What kind of car/vehicle do you have?”. There are relatively few occurrences of the keywords, except for the keyword *truck*. These keywords are also used for arranging further contact purposes in a subset of the concordance lines by providing vehicle descriptions when arranging a meeting with the target (*well i will be in a big red truck*). There is a sub-group of keywords, which mainly contribute towards developing deceptive trust via small talk, similar to examples 10-15, but that are also used for seeking sexual gratification purposes in a number of occurrences. The following are illustrative examples:

- (16) G: just think you could suck on that **snack** for ever and don't have to get another one
- (17) G: lol like some guys like it when girls **pee** on them
- (18) G: in case your interested some time love to **shower** with u
- (19) G: i 'm a little nervous too you will be the **youngest** girl ever for me
- (20) G: I had to get up and immediately get **dressed** and take them out

In most cases, *snack* is used in small talk (*so what did u get for snack*), but in around a third of the occurrences, *snack* is used to suggest a sexual act (see example 16). The same is true for *shower*, which is used to express the desire to shower with the target in example 18. This use for seeking sexual gratification purposes is present in around a third of the occurrences. Similarly, *pee* is usually used to excuse leaving the conversation temporarily (*be right back got ta pee*) but is also used to explain sexual

preferences (example 17) or to “educate” the target about human anatomy (*yes where your pee comes out*) and sex (*a female when she cum is like she has to pee*).

In example 19, the groomer discloses his own nervousness, which contributes towards building deceptive trust, but also implies that he has not had sex with a minor before. This is simultaneously oriented towards building deceptive trust, seeking sexual gratification, and negotiating power dynamics and shows the interconnectedness of these grooming intentions. *Youngest* was used to advance sexual gratification in 36% of the occurrences. *Dressed* is used for small talk in example 20 but is also used in about 30% of the concordance lines to ask if the target is dressed or what they are wearing, contributing towards the seeking sexual gratification grooming intention. Lorenzo-Dus et al. (2016) found a link between sociability, which is mainly realised through small talk, and implicit desensitisation, mental isolation, and reframing. The examples above are mainly explicit desensitisation; Mental isolation and reframing are not key words in this analysis, which might be due to the specific reference corpus being used.

Several keywords are used in a variety of ways to maintain the conversation via small talk, discussions of activities and other things that do not seem to have another communicative function but are linked to sociability. These keywords do not contribute towards developing deceptive trust in their own right but are examples of sociability:

- (21) G: i always get up to damn late got **ta** start getting to bed earlier
- (22) G: now go 2 bed **k**, ill see you tomorrow
- (23) D: no ur not old at all
G: well **ty**

Ta is a misspelling of *gotta*, meaning *got to* and is used to fill the conversation. *K*, an abbreviation of okay is also used similarly. It is used to agree, to acknowledge and as a backchanneling device, asking for acknowledgment, like in example 22. *Ty* is an abbreviation of *thank you* and is mostly used by the groomer to thank the target for a compliment (example 23). Some keywords are more versatile than the conversation fillers in examples 21-23, while still mainly being used for small talk:

- (24) G: thats not true i **fell** asleep after i got out of the shower im sorry

- (25) G: Am going to **disney** tomorrow
- (26) G: what u doing **tonight**
- (27) G: lol no i meant , whatd you do to get **grounded**
- (28) G: then watched some tv , did some work , cleaned **house**

Fell is mostly (82%) used for small talk, see example 24, but is also used to bond with the target (*i hope i can make you fell like you can tell me anything or ask anything*) and to gain sexual gratification (16%) (*show yo how it would fell and teach you about your self and me with you*). It is a misspelling of *feel* in these cases. *Disney* is used by the groomer to tell the target what they are doing, future plans they have (example 25), their past experiences and to a lesser extent making plans with the target. *Tonight* is also used to talk about current activities and planned activities to maintain the conversation (example 26). Additionally, it is used to arrange further contact, primarily online (*i really hope u come on tonight*) in 20% of the concordance lines. Furthermore, the groomer uses it to arrange offline meetings (*can u go out tonight*) and gain sexual gratification (*god why am i so horny tonight*). *Grounded* (example 27) is used for small talk but is also used for assessing risks and isolating purposes (*u get grounded a lot*) in 26% of the concordance lines. *House* is another example of a keyword that is used for multiple grooming intentions. The most frequent one (50%) is developing deceptive trust via small talk, discussions of activities (example 28), eliciting personal information and negotiating relationships. However, groomers also use it to arrange meetings with the target (*will we be able to hang at your house*) in 32% of the concordance lines. Assessing risks and isolating is a third intention it is used for (*how far does your mom work from your house*).

Nosy is a very versatile keyword and contributes to four grooming intentions (developing deceptive trust, sexual gratification, negotiating power dynamics and assessing risks and isolating). It is mostly used by the groomer to justify his curiosity (*so where u been and doing yes nosy again*) while extracting personal information, which is disguised as discussions of activities. The groomer subtly asks the target about their current or planned activities and then adds *if I can be nosy* and acknowledging his nosiness, as in the above example. He puts his potential intrusiveness on the record, pre-empting and mitigating the perceived negative face threat. The same strategy is also employed when *nosy* is used for seeking sexual

gratification purposes (*did u and ur ex_b/f do it if that's not to nosy*). This is also true for assessing risks and isolating (*that's cool so not to sound nosy but where is ur mom at right now*). Being nosy is attributed to other people in the case of negotiating power dynamics: *sure your neighbors wont say nothing there is a lot of nosy people*. The groomer adopts low-risk behaviour and asks the target if they are sure the neighbours will not suspect anything. This serves several purposes: it emphasises the secrecy of the groomer-target interaction, the groomer pretends to be concerned for the target's welfare and the groomer can make sure not to get caught.

A small number of keywords (*dreams, nite*) are used to say goodbye, which is a conversational ritual of ending a conversation in interactions between individuals who are interpersonally close. Simultaneously it decreases the social distance between the groomer and the target. This is similar to the sociability strategy in the OGDM and is a specific linguistic realisation of this grooming intention:

(29) G: see you latter sweet **dreams**

(30) G: ok babe good **nite**

It builds trust between the groomer and the target and does not seem to have any additional conversational value.

4.3.1.3 Bonding

Another strategy within developing deceptive trust is related to expressing how the groomer feels about the target. This is related to the relationship strategy of the OGDM and is a linguistic realisation of negotiating the groomer-target relationship by talking about their relationship and feelings towards each other. This strategy consists of the following keywords: *Kindness*²², *smiling, gifts, dreaming, smile, mad, miss, aww, promise*. The following are illustrative examples of how groomers express feelings for the target and convey that they (the groomers) will look after them (the targets):

(31) G: cause I wan na make you **smile**

(32) G: ill always try to keep u **smiling** bi doing or saying something

²² Kindness had fewer than 10 occurrences and was not analysed further.

- (33) G: whatever u want babe i wont get **mad** at u ever §_HAPPY-SMILEY_§
- (34) G: i wuldd need to get you lots of **gifts** because u are my princess
- (35) G: i'll be **dreaming** of you
- (36) G: hi baby i **miss** u
- (37) G: Its nothing u have done wrong i **promise**

The groomer uses concern and affection to strengthen the bond with the target. This is done by saying he will make the target smile (examples 31-32), reassuring the target that he will not get mad at them (example 33) or stating that the target has done nothing wrong (example 37) and promising gifts to the target (example 34). *Dreaming* is also used to create emotional closeness with the target (example 35), telling them that they are always on the groomer's mind, even at night. This remains innocent and romantically focused in most of the concordance lines (68%). However, a small subset (24%) is more sexual (*yep in my dreams u make so horny i was dreaming that i was liking ur puss*). *Miss* is used to express strong emotions for the target (example 36), which amplifies the romantic nature of their relationship and strengthens the deceptive bond. *Promise*, *smile*, and *smiling* are a way for the groomer to present himself as a person who has the target's best interest at heart and will do anything to make them happy. They can contribute to all grooming intentions and act as a booster to trustworthiness as part of romantic love. *Promise* is used for all grooming intentions, but predominantly (53%) developing deceptive trust (*that wo n't happen hun I promise*). Seeking sexual gratification (*oh my god i cant wait it is fun , i promise*) and negotiating power dynamics (*U have to promise not to tell*) make up about a third of the concordance lines (22%, 15%, respectively). Arranging further contact (*i promise im going to bring some protection*) and assessing risks and isolating (*you promise your on your computer and you are private*) are boosted by *promise* less frequently (eight % and one % respectively).

The remaining keywords (*comfortable*, *comfy*, *blush* and *aww*) do not fit into the above strategies while being used for bonding purposes. The keywords *comfortable* and *comfy* are used by the groomer for both developing deceptive trust and seeking sexual gratification purposes. Comfort in these cases refers to being comfortable enough to

talk to the groomer about difficult things, so emotional comfort, (example 38), as well as physical comfort (example 39):

(38) G: good I want you to feel **comfortable** with me like that

(39) G: are u going to be **comfortable** getting naked and having sex with someone u just met

Comfy is used predominantly for developing deceptive trust via showing concern for the target (*u all comfy*) and making small talk. It is also used for seeking sexual gratification, most often to ask about the target's underwear (*what color is your comfy pants*), which likely leads to sexual fantasising about the target. *Blush* is used by the groomer to pay the target compliments or receiving compliments that seemingly made them (the groomers) blush. Deceptive trust is built by sharing these emotions. It is also used to a lesser extent to refer to sexual activities between the groomer and target, which causes either or both of them to blush. *Aww* is also used by groomers to react to compliments from the target or something else the target said and to express emotions towards the target, thus bonding with them:

(40) G: **aww** , i wish i could take care of ya

It is predominantly used to develop deceptive trust (84%), but also contributes to the grooming intention of seeking sexual gratification.

4.3.1.4 Exact Location

The fourth strategy within developing deceptive trust is connected to the target's exact location and consists of the following keywords: *Detroit, addy, beach, house*. These are illustrative examples:

(41) G: ok so how close are u to **detroit** ?

(42) G: so when did u move to long **beach**

(43) G: I need your **addy** 4 directions lol

As can be seen in example 41 *Detroit* is used to gauge roughly where the target is. It is also used to work out how to get to them (*so i can reach to detroit and then i can*

take taxi right) for a face-to-face meeting. The groomer also uses it to disclose his own rough location (*i am north west of detroit*). *Beach* refers to Long Beach (example 42) or Daytona beach and is used in much the same way as *Detroit*. The PJ foundation set up several sting operations in distinct locations, e.g., Detroit and Long Beach. Detroit was used in 25 chat logs and Long Beach, or Daytona Beach appeared in 11 grooming interactions. Additionally, *beach* is used for small talk (*have you been out much this summer , beach or anywhere*) and hypothetical scenarios (*i was thinking of us walking down the beach while u sipped your manhattan tea*). *Addy* is used to extract personal information from the target, e.g., their email address (*lol i need either ur email addy or ur last name to add u*), address of nearby landmarks (*what was the addy for 7_11 again*) or their actual address (example 43). This is a specific linguistic realisation of the exchanging personal information strategy in the OGDM and is connected to the arranging further contact grooming intention.

The above analysis identifies specific linguistic realisations of developing deceptive trust strategies originally developed in the OGDM by using a different methodology (CADS). The groomer-target relationship is built and reinforced by using explicit relationship terms (*girlfriend, sweetie*), which are strategically interspersed with other developing deceptive trust strategies to maximise the success of gaining the target's trust. In contrast, *precious* and *adventure* are used by groomers to avoid using explicit relationship terms and mitigating the perceived threat to the target's negative face. The interconnectedness of seeking sexual gratification and developing deceptive trust becomes apparent in the groomers' use of these implicit words, as well as the collocational pattern *thinking + about*, which is used to imagine sexual hypothetical scenarios involving the target. Small talk is another strategy with specific language examples identified above. *Laundry, nada, bills, chattin* are all used for this purpose by groomers. Another pattern identified above also shows that groomers use some words (*snack, pee, shower*) for specific seeking sexual gratification purposes in some contexts, while just making small talk in other contexts. One specific language realisation within the relationship strategy is identified as bonding and is related to expressing feelings and affection towards the target. This is done by using keywords such as *smile, smiling, mad dreaming* and *promise*. The exact location keywords (*Detroit, beach, addy, house*) are a specific linguistic realisation of the exchange of

personal information strategy and is also closely connected with arranging further contact with the target.

Overall, a different focus from that of Lorenzo-Dus et al. (2016) emerges in terms of developing deceptive trust. Relationship terms and terms of endearment such as *bf/gf* and *hun*, which fall under the relationship label as defined in the model, are highlighted to negotiate the groomer-target relationship. Compliments, although present in the concordance lines, did not form their own strategy within developing deceptive trust, while being quite prominent in Lorenzo-Dus et al. (2016), which is likely due to the fact compliments are formulaic and would not appear on the one-word unit level, such as keywords. A focus on the exact location of the target also emerges from the keywords. This ranges from the area (*detroit, beach*) to the target's address (*addy*). A number of keywords are used to contribute towards developing deceptive trust, specifically small talk, as well as seeking sexual gratification, which reinforces Lorenzo-Dus et al.'s (2016) link between sociability and implicit desensitisation.

4.3.2 Seeking Sexual Gratification

Seeking sexual gratification is the second most prevalent grooming intention in the keywords, and it is also often interspersed with developing deceptive trust, as discussed in section 4.3.1. This grooming intention is divided into implicit and explicit desensitisation in the OGD. Using CADS in this thesis allows the analysis to give a more detailed insight into how this implicitness and explicitness are interconnected and to identify specific linguistic realisations of the strategies within seeking sexual gratification. Table 12 shows the five strategies, and keywords, within the seeking sexual gratification intention. The strategies are sex acts, physical contact, clothes/appearance, contraception and other. They will be looked at in more detail next.

Table 12: Five strategies and their keywords in the seeking sexual gratification

Strategy	Keywords
Sex act(s)	Doggie, tingly, gentle, massage, foreplay, excited, oral, swallow, sexually, orgasm
Physical contact	Snuggle, kisser, hugging, <i>tickles</i> , kisses, cuddle, cuddling, tickle, kissing, kiss, hug, kissed

Clothes/appearance	Pyjamas, thongs, bras, cloths, panties, cotton, bikini, shave, DAM
Contraception	Pill, condoms, wear
Other	<i>Privates</i> , blanket, embarrassed, tummy

4.3.2.1 Sex Acts

The first of these strategies is the description of sex acts, which ranges from more implicit keywords (*tingly, excited, massage, gentle*) to more explicit ones (*foreplay, doggie, orgasm, oral, swallow, sexually*). The following are examples of a more implicit approach, which contributes towards implicitly desensitising the target:

- (44) G: it was a good **tingly** feeling - right ?
- (45) G: im getting more **excited**
- (46) G: i could give you a **massage** and some kisses all over to make up for me not making it last week
- (47) G: thankyou for letting me know , i am curious though when we get in the hot tub and we are both turned on are you willing to make out tonight , i would love to be your first i am very **gentle** and careful with you it 's up to you

In the above examples, vague language is used to mitigate the sexual content. Vague language in online child sexual grooming has been analysed by Lorenzo-Dus & Kinzel (2021). Vague language refers to ‘a linguistic unit (word, phrase or utterance) that has an unspecified meaning boundary, so that its interpretation is elastic in the sense that it can be stretched or shrunk according to the strategic need of communication’ (Zhang, 2013: 88).

In the above examples, *it*, a vague category identifier (Zhang, 2013) is used instead of naming the specific action that would give the target a tingly feeling or is making the groomer excited, namely sexually touching the target (example 44) and imagining having sex with the target (example 45). This simultaneously introduces the target to sexual content, while minimising the perceived face threat to the target. In example 46, *massage* is used as a de-intensifier (Zhang, 2013) to mitigate discussing sexual content. A massage and kisses could lead to sex. However, the groomer decides not to

disclose this and instead leaves it vague. Finally, in the last example, the groomer also does not mention sex, but by saying *we are both turned on* and *I would love to be your first* implies that he wants to have sex with the target, whom he presumes to be a virgin. *Gentle* and *careful* are also used as downtoners to lower the intensity of speech and avoid mentioning sex explicitly. The remaining keywords in this strategy are sexually explicit, which the following illustrative examples demonstrate:

- (48) G: but ill teach you about **foreplay** and other stuff if you wan na learn that
- (49) G: i like to bend u over **doggie** style n give it to u n reach under u n play with your nipples
- (50) G: the best **orgasm** for a woman is from **oral** sex ...
- (51) G: ill teach u to suck dicks and **swallow** cum
- (52) G: you'd let me go all the way **sexually** with you then

In the first two examples, vague language is used. In example 48 *and other stuff*, another vague category identifier, is used to describe sex, which occurs after an explicit term, *foreplay*. It is used similarly in example 49, despite the explicit description of a sex act (*doggie style*). This shows that vague language also features in instances of explicit desensitisation. The groomer uses a push-pull structure (Montgomery, 2006) here, which combines assertiveness and tentativeness. This was found prominently in Lorenzo-Dus and Kinzel (2021) in which groomers used sexually explicit words strategically in conjunction with vague language terms to advance the seeking sexual gratification grooming intention by explicit desensitisation and the developing deceptive trust intention by ‘framing sexual activity in non-sexual terms’ (Lorenzo-Dus & Kinzel, 2021: 203). Example 51 also features specific roles the groomer assigns to himself (teacher) and to the target (pupil) to reframe sex as beneficial. Overall, this is a way to explicitly desensitise the target.

4.3.2.2 Appearance/Clothes

Another strategy within the seeking sexual gratification grooming intention pertains to physical appearance and clothes. The keywords are used by the groomer to ask about the clothes the target is wearing and implicitly seeking sexual gratification:

- (53) G: to bad you cant stop over and let me take your **cloths** off you
- (54) G: do you wear **thongs** ?
- (55) G: Are your **bra** and **panties** silk **cotton** or lace ?
- (56) G: do n't get mad , but i 'd like to see a pic of u in a **bikini**
- (57) G: are you in your **pyjamas** now
- (58) G: do u **shave** your pussy area
- (59) G: u will alwaybe **dam** sexy to me

Example 53 shows that groomers imagine hypothetical scenario and fantasise about sexual scenarios with the target. The concepts of ‘shared fantasy construction’ (Mortensen, 2017: 582) and ‘postponement of pleasure’ (Mortensen, 2017: 594) were also found to be a feature in a study of flirting in online dating (Mortensen, 2017). Similarly, examples 54, 55, 56, 57 and 58 also seem to lead to fantasies and thus seeking sexual gratification. Groomers ask about the target’s clothes, underwear, and requests pictures in minimal clothing (example 56), pre-empting this request with *don’t get mad*, which suggests a level of awareness of the inappropriateness of the situation. Shaving is another thing groomers enquire about (example 58). Example 59 shows that groomers intersperse developing deceptive trust with seeking sexual gratification, specifically praising the target with sexual gratification, by making the compliment sexual. All of these are linguistic realisations of the implicit desensitisation grooming intention in the OGDM.

4.3.2.3 Physical Contact

The strategy of physical contact with the target includes 12 keywords, which are different forms (e.g., other verb tenses) of *snuggle*, *kiss*, *hug*, *cuddle*, and *tickle*. They are not overtly sexual, but still connected to the seeking sexual gratification intention, as they are inappropriate physical contact between an adult and a child unknown to each other. The following are illustrative examples:

- (60) G: cant wait to meet you either to hold you and **kiss** you
- (61) G: are u a good **kisser**?
- (62) G: So we can **cuddle** all night?
- (63) G: I would love to **hug** u right now

- (64) G: because I'd rather **snuggle** with you
- (65) G: ok it might **tickle** you at first

Example 60 refers to a meeting between the groomer and their target. The following extract 4.1 is a longer extract of this interaction:

Extract 4.1

[Context: The interaction starts roughly 440 conversational turns into the overall conversation. The groomer and target are discussing missing family members when they move. The target has just referred to an email from their brother who moved away.]

Turn	Speaker	Message
1	g	yeah , you tend to miss family when they move
2	t	ya
3	g	you will probably get homesick when you go to collage
4	t	i guess
5	t	but it will be cool if I was with you
6	g	yeah
7	g	that will be nice
8	t	ya
9	t	I ca n't wait to meet you .
10	g	I cant wait to meet you either
11	g	to hold you and kiss you
12	t	Yeah , that is gon na be so nice
13	g	to walk with you , sit down and talk with you
14	t	Yeah , that will be orig: soooo so nice
15	g	yes it will
16	g	and maybe you can sing for me ?

The groomer and target discuss missing family members after they move away in turn one, contributing to the developing deceptive trust grooming intention. In turn three, the groomer suggests that the target will feel homesick when they go to college. The target uses this as a segue to introduce wanting to be with the groomer in turn five (*but it will be cool if I was with you*). They reiterate that they cannot wait to meet the groomer in turn nine. The groomer agrees in turn 10 (*I cant wait to meet you either*) and then uses a quality approximator *hold you and kiss you* to introduce implicitly sexual themes while also contributing to the arranging further contact grooming intention. He uses a push-pull structure here, as turn 13 is much more focused on other planned activities, such as walking and talking, driving the developing deceptive trust intention forward, rather than the seeking sexual gratification intention.

This is more evidence for the interconnectedness of seeking sexual gratification and developing deceptive trust, also connected to arranging further contact. The other examples also show implicit desensitisation, as the groomers use inappropriate physical contact with the target as implicit desensitisation. *Cuddle all night* (example 62) expresses the desire to physically touch the target but also implies a close bond with the target since people generally only want to touch or kiss a person they know and trust and this behaviour is more typical of a romantic relationship, where love rather than lust is emphasised. The same is true for *snuggle*, which also expresses this inappropriate desire to touch the target (example 64). All of these words have an element of developing deceptive trust to them and show how the groomer strategically uses implicitness to desensitise the target and normalises the desire to touch them. The interconnectedness of seeking sexual gratification and developing deceptive trust is at play again.

4.3.2.4 Contraception

The fourth broad strategy is connected to contraception, which is part of the explicit desensitisation strategy groomers use to gain sexual gratification. This is used to either introduce contraception into the conversation (example 66) or directly related to a planned offline meeting with the target involving sex (examples 67-68).

- (66) G: you're not on the **pill** ?
- (67) G: do i need to bring **condoms** tonight
- (68) G: i will **wear** condom just to be safe

Wear also refers to the target's clothes (top collocates are skirt, bra, panties, jeans, shorts, thongs) specifically what they will wear when meeting the groomer (*U gon na wear your red thong tomorrow ?*), also contributing to seeking sexual gratification and arranging further contact. These discussions about contraception are another specific linguistic realisation of the explicit desensitisation intention.

4.3.2.5 Other

The last four keywords within the seeking sexual gratification intention are *privates*, *blanket*, *tummy* and *embarrassed*. The first of these is used to educate the target about genitalia and sex. It is used by fewer than four groomers and will not be analysed further. *Blanket* is used by the groomer to talk about a planned offline meeting with the target that involves sex:

(69) G: We would have a **blanket** over us while we make out and make love

Tummy is used to describe sex acts the groomer plans on performing on the target, descriptions/instructions to masturbate, and feelings of arousal:

(70) G: okay and your **tummy** is still all jittery

The use of *tummy* instead of *stomach* is interesting, as it shows the groomer adapting their language so that the target understands, almost infantilising them. It is reminiscent of how an adult would talk to a child and the concept of ventriloquizing (Goffman, 1974), which means clear discourse roles have been assigned; the groomer is the teacher, and the target is the pupil.

Embarrassed is used to mitigate discussions of sexual content. The groomer states that there is no need to be embarrassed or asking the target if they are embarrassed. This addresses the target's negative face needs of not wanting to be imposed upon. By talking about sex, the groomer threatens the target's negative face. Discussions about sex are impositions, as this is highly personal and sensitive. The groomer tries and softens this perceived negative face threat by using vague language or hedging, which strengthens the deceptive bond between the two. It also conveys a camaraderie, which contributes towards mental isolation and shows a connection between seeking sexual gratification and assessing risks and isolating. The groomer attends to the target's positive face needs of wanting to be accepted and liked and provides sympathy, while attempting to sever the target's ties with their support network. The following is a typical example of the former:

(71) G: baby, I don't want you to be **embarrassed** to talk like that ok

The groomer addresses both negative and positive face needs in example 71. He increases his interest in the target by calling them *baby*. In this case, the vague expression *talk like that* is used to describe discussions about sex and is a minimisation of the imposition, a negative politeness strategy. The groomer reassures the target and adds a tag question at the end of the sentence to emphasise his concern, fulfilling the target's positive face needs.

Overall, the seeking sexual gratification grooming intention is well represented in the keywords. Four strategies emerge, which show how seeking sexual gratification is realised by groomers. These are sex acts, physical contact, clothes/appearance, and contraception. The analysis shows that groomers strategically use both implicit and explicit desensitisation to gain sexual gratification and introduce sexual themes. Some implicit keywords are *tingly, excited, massage, gentle* and more explicit ones are *foreplay, doggie, orgasm, oral, swallow, sexually*. Groomers also gain sexual gratification as part of using a push-pull structure, in which they use an explicit term alongside a vague language term. This simultaneously advances implicit and explicit desensitisation and often features developing deceptive trust. The two grooming intentions are the strongest interlinked intentions in this study. Groomers also assign discourse roles to themselves (teacher) and the target (pupil) when presenting sex as beneficial and reframing the sexual content. Physical appearance and clothes the target wears are used to gain sexual gratification by fantasising about the target and creating hypothetical future scenarios in which the groomer and target spend time together and have sex. This imagined shared fantasy and postponement of pleasure are a strategy used in online flirting (Mortensen, 2017). Groomers request photos of the target wearing minimal clothing, ask them if they shave and use praise and sexual compliments, which contributes to both developing deceptive trust and seeking sexual gratification and are specific linguistic realisations of the sexual gratification grooming intention.

The third strategy is related to inappropriate physical contact and touch between the groomer and his target. *Snuggle, kiss, hug, cuddle, and tickle* are all used by groomers to emphasises the (deceptive) romantic relationship between them and their targets. People who like each other are more likely to kiss, hug or cuddle. This strategy is

connected to discussions about family members moving away and homesickness in one instance, prompting one target to state they wish to be with the groomer. The groomer uses a push-pull structure to simultaneously advance developing deceptive trust and sexual gratification. In this instance, physical contact is linked to both seeking sexual gratification and arranging further contact between the groomer and his target. All of the words have an element of inappropriateness attached to them, but they are implicit, as they are not explicit sex acts. Groomers strategically normalise the desire to touch their targets.

Contraception is the fourth strategy that emerges within seeking sexual gratification. This is part of explicit desensitisation and is used to introduce contraception into the conversation or discuss a planned offline meeting which involves sex. These discussions are specific language realisations of the sexual gratification grooming intention. Four more keywords are also found to contribute towards seeking sexual gratification without belonging to the four strategies above. They are *privates*, *blanket*, *tummy* and *embarrassed*. *Privates* is only used by four groomers to educate their targets, again showing the assigned discourse roles of teacher – pupil. *Blanket* is used to talk about a planned meeting involving sex. *Tummy* is used by groomers to describe sex acts and feelings of arousal and to instruct the target to masturbate. Using *tummy* instead of stomach shows that groomers infantilise targets, ventriloquize them and assign teacher roles to themselves and pupil roles to their targets. *Embarrassed* is used to negotiate discussions about sex and address the target's negative and positive face by using vague language and positive politeness strategies. All of these keywords add nuance to the grooming intentions and are specific language realisations of the seeking sexual gratification intention.

4.3.3 Arranging Further Contact

There are four strategies within the arranging further contact grooming intention (see Table 13). This intention is similar to the approach grooming intention in the OGDM. The most prevalent of these strategies is establishing offline contact, with 15 keywords. This is likely due to the nature of the data, which was collected by PJ volunteers acting as decoys and chatting to online groomers that specifically arranged offline meetings with these decoys and showed up at the decoy's supposed address.

The further contact by phone strategy has five keywords. Referring to a specific day or time for further contact has four keywords and online further contact is represented by one keyword. The strategies will be explored in more detail next.

Table 13: Four strategies and their keywords in arranging further contact

Strategy	Keywords
Offline	Mapquest, Daytona, motel, cedar, mikes, mall, bowling, Park, Theater, gas, nervous, Sat, noon, thur, thurs,
Phone	313, 714, cal, voice, anytime
Day/Time	Tues, 9pm, Sunday, Saturday
Online	Yahoo.com

4.3.3.1 Offline

The first and most prevalent of the strategies is offline communication with the following keywords: *Mapquest, Daytona, motel, cedar, mikes, mall, bowling, Park, Theater, gas, nervous, sat, noon, thur, thurs*. This strategy is closely aligned with the approach part of the discourse model of online grooming, as it relates to discussions about a face-to-face meeting with the target. It is also related to the exact location of the target and groomer, which is part of exchanging personal information within the developing deceptive trust grooming intention, as the following illustrative examples show:

- (72) G: I need your address and stuff so i can **mapquest** my way up there
- (73) G: Which side of **Daytona** do you live on ?
- (74) G: im south between **bowling** green and hopkinsville
- (75) G: what part of **cedar** ?
- (76) G: that would work meet at the **gas** station or the park or something

The examples show how arranging further contact and developing deceptive trust are connected. In the original OGD, developing deceptive trust positively correlated with all other intentions. Other keywords are more closely connected to details of the planned offline meeting, such as the venue:

- (77) G: or you can come to the **motel** what ever you want to

(78) G: let 's say , for example that I came down there last weekend ok?
and we met at the **mall**

(79) G: i just want to see ugo for a walk in a **park** or something ...

The groomers also discuss planned activities at these offline meetings (examples 80-81):

(80) G: i would like to go eat n go to the **theatre**

(81) G: yea we could just hang out go **bowling** or just chill just hang and be
or even just snuggle or kiss or whatever

Other considerations are also mentioned (example 82-84):

(82) G: Where can i **park** my truck out of sight ?

(83) G: ok i will buy u **mikes** hard lemonade then

(84) G: i get it I 've been **nervous** ever since you said they were going away
fro their anniversary

Lastly, time and weekdays connected to offline meetings are discussed (examples 85-88):

(85) G: i won't get there till **sat**

(86) G: and it will be on a **thurs** and i could stay a lil longer in the morning

(87) G: ok let me know if we can go for **thur**

(88) G: how about 1200 **noon** so u can have ur lunch

In addition to groomers seeking face-to-face meetings with the target, they attempt to establish other contact with the target as well, which is a new aspect of the arranging further contact grooming intention.

4.3.3.2 Phone

One of these other means of further contact is talking on the phone.

This strategy is arranged via the following keywords: *313*, *714*²³, *cal*, *voice*, *anytime* as in these illustrative examples:

- (89) G: the wife left today so if you want call me later **313** 461 4712
- (90) G: ok **714** 318 1285 tell me when u calling
- (91) G: u can **cal** to my mobile also
- (92) G: well it would be nice to hear your **voice**
- (93) G: u can call **anytime** u want to baby

The groomer provides a phone number (examples 89, 90) and different options of calling the target, e.g., calling them on their *mobile* (91) and *anytime* they want to (93). He also give reasons for wanting to talk on the phone (example 92) and instructs the target to let him know in advance when the target will be calling (example 90). *Anytime* specifically is used to reassure the target and create deceptive trust, while arranging further contact via phone (see above) or offline (*you can come here anytime you want to*).

4.3.3.3 Day/Time

The next strategy pertains to weekdays and times, which is related to planning online, offline, and further contact with the target on the phone, such as:

- (94) G: you can call me anytime on **tues**
- (95) G: I am usually on after **9pm**
- (96) G: and if I came next weekend on a **saturday** or **Sunday** ?

Examples 94-96 show that the groomer tries to give the target control of the situation and as much freedom as possible, addressing their negative face needs, leaving times and days vague, such *as anytime on tues* and *after 9pm* or giving multiple options like *Saturday or Sunday* for the target to choose from. This is similar to arranging further contact by phone above and seems to be connected to negotiating power dynamics.

²³ The two area codes 313 and 714 are for Detroit (Michigan) and Anaheim (Michigan). The PJ Foundation carried out several sting operations in the Michigan area in 2008. For these, physical addresses were rented, where the groomers thought they were meeting their targets but instead met PJ volunteers and were arrested.

4.3.3.4 Online

The last strategy only has one keyword (*@yahoo.com*), which is likely due to the fact the entire interaction is taking place in an online environment. This keyword is about exchanging email addresses for further online contact, such as exchanging photos (see longer extract 4.2) and other personal information (example 97):

Extract 4.2

[Context: The groomer and target are exchanging photos and email addresses]

Turn	Speaker	Message
1	g	found one other pic
2	t	whats u email
3	g	[groomer username] @ yahoo.com
4	t	k
5	t	k sent

In Example 97 the groomer mitigates the perceived face threat of a command of sending an email by adding *no pressure* and the tag question *ok*:

(97) G: send email to me at [groomer username] @ **yahoo.com** with ur address ok no pressure

It also shows the groomer convincing the target to use multiple online platforms and means of communication.

As mentioned, language realisations of the arranging further contact grooming intention have not been analysed in detail since the approach intention in the OGDM only referred to instances of offline contact. The above has shown that in addition, groomers attempt to keep the conversation going by talking to the target on the phone (*313, 714, cal, voice, anytime*), arranging to talk further online (*@yahoo.com*) and across different platforms. There might be an element of ensuring that the conversation cannot be traced easily, because the target and groomer are using multiple platforms, which shows a connection between arranging further contact and assessing risks and isolating. When arranging to talk on the phone and specific days and times, the groomers give the targets several options to choose from and try to give the target

control of the situation, which is an element of negotiating power dynamics. This suggests that all three grooming intentions have a connection. Offline meetings are discussed by asking for the target’s exact location (*Mapquest, Daytona*) and asking about specific landmarks near the target (*bowling, cedar, gas*). Planned activities at these offline meetings are also discussed (*motel, bowling, theater, mall*).

4.3.4 Assessing Risks and Isolating

The assessing risks and isolating intention has six keywords and four strategies associated with it (see Table 14). Secrecy and supervision are represented by two keywords each, while the patterns of replacing and criticising a parental figure have one keyword each. The former two can be connected to the physical isolation intention in the OGDm and the latter two are examples of mental isolation. The strategies show how these grooming intentions are linguistically realised by groomers and will be analysed in more detail next.

Table 14: Four strategies and keywords in assessing risks and isolating

Strategy	Keywords
Secrecy	Sneak, neighbors
Supervision	Grandma, aunt
Replace parental figure	tuck
Criticise parental figure	moms

4.3.4.1 Secrecy

The first of these strategies is that of secrecy with two keywords (*sneak, neighbors*), which are used by the groomer to emphasise that the groomer-target relationship should stay a secret. This is connected to planned offline meetings and confirms that there is a connection between arranging further contact and assessing risks and isolating:

- (98) G: u sure babe wont your **neighbors** say something to grandpa's ?
- (99) G: what if ur **neighbors** see me ?
- (100) G: so would you **sneak** out when am there
- (101) G: could u **sneak** out without being caught

Examples 98 and 99 show the groomer adapting low-risk behaviour and making sure they are not caught by asking the target specifically about their neighbours. The groomer attempts to convince the target to sneak out to meet them (examples 100-101), implying that they better not wake up anyone else in the house and emphasising the secrecy of the planned meeting. These examples add more detail to the assessing risks and isolating intention because neighbours seeing the groomer could be a risk factor. The groomer asking the target to sneak out is also a new linguistic realisation or strategy of the physical isolation strategy.

4.3.4.2 Supervision

The second strategy relates to supervision and making sure the target is unsupervised while chatting with the groomer and keeping the interaction a secret. The following are illustrative examples:

- (102) G: so when does your **grandma** come home
- (103) G: so u better erase it before **aunt** gets home
- (104) G: do u know about ur **grandma** schedule ?

The first example (102) shows that groomers make sure the target is unsupervised. Groomers also ask about the target's parents' or guardian's schedule (104) to arrange alone time with the target and further isolate them mentally from their support network. Example 104 shows the groomer's need to make sure the conversation is kept from parents or guardians who could become suspicious and alert the authorities. This realisation is very similar to examples in the OGDGM but adds more nuance as to who the support network of the target might be. In the above cases, it is not just parents or guardians but other family members (aunt, grandmother) that are considered.

4.3.4.3 Replace Parental Figure

Another keyword that is connected to assessing risks and isolating is *tuck*. This is used by the groomer to express the wish to tuck the target in at night (*need me to tuck you in ?*). This expresses a sense of caring for the target, so building deceptive trust, while being connected to the strategy of assessing risks and isolating in the form of desiring to replace a parental figure in the target's life. This is a form of emotional isolation,

whereby the groomer attempts to replace and become the target's support system. The following are illustrative examples:

- (105) G: aww need me to **tuck** you in?
- (106) G: just have to come **tuck** you in lol
- (107) G: want me to cum **tuck** u in ?? lol

The *need* in example 106 shows that the groomer thinks the target is dependent on him, helpless without him and he assumes a paternal role. Example 107 suggests that the groomer may also gain sexual gratification from the idea of tucking in the target, although the typo of *cum* instead of *come* might not be intentional. *Tuck* is a newly identified linguistic realisation of the mental isolation grooming strategy.

4.3.4.4 Criticise Parental Figure

The last keyword in the assessing risks and isolating grooming intention is *moms*, which is used for different isolation purposes, such as criticising the target's parental figures (*who cares what her bf wants u should be ur moms # 1 in everything*) and finding out their schedule (*whats your moms schedule*) to arrange alone time with the target. The groomer also talks about their own mom, contributing towards developing deceptive trust and to a lesser extent uses *moms* for sexual gratification purposes (*ever see dad kiss moms tits*) showing that groomers use assessing risks and isolating in conjunction with both developing deceptive trust and sexual gratification.

The analysis shows that four strategies emerge within the assessing risks and isolating grooming intention. Secrecy and supervision are part of the physical isolation grooming strategy, whereby the groomer ensures the target is not supervised, the interaction is kept secret and the groomer arranges alone time with the target. This physical isolation is done by explicitly worrying about the neighbours and asking the target to sneak out to meet the groomer, ensuring privacy and alone time. This is connected to and adds further nuance to the arranging further contact grooming intention in the OGDM. The groomer also adapts low-risk behaviours, which suggests negotiating power dynamics is connected to assessing risks and isolating and arranging further contact. Supervision is the second strategy, which is also used to physically isolate the target. Groomers make sure the target is unsupervised during

their conversation, further isolating them and ask about the schedules of the target’s parents, guardians, and other family members around them to ensure the conversation is kept secret. Mental isolation, through which the groomer tries to separate the target from their support network and even tries to replace it and prevents being caught by convincing the target to keep the interaction a secret, is connected to the other two strategies, replacing, and criticising parental figures. *Tuck* is used by groomers to express they would like to tuck in the target at night. This shows they wish to replace the parental figure and think the target is dependent on them. There is evidence that groomers might also gain sexual gratification from imagining tucking their target in at night, showing a connection between mental isolation and seeking sexual gratification. The strategy connected to criticising the parental figure. The keyword, *moms* is quite versatile and is used for both mental and physical isolation. Additionally, *moms* is used to advance the developing deceptive trust and seeking sexual gratification grooming intentions.

4.3.5 Negotiating Power Dynamics

Lastly, there are three keywords contributing towards the negotiating power dynamics intention (Table 15), which are connected to one strategy of legality. This grooming intention emerges differently in Lorenzo-Dus et al. (2016), where it is called Compliance Testing (see chapter two, section 2.2.1.2) as it is not connected to the concept of legality, but rather to seemingly giving the target control, adopting low-risk behaviour, and challenging the target to overall gauge how likely the target is to engage in behaviour proposed to them by the groomer, which do not appear in the keywords. This finding suggests that negotiating power dynamics might be realised differently than the other grooming intentions and warrants a closer look at how it is achieved and what specific linguistic realisations are done to drive it forward. This will be done in chapter six (section 6.1).

Table 15: Strategy and keywords in negotiating power dynamics

Strategy	Keywords
Legality	Jail, cops, cop

4.3.5.1 Legality

The strategy within negotiating power dynamics is connected to legality, more specifically legal repercussions, and possible undercover sting operations. The following are illustrative examples:

- (108) G: because i can go to **jail** just for talking to u sweetie
- (109) G: and will feel better once I know **cops** wont be hiding in ur house
- (110) G: ur not a **cop** of anykind

In example 108, the groomer adopts a low-risk attitude and warns the target that he can go to jail if they continue talking and are not careful. Examples 109 and 110 are related to undercover sting operations. The groomer believes the target to be working for the police, which explains their low-risk behaviour. This low-risk behaviour is a linguistic realisation of the strategy of role reversal, as identified in the OGDM. Groomers also repeatedly accuse the target of being a cop or seek reassurance that they are not (110). This is broadly connected to making sure the target is actually underage. This specific realisation of negotiating power dynamics did not emerge in the original OGDM. A CADS analysis shows that online groomers mention sting operations and the possibility of the targets being undercover police officers or working with the police while two of the three strategies in the OGDM (reverse psychology and strategic withdrawal) do not appear in the keywords.

The analysis above shows that negotiating power dynamics is realised differently than it was in the OGDM, and possibly differently to the other grooming intentions. It is connected to the strategy of legality with three specific keywords used by groomers: *jail*, *cops*, *cop*. These are related to the repercussions if the groomer fails to keep the interaction secret and gets caught. Groomers thus adopt low-risk behaviour, which is an example of role reversal from the original OGDM. Sting operations and the possibility that the target might be an undercover police officer are also mentioned by groomers.

4.4 Conclusions on the Features of Online Grooming

All in all, groomer language emerges as distinct compared to a general digital language corpus with all grooming intentions identifiable at the keyword level. The top keywords are highly frequent compared with PAN2012 and they display online grooming intentions. This suggests that the OGDM is a robust model with the same grooming intentions identifiable using a larger dataset and different methodology. Some specific linguistic realisations of grooming intentions also appeared in the above analysis, which add nuance and detail to the grooming intentions and strategies originally identified in the OGDM.

In terms of developing deceptive trust, four strategies are identified: relationship terms, small talk, bonding, and exact location. A different focus to the intention in the OGDM emerges. Rather than sexual and non-sexual compliments, which are difficult to identify on the one-word level, groomers use relationship terms (*girlfriend, sweetie*) and terms of endearment to negotiate the groomer-target relationship. These terms are often interspersed with other developing deceptive trust strategies to maximise the success of gaining the target's trust and building a deceptive bond with them. *Precious* and *adventure*, on the other hand, are used to avoid using explicit relationship terms and to address the perceived threat to the target's negative face.

A specific language realisation of the relationship strategy is identified in the form of bonding with the target. Keywords like *smile, smiling, mad dreaming, and promise* are used by groomers to express their feelings and affection towards the target to advance their relationship, which also addresses and fulfils the target's positive face needs of wanting to be liked by others. A few keywords have been identified as linguistic realisations of small talk (*Laundry, nada, bills, chattin*). Some keywords (*snack, pee, shower*) are used by groomers for small talk purposes but are used to gain sexual gratification in specific circumstances. Additionally, within developing deceptive trust, a specific focus on the exact location of the target emerges. This is one linguistic realisation of the exchange of personal information strategy within Developing deceptive trust and is done by using keywords *Detroit, beach, addy, house* to ask for the target's exact location. Developing deceptive trust and seeking sexual

gratification are closely linked, which the collocational unit *thinking + about* shows. The groomer imagines hypothetical scenarios with the target that are sexual.

Seeking sexual gratification emerges as prominent in the keyword analysis, compared with general digital language. Four strategies emerge within this grooming intention: sex acts, physical contact, clothes/appearance, and contraception. Groomers use both implicit (*tingly, excited, massage, gentle*) and explicit desensitisation keywords (*foreplay, doggie, orgasm, oral, swallow, sexually*) strategically to gain sexual gratification and introduce sexual themes into the conversation. A push-pull structure emerges in some of these keywords, which entails the groomer being explicit (push) and implicit (pull) in the same turn or subsequent turns. This push-pull structure also features the developing deceptive trust, showing how the two intentions are interlinked. Reframing is also present in the keyword analysis above. Groomers present sex as beneficial to their targets and assign clear discourse roles: Teacher (groomers) – pupils (targets).

The strategy of inappropriate touching is used to emphasise the deceptive relationship with the target, rather than the sexual side while normalising the groomer's desire to touch the target inappropriately. This is done by using the keywords *snuggle, kiss, hug, cuddle, and tickle*. References to the target's appearance using the keywords *pyjamas, thongs, bras, cloths, panties* are also used to gain sexual gratification by fantasising about the target and imagining future meetings with the target that involves sex. This shared fantasy and pleasure postponement are also strategies found in online flirting (Mortensen, 2017). Groomers use the keywords connected to the target's clothes and appearance to request photos of them wearing minimal clothing. Sexual compliments and praise also appeared within the seeking sexual gratification grooming intention. The last strategy within seeking sexual gratification is contraception with the keywords *pill, condoms, and wear*, which are used to discuss future meetings involving sex and to introduce contraception into the conversation. These keywords are a specific linguistic realisation of the explicit desensitisation strategy identified in the OGDM. Four more keywords that are not aligned to the four strategies discussed above are *privates, blanket, tummy* and *embarrassed*. *Blanket* is used to discuss an offline meeting involving sex with the target, which is linked to the arranging further contact grooming intention, while *tummy* is used to describe sex acts

and instruct the target to masturbate. These two keywords are both linguistic examples of the explicit desensitisation strategy within the OGDM.

Arranging further contact is an addition to the approach intention in the OGDM, which focused on arranging offline meeting. The keywords in the above analysis show that there are other elements to this grooming intention, which can be sorted into three additional strategies: phone, day/time, online. Groomers keep the conversation going by arranging to talk to the target on the phone (*313, 714, cal, voice, anytime*), switching platforms (*yahoo.com*) and arranging the logistics of further contact offline and online (*Tues, 9pm, sunday, saturday*). They give their targets vague date and time arrangements and options to choose from to supposedly give them control over the situation, which is an aspect of negotiating power dynamics. Groomers also might want to switch to other platforms to make the conversations with their targets more difficult to trace and avoid being caught, which suggests physical isolation is also connected to this arranging further contact grooming intention. Offline meetings are discussed by asking for the target's exact location (*Mapquest, Daytona*) and specific landmarks near the target (*bowling, cedar, gas*) and planned activities (*motel, bowling, theater, mall*).

Four strategies also emerge within the assessing risks and isolating grooming intention. These are: Secrecy and supervision as part of the physical isolation strategy and replacing the parental figure and criticising the parental figure, which are connected to the mental isolation strategy. The groomer uses the keywords *sneak* and *neighbors* to make sure their time with the target is not observed and to convince their target to sneak out to meet them. This is also closely connected to the arranging further contact grooming intention and the negotiating power dynamics intention, as the groomer adopts low-risk behaviour. Keywords *grandma* and *aunt* are used to ensure the target is not supervised while the groomer and target are interacting and to make sure the interaction is kept secret. The mental isolation grooming strategy is done by using the keywords *tuck* and *moms*. *Tuck* refers to groomers wanting to tuck their target in at night, replacing their parental figure. Groomers also express that the target is dependent on them, and they might gain sexual gratification from this fantasy of tucking the target in, suggesting a connection between mental isolation, and seeking sexual gratification. The keyword *moms* is used for both mental and physical isolation

purposes. The groomer asks about the schedule of the target's mother, talks about his own mother, and criticises the target's mother. Developing deceptive trust also seems to be connected to mental isolation. The groomer increases his own bond with the target while isolating them and weakening other relationships.

Negotiating power dynamics behaves differently to the other grooming intentions. Only one of the three strategies identified in the OGDM, namely role reversal, appears in the above analysis. Negotiating power dynamics is connected to the strategy of legality with three keywords: *jail*, *cops*, *cop*. This strategy is connected to the repercussions of the groomer getting caught. In the concordance lines, groomers also accuse targets of being part of undercover sting operations and being undercover cops wanting to catch and arrest them. Groomers adopt low-risk behaviours here, which is the one example of role reversal within negotiating power dynamics. The grooming intention might be realised differently, and it might not be possible to trace and identify it on the one-word level. This is something that will be further investigated in chapter six.

This chapter has compared a corpus of groomer language with a general digital chat language corpus to find out what the features of online groomer language are and whether they are distinct. It first analysed the top keywords of this comparison which were highly frequent and showed evidence of online grooming intentions, which suggests that online groomer language is distinct. The chapter then moved on to examining the keywords in more detail by sorting them into online grooming intentions from the OGDM. This analysis showed that all online grooming intentions are identifiable at the keyword level and that a CADS approach can add nuance to the OGDM by providing specific linguistic realisations of the intentions and strategies. The analysis will now move on to grooming duration in the next chapter.

5. Chapter 5: Duration of Online Grooming and its Impact on Discourse

As noted in the literature review chapter (chapter two, section 2.2.2), the duration of online grooming has not been studied systematically. Duration refers to the overall interaction time between a groomer and a target, measured in minutes or hours. Duration or ‘degree of contact’ (Kloess et al., 2017a: 16) has been mentioned in a few studies, which point out that there is a wide range (Briggs et al., 2011; Wolak et al., 2014; Kloess et al., 2017; Lorenzo-Dus & Izura, 2017). The duration of grooming interactions ranges from anywhere between ten minutes (Kloess et al., 2017a) to more than 12 hours (Lorenzo-Dus & Izura, 2017). Lorenzo-Dus and Izura (2017) analysed compliments in different online grooming speed categories. They divided their data into three grooming speed categories (fast, average, and slow) depending on the duration of the interaction and the number of times the groomer logged into the chat room. These speeds corresponded to the following groups: fast (under four hours), average (between five and 11 hours) and slow (over 11 hours). Lorenzo-Dus and Izura (2017) isolated all the compliments from 68 chat logs, taken from the PJ website, to find out how online groomers used the praise strategy within the developing deceptive trust process of the online grooming communication model. More specifically, they sought to answer whether compliment topics differed across groomers and what syntactic structures were used most and least often. The findings showed that slow groomers used more compliments than fast groomers with a prevalence of compliments concerning the physical appearance of the targets (Lorenzo-Dus & Izura, 2017). Groomers’ compliments were also found to support other online grooming processes, such as assessing risks and isolating and seeking sexual gratification in the form of desensitization. Importantly, Lorenzo Dus and Izura (2017) found a link between groomer speed and the number and type of compliments they use, indicating that the language of groomers who spent more time grooming differed from the language of groomers who spent less time.

Duration of online grooming has not been analysed further, showing an important gap in the literature, which might show different vocabulary use and strategy use or point to different models for different groomers. Lorenzo-Dus and Izura’s (2017) study also highlights the importance of a Discourse Analysis-based approach to online grooming communication research that can further develop the OGDm. Whether a short

grooming duration compared to a long grooming duration changes the language of online groomers e.g., the presence and salience of specific grooming intentions or strategies used has not been studied extensively. In other contexts, for example advertising language, language has to be modified when producing short advertisement clips that are up to 120 seconds long, compared to infomercials that are typically 30 to 60 minutes long. Advertisers use more varied tactics (e.g., a mixture of verbal and non-verbal techniques) and establish a personal relationship with consumers (Komar, 2015). It stands to reason that online groomers also employ different tactics with less time spent interacting with their target. Groomers might also have different motivations, such as instant gratification compared with fantasy re-enactment, which will lead them to either spend more or less time grooming their targets (O’Connell, 2003; Brigg et al., 2011; Broome et al. 2018). This chapter will therefore explore the relationship between duration of online grooming and the language use to answer the following questions:

2. Does duration of grooming influence the grooming process/intentions?
 - a. Is usage of specific words/specific grooming intentions associated with different duration of grooming?
 - b. Can different duration-based grooming profiles be established, and if so, what are the duration cut-off points?

The organisation of the chapter is as follows: The chapter starts with an overview of the corpus in terms of duration and describing the data segmentation approach in sub-section 5.1. The duration sub-corpora will also be described in the same sub-section. It will then move on to keyword analyses of these duration sub-corpora in sub-sections 5.2 to 5.4. Conclusions will be drawn in sub-section 5.5.

5.1 Duration of Online Grooming – an Overview of the Corpus

After the data was collected and prepared for analysis, as outlined in chapter three (sections 3.3.1-3.3.3), different research avenues and ways to segment the data were considered. The first one was based on the hypothesis that groomers who spend more time interacting with their targets introduce explicitly sexual topics later in the

conversation than groomers who spend less time interacting with their targets. As explicit sexual terms are arguably easier to identify than other grooming intentions and there is some evidence of different approaches, e.g., a ‘more gentle approach’ (Kloess et al., 2017a: 9) and a ‘very direct, highly sexualized approach’ (Kloess et al., 2017a: 9), the proposed procedure was to compile a list of sexual topic items and conduct a pilot study using 10% of the corpus to identify when groomers introduce the item in the different chat logs using concordance plots. Based on this, the data could then be segmented into different groups. This option was rejected because the sample size of the pilot study would have been comparatively small. There is also not enough evidence to suggest that the use of sex related terms is different across groomers. The literature also reports that sexual topics are often introduced early on with one piece of research citing 35% of their sample (n=100 chat logs) introducing sexual topics in the first 20 lines of chat (Gauz, 2014). The second option that was considered was sampling and analysing one very short chat log and one very long one qualitatively, conducting a Content Analysis and paying close attention to vocabulary and grooming intentions, which could then be used to segment the data. However, this option was also discarded, because the extant literature into online grooming is already based on qualitative approaches, so conducting a new analysis would not add anything new to the literature.

The third option, which was considered and chosen, was experimentation. The data was segmented into different duration groups, divided by one or more defined grooming duration cut-off points, and cross comparisons in the form of keyword analyses were carried out to examine a keyword list of the two or more different duration groups. As outlined in the methodology chapter, a keyword analysis is a good starting point for a CADS analysis. It helps narrow the analysis down to a representative subset, indicated by the “aboutness” of a text in the form of keywords, which can then be analysed qualitatively. If no or not enough differences were found, this process could be repeated with different cut-off points until differences were noticeable. This approach was further modified to include two duration groups (long duration and short duration) to start with. This essentially allowed for a comparative analysis of the language groomers use, their grooming intentions and how they achieve these through language.

The process for determining the duration will be explored next. A duration overview of the number of chat logs²⁴ in 100 minutes segments was produced (see Figure 21). Figure 21 was created using the metadata table, which includes manually collected information about the groomer and decoy usernames, duration of grooming interaction, time span (from first contact to arranged meeting) in days, and chat platform used, as outlined in the chapter three (section 3.3.1). As can be seen, the range of duration was quite large with interactions lasting from 17 minutes to 10,597 minutes (~177 hours). Several chat logs lasted between 17 and 100 minutes (n=55). Most chat logs in the 100-minute segments (n=101) seemed to last 100 to 200 minutes (1.6 to 3.3 hours). The range of 200-300 minutes (3.3 to 5 hours) also had a large number (n=89) of grooming cases associated with it. After this first peak, the number of interactions per 100-minute step falls steadily until the 700 to 800 range (n=14), where it very slowly increases until 1000 to 1100 minutes (n=17). In fact, 50% of interactions fall into the 0 to 400-minute range. After the second incline, the number of interactions per 100-minute steps slowly descends with a few outliers between 1500 to 1600 minutes, 2100 to 2200 minutes, 2600 to 2700 minutes, and 2900 to 3000 minutes.

Between 3000 to 3100 minutes and 10,600 minutes there are barely any interactions left. This makes up only three percent of the corpus. Duration groups could not be defined by the average duration of grooming as the standard deviation (SD=1246.46) was too high to justify selecting the average as one of the cut-off points for analysis, as Lorenzo-Dus and Izura (2017) did. It was concluded that the data could not be divided up using statistics. Instead, the average time teenagers were spending on Instant Messenger services daily during the period that the data was collected by the PJ Foundation (2004-2016) was used as an indicator for long/short grooming duration.

²⁴ 15 chat logs had to be excluded, as they followed a different chat structure and used a mixture of newer platforms and text messages making it impossible to establish the duration of the interaction.

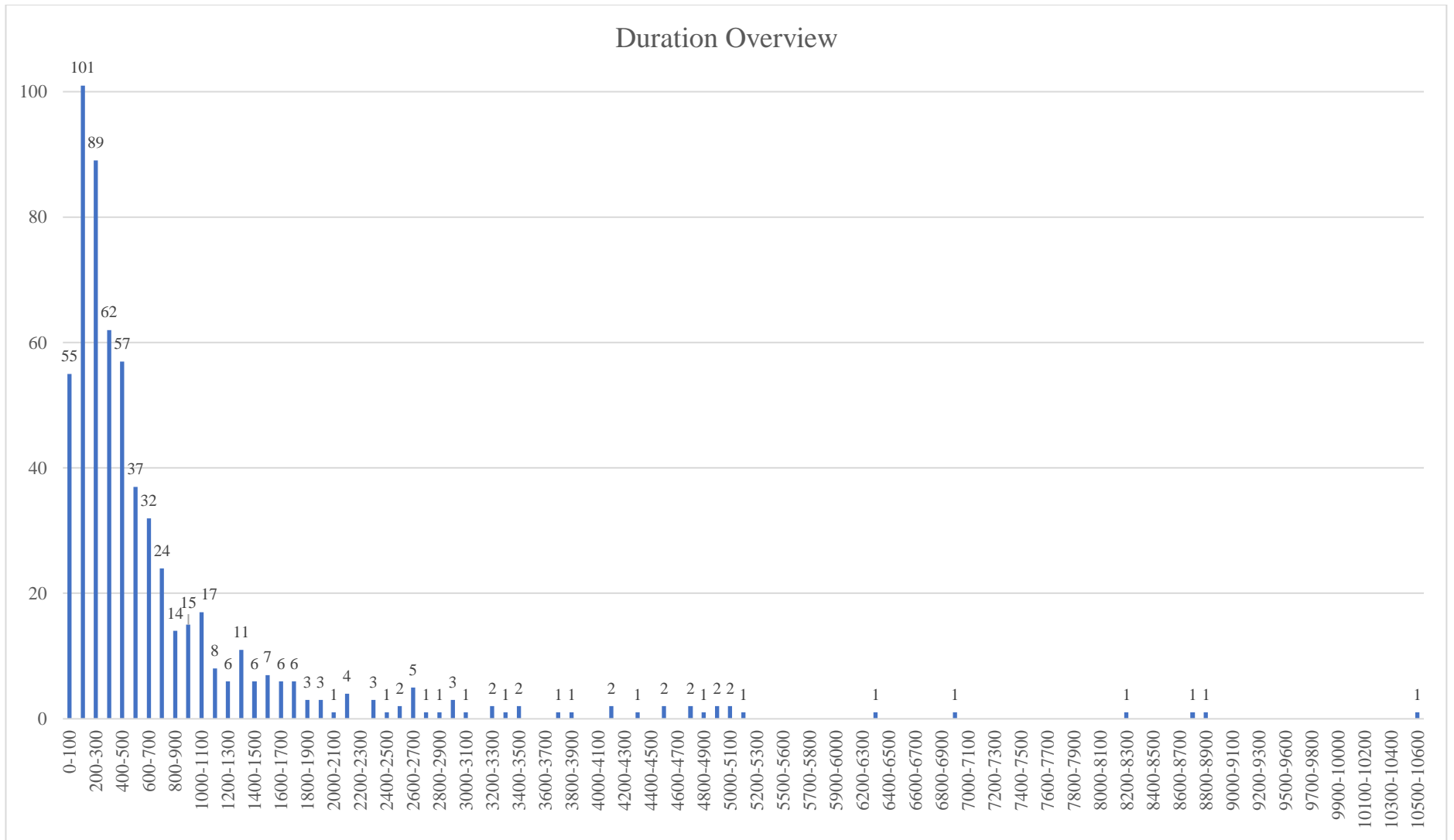


Figure 21: Duration overview - Number of chat logs per number in 100-minute steps

The average served as a baseline to compare the duration of grooming against. Anything above this average time would signal more involvement/effort by the target and the groomer to communicate. As such, anything below this cut-off point would be regarded as a short grooming duration and anything above it would be a long grooming duration.

As mentioned above, the time span of the chat logs in days was collected in the metadata table. This ranged from one day (any chat log taking place on the same calendar day, regardless of the length of time spent chatting) to 472 days (~1.2 years) with an average of 25 days and a standard deviation of 46.1. Next, a suitable daily average for US teenagers²⁵ chatting online needed to be considered, that resembled the date range (2004-2016) of the Perverted Justice website archive. A 2010 report that was conducted to find out how media affects children and how to best educate young people on media use, was chosen. It surveyed more than 2000 teenagers (ages 8-18), looking at different media types, accounting for multitasking and ‘document[ed] changes in children’s media habits since the first two waves of the study, in 1999 and 2004’ (Rideout et al., 2010: 1). Importantly, the report spanned the years 2004 to 2009, which partially covers the Perverted Justice website time span. The report stated that in 2004, teenagers spent an average of 17 minutes daily on Instant Messenger. This decreased to 11 to 14 minutes on average in 2009 (Rideout et al., 2010). The average of 11 to 14 minutes from 2009 was taken as most representative of the data and was multiplied by the average time span of 25 days (see above), which resulted in 275 to 350 minutes, which was rounded down to up to 300 minutes as the first cut-off point for the analysis. A second cut-off point of 100 minutes was chosen by re-examining the duration overview and identifying the first peak of number of grooming interactions between 100 to 200 minutes. This (100-299 minutes) was compared with the first 55 groomers (0-99 minutes). A third analysis examined two more sub-corpora with the parameters of 300-799 minutes and 799-10,600 minutes, as the number of grooming cases declines steadily after 800 minutes, as stated above. See Table 16 for an overview of the duration sub-corpora and analyses that were carried out in this chapter.

²⁵ The data taken from the Perverted Justice website was produced in the US and the chat logs use American English

Table 16: Overview of duration sub-corpora and analyses

Sub-corpus	Time span	Analysis
G1	0-99 minutes	2
G2	100-199 minutes	2
G3	0-299 minutes	1
G4	300-10,600 minutes	1
G5	300-799 minutes	3
G6	800-10,600 minutes	3

As mentioned in the methodology chapter the keywords were calculated using Log ratio with a Log Likelihood filter and a minimum of three occurrences in both corpora (chapter three, section 3.3.5.1). Dice-coefficient (Baker & Levon, 2015) was used for the computation of collocation.

5.2 One Corpus with Two Grooming Approaches?

Based on the above data segmentation, the first analysis looked at group three (0-299 minutes) and group four (300-10,600 minutes). The details of these two sub-corpora G3 and G4 can be found in Table 17.

Table 17: Details of duration sub-corpora G3 and G4

Sub-corpus	No. of usernames	No. of Words	Duration in minutes
Groomers 3(G3)	240	308,098	0-299 min
Groomers 4 (G4)	393	3,296,828	300-10,500
total	633	3,604,926	0-10,500

Table 18 shows the top 50 keywords of G3 and G4. Negative keywords, which are keywords that are more prevalent in G4, are shaded in grey. The full list of keywords (n=134) can be found in the appendix (section 9).

Table 18: Top 50 keywords of G3 compared with G4

#	Word	G3	G4	Log ratio	LL
		Freq (per mill)	Freq (per mill)		
1	l	490.1	14.01	5.13	260.23
2	muah	282.38	8.75	5.01	147.95
3	w/	129.83	7	4.21	60.54
4	wanna	460.89	26.26	4.13	211.84
5	princess	51.93	808.84	-3.96	292.12
6	you're	201.23	14.01	3.84	87.19
7	that's	94.13	7	3.75	39.9
8	ne	162.29	14.01	3.53	65.24
9	asl	116.85	12.26	3.25	43.34
10	gonna	210.97	22.76	3.21	77.27
11	}	29.21	238.1	-3.03	68.76
12	ru	113.6	17.51	2.7	34.36
13	prob	181.76	28.01	2.7	54.99
14	chill	120.09	19.26	2.64	35.42
15	child	38.95	234.6	-2.59	57.84
16	x	81.14	444.69	-2.45	103.3
17	missed	64.91	355.4	-2.45	82.5
18)	1723.48	315.13	2.45	464.86
19	ooh	152.55	28.01	2.45	40.99
20	thanx	142.81	26.26	2.44	38.33
21	address	938.01	173.32	2.44	250.96
22	zip	146.06	28.01	2.38	37.99
23	condom	194.74	38.52	2.34	49.45
24	(1515.75	313.38	2.27	371.69
25	edit	149.3	33.26	2.17	34.33
26	;	165.53	731.81	-2.14	145.12
27	dream	45.44	199.58	-2.13	39.35
28	alright	331.06	78.78	2.07	71.68
29	miss	253.17	1048.69	-2.05	196.77
30	mins	227.2	59.53	1.93	44.71
31	ain't	178.51	49.02	1.86	33.4
32	number	960.73	274.87	1.81	171.73
33	--	688.09	218.84	1.65	108.07
34	virgin	298.61	96.29	1.63	46.05
35	profile	256.41	84.04	1.61	38.7
36	<	262.9	770.33	-1.55	99.34
37	§_KISS-SMILEY_§	597.21	1666.7	-1.48	201.19
38	n	408.96	1120.47	-1.45	131.69

39	busy	223.95	611.01	-1.45	71.35
40	honey	334.31	127.8	1.39	40.17
41	nite	240.18	616.26	-1.36	65.55
42	'm	1931.2	765.07	1.34	218.88
43	finger	331.06	131.31	1.33	37.41
44	...	9928.66	4007.44	1.31	1095.22
45	mad	204.48	498.96	-1.29	48.87
46	cell	460.89	189.08	1.29	49.09
47	her	675.11	1626.44	-1.27	155.95
48	went	207.73	497.21	-1.26	47.1
49	m	334.31	140.06	1.26	34.27
50	drink	480.37	201.34	1.25	49.2

At first glance, there seem to be quite a few differences between the two corpora, which is apparent by the number of keywords (n=134), which seems comparatively high. As discussed in chapter three (section 3.2.2.1), there is no universally agreed Log Ratio cut-off point for keyword analyses. For this analysis only words with an LR higher than 0.55 will be considered. The majority (n=43) of keywords have an LR value between 0.55 and 0.99, meaning they are just under two times more common in one corpus compared with the other corpus. There are 32 keywords that are twice to three times more common in one corpus (LR value between 1.02 and 1.93) than the other corpus. Thirty keywords have an LR value between 0.11 and 0.53, which suggests they have a comparable frequency in both corpora. Eighteen keywords have an LR between 2.05 and 2.7, indicating words are around four times more common in one corpus than the other. A minority of keywords (n=11) have an LR higher than 3, indicating they are eight times as or more frequent in one corpus. The negative LR indicates that keywords belong to the corpus that occupies the reference corpus slot in this keyword analysis. If the order of the corpora were switched keywords belonging to G3 would be negative instead. This also applies to the second and third analyses in this chapter. The 104 keywords above the log-ratio cut-off point were sorted into grooming intentions by exploring the first 10 collocates and extended concordance lines, (see Table 19). They were then further analysed within their assigned grooming intention.

Table 19: Keywords of G3 and G4 sorted into grooming intentions

Grooming intentions	Keywords G3 (short)	Keywords G4 (long)
Developing deceptive trust	muah, asl, address, zip, number, profile, honey, cell, m, call, pics, cam, live, pic, house, phone, where, prob, thanx, give, tomorrow	princess, x, missed, miss, §_KISS-SMILEY_§, sweet, baby, happy, love, dream, nite, sleep, bed, went, doing, hope, wish, busy, mad
Arranging further contact	come, meet, chill, hang, number, cell, call, phone, tomorrow	-
Seeking sexual gratification	condom, virgin, finger, sex, wanna	sexy
Assessing risks and isolating	older	-
other	/w, gonna, ru, drink, you're, ur, ill, alright, l, ain't, anything, mins, ne, ooh, whatever, hmm, ta, that's, whats,) (-- ... : - ? . ,	Thinking, think, told, has, was, been, am, her, him, she, them, n, much, day, thing, this, that, the, §_SAD-SMILEY_§, } ; < >
not related to chat log	edit ²⁶	child ²⁷

As can be seen, developing deceptive trust is quite dominant in both G3 and G4. Seeking sexual gratification also seems prevalent in G3, but not in G4, with only one keyword (*sexy*) directly associated with this grooming intention. Arranging further contact, while prevalent in G3, does not seem to be salient in G4. Assessing risks and isolating is identifiable at the keyword level in G3 but does not seem to occur in G4.

The “other” category contains words that do not easily align with the grooming intentions considering their top ten collocates and their expanded concordance lines. This category comprises auxiliary verbs (*was, been, am*), abbreviations and contractions (*w/, ur, l, mins, n, ru, ne, gonna, you're, that's, ill, whats*), units belonging to words (*ta* as part of *gotta*), symbols, punctuation marks and emoticons ((-...:-

²⁶ This was used by the PJ volunteer to indicate that something in the chat log had been edited, e.g. personal information, such as addresses, phone numbers and names were taken out of the final transcript.

²⁷ In most occurrences, this was used to protect a minor’s identity and the PJ volunteer added a comment in the chat log “Child’s name removed”, which skews the keyword analysis. Only 13% of occurrences make up other instances of the word *child*.

?,§_SAD-SMILEY_§ } ; <>)), function words (*this, that, the, much, anything, whatever, alright*) and pronouns (*his, her, she, them*). Some content words (*thinking, think, told, day, thing, drink*) were also included in this category. These did not have emerging patterns, while being aligned to different online grooming intentions, as there was no one prevalent online grooming intention they were contributing towards. *Told*, for example, is used for all grooming intentions: assessing risks and isolating (*thats cool , have you told any of your g/f about me*), seeking sexual gratification (*i told u that i want to have sex*), developing deceptive trust (*u never told me ur b_day*), arranging further contact (*what if i told u im coming to gr tomorrow*) and negotiating power dynamics (*I just want you to tll me what u would like to instead of just saying what ever I told u to do*) without any overarching intention. *Drink* is also used for several grooming intentions without any clear patterns: negotiating power dynamics (*u are to young to drink*), developing deceptive trust (*I want something to drink , but I ca n't decide what ?*), seeking sexual gratification (*drink and have sex .its that ok ?*), arranging further contact (*sweet we could drink and play*). The key words aligned to the other grooming intentions will be analysed in more detail next.

5.2.1 Developing Deceptive Trust

In developing deceptive trust, four strategies emerge in the short grooming duration sub-corpus G3 (Table 20) which are exact location with seven keywords, visual information with four keywords, mitigation/negative politeness and relationship strengthening with two keywords each.

Table 20: Four strategies in G3 within developing deceptive trust

Strategy	Keywords
Exact location	asl, address, zip, house, live, where, give
Visual information	profile, pics, cam, pic
Mitigation/Negative politeness	prob, thanx
Relationship strengthening	muah, honey

Three strategies appear in the keywords of the long grooming duration sub-corpus G4 (Table 21), which are relationship strengthening with 11 keywords, activities with three keywords and sociability with five keywords.

Table 21: Three strategies in G4 within developing deceptive trust

Strategy	Keywords
Relationship strengthening	princess, x, missed, miss, §_KISS-SMILEY_§, baby, mad, happy, love, hope, wish
Activities	busy, went, doing
Sociability	sweet, dream, nite, sleep, bed

5.2.1.1 Exact Location

The first strategy in the short grooming duration sub-corpus G3, within developing deceptive trust is personal information sharing, focused on the exact location of the target. The following keywords contribute to this pattern: *asl*²⁸, *address*, *zip*, *house*, *live*, *where*, *give*. All of them appear in the first 25 keywords (see Table 18) and have an LR score between 2 and 4, meaning they are four or eight times as frequent in G3 compared with G4. Their top ten collocates (eight in the case of *asl*) confirm that these words are used to elicit location related information, which can also be gleaned from the following illustrative examples, in which G indicates the groomer and T the target talking²⁹. Short duration groomers seem to be very focused on knowing the exact location of their targets at different points of the conversation:

(1)

G: good and u? ... **asl** again

T: 13 f Michigan

(2) G: i need an **address** so I know how to get there

(3) G: what's ur **zip** code ?

²⁸ *Asl* is an abbreviation for “age, sex, location?”. It is frequently used at the beginning of a chat log to elicit personal information (Chiang & Grant, 2018).

²⁹ This will also be the case in all subsequent examples.

Example one illustrates how a groomer typically starts a conversation with a target. In this case, he reinitiates contact about two weeks after first chatting to the target and asks about their age, sex, and location again. It is important to note that groomers are aware of the fact targets are underage and this is established very early on in the interaction. The statement in example two, which implies that the groomer has already arranged a meeting with the target, is worded as a polite request: the groomer provides a grounding move (*so I know how to get there*) in support of his request, thereby minimising the perceived imposition on his interlocutor's face needs. However, this type of information elicitation varies vastly in terms of both politeness and directness in the corpus. Groomers use questions and demands with different levels of insistence, which threatens the target's negative face of not being imposed upon. One groomer asks their target for their address nine times in the space of 300 lines of conversation, which is an unmitigated threat to the target's negative face needs. The eighth occurrence becomes slightly more assertive, while the tenth includes intimidating language that borders on a threat:

(4) Now TELL ME YOUR ADDRESS jonalyn and NO EXCUSES!!!

In example four, the groomer uses two consecutive imperatives (*tell me* and *make no excuses*), capitalises most of his message, which is the equivalent of shouting in verbal language, and uses exclamation marks to underline his message. Additionally, he uses the target's first name, which might evoke an authority figure's reprimand, e.g., a mother or a teacher using a first name rather than a nickname to indicate the severity of the situation. The third example above is similar to the first one, in terms of politeness. The remaining keywords (*live, give, house, where*) are not as characteristic of the language of short grooming duration (G3), as their LR is between 0.85 and 0.57. However, they appear up to twice as often in G3 compared with G4 and continue the above pattern of eliciting exact location. The following are illustrative examples:

(5) G: Ur not stupid, but I need to know where you **live**

(6) G: Remember u still got ta **give** me the address so i can get there

(7) G: So back to the location at ur **house**

(8) G: I have to first find out **where** u live lol

Live is used by the groomer to identify the exact location of the target in example five, but also which type of accommodation they reside in e.g., a house or flat. Additionally, groomers attempt to find out with whom the target lives, which also contributes to the intention of assessing risks and isolating. The keyword *house*, while being used by groomers to elicit location details from the target, is also closely linked to arranging further contact with the target (*Yea, but not at your house. Somewhere else*). The keyword *give*, is also used to extract other personal information from the target (e.g., their phone number) and is connected to seeking sexual gratification with *head* and *blowjob* being among the top ten collocates. *Where* is the least straightforward keyword out of the four, however it clearly is connected to the location elicitation pattern, as example eight shows. The examples above echo the persistence of groomers to find out the location of the target. Short duration groomers tend to use fixated discourse about this specific aspect of the developing deceptive trust intention. Fixated discourse was found by Bogdanova et al. (2014).

5.2.1.2 Relationship Strengthening

In contrast, the focus on the exact location of the target seems to be largely absent in the keywords of long duration groomer language (see Table 21). For instance, in G4 *asl* does not have any collocates and only appears seven times. While the other keywords have similar collocates and seem to be used somewhat similarly to the way they are used in G3, they are much less frequent. Instead, there seems to be a focus in the G4 corpus on relationship strengthening, which is similar to the bonding strategy in chapter four (section 4.3.1.3). Both are used to strengthen the relationship, but bonding is a more specific linguistic realisation of this. The relationship strengthening strategy in this chapter results from the following keywords: *princess*, *x*, *missed*, *miss*, *\$_KISS-SMILEY_\$*, *baby*, *mad*, *happy*, *love*, *hope*, *wish*. Of these, the first four have an LR higher than 2, meaning they appear over four times as frequently in G4, compared with G3. The following are illustrative examples of how they are used:

- (9) G: u r my little **princess**
- (10) G: i **missed** you too baby I cant wait to see you
- (11) G: I **miss** my princess so much

As can be seen in examples nine and 11, *princess* is used as a term of endearment. This is also the case for *baby*. *Missed* and *miss* are used to express feelings towards the target and to state that the groomer missed the target the previous day, because they were not online at the same time. *X* (LR: 1.48) is used as a kiss emoticon, which conveys feelings of love and also strengthens the relationship. The remaining seven keywords have an LR between 1.29 and 0.64 in G4 with most of them being close to 1. They reinforce the above pattern of relationship strengthening:

(12)G: I could never be **mad** at u

(13)G: you make me **happy**

(14)G: I **love** u so much

Love is used very broadly, has over 1,000 occurrences and is used to exchange love statements and affirmations (as in example 14). It is also used for seeking sexual gratification purposes (*ill make love to you there*). *Wish* is also used in a variety of ways, but mostly to express how the groomer wished they were with the target (*yup wish you were here*) or other wishful thinking related to the target (*yes it was I wish I could take you away too*). It is important to realise that although these words are used to strengthen the relationship between groomer and target, they are also connected to implicit sexual content, as romance is used by groomers to disguise their ulterior motive of sexually abusing their targets. G4 groomers, who spend more time interacting with their targets than those in sub-corpus G3, seem to do this by more implicit means, which is identifiable at the keyword level.

There are two keywords (*muah*, *honey*) in the short duration sub-corpus G3 that are focused on strengthening the relationship between the groomer and the target. Although *muah* is the second keyword with a LR of 5.13, it is primarily used by one groomer and only 2% of the occurrences occur with a different groomer. This is therefore not representative of the groomers in G3 overall and is discarded for further analysis. *Honey* (LR: 1.39), on the other hand, is used as a term of endearment by many groomers, similarly to how *princess* is used by long duration groomers (G4). This shows that while there is relationship building present in the G3 keywords, it emerges as more prevalent in G4:

(15)G: anything you want **honey** i like talking with you

5.2.1.3 Activities

In the keywords of long duration groomers (G4) another pattern connected to activities (*busy, went, doing*) emerges as salient. *Busy* and *went* have an LR above 1, while *doing* has one of 0.58. The following examples show how they are used:

(16)G: u seem **busy**

(17)G: Yup sure did and **went** swimming too

(18)G: what are u **doing**

Examples 17 and 18 are used to discuss current activities, but also seem to be a way to re-engage the target and regain their full attention. *Went* is also sometimes used similarly (*I thought u went to sleep*). These keywords contribute towards the activities and sociability strategies within the grooming intention of developing deceptive trust in G4.

5.2.1.4 Visual Information

A third strategy within the developing deceptive trust in G3 (short duration) – when compared with G4 (long duration) – is seeking and checking visual information (*profile, pics, cam, pic*). Out of these four keywords, *profile* (LR: 1.61) is the only keyword with a Log Ratio above 1. The following are its top ten collocates: *look, pic, on, my, one, ur, your, at, in, k*. This is used to discuss social media/Instant Messenger profiles, specifically referring to profile pictures. The short duration groomers (G3) use this keyword to confirm what the target looks like, their age and other personal information about them. It is also used to pay them compliments (*I like that pic u got on ur profile u look beautiful n sexy*), and to direct the target to the groomers' own profiles (*did you look at my profile*). *Pics* (LR: 0.91), *cam* (LR: 0.87) and *pic* (LR: 0.69) are less distinctive, but still important keywords in G3. They are about twice as frequent in G3 (short duration) compared with G4 (long duration). They continue the strategy of visual information sharing/checking, which can be seen in the following examples:

(19) G: maybe get some sexy **pics** of you

(20) G: I'm looking at your **pic** on myspace and you have a nice looking body

(21) G: u have a web **cam**?

These keywords are used more generally to ask for pictures from the target, share pictures of the groomer and verify that the target is underage, which is linked to the assessing risks and isolating grooming intention. Example 20 shows the underlying introduction of implicit sexual content, while example 19 introduces explicit sexual themes. This is an important strategy to highlight, as it shows the interconnectivity and complexity of the different intentions that comprise the online grooming process (see chapter two, section 2.2.1.2, and chapter four section 4.3.1.4). Groomers in G3 use visual information checking or sharing to further developing deceptive trust, assessing risks and isolating and seeking sexual gratification concomitantly.

5.2.1.5 Sociability

In G4 (long duration), sociability is also present in the keywords (*sweet, dream, nite, sleep, bed*) and it is more prevalent than in G3 (short duration). The keywords are all connected to the semantic domains of going to bed, sleeping, and dreaming and appear around twice to four times as often in G4 compared with G3. They all seem to follow the same broad patterns, which can be seen below:

(22) G: **sweet dreams** ok

(23) G: **night** my beautiful princess

(24) G: **sleep** tight sweet dreams

(25) G: go to **bed** my princess

These keywords also contribute towards relationship strengthening, as the examples show, further emphasising the relationship and trust focus of groomers who spend more time chatting with their targets. Although *sleep* and *bed* could be used for seeking sexual gratification, only one of the top ten collocates of *sleep* and two of *bed* denote sexual content (*would u sleep naked with me?*). Broadly speaking, these

keywords are used to focus on romance and relationship, i.e. implicit sexual content, rather than explicit content in G4.

One final pattern within the sociability strategy in G3 is related to the key words *prob* and *thanx*. *Prob* is most often a short form of *problem* and appears most frequently as *no problem* – that is used as a positive politeness strategy by showing concern and interest in the target and addressing their positive face. These are two illustrative examples of *prob* and *thanx*:

(26) G: i can call you when i get home .. no **prob**

(27)

T: Ur cute

G: **thanx**

These keywords are used by the groomer in combination with positive politeness strategies, which contributes to developing deceptive trust. In example 26, the groomer attends to the target's needs by saying he can call the target when he gets home and that this is not a problem for him, addressing their positive face needs. He also includes both the target and him in the activity, which is another positive politeness strategy. In example 27, the groomer thanks the target for a compliment, which shows engagement with the target, i.e., attending to their interests, another positive politeness strategy, oriented to the target's positive face wants.

5.2.2 Arranging Further Contact

Arranging further contact only emerges as salient in the short grooming duration sub-corpus G3 and it is the second most prevalent online grooming intention. Three strategies (online, offline and time) arise, which can be seen in Table 22.

Table 22: Three strategies in G3 within arranging further contact

Strategy	Keywords
Offline	hang, come, meet, chill
Phone	number, cell, call, phone
Time	tomorrow

5.2.2.1 Offline

The first strategy with the keywords *hang*, *come*, *meet*, *chill* contribute towards this arranging further contact intention. *Chill* and *hang* are twice to four times as frequent and *meet* and *come* a little bit more frequent in G3 compared with G4. The following are examples of the first two keywords:

(28) G: yea we could just hang out go bowling or just **chill**

(29) G: want to do it as soon as im there or **hang** out a bit first

Chill seems to be used as euphemisms for having sex, especially the minimiser *just* in example 28 highlighting this. Example 29 shows that *hang[ing] out* is differentiated from having sex, which is expressed using implicit terms (*do it*). In general, these two keywords are used to describe plans for an eventual meeting between the groomer and their target, which also advances the groomer's sexual gratification and shows that arranging further offline contact is connected to the grooming intention of seeking sexual gratification. The other two keywords discuss different aspects of the planned meeting, focussing on the logistics (venue/time) of the offline meeting (*ill come over about that time*), describing possible scenarios if a meeting occurs, which is connected to assessing risks and isolating (*I could come over before she comes home*), asking if the target wants to meet (*would you wan na meet me so soon?*) and planned activities (*and what would you enjoy doing when I come see you?*). As with some of the developing deceptive trust strategies, arranging offline meetings is lacking in G4 (long duration) at the keyword level. The above keywords *do* appear in the corpus, but much less frequently except for *meet* and *come*, which seem to be used similarly to the way they are used above. However, the overall grooming intention does not emerge saliently in G4.

5.2.2.2 Phone

The second strategy in arranging further contact pertains to contact by telephone. The following keywords belong to this strategy: *number, cell, call, phone, tomorrow*. The first two of these have a Log Ratio score above 1, which means they are approximately twice as frequent in G3 (short duration) compared with G4 (long duration). The groomer uses these keywords to ask and confirm if the target has his number. These are two typical examples taken from G3:

(30) G: Yes I gave u my **number** yesterday

(31) G: u have my **cell** number right

In examples 30 and 31 it is the groomer that either makes sure the target has his cell phone number or reminds them that they have it. These keywords are also used to elicit the target's cell phone or house phone number and the collocate *wrong* is used as a risk assessment and isolation strategy to ward off questions of a curious parent (*if ur mom asked tell her someone got the wrong number*). The number exchange is typically established to either instruct the target to call the groomer or the other way around, in short, to arrange further contact via telephone calls. Two of the remaining keywords (*call, phone*), which are less characteristic of short duration groomers (G3), are used to do just that, which can be seen in the following examples:

(32) G: ok **call** me when u get home

(33) G: Can you talk on the **phone** now?

5.2.2.3 Time

The last keyword in the grooming intention of arranging further contact in G3 is *tomorrow* (LR 0.59), which is used to arrange further contact by Instant Messenger (*talk to u tomorrow*), phone call (*well call me tomorrow morning at 9 ok*) and face-to-face meeting (*what time do u want me over tomorrow*). It is also used to make general plans (*im free all day tomorrow*). This is used very similarly in G4.

5.2.3 Seeking Sexual Gratification

Table 23: Three strategies in seeking sexual gratification in G3

Strategy	Keywords
Explicit	condom, virgin, sex
Used for sexual content	finger
Sexual collocates	wanna

Regarding the seeking sexual gratification grooming intention, there is only one explicit keyword in G4 (*sexy*), which has a Log Ratio of 0.63. *Sexy* is used as a term of endearment by the groomers, similar to *princess* and *baby* (*hey sexy whats up*). It is also used as a descriptive adjective that serves to pay compliments, which simultaneously introduces sexual content (*thinking of ur hot sexy little body*). One collocational pair *sweet dreams* that invokes a farewell sociability strategy that is connected to interpersonal closeness, e.g. parents and children saying good night, is used with *sexy*. The formula is thus creatively modified by inserting the adjective *sexy* and introduces the sexual desensitisation grooming intention (*well sweet sexy dreams*), thus showing how intertwined the two grooming intentions can be and again underlining the focus on implicitness long duration groomers in G4 seem to have. In contrast, the seeking sexual gratification intention in G3 (short duration) features three sexually explicit keywords (*condom*, *virgin*, *sex*), one keyword that is exclusively used to discuss sexual content (*finger*) and one keyword with strong sexual collocates (*wanna*). *Wanna* may not appear to be sexual at first glance, however its strongest collocate is *fuck*. The following are some examples of the keywords:

- (34) G: would u **wanna** fuck Monday
- (35) G: do I have to wear a **condom** when we have sex
- (36) G: are you still a **virgin**
- (37) G: did u **finger** your pussy for daddy

These are all explicitly used to seek sexual gratification. *Condom*, as shown in example 35, is used to discuss whether the groomer will use a condom with the target, and also to ask about the target's sexual past (*did he cum in you or only use condom*

?), which *virgin* is also used for (see example 36). The focus with *virgin* tends to be on the target, although it is also used to self-disclose in one instance (*im a virgin lol*), which may not be genuine considering the use of *lol*. *Lol* could also be used by the groomer because he is embarrassed about this self-disclosure. The keyword *finger*, which is exclusively used to denote sexual content in the data is also used to focus on the target. The groomer focuses on instructing or asking the target to masturbate, like in example 37. A few instances of the use of the verb *finger* also refer to the activities the groomer plans on doing when a meeting takes place. Despite the focus on the target, the groomer likely gains sexual gratification from the discussion of masturbation and teaching the target (Lorenzo-Dus et al., 2016). *Sex* (LR: 0.79), although less frequent than the rest of the keywords, is also a keyword in G3 (short duration) that denotes sexual content. It is used by the groomer to discuss the target's sexual past (*when was the last time you had sex*) and experience levels, self-disclose, discuss the groomer and target having sex and introduce sexual content into the conversation (*kids ur age are having sex all the time*) featuring assessing risks and isolating. Overall, this grooming intention emerges explicitly and frequently in G3, while it is more subtle and implicit in G4.

5.2.4 Assessing Risks and Isolating

The last grooming intention in G3 (short duration) is assessing risk and isolating, which has only one keyword (*older*) attached to it. Its Log Ratio is 0.85, meaning it is almost twice as frequent in G3 compared with G4. It is used to ask the target if they like/have been with older guys, seeking sexual gratification, but primarily used to test the waters asking about the target's parents (*u parents know u be chatting to older guys?*) and making references to police undercover sting operations (*where they catch older guys hooking up with younger guys*). Groomers also use it to highlight the age difference (*I'm way older than you*) and possible (legal) repercussions (*im older than 18..that means its against the law I could go to jail*). Another use is to compliment the target and state that they look older than they are and describing hypothetical scenarios, which could be used to justify and defend the groomer's subsequent actions. Overall, *older* is used to gauge the target's reaction, making sure they are indeed underage, pointing out possible dangers of proceeding and justifying the groomers'

actions to themselves. This keyword does not appear saliently in the long grooming duration sub-corpus (G4), though some of the collocates look similar to the ones in G3.

As the above discussion has shown, groomers who spend less time interacting with their respective targets, have four grooming intentions that are identifiable at the keyword level, while groomers who spend more time with their targets have two intentions. This seems reminiscent of how advertising language makes use of more varied tactics in 120 second ads compared with longer infomercials (Komar, 2015), drawing on visual, audible, and written cues, reminding the consumer about the benefits of the product through repetition and enumeration and using statements, imperatives, and exclamations. The less time advertisers and groomers have, the more resources they draw on to really convince the consumers and targets and persuade them.

Groomers with short grooming durations are very focused on the exact location of their target, arranging further contact with them and checking/sharing visual information, which is used to simultaneously further the developing deceptive trust and seeking sexual gratification intentions. At the keywords level, they do not seem to be focused on developing and strengthening the relationship, only using one term of endearment (*honey*), while groomers with a longer grooming duration use two terms of endearment (*princess*, *baby*). The language of groomers with a short grooming duration (G3) is also highly sexualised with five keywords that are used for explicitly sexual content and a strong focus on the target, while occasionally self-disclosing. In contrast, groomers who spend longer interacting with their targets (G4) only use one explicitly sexual keyword (*sexy*) and seem to be using a more implicit approach, using *sexy* similarly to terms of endearment like *princess* and *baby*. *Sexy* is also used to add a sexual element to the farewell formula of sweet dreams, connecting developing deceptive trust and seeking sexual gratification. Two additional intentions were identified with short duration groomers (G3): Arranging further contact and assessing risks and isolating. Groomers use implicit sexual content in connection with arranging meetings, while also assessing the risks and justifying their choices and actions to themselves. These intentions did not appear saliently in the language of long duration groomers (G4). Instead, groomers with a longer grooming duration were very focused

on building the foundation of relationship and trust, while not being very focused on explicit sexual gratification. Romance – in the form of more vague and implicit sexual language, masquerades their ulterior sexual abuse motives.

Overall, in a first attempt to answer the research questions, grooming duration *does* seem to influence the grooming process and the specific grooming intentions of the groomers. Some words, such as *princess* and *baby* for G4 and *honey* for G3 can be related to the two different grooming duration groups. There are clearly two different groomer language approaches present. However, while language difference is apparent, it is unclear whether there might be more language nuances within these two groups leading to more than two groups, and hence different duration-based cut-off groups. For this reason, it was decided to delve deeper into the language of the short duration groomers (G3) and see if a further distinction could be drawn within this group.

5.3 Three or Two Grooming Duration approaches?

To find out whether there were further differences in the first group analysed above, a second analysis was conducted, using 100 minutes as the cut-off point. Two more sub-corpora were defined, the details of which can be seen in Table 24.

Table 24: Details of G1 & G2

Sub-corpus	No. of usernames	No. of Words	Duration in minutes
Groomers 1 (G1)	52	22,931	0-99
Groomers 2 (G2)	188	285,167	100-299
total	240	308,098	0-300

Table 25 shows the 17 keywords of G1 and G2. Keywords appearing more often in G2 are shaded in grey.

Table 25: Keywords of first two sub-corpora

#	Word	G1	G2	Log ratio	LL
		Freq 1 (per mill)	Freq 2 (per mill)		
1	quest	436.09	18.02	4.6	33.16
2)	130.83	1895.26	-3.86	62
3	:	479.7	4410.27	-3.2	125.51
4	.	566.92	4493.14	-2.99	120.43
5	...	1831.58	10852.72	-2.57	252.6
6	babe	2180.45	446.79	2.29	68.32
7	m	1177.45	255.82	2.2	34.85
8	please	2093.24	558.49	1.91	49.59
9	,	4012.04	11836.39	-1.56	151.88
10	§_HAPPY-SMILEY_§	1613.54	4233.72	-1.39	46.04
11	meet	3009.03	1351.19	1.16	31.06
12	call	5887.23	2644.72	1.15	60.83
13	k	5145.87	2623.1	0.97	39.32
14	u	52025.64	31920.41	0.7	230.57
15	want	10553.4	6727.1	0.65	39.02
16	me	18272.21	12690.34	0.53	45.91
17	ok	16004.54	11216.64	0.51	38.29

In terms of Log Ratio, the majority (n=5) of keywords have an LR between 1 and 2, meaning they are two to three times more common in one corpus. Four keywords are four to five times more frequent (LR: 2-3). Three keywords have an LR above 3, meaning they are eight or more times more common in one corpus. The minority (n=2) of keywords (*me*, *ok*) fall under the chosen cut-off point of 0.55 and will not be considered in the analysis below. On the surface, these keywords do not seem to indicate much difference between the two sub-corpora, especially as there are only 17 of them. Interestingly, all symbols and punctuation marks are keywords of G2, while the other words are keywords of G1. As with the first analysis, collocates of these keywords were calculated, which were analysed by looking at extended concordance lines and sorting the keywords into their respective grooming intentions (see Table 26).

Table 26: Keywords of G1 and G2 sorted into grooming intentions

Grooming intentions	Keywords G1	Keywords G2
Developing deceptive trust	Babe, m, k, quest	
Arranging further contact	Meet, call	
Negotiating power dynamics	Please, want	
other	u) : . , §_HAPPY-SMILEY_§

Developing deceptive trust seems to be the most prominent grooming intention in G1, with four keywords (*quest*, *babe*, *m*, *k*). Arranging further contact has two keywords (*meet*, *call*) and there are two keywords associated with negotiating power dynamics in G1 (*please*, *want*). The “other” category in G1 only consists of one word (*u*) which does not display any clear-cut patterns, instead independently driving all grooming intention forwards. None of the keywords in G2 are aligned with the grooming intentions, as they are mostly punctuation marks that may be used as emoticons and an emoticon (§_HAPPY-SMILEY_§). Although emoticons have meaning making potential and there is a body of work investigating them in Computer Mediated Discourse Analysis (see for example Dresner & Herring, 2010; Garrison et al., 2011; Vandergriff, 2013; Thompson & Filik, 2016), this thesis will not focus on them. Only three types of emoticons were transliterated, and the remaining emoticons have a large variety of different spellings. In G2 none of the grooming intentions emerge as salient or key. The keywords appearing in G1 will also be taken into consideration in G2 to see if the language differs.

5.3.1 Developing Deceptive Trust

Within the developing deceptive trust intention in G1 tentative strategies of extracting exact locations (*quest*), strengthening the relationship (*babe*), exchanging information (*m*), and sociability (*k*) arise. The first three of these keywords are between four and 16 times as frequent (respective LR: 4.6; 2.29; 2.2), while the last keyword is just under twice as frequent (LR: 0.97) in G1 compared with G2. The following are illustrative examples:

(38) G: and do you want to give me ur address so I can look on map **quest** to see
how to get there

(39) G: hey u there **babe**?

(40)

G: asl

T: 13 f mi

G: o

T: and u

G: 20 **m** det

(41) G: **k** one sec

Example 38 demonstrates that *quest* is exclusively used with its collocate *map*, referring to a free online mapping service *MapQuest*³⁰ comparable with *Google Maps*, which is being used by the groomers to elicit the exact address of the target and to map their distance/route to the target, which is part of the sharing/eliciting information strategy. This is similar to how the groomer and target share their age, sex, and location in example 40. *M* refers to male and is used by the groomer to describe their sex as well as their location and age. Example 39 shows how *babe*, is used as a term of endearment, similar to *honey* and *princess* in the previous analysis. It strengthens the deceptive trust and groomer-target relationship. *K*, which is always used as an abbreviation of *okay*, is mostly used to signal agreement, as the above example (42) shows. It sometimes occurs in a sequence where the groomer echoes the target and *k* seems to cluster:

(42) T: yah lem me call u wen i get outta shower k ?

G: k

T: k

In G2, these keywords are used very similarly. However, they do not emerge as salient and are much less frequent:

(43) G: can i at least see where you live in map **quest** i won't go until you tell me
to go

³⁰ <https://www.mapquest.co.uk/>

(44)

T: hi asl?

G: 23 **m** fl

(45)

T: be right back

G: **k**

T: k

G: k

(46) G: **babe** i love you

Examples 43, 44 and 45 show that these keywords are used in the same way. One small difference shown in the usage of *babe* (46) is that in G2 there is a slightly stronger connection to feelings, shown by the collocates *love* and *muah*.

5.3.2 Arranging Further Contact

Meet and *call* are used in G1 to arrange and talk about future face-to-face meetings with the target and to arrange further contact by telephone. *Meet* is partly connected to seeking sexual gratification, as the target and groomer engage in meeting planning which includes clarifying whether they will have sex. The following are illustrative examples of this meeting planning:

(47) G: so you wan na talk alittle when we **meet** and then have sex?

(48) G: would u like to **meet** m?

(49) G: so cant we **meet** somewhere and get a hotel room

(50) G: think about were we can **meet** on sat

This meeting planning relates to activities at the meeting (example 47), the location of the meeting (examples 49-50) and whether the target would like to meet (example 48), seemingly giving control to the target, which is related to negotiating power dynamics. The keyword also occurs in the phrase *nice to meet you*, which is related to developing deceptive trust, specifically sociability. *Call* is used by groomers to arrange further contact by telephone, to elicit the target's telephone number, to confirm an offline

meeting is still happening, to ask the target to call or offer to call the target. The following are illustrative examples:

(51) G: so how about u give me ur number and ill **call** u right now

(52) G: so **call** me tomorrow by 3 to let me know for sure if we are meeting .k ?

(53) G: **can** u call me ?

(54) G: babe can i **call** u on the phone please ?

In example 52, the groomer instructs the target to call him at a specific time the next day to confirm they are still meeting face-to-face. He uses a backchanneling device to ask for confirmation in the same turn. This example shows that arranging further contact via phone is linked to arranging offline contact in some cases. Example 54 is another example of the groomer using positive politeness (term of endearment and please) to arrange a call with the target, which simultaneously contributes to arranging further contact and developing deceptive trust. In G2 the keywords *meet*, and *call* are used very similarly. The *nice to meet you pattern* is slightly more prevalent than in G1 and other patterns emerge as well as the meeting planning one:

(55) G: nice to **meet** u emma

(56) G: would you ever **meet** someone from yahoo in person for sex

Interestingly, *sex* is one of the collocates of *meet* in G2, suggesting that groomers in this sub-corpus are more explicit about bringing up sensitive topics when discussing meetings. As Example 56 shows, this could also relate to hypothetical scenarios to gauge the target's reaction and assess the likelihood of them going along with the proposed activity, which is related to assessing risks and isolating and seeking sexual gratification.

5.3.3 Negotiating Power Dynamics

The two keywords in G1 that are used to contribute to the grooming intention of negotiating power dynamics are *please* and *want*. *Please* is around four times more

frequent in G1 compared to G2 (LR: 1.91). It is also connected to begging, insisting and repetitions. The following are a few examples of this:

(57) G: please understand me please **please**

(58) G: **Please** please I can take care of everything for you dear

(59) G: More pics **please**

Examples 57 and 58 show this insistence and clustering of *please* to convince the target of something and attempt to make them feel guilty if they refuse the groomer, which is a clear example of the groomer trying to enact power over the target. One groomer also tries to convince the target to come to his house on five separate occasions (*yes but **please** come here dear*). The usage of *please* in G2 is slightly different, however the begging and repetition can also be observed. Additionally, there is a focus on assessing risks and isolating, eliciting personal information, and arranging further contact, which the following examples show:

(60) G: I hear ya sweetiejust **please** please please keep me a secret

(61) G: **Please** just tell me babe

(62) G: then address **please**

(63) G: 1 time throw this number way **please**

As example 60 shows, the groomers in G2 also beg targets to convince them of something, in this case keeping them a secret. Clustering of the word *please* can again be seen to emphasise the message and influence the target. This politeness formula of seeking avoidance of imposition through emphasis (*please*) ends up face-threatening, rather than polite through repetition. In example 63 the groomer asks his target not to leave any trace of their interaction, which shows a focus on assessing risks and isolating. Example 62 is a subtle way to elicit information from the target. The other keyword directly related to negotiating power dynamics in G1 is *want* (LR: 0.65). This is also connected to the other grooming intentions, which the following examples will show:

(64) G: u really **want** me to fuck u?

(65) G: I don't know if u **want** to see me naked

(66) G: If u **want** I can call u since ur maybe having problems?

(67) G: but i **want** u to feel good

(68) G: I don't **want** to go to jail

(69) G: do u know what time u **want** me over tomorrow

While examples 64, 65 and 67 are directly contributing towards the grooming intention of seeking sexual gratification, they are also seemingly putting the target in control by adding (*if*) *u want*. Example 66 is related to developing deceptive trust; The groomer shows concern for the target and their problems and wants to arrange further contact by calling them. The intention of assessing risks and isolating is also present in example 68, where the groomer points out the risk of going to jail due to the illegal nature of the interaction. While most of the *want* concordance lines are target focused, some of them are groomer focused. Example 69 shows negotiating power dynamics in combination with arranging further contact. The groomers in G2 use *want* in very similar ways and its Log Ratio (0.65) suggests it has roughly the same frequency in both corpora:

(70) G: yes, only if you **want** to give a blowjob it take like chocolate do you know what

(71) G: u **want** to see my penis

(72) G: if u **want** me to call u 1 time I will

(73) G: love to touch you anything you **want** to know

(74) G: I know – I just **want** you to be aware

(75) G: do u really **want** to see me this weekend

As above, examples 70, 71 and 73 are connected to seeking sexual gratification, while supposedly putting the target in control of the situation by adding (*if*) *u want*. The groomers in G2 also offer to call the target (72). In example 74 the groomer warns the target that he could get in trouble if they got caught, which is similar to example 68 above, except that the groomer puts responsibility on the target. Example 75 is connected to negotiating power dynamics and arranging further contact.

Overall, the two groups seem to be more similar than different, as the above analysis has shown. There might be a few small differences in terms of the specific strategies

used in G1 compared with G2, such as a greater focus on feelings in *babe* concordance lines and eliciting information using *please* in G2. Overall, the language is used in a similar way to attain the communicative intentions of the groomers. Even though at the keyword level no grooming intentions were identified for G2, the analysis shows that all four grooming intentions that are present in G1 (developing deceptive trust, seeking sexual gratification, arranging further contact, negotiating power dynamics) are also present in G2. To further answer the research questions, it does not look like grooming duration makes a difference below the cut-off point of 300 minutes. Instead, a similar approach is used by both groups, which includes a slight focus on developing deceptive trust, specifically eliciting an exact location, and exchanging information with the target. Seeking sexual gratification also features. However, it is not as apparent at the keyword level as it was in the first analysis. Instead, some keywords are used to simultaneously contribute towards seeking sexual gratification and other grooming intentions. In this analysis, negotiating power dynamics also emerges as its own grooming intention at the keyword level. Based on this, it was decided to examine the long grooming duration sub-corpus (G4) in more detail to further explore the questions of how many different grooming approaches are linked to different durations of grooming.

5.4 Long or short Grooming? Investigating Long Duration Grooming

To figure out whether there are two distinct groups, or three distinct groups based on the duration of grooming, a third cut-off point was drawn between 300-799 minutes and 800-10,600 minutes, the details of the fifth and sixth sub-corpus can be seen in Table 27.

Table 27: Details of fifth and sixth sub-corpora

Sub-corpus	No. of usernames	No. of Words	Duration in minutes
Groomers 5(G5)	220	889,358	300-799
Groomers 6 (G6)	173	2,407,470	800-10,600
total	393	3,296,828	300-10,500

Table 28 shows the top 50 keywords of G5 and G6. All 64 keywords can be seen in the Appendix (section 9).

Table 28: Top 50 keywords of G5 compared with G6

#	Word	G5 Freq 1 (per mill)	G6 Freq 2 (per mill)	Log ratio	LL
1	i am	1380.97	20.99	6.04	228.86
2	-	40.62	729.26	-4.17	69.65
3	@	54.16	613.83	-3.5	51.72
4	;	148.93	1668.37	-3.49	140.09
5	n	1394.51	125.91	3.47	155.44
6	'm	1313.28	131.16	3.32	140.16
7	special	54.16	524.65	-3.28	41.81
8	dreams	54.16	503.66	-3.22	39.51
9	daddy	324.93	36.73	3.15	32.7
10	..	676.95	78.7	3.1	67.17
11	's	1800.68	230.84	2.96	169.73
12	x	94.77	634.82	-2.74	42.83
13	s	392.63	62.96	2.64	32.27
14	bye	1001.88	162.64	2.62	81.7
15	girlfriend	108.31	613.83	-2.5	37.6
16	question	135.39	739.75	-2.45	44.27
17	fuck	717.56	136.41	2.4	52.15
18	sunday	622.79	120.67	2.37	44.59
19	b	487.4	94.44	2.37	34.89
20	sexy	446.79	2266.47	-2.34	129.12
21	dick	744.64	146.9	2.34	52.55
22	i see	1110.19	220.35	2.33	77.98
23	horny	663.41	178.38	1.89	34.95
24	busy	243.7	891.9	-1.87	38.62
25	suck	839.41	230.84	1.86	43.13
26	sex	2477.63	682.04	1.86	127.28
27	sweet	541.56	1872.98	-1.79	76.55
28	.	988.34	3273.79	-1.73	127.74
29	!	1299.74	398.73	1.7	58.61
30	lick	812.34	251.83	1.69	36.14
31	<	324.93	1033.55	-1.67	38.43
32	sweetie	1570.52	498.41	1.66	67.8
33	:	866.49	2686.18	-1.63	96.96
34	later	1056.04	346.27	1.61	43.62

35	§_KISS-SMILEY_§	1015.42	3032.45	-1.58	104.5
36	girl	2139.15	718.76	1.57	85.45
37	aww	473.86	1369.32	-1.53	45.15
38	miss	514.48	1463.76	-1.51	47.27
39	pussy	1489.28	556.12	1.42	50.73
40	am	1814.22	4847.72	-1.42	143.54
41	r	2437.01	949.61	1.36	77.42
42	pic	1123.73	445.95	1.33	34.57
43	thing	731.1	1815.27	-1.31	47.83
44	cum	1313.28	529.89	1.31	39.24
45	id	1299.74	524.65	1.31	38.8
46	>	636.33	1453.27	-1.19	33.07
47	i'll	1584.06	750.24	1.08	34.35
48	mom	1570.52	750.24	1.07	33.4
49	n't	3168.11	1521.47	1.06	66.63
50	ur	5740.51	2770.13	1.05	119.6

Most keywords (n=28) are two to three times more common in one corpus, while 12 keywords are at least four times more frequent in one corpus (Log Ratio between 2 and 3). Ten keywords have an LR value of 3 or more, meaning they are at least eight times more common in one corpus, while nine keywords with an LR value under one but more than 0.55 appear just under twice more frequently in one corpus. The minority (n=6) of keywords (*yes, can, do, u, just, ?*) will not be considered below, as their LR values are under 0.55, the chosen LR cut-off point. The remaining keywords (n=59) were sorted into grooming intentions, which can be seen in Table 29.

Table 29: Keywords of fifth sub-corpus sorted into grooming intentions

grooming intention	keywords G5	keywords G6
developing deceptive trust	sweetie, girl, pic, bye, later, yeah, ok, oh	special, x, girlfriend, sweet, §_KISS-SMILEY_§, baby, aww, miss, busy, dreams, question
seeking sexual gratification	fuck, dick, horny, suck, sex, lick, pussy, cum	sexy
assessing risks and isolating	daddy, mom	-
arranging further contact	Sunday, here	-
other	see, will, b, I am, I see, I'll, they, n, 'm, 's, s, n't, .., !, r, id, ur	-, @, :, ,, <, :, > am, think, thing

As Table 29 suggests, the most prevalent grooming intentions in G5 are developing deceptive trust and seeking sexual gratification with eight keywords each. In contrast, developing deceptive trust seems to be the single most prevalent grooming intention in G6 with 11 keywords. Elements of seeking sexual gratification are also present in G6 with one keyword (*sexy*). Assessing risks and isolating and arranging further contact have two keywords each in G5. They do not seem to appear in G6, at least not on the keyword level. The “other” category comprises words that are not readily identifiable as belonging to one of the groomer intentions while looking at the collocates and briefly examining the concordance lines. These are punctuation marks and symbols (-, @, :, ., <, :, >, .., !), misspelled words that have been normalised (*I am, I see, I'll*), fragments of words that are misspelled (*'m, 's, s, n't*), words that are purposefully abbreviated (*n, b, r*), auxiliary verbs (*will, am*) pronouns (*they*) and a few content words that are used very broadly and contribute to all grooming intentions without any overarching patterns or dominant grooming intentions (*think, thing, see*).

5.4.1 Developing Deceptive Trust

Within the developing deceptive trust intention in G5, the focus seems to be on sociability and features positive and negative politeness (*bye, later, yeah, ok, oh*). The first of these keywords appears around four times more frequently in G5 compared with G6 (LR: 2.62), while *later* is around twice as frequent. The remaining keywords have a similar frequency in G5 and G6 (LR: 0.81; 0.71; 0.62). The following are illustrative examples:

- (76) G: **bye** sweetie
- (77) G: k talk **later** then
- (78) G: **yeah**, do you feel the same way
- (79) G: **ok** I will try hun
- (80) G: **oh** ok, that would be cool

As examples 76 and 77 show, *bye* and *later* are used by the groomer at the end of a conversation to say goodbye. *Bye* is often used with a kiss smiley, happy smiley, or term of endearment, which strengthens that it contributes to developing deceptive

trust, while *later* has an added element of making plans to talk (on the phone or via Instant Messenger) in the future, simultaneously advancing the arranging further contact and developing deceptive trust intentions. *Yeah* and *ok* are used very generally by groomers in G5 to signal agreement. *Yeah* is also used as a discourse marker or to introduce a change of topic (*yeah, how was your day?*). *Yeah* occurring with the collocate *it* (dice co-efficient: 0.091) is also used to introduce sexual acts, porn, and other explicit content. The groomer and target discuss the target's experience levels, possible sexual acts they are willing to perform, and the groomer gauges the target's reaction. The collocate *but* (dice co-efficient: 0.091) seems to minimise the agreement with the target or a specific fact (*yeah but there are some in this world who are bad people*). It is also used to introduce and mitigate risks (*yeah but your 14 and I am 31 lol*). Overall, *yeah* seems to be contributing towards developing deceptive trust by conveying and minimising agreement and to a lesser extent towards seeking sexual gratification and assessing risks and isolating. As well as being used to signal agreement, *ok* is used as a way for the groomer to make promises (example 79) and give reassurance (*its ok*), which are positive politeness strategies. *Oh* (example 80) co-occurs with *ok* quite frequently (dice co-efficient: 0.282), which can convey e.g., surprise or disappointment, but is generally used by G5 groomers to express emotions and keep the conversation with the target going.

This sociability strategy seems to be largely absent in the keywords of G6. There are only three keywords (*dreams, busy, sweet*) that seem to advance it, which are four to eight times as frequent compared with G5 (LR: -3.22; -1.87; -1.79):

(81)G: nite **sweet dreams**

(82)G: is my princess **busy**

Sweet mostly co-occurs with *dreams* (dice co-efficient: 0.31), as example 81 shows, and is used to say goodbye and as a compliment (*u are so sweet*), which are both used by groomers in G6 to develop deceptive trust, by being sociable and using positive politeness strategies. *Busy* is used to keep the conversation going, contributing to the sociability and activities strategy within developing deceptive trust, and is sometimes used with a term of endearment (see example 82) to strengthen the groomer-target relationship. However, the main developing deceptive trust intention focus of

groomers who spend a longer time interacting with their targets (G6) seems to be on building a trust foundation (*special, baby*), strengthening the relationship (*girlfriend, §_KISS-SMILEY_§, x*) and expressing emotions (*miss, aww*). *Special* (LR: -3.28), *x* (LR: -2.74) and *girlfriend* (LR: -2.5) are four to eight times more frequent; *§_KISS-SMILEY_§* (LR: 1.57), *aww* (LR: -1.53) and *miss* (LR: -1.51) are just under twice as frequent in G6 compared with G5. *Baby* is not a feature that is as distinct of the G6 groomer language. The following are illustrative examples of how groomers strengthen the trust foundation:

(83) G: I like u as a very **special** friend

(84) G: I wanna make love to u **baby**

Special is used to emphasise the relationship between the target and the groomer (example 83), but it is most frequently used as part of a compliment, which flatters the target and builds a deceptive foundation of trust. This is also a positive politeness strategy, which addresses the target's positive face needs. *Special* is also used to a lesser extent for seeking sexual gratification (*I wanna make your nite a special one too*). *Baby* is used similarly to other terms of endearment (*sweetie, honey*). However, it has strong sexual undertones, which example 84 exemplifies. Additionally, it is used to elicit personal information from the target (*what is your address baby*) and express the groomer's feelings. *Girlfriend* is used by groomers in G6 to negotiate and strengthen the relationship. It takes on a similar role as *baby*:

(85) G: Hey sexy **girlfriend**

Additionally, it is used to explore what being a girlfriend means (*Good thats what it takes to be bf and girlfriend*). The kiss emoticon and *x* also seem to be used to create closeness and reinforce the groomer-target relationship. Their collocates, which consist of other punctuation marks and emoticons, show an accumulation of emoticons particularly at the end of individual conversations. Finally, within the developing trust intention in G6, groomers use *aww* and *miss* to express emotions:

(86) G: **aww** ty but ur just saying that. i don't think im cute

(87) G: I **miss** u so much

Aww is mostly a reaction to something the target says, e.g., a compliment (as in example 86), or having to leave, to which the groomer expresses a sad reaction. Similar to this, *miss* is used to convey strong feelings of missing the target (example 87). It is also used to state that the groomer and target missed the opportunity to speak to each other, because they were not online at the same time. These two keywords contribute towards developing deceptive trust intention of groomers spending more time interacting with their targets (example G6). More specifically the emotional side of the groomer-target relationship.

Relatedly and focussing on the other sub-corpus, the only two keywords of groomers in G5 that relate to this close trust-based relationship are *sweetie* and *girl*. They both have a similar Log Ratio score (1.66; 1.57) and appear about twice as frequently in G5 compared with G6. Groomers use *girl* to address the target (*Hey girl*), in instances connected to seeking sexual gratification (*yes i love sex with a girl*) and to praise the target (*good girl*). There are sexual undertones in its usage, especially coupled with adjectives such as shy, bad, naughty, nice, hot, petite, sexy and smart (*i wouldn't miss a chance to meet a sweet sexy smart girl like u*). *Sweetie*, on the other hand is most often used as a term of endearment in greetings or goodbyes. It is also used to check that the target is still there, to keep the target talking (*busy still sweetie?*) and in small talk (*hows my lil sweetie today?*), which contribute to developing deceptive trust through sociability.

The last keyword in G5, which does not seem to belong to a broader pattern is *pic*. This keyword seems to both contribute to eliciting personal information and seeking sexual gratification. The groomer asks for general pictures of the target, naked pictures of the target, pictures of body parts of the target and groomer and pictures of the target's surroundings (*u have a pic of ur room*). The groomer also refers to sharing pictures of him, his body parts (*does the pic of my cock make u horny?*) and his surroundings, but to a lesser extent.

5.4.2 Seeking Sexual Gratification

A few of the above keywords have already been connected to seeking sexual gratification. However, there are eight very explicitly sexual keywords (*fuck, dick, horny, suck, sex, lick, pussy, cum*) in G5, which will be examined below. *Fuck* (LR: 2,4) and *dick* (LR: 2.34) occur around four times more often compared with G6, while *horny, suck, sex, and lick* (LR: 1.89; 1.86; 1.86; 1.69) appear just under four times more frequently in G5. *Pussy* (LR: 1.42) and *cum* (LR: 1.31) are more than twice as frequent. The following are illustrative examples:

(88) G: u can watch me **fuck** ur pussy

(89) G: do u still want to suck my **dick**

(90) G: kinda makes me **horny** lol

(91) G: u will **suck** my cock

(92) G: Can we have **sex**

(93) G: and let me **lick** ur pussy while you do it

(94) G: try rubbing your **pussy** as I cum

(95) G: can I **cum** in your mouth???

As these examples show, groomers in G5 use highly sexualised language, which is identifiable at the keyword level. There is a focus on the target performing the sexual act (examples 89, 91, 94). In example 88, the groomer places focus on the target as watching the activity he performs. These keywords are used to introduce sexual language, describe sexual acts, ask the target about their sexual history and experience levels and what they are willing to do. There is a specific focus on the groomer's sexual organ and arousal in the usage of *dick*. Planned sexual activity, anxieties, and legality around sex with the target are discussed using the keyword *sex*. The groomer also self-discloses, however, more often the focus is on the target. Additionally, as example 92 shows, the groomer asks if the target wants to have sex with him. The groomer also tells the target he wants to have sex with them (*I want sex now*) and asks if they changed their minds or states that he will only have sex if the target still wants to. This contributes to negotiating power dynamics by supposedly giving the target control, as well as seeking sexual gratification. The groomer also constructs togetherness (*we have sex n e time we want then*) and compares sex to other things

(*once u have full sex its like drugs u cant get enough*), which could be used to self-justify their pursuit of a minor. Example 93 shows a particular pattern connected to the keyword *lick* and its collocates, which appear in the following order: *lick your/ur pussy*. This keyword is about detailed accounts of what the groomer will do or instructions to the target. *Pussy* (example 94) is connected to discussions of the target's arousal (*does your pussy get wet thinking of it*), other sexual acts (*I will eat ur pussy*), masturbation (*can u finger ur pussy for me?*) and the target's anatomy (*I bet your pussy is real tight*). This keyword is also exclusively sexual. Interestingly, *lol* is one of the collocates of *horny* (example 90), which suggests both target and groomer use *lol* to mitigate sexual content, downplaying the seriousness of the topic and potential face threat. This is not only the case with this particular keyword (*Should I say suck my dick? LOL*). The ellipsis in connection with *lol* seems to play a similar role (*I am horny ... lol*).

In contrast, there is only one sexually explicit keyword in G6 (*sexy*), which has a Log Ratio of 2.34 and is thus more than four times more frequent in G6 compared with G5. Its strongest collocates (*hey, aww, lady, ty*) suggest that it may be used like a term of endearment:

(96) G: I love u **sexy**

It is also used to give compliments about the target's clothes, mind, face, smile and other attributes (*Hey sexy little lady*) and to mitigate sexual content (*how bout some sexy underwear lol*). The fact that sexual explicitness is not identifiable at the keyword level suggests that sexual content is introduced more implicitly by groomers who spend more time interacting with their respective targets (in this case, G6, compared with G5).

5.4.3 Assessing Risks and Isolating

The two words contributing towards the assessing risk and isolating intention in G5 are *daddy* and *mom*. *Daddy* (LR: 3.15) occurs more than four times as frequently in G5, while *mom* (LR: 1.07) is around twice as frequent compared with G6. The

collocates of *mom* suggest the groomer is trying to find out the schedule, location and job of the target's mother and asking other questions about her, indicating a goal of assessing risks and isolating. The collocates of *daddy* seem less focused on assessing risks and isolating. The concordance lines reveal that the groomer wants the target to call him *daddy* (*id be ur missing daddy if ya wanted*), which suggests an element of seeking sexual gratification, as well as assessing risks and isolating. Nevertheless, assessing risks and isolating can be discovered as a feature of G5. In G6 *daddy* and *mom* do not appear to be salient keywords. In fact, *daddy* exclusively occurs in one groomer-target pair. It is used in a similar way as in G5 (*do you want daddy to spank you too*). *Mom* in G6 also seems to be used for assessing risks and isolating (*what if your mom calls the house and you're not there?*) and assessing risks and isolating (*do you forgive your mom for she has done to you and your brother?*). The groomer also talks about his mother (*could bring you to meet my mom*).

5.4.4 Arranging Further Contact

The arranging further contact grooming intention consists of two keywords (*Sunday*, *here*) in G5. *Sunday* (LR: 2.37) is four times as frequent, while *here* (LR: 0.82) is under twice as frequent in G5. The concordance lines for *Sunday* suggest this is indeed a keyword that is about discussing details about arranging further contact, specifically an offline meeting between the target and the groomer. *Here* is a little different, as the strongest collocates (*on, i, ?, you, be*) and concordance lines suggest this refers more to a virtual meeting and arranging to talk later or reassuring the target that the groomer is indeed in front of their computer (*im here now*). It also refers to hypothetical scenarios (*wish you was here right now*). Only some of the concordance lines of *be* refer to an offline meeting (*i'll be leaving here at 2*). Overall, though, this does contribute to the grooming intention of arranging further contact with the target. *Sunday* is used to discuss general Sunday plans and activities, rather than specific details of a groomer-target face-to-face meeting (*i actually said a prayer for you on Sunday*) by groomers in G6. The usage of *here* seems quite similar to that in G5. It is used for hypothetical scenarios (*yup wish you were here*), reassuring the target (*Am here cutie*) and talking about the online space of interaction (*so you meet any one new on here*). This grooming intention is not very salient in this corpus judging from the

keyword list alone. It is present in G5, whereas it does not seem to be a salient feature of groomer language in G6.

Overall, the groomers spending less time interacting with their targets (G5) seem to be more focused on sociability, using positive politeness strategies as part of this, when it comes to developing deceptive trust, while groomers who spend more time interacting with their targets (G6) have a stronger focus on building a relationship and trust. In terms of seeking sexual gratification, the language of groomers spending less time interacting with their targets (G5) is highly sexualised with eight explicitly sexual keywords. In contrast, G6 only has one sexually explicit keyword (sexy), which is mostly used like a term of endearment and thus contributes towards developing deceptive trust. Sexual language seems to be more implicit in this sub-corpus and driven by other not overtly sexual grooming intentions. Additionally, assessing risks and isolating and arranging further contact are prevalent in language of groomers spending less time grooming (G5), while they do not feature in the keywords of groomers spending more time grooming (G6). Despite not being salient features, the keywords are used in a similar way by G6 groomers.

5.5 Towards a First Mapping of Two Grooming Duration Approaches

There is evidence to suggest that duration of grooming makes a difference and influences the language and strategies online groomers use, as has previously been pointed out by Lorenzo-Dus and Izura (2017), who found that the number and type of compliments groomers give was related to how much time they spent interacting with a target. The longer the grooming process lasted, the more compliments were given, and the topic orientation changed as well.

Overall, the analyses conducted in this chapter show that the more time groomers spend with their targets the fewer grooming intentions are identifiable at the keyword level. For instance, while groomers in G3 (0-299 minutes) make use of four grooming intentions, groomers in G6 (800-10,600 minutes) only make use of two and use the one sexual keyword to advance developing deceptive trust (DDT).

This difference in language can be seen most obviously in the first analysis, in which the groomers who spend more time grooming the target (G4) seem to focus more on building and strengthening the foundation of the deceptive trust they are building with the target, as Figure 22 shows. They use terms of endearment, express feelings, and emotions, use emoticons and sociability to give the target the impression they are in a loving and caring relationship. Only one sexually explicit keyword (*sexy*) is identified.

G3 (short)	G4 (long)
<ul style="list-style-type: none"> • Developing Deceptive Trust <ul style="list-style-type: none"> • exact location • further contact • checking visual info • Seeking Sexual Gratification <ul style="list-style-type: none"> • target history • experience levels • masturbation 	<ul style="list-style-type: none"> • Developing Deceptive Trust <ul style="list-style-type: none"> • building relationship • trust foundation • terms of endearment

Figure 22: General patterns in first analysis

In contrast, the groomers who spend a shorter amount of time grooming the target (G3) make use of a variety of grooming intentions identifiable at the keyword level (developing deceptive trust, seeking sexual gratification, arranging further contact, assessing risks and isolating). Although developing deceptive trust is the most prevalent intention, it has a different focus, which is the elicitation of personal information, sharing of location, arranging further contact and meetings and seeking out visual information via photos and the use of web cams. G3 groomers also seem to be more focused on discussing sexually explicit themes with the target. This is mostly focused on the target's sexual history, experience and introducing the target to masturbation.

These two corpora represent the two distinct approaches to grooming emerging in the above analysis, which are presented in Figure 23. While there are four different grooming intentions in short duration grooming, only two can be identified in long duration groomers.

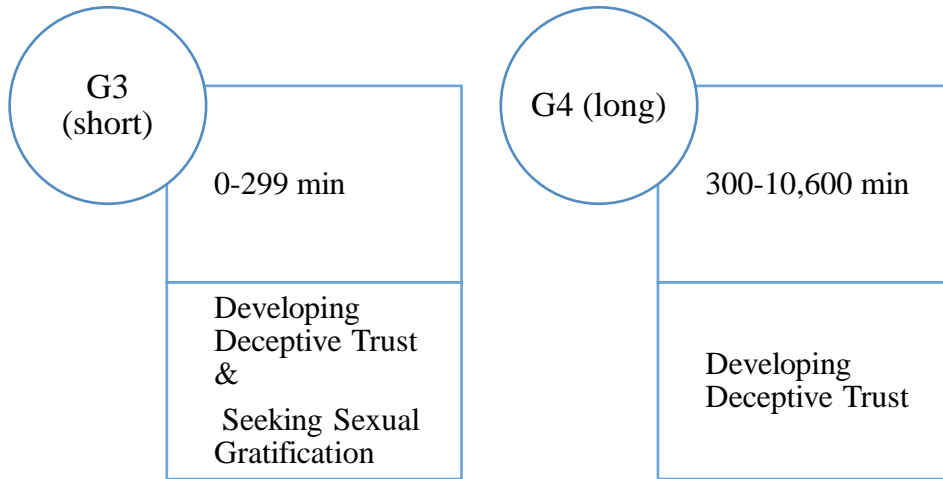


Figure 23: The two grooming approaches

The second analysis suggests that the first two sub-corpora G1 (0-99 minutes) and G2 (100-199 minutes) are fairly homogenous and do not show much difference in language use while being quite focused on developing deceptive trust with the target. Three grooming intentions (developing deceptive trust, arranging further contact, negotiating power dynamics) are identified on the keyword level in G1. The analysis shows that they are also present in G2, however less saliently, demonstrated in Figure 24.

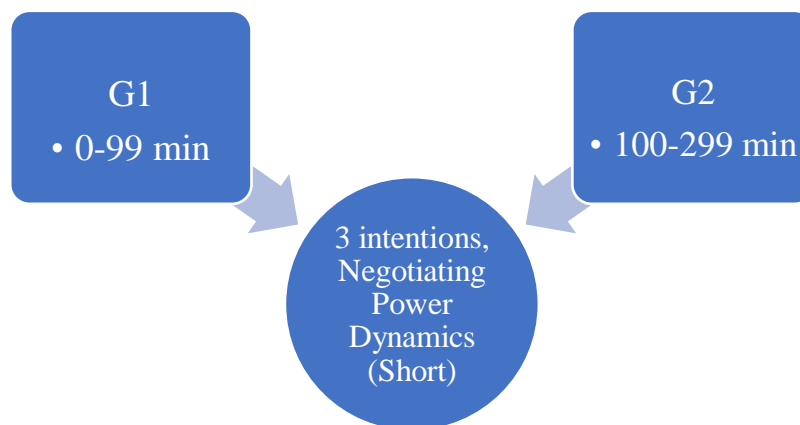


Figure 24: Short grooming duration similarities

The analysis also revealed that grooming duration does not seem to make a difference below the cut-off point of 300 minutes. However, negotiating power dynamics by clustering positive politeness that becomes impolite as a result, begging and guilt

tripping the target, emerges as a grooming intention unique to groomers who interact with their targets for less than 300 minutes. Seeking sexual gratification is mostly attained by combining other grooming intentions, as it does not emerge as its own salient grooming intention in the keywords.

The third analysis echoes the general patterns of the first analysis. It shows that groomers who spend longer interacting with their targets seem to be split into two groups (Figure 25): In terms of the grooming intentions, the focus of groomers who spend less time interacting with their targets seems to be developing deceptive trust and seeking sexual gratification, but they also make use of the grooming intentions of assessing risks and isolating and arranging further contact, which are identifiable on the keyword level. These are the same grooming intentions found in the third sub-corpus (G3), a comparison can be seen in Figure 25. The focus within the developing deceptive trust intention is slightly different, however. In G5 there is a focus on sociability, rather than exact location, visual information and arranging further contact. The sub-corpus has the highest number of sexually explicit keywords. This suggests that the language of groomers in G5 is more similar to that of groomers in G3 and is part of the short duration grooming approach.

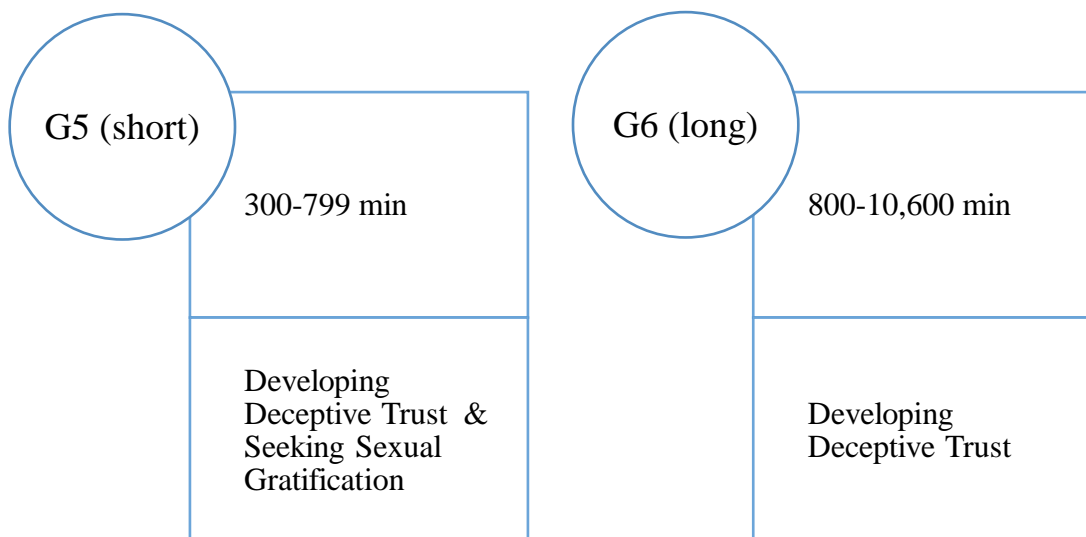


Figure 25: Comparing G5 and G6

The sixth sub-corpus, on the other hand, focuses on developing deceptive trust above all other grooming intentions. The groomers compliment the target, use terms of

endearment, express their emotions, and use relationship terms to negotiate and strengthen the foundation of the deceptive trust. The same sexually explicit keyword (*sexy*) that was identified when comparing them to the third sub-corpus, emerges again as the only sexually explicit keyword. No other grooming intentions can be identified on the keyword level.

G3 (short)	G5 (short)	G4 (long)	G6 (long)
<ul style="list-style-type: none"> • DDT <ul style="list-style-type: none"> • location, visual info, contact • SSG <ul style="list-style-type: none"> • 5 KWs 	<ul style="list-style-type: none"> • DDT <ul style="list-style-type: none"> • sociability • politeness • SSG <ul style="list-style-type: none"> • 8 KWs 	<ul style="list-style-type: none"> • DDT <ul style="list-style-type: none"> • relationship • trust <ul style="list-style-type: none"> • terms of endearment 	<ul style="list-style-type: none"> • DDT <ul style="list-style-type: none"> • relationship • trust • SSG <ul style="list-style-type: none"> • advances DDT

Figure 26: Comparison of short and long grooming analyses

Seeking sexual gratification seems to be hidden behind less overtly sexual grooming intentions and be more implicit than in the seemingly highly sexualised approach of groomers who spend less time interacting with their targets. This suggests that while two distinct grooming approaches with different grooming intentions and individual strategies can be observed, the cut-off point seems to be at the sixth sub-corpus, and thus around 800 minutes.

Groomers who spend 0-800 minutes interacting with their targets, use similar language and grooming intentions, while there is a clear difference to groomers who spend longer than that (>800) minutes interacting with their targets. Nevertheless, a clear duration cut-off point cannot be defined, as duration lies on a continuum. The keywords suggest the more time groomers spend with their targets, the more they focus on developing deceptive trust and building a relationship with the target over time, while seeking sexual gratification seems to be a secondary goal and not identifiable at the keyword levels. Sexual explicitness is at the other end of this continuum with shorter duration groomers making use of more grooming intentions and simultaneously having a clear focus on seeking sexual gratification. It is acknowledged that the duration cut-off points in this chapter could impact on the

results seen. Especially in the chat logs that are only a minute under or over the next cut-off point e.g. a chat log that is 299 minutes long or 799 minutes long³¹.

This chapter has explored duration of grooming by first looking at segmenting a corpus of grooming according to time and then carrying out three keyword analyses based on six duration based sub-corpora. The analysis shows that duration does make a difference and influences the number and type of grooming intentions used. Groomers who spend less time interacting with their targets use more grooming intentions and highly sexualised language, while groomers who spend a longer time grooming their targets take more time to develop the deceptive trust. Grooming duration of less than 300 minutes is homogenous and represents one grooming approach. The language changes around the 800 minutes mark but it is not possible to draw a clear duration cut-off point, as duration lies on a continuum.

Grooming is about manipulation, and it seems the less time groomers spend interacting with their targets, the more resources they use to manipulate them (see section 5.2.4). Previous findings have also suggested that the negotiating power dynamics grooming intention operates on a different level from the others. The next chapter will investigate manipulation, looking specifically at a speech act, requests, to analyse markers of power dynamics and manipulation in groomers' requesting behaviour and to see if duration influences this, as it did the other grooming intentions.

³¹ Future research into duration could choose duration cut-off points that do not border one another, e.g. instead of choosing 0-299 minutes and 300-10,600 minutes, 0-299 minutes and 800-10,600 minutes could be chosen to address the issues of chat logs that are near the duration cut-offs and that could belong to either duration group.

6. Chapter 6: Power and Manipulation in Online Groomer Requests

As previous empirical chapters have shown, online grooming language differs from the language used in other digital contexts, including that of PAN2012, which was taken from general technical online support chats, as well as Omegle chats, which featured discussions about sex and dating (chapter four). Grooming duration also makes a difference to groomers' grooming intention orientation and level of directness with short duration groomers using more sexualised language and being more direct (asking for exact location, visual information and arranging further contact) than long duration groomers, who emphasise relationship and deceptive trust (using terms of endearment); Directness is a variable in different duration approaches to grooming (chapter five).

In a broad sense, online grooming is concerned with attempting to manipulate the target to do something using communicative grooming intentions (engage in/discuss sexual behaviour online, meet groomer offline, and so forth). Discursive manipulation is difficult to define and there is no 'consistent theoretical model' available (Maillat & Oswald, 2009: 348). As van Dijk (2006) points out, there are also no discursive elements that are only used for manipulative purposes. Instead, 'the same discourse structures are used in persuasion, information, education and other legitimate forms of communication.' (Van Dijk, 2006: 375). Manipulation is achieved when the interlocutor does not detect, avoid, or resist it due to incomplete or lack of relevant information, fundamental values that cannot be denied, strong emotions or their social position that will prevent them from disagreeing with the discourse put forward (van Dijk, 2006). Power and covertness are therefore crucial features of manipulation, specifically an unbalanced power dynamic in which one interlocutor has significantly more power over the other one leaves the other interlocutor vulnerable to being manipulated. In the case of online grooming, the groomer's ulterior motive is to sexually abuse his target either online or offline, which he tries to hide from his target, thus covertness exists in this context. Groomers are also adults whereas targets are believed to be minors, which leaves the targets further vulnerable to being manipulated, because targets will see the groomer as an authority figure and will likely go along with what he proposes. Groomers use this imbalanced power to their advantage. Furthermore, groomers engage their targets in discussions about emotions

and feelings and attempt to build a (deceptive) relationship with them, which will leave targets more vulnerable to being manipulated, as they are more likely to reciprocate if a groomer has already disclosed information to them and there is mutual trust.

This chapter will look at how groomers manipulate their targets in relation to a particular speech action that the groomers regularly perform, namely requesting something from the targets. The topic of the request varies, depending – amongst other factors – on the grooming intention/s being pursued at a particular point in the grooming interaction: seeking personal information, asking for an offline meeting etc. The particular realisation of the speech act of requesting, as performed by groomers, will also yield insights into the power dynamics involved in online grooming, as requesting is considered face-threatening by Brown and Levinson (1987) (see chapter three, sections 3.2.2.3 and 3.3.5.2.).

The chapter will first outline the specific methodological approach taken in the following analysis by providing a definition of what constitutes a request, its component parts (head act, support move) and the methodological steps followed in section 6.1. Section 6.2 will show the results of the analysis of three head act types with specific foci and section 6.3 will draw conclusions of requesting behaviour of G3 (0-299 minutes) and G4 (300-10,600 minutes) groomers.

6.1 Approach to Examining Requests in Online Grooming

A request in its most basic form is ‘getting another person to do something’ (Zinken, 2015: 23). This can be a request for information, for action, for objects/products or for assistance. In the context of online grooming, requests for information and action are particularly pertinent, as both are used to contribute to various grooming intentions, e.g., exchanging details about general likes and dislikes to build trust. Requests for actions can include anything from participation in sexual activities to deleting evidence (e.g., chat archives). In the PJ online grooming data used in this study, a request is any instance in which the groomer attempts to get the target to do something. This includes a broad number of things, such as giving the groomer personal information, participating in discussions about sensitive topics, meeting the groomer

offline, turning off chat archives, or calling the groomer, etc. Requests are generally regarded as being face-threatening acts because they hinder the interlocutor's negative face want of not being impeded (Brown & Levinson, 1987).

Requests can be made directly or indirectly. Indirect requests mitigate some of this potential face-threat, while direct requests are at risk of being interpreted as a face-attack. Blum-Kulka, House and Kasper (2011) developed nine head act types (six of which were outlined in chapter three, section 3.3.5.2) that are on a scale of directness (see Table 30). They range from the most direct requests made by using imperatives (*do it*) to the least direct way of requesting, which is making strong or mild hints.

Table 30: Request strategy types taken from Blum-Kulka et al. (2011: 349-350)

Request strategy type	Definition
mood derivable	The grammatical mood of the verb in the utterance marks its illocutionary force as a request.
performatives	The illocutionary force of the utterance is explicitly named by the speakers.
hedged performatives	Utterances embedding the naming of the illocutionary force.
obligation statements	The illocutionary point is directly derivable from the semantic meaning of the locution.
want statements	The utterance expresses the speaker's intentions, desire or feeling vis a vis the fact that the hearer do X.
suggestory formulae	The sentence contains a suggestion to X.
query preparatory	Utterance contains reference to preparatory condition (e.g. ability or willingness, the possibility of the act being performed) as conventionalized in any specific language.
strong hints	Utterance contains partial reference to object or to elements needed for the implementation of the act (directly pragmatically implying the act).
mild hints	Utterances that make no reference to the request proper (or any of its elements) but are interpretable through the context as requests (indirectly pragmatically implying the act).

As mentioned at the end of the last chapter, groomers use more online grooming intentions and draw on more manipulative resources the less time they spend with their target. This is particularly apparent in the under-300-minute group (G3) compared with the over-300-minute group (G4). In this chapter it will be investigated whether

duration also influences the requesting behaviour of online groomers, drawing on the same sub-corpora.

As mentioned in chapter three (section 3.3.5.2) it can be difficult to query a corpus using a pragmatic concept that is not just a one-word unit. There is no automatic method for identifying requests using Corpus Linguistics, unlike there is for keywords and collocates. To mitigate this, first, a general literature search into Corpus Pragmatics, analysing requests using corpora, was conducted. Only a limited number were found, and they involved pragmatically tagging the corpus (John, Brooks, & Schriever, 2019; Culpeper & Tantucci) or a detailed reading of the entire corpus (Garcia Mcallister, 2014) which was not feasible in this thesis.

Requests are multi-word units that do not follow a set of syntactic structures, they are open-ended. Using imperatives alone, one could draw an almost endless list of requests in a corpus. The requests needed to be narrowed down to particular types of requests that were corpus searchable queries. In order to do so, a focused literature search for qualitative Pragmatics analyses of requests was carried out and yielded results. Indeed, requests are one of the most studied speech acts, across different discourse genres and data source types, such as: product requests (e.g., Gagne, 2018), phone calls (e.g. Curl & Drew, 2008), video recordings (e.g. Zinken, 2015; Wootton, 2005), data from written discourse completion tests (Blum-Kulka & Ohlstein, 1984) and emails (Merrison, Wilson, Davies and Haugh, 2012; Lorenzo-Dus & Bou-Franch, 2013; Savić, 2018). No studies of request structures in chat log interactions were found in the literature. From the reviewed studies, 14 formulaic request head act type structures, that followed a set word structure and thus became searchable, were obtained, and turned into 24 search queries to extract requests from G3 and G4 (see Table 31). Spelling varieties were added to account for misspellings or common ways to spell specific words. These search queries were then sorted into the nine head act types of requests.

Table 31: The formulaic request head act types and corpus search queries

Request structure	Corpus search queries	Head act type
-	-	Mood derivable
asking you	ask* ³² you, ask* u	performatives
would like you	would like you, would like u	hedged performatives
want you to	want you to, want u to	want statement
wish you would	wish you would, wish u would	
wish you could	wish you could, wish u could	
you will have to	you will have *, u will have *	obligation statement
how about	how about	suggestory formulae
what about	what about	
would you mind	would you mind, would u mind	query preparatory
can you	can you, can u	
could you	could you, could u	
can I	can I	
could I	could I	
wonder if	wonder if ³³	-
-	-	strong hints
-	-	mild hints

No formulae could be found for the mood derivable requests, strong hints, and mild hints via the above corpus searches. These are the most and least direct of the request head acts and highlights the limitation of Corpus Linguistics of identifying formulaic language mapped onto a specific speech act and searching for it using a corpus. Since speech acts are not always inherently linked to function, it can be difficult to pin down less formulaic speech acts using a corpus:

The use of language corpora in pragmatics, and more specifically in the investigation of speech acts, is, however, problematic to some degree. While the starting point in corpus linguistics is always a linguistic form that is to be searched for in a corpus, pragmatics often takes a functional perspective.

³² The * denotes a wild card character, where any character can follow, such that: “ask, asking, asks” are all considered in the search query “ask*”

³³ Upon examination of the requesting sequences in which this structure was embedded in the data, *wonder if* was found to serve as a support move, rather than a request head act, so it was excluded from the search queries.

Language functions, however, do not lend themselves to searches in language corpora per se. (Flöck, 2011: 70).

This is one of the reasons why a CADS approach was used to analyse requests in online grooming, as it can follow a quantitative analysis up with a more qualitative analysis, which can uncover other less form-bound language functions. In fact, during the qualitative analysis seven mood derivable requests were identified. They had the following imperative forms: name it, tell me, call me (twice), don't turn me in, remember and stay on. These are by no means the only mood derivable requests in the corpus or an exhaustive list as it is not possible to extract all the imperative forms of verbs from the corpus and emerged as part of the analysis. The only unidentified head acts remain hints, which are impossible to identify in a corpus using a set of formulaic structures as the ones above.

Concordance lines were extracted for the seven identified head acts, incorporating all 24 search queries (one query, *wonder if*, was excluded from the list of search queries, see footnote 31). If there were more than 50 concordance lines, a sample of 50 random concordance lines were extracted and analysed. Non-requests such as statements and indirect speech/reported speech containing these formulae were discarded, resulting in 571 concordance lines from G3 and 1,006 from G4 all containing a request (including mood derivable concordance lines that were identified later). A breakdown of the number of concordance lines per head act type can be seen in Table 32.

Table 32: Breakdown of number of request concordance lines per head type act

Head act type	Requests (G3)	Requests (G4)
performative	73	99
query preparatory	212	465
suggestory formulae	99	100
hedged performatives	5	46
want statement	100	100
obligation statement	78	193
mood derivable	4	3
Total	571	1006

The request concordance lines were subsequently coded manually for grooming intentions at the macro level. As one request could – and did – contribute towards

more than one grooming intention, multiple codes could and were assigned to one concordance line. The concordance lines were also coded for whether they included one or more support moves (see section 6.2.1.2).

6.2 Results

6.2.1 Overview

The chapter will now move on to the results of the analysis by first looking at requesting behaviour in online grooming in general and the distribution of head act types and online grooming intentions (sub-section 6.2.1.1). Then, the general support move functions will be defined in sub-section 6.2.1.2 and compared in general terms within the online grooming corpus.

6.2.1.1 Requesting Behaviour

An overview of the online grooming intention per head act type in G3 and G4 (in percentage) can be seen in Tables 33 and 34 (mood derivable concordance lines are not included here as there were too few of them). As mentioned above, the requests were extracted while already classified into the six head act types. These concordance lines were then analysed qualitatively and sorted into contributing towards one or more online grooming intentions (percentages are shown in the Tables below)³⁴.

Table 33: Online grooming intentions in request head act types in G3

Head act type	G3					
	Requests (raw)	DDT (%)	SSG (%)	NPD (%)	ARISO (%)	AFC (%)
performative	73	63.0	63.0	17.8	11.0	24.7
query preparatory	212	35.4	38.7	0.5	5.7	56.1
suggestory formulae	99	44.4	33.3	3	23.2	58.6
hedged performatives	5	0	60	0	20	40
want statement	100	29.0	60.0	18.0	18.0	51.0
obligation statement	78	39.7	39.7	21.8	15.4	29.5
mood derivable	4	100	50	25	0	75

³⁴ In both tables request numbers are raw numbers and grooming intentions are shown in percentage.

Table 34: Online grooming intentions in request head act types in G4

Head act type	G4					
	Requests (raw)	DDT (%)	SSG (%)	NPD (%)	ARISO (%)	AFC (%)
performative	99	70.7	52.5	14.1	21.2	23.2
query preparatory	465	85.4	42.4	2.6	6.7	51.4
suggestory formulae	100	81	41	6	14	34
hedged performatives	46	54.3	65.2	10.9	0.0	65.2
want statement	100	65	50	17	8	53
obligation statement	193	77.7	40.9	9.3	12.4	49.2
mood derivable	3	100	33.3	0	33.3	0

Looking at G3, out of the 73 performative requests, 63% contribute towards developing deceptive trust (DDT) and seeking sexual gratification (SSG), 17.8% of them are aligned with negotiating power dynamics (NPD), 11% contribute towards assessing risks and isolating (ARISO) and 24.7% to arranging further contact (AFC).

Out of the 212 query preparatory requests in G3, over half contribute to arranging further contact (56.1%) and just over a third to seeking sexual gratification (38.7%). Developing deceptive trust (35.4%) also seem to be quite prevalent, while only around five percent of the requests are oriented towards assessing risks and isolating and under one percent towards negotiating power dynamics. The suggestory formulae requests (n=99) in G3 contribute mostly to arranging further contact (58.6%) and developing deceptive trust (44.4%). Seeking sexual gratification also features in a third of these requests (33.3%). Assessing risks and isolating is less represented (23.2%) and once again negotiating power dynamics is the least frequent online grooming intention in these requests (three percent). The focus of the five hedged performative requests in G3 seems to be on contributing towards seeking sexual gratification (60%) and arranging further contact (40%). Assessing risks and isolating is also present (20%), while developing deceptive trust and negotiating power dynamics are not promoted by these requests.

Want statement requests in G3 (n=100) are oriented towards seeking sexual gratification (60%) and arranging further contact (51%). Developing deceptive trust occurs in just under a third of want statement requests (29%) analysed while 18% of

concordance lines contribute towards negotiating power dynamics and assessing risks and isolating. Finally, obligation statement requests (n=78) in G3 are mostly aligned with developing deceptive trust (39.7%) and seeking sexual gratification (39.7%). Just under a third of them (29.5%) are oriented towards arranging further contact while a minority contribute towards negotiating power dynamics (21.8%) and assessing risks and isolating (15.4%).

In G4, the 99 performative requests are mainly oriented towards developing deceptive trust (70.7%). Just over half of them (52.5%) also contribute to seeking sexual gratification while just under a fourth (23.2%) of them feature arranging further contact. Assessing risks and isolating (21.2%) and negotiating power dynamics (14.1%) were less common in these requests.

The 465 references to preparatory condition requests in G4 also mainly feature developing deceptive trust (85.4%). Half of them (51.4%) are aligned with arranging further contact and 42.4% contributed towards seeking sexual gratification. Assessing risks and isolating (6.7%) and negotiating power dynamics (2.6%) are far less frequent. Suggestory formulae requests (n=100) are mostly oriented towards developing deceptive trust (81%). Seeking sexual gratification (41%) and arranging further contact (34%) are also prevalent in these requests while 14% of the concordance lines contribute towards assessing risks and isolating. Negotiating power dynamics is not very common in the requests in G4 (6%). Want statement requests (n=100) in G4 mainly contribute towards developing deceptive trust (65%) and arranging further contact (53%). Half of them (50%) also feature seeking sexual gratification while negotiating power dynamics (17%) and assessing risks and isolating (8%) are less common. The 193 obligation statement requests in G4 are mainly aligned with developing deceptive trust as well (77.7%) while just under half of them (49.2%) feature arranging further contact and 40.9% of them contribute towards seeking sexual gratification. Assessing risks and isolating (12.4%) and negotiating power dynamics (9.3%) are less frequent in the requests in G4.

Notably, in G3, the frequency of use of different request head act types aligned to developing deceptive trust range from no occurrences in hedged performatives to 63% in performatives. In contrast, the equivalent percentage in G4 aligned to developing

deceptive trust start from 54.3% in hedged performances and reach 85.4% in references to preparatory conditions and they are prevalent across head act types. The percentages for seeking sexual gratification, negotiating power dynamics, and arranging further contact in the head act types seem to be similar in G3 and G4. Assessing risks and isolating is more varied ranging from 5.7% in query preparatory requests to 23.2% in suggestory formulae in G3 and zero occurrences in hedged performative requests to 21.2% in performative requests in G4.

As the overview above left some questions unanswered, I next identified and examined the number and percentage of requests aligned to each online grooming intention in G3 and G4 (Figure 27) for all head act types combined (mood derivable requests are not included, because only a few were identified by different means than the other requests).

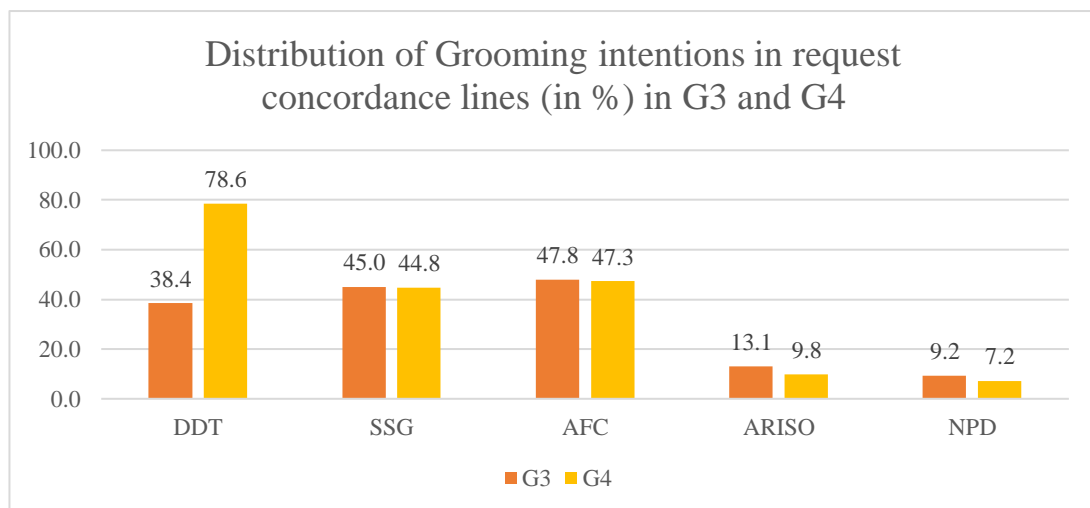


Figure 27: Distribution of online grooming intentions in request concordance lines

The percentages in Figure 27 show that developing deceptive trust is the prevalent online grooming intention in G4 (78.6%) as already suspected. In contrast, it is only the third highest online grooming intention in G3 (38.4). The percentages for assessing risks and isolating in the request concordance lines also differ, with 13.1% in G3 and 9.8% in G4. The figures for seeking sexual gratification and arranging further contact seem to be on par in G3 and G4. Negotiating power dynamics, the way it is currently defined in the OGDm is least found in the request concordance lines in both G3 (9.2%)

and G4 (7.2%). This suggests some difficulty in analysing this particular online grooming intention the way it is defined in the original model.

To find out more about the support of these requests, Figures 28 and 29 were produced. They show that in both sub-corpora more requests are supported than left unsupported in general. Specifically, 55% of request concordance lines in G3 receive support, while 45% are made without support moves. In G4, 70% of requests are accompanied by one or more support moves while 30% are unsupported.

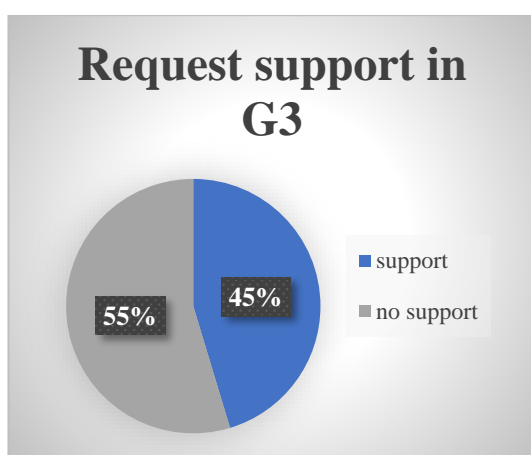


Figure 28: Request support in request concordance lines (n=567) in G3 in percentage

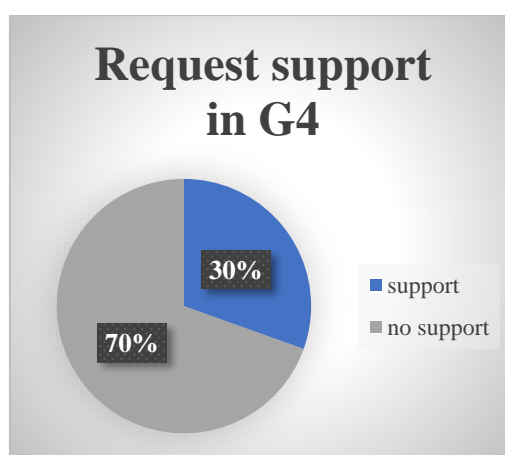


Figure 29: Request support in request concordance lines (n=1004) in G4 in percentage

Table 35 shows this in more detail. Broadly speaking, the preference of giving support or no support to specific head act types in G3 and G4 for the individual requests is similar. Performative requests (68.5% in G3 and 53.5% in G4), hedged performative requests (100% in G3 and 97.8% in G4) receive more support than no support. Query preparatory requests (56.1% in G3 and 72.3% in G4) and suggestory formulae requests (84.8% in G3 and 92% in G4) do not receive support more than they do. One exception to this is want statement requests, where 57% of the requests receive support in G3 but 56% do not receive support in G4. Another exception is the obligation statement, where more requests are supported in G3 (52.6%) than not supported while they receive no support (86.1%) in more cases than they are supported in G4.

Table 35: Support (in %) in G3 and G4 per head act type

Head act type	G3		G4	
	Support (%)	No support (%)	Support (%)	No support (%)
performative	68.5	31.5	53.5	46.5
query preparatory	43.9	56.1	27.7	72.3
suggestory formulae	15.2	84.8	8	92
hedged performatives	100	0	97.8	2.2
want statement	57	43	44	56
obligation statement	47.4	52.6	13.5	86.5
mood derivable	100	0	100	0

As the above still left questions unanswered about the specific differences in grooming intention distribution and how groomers use support moves to help their requests, I decided to delve further into the patterns and focus on specific online grooming intentions and head act types. Since the biggest difference between G3 and G4 is in developing deceptive trust, this was chosen as the first head act type. More specifically, suggestory formulae and obligation statements that featured developing deceptive trust were chosen. Suggestory formulae with a focus on developing deceptive trust were chosen due to the big difference in G3 (44.4%) and G4 (81%) seen in Tables 33 and 34 and having low support in both G3 (15.2%) and G4 (8%) (see table 35). Obligation statement requests with a focus on developing deceptive trust were chosen because there is also a noticeable difference in developing deceptive trust in obligation statement concordance lines in G3 (39.7%) and G4 (77.7%). However, in this case, the support is low in G4 (13.5%) and high in G3 (47.4%). The second online grooming intention to focus on was assessing risks and isolating as this showed the second biggest difference in G3 and G4. The specific head act type chosen was want statement, which has high support in both G3 (57%) and G4 (56%).

Therefore, the concordance lines that fit the following criteria were extracted from both G3 and G4:

- 1) First analysis
 - a. Head act type: Suggestory formulae
 - b. Online grooming intention: DDT

- c. Support: with support
- 2) Second analysis
 - a. Head act type: Obligation statement
 - b. Online grooming intention: DDT
 - c. Support: with support
- 3) Third analysis
 - a. Head act type: Want statement
 - b. Online grooming intention: ARISO
 - c. Support: with support

Table 36 is an overview of the number of concordance lines and their head act structures from both G3 and G4 that were selected for further qualitative analysis. These concordance lines were analysed in their context including the full requesting sequence and context, which included up to 23 conversational turns in context. The qualitative Discourse Analysis was informed by Blum-Kulka et al.'s 2011 framework on support move types, im-politeness frameworks (Brown and Levinson, 1978, 1987; Culpeper 1996, 2005, 2011), and the OGD (Lorenzo-Dus et al., 2016), which also drew upon relational work concepts (Locher and Watts, 2005, 2008).

Table 36: Selection of head acts and number of concordance lines analysed

Head act type	Head act type structure	G3	G4
Suggestory formulae	how about	2	7
	what about		
Obligation statement	you have to	8	15
	u have to		
	you will have (to)		
	u will have (to)		
Want statement	want you to	10	3
	want u to		
Total	-	20	25

Overall, 20 concordance lines were analysed qualitatively in G3 and 25 in G4, which were split into the three head act types. Additionally, four concordance lines in G3, and three in G4 featuring mood derivable requests, were analysed after being identified as part of the qualitative analysis of obligation statement and suggestory formulae requests. The specific support move functions within their head act types

will now be defined and then examined in more detail using extended examples of concordance lines showing groomer-target interactions in context.

6.2.1.2 Support Move Functions

As mentioned above, the concordance lines were coded for the type of support the requests received. The four support move functions (Table 37) were informed by Blum-Kulka et al.'s (2011) framework, originally developed in Blum-Kulka and Ohlstein (1984), which is 'a pragma-linguistic one, i.e., it attempts to capture the pragmatic functions played by various linguistic elements in the discourse' (Blum-Kulka & Ohlstein, 1984: 203) and the Pragmatics request literature. The support move functions were adapted from this framework and modify the head act internally (within the head act) and externally (within the requesting sequence).

Table 37: Support move functions in online grooming request concordance lines

Support move function	definition	Examples
minimising	The support move function softens/minimises the imposition on the target's negative face by e.g., using hedges, discourse markers and vague language	<ul style="list-style-type: none"> - i just want you to know how serious it could be - what about the friend thingy? - oh well i guess you have to make some huh
intensifying	The support move function intensifies/boosts the request force through devices such as boosting adverbials, request reinforcement and tag questions	<ul style="list-style-type: none"> - you have to always know that i love you - if you want me to trust you you have to trust me right - how about desert , whats your favorite desert ?
giving grounds	The support move function gives grounds for a request via speech acts such as explanation, condition, or threat	<ul style="list-style-type: none"> - you have to promise to wear it though or i'll feel really bad - i don't want you to hurt her but i want to see you tonight - i don't want you to get in trouble ... will she find out ?

enhancing interpersonal relations	The support move function enhances the positive face needs and emphasises interpersonal relations by using e.g., terms of endearment and promises	<ul style="list-style-type: none"> - i want u to be my baby girl - any thing u want all u have to do is ask
-----------------------------------	---	---

In the online groomer request structures, minimising support move functions are used to soften or minimise the imposition on the target and mitigate the perceived threat of the negative face. This is done by using hedges (*just, lol*), vague language (*friend thingy*), and discourse markers (*well, okay*), which introduce tentativeness, possibility, and implicit sexual intention. These were called understaters in Blum-Kulka & Ohlstein (1984) and downgraders in Blum-Kulka et al. (2011). However, minimising was deemed to be more accurate, as they minimise the imposition of the perceived face-threat towards the target’s negative face. In the context of online grooming, this strategic use of vague language applies to sexual content and can be a way of ‘approximating, avoiding and downtoning sexual intent’ (Lorenzo-Dus & Kinzel, 2021), highlighting developing deceptive trust over seeking sexual gratification.

Another support move function is intensifying, which has the opposite effect to that of minimising, and the request force is boosted. This is done by using tag questions (*right?, okay?*), which asked the target for confirmation or acknowledgment, repetitions of the head act type either in the same turn or separate turns, also called ‘request reinforcement’ (Lorenzo-Dus and Bou-Franch, 2013), and boosting adverbials (*always, ever*), which provide certainty and confidence. Request reinforcement has been found in other digital contexts, namely email messages in contexts of power asymmetry, specifically lecturer-student (lecturer + power; student – power) (Lorenzo-Dus and Bou-Franch, 2013). In the online grooming context, request reinforcement is used to intensify the request force by repeatedly asking the same question, which borders on harassment in some cases.

The third support move function, giving grounds, is used to provide reasons for making the request, which is supported by a condition or a threat. This was called grounder in Blum-Kulka & Ohlstein (1984) and Blum-Kulka et al. (2011). In the online grooming context, some of these giving grounds functions are linked directly

to online grooming intentions, such as not wanting to be caught, not wanting the target to get into trouble, or to know exactly where the target lived.

Lastly, one support move function is used to enhance the interpersonal relations between groomer and target by using terms of endearment (*hun sweetie, baby girl*), implying a certain closeness that is usually reserved for close friends, family members and significant others, as well as making promises, which addresses the positive face needs of the target and strengthens the bond between groomer and target.

The overall number of support move functions in the three chosen head act types (see section 6.2.1.1. for details) can be seen in Table 38 (n=55). Twenty-eight of these 55 support move functions are identified in G4, while the other 27 belong to G3³⁵. The most frequent support move function across corpora in the three head act types is giving grounds (34.5%) followed by both minimising (25.5%) and intensifying (23.6%). The least frequent support move function is enhancing interpersonal relations (16.4%).

Table 38: Overall number of support move functions

Support move function	#	%
Giving grounds	19	34.5
Minimising	14	25.5
Intensifying	13	23.6
Enhancing interpersonal relations	9	16.4
Total	55	100

A comparison of the support move functions found in suggestory formulae, obligation statement and want statement requests are shown in Figure 30 (G3) and Figure 31 (G4). The support move functions in the three chosen head act types show differences in G3 and G4. Giving grounds (37%) is the most common support move function in G3, while intensifying (39.3%) is the most used one in G4. Giving grounds (32.1%) is the second most used support move function in G4 while intensifying is the least used one in G3 (11.1%). This suggests that long duration (G4) groomers prefer to intensify their requests more than short duration (G3) groomers across the three head

³⁵ Note that more than one support move could occur in one concordance lines, and they were all counted here, which is why these numbers are higher than the number of extracted concordance lines in Table 36. Mood derivable requests were treated separately and are not included here.

act types under consideration. Minimising is used more (29.6%) in G3 than in G4 (17.9%) in these three head act types, suggesting a preference for minimising the imposition in G3 with these specific head act types. Enhancing interpersonal relations (22.2%) is the third most used support move in G3 in these types of requests while it is the least used one (10.7%) in G4. This is surprising given developing deceptive trust is the most used online grooming intention in G4, as shown in Figure 27.

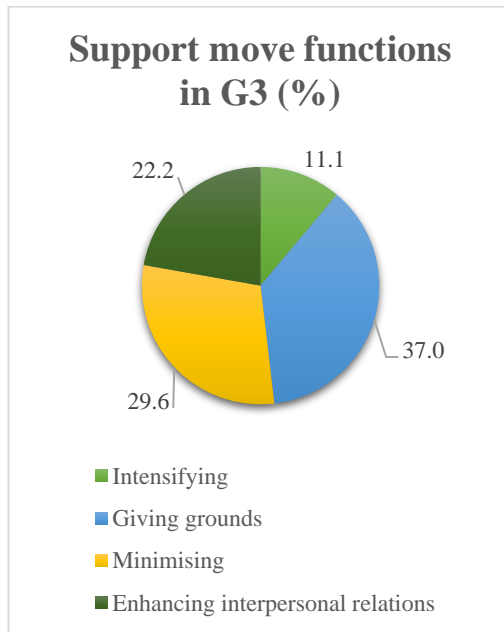


Figure 30: Support move functions in three head act types in G3 (in percentage)

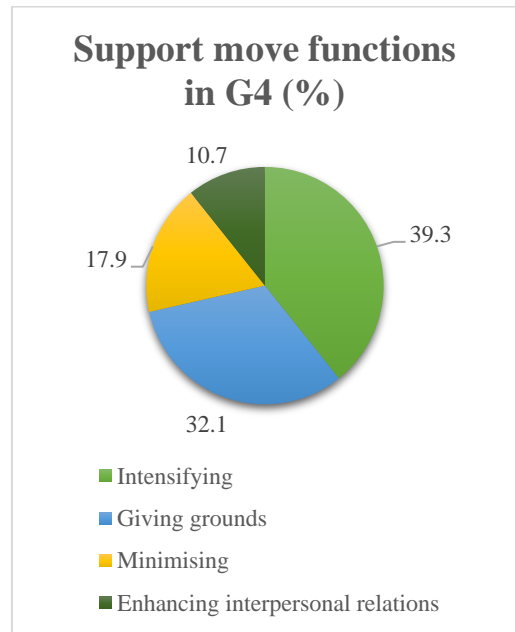


Figure 31: Support move functions in three head act types in G4 (in percentage)

The above defined support move functions will now be analysed in more detail, paying close attention to the similarities and differences between G3 and G4 and looking at relational work (politeness and impoliteness) in requesting sequences within online grooming interactions.

6.2.2 Requesting Behaviour in Online Grooming

As mentioned above, three head act types were chosen for further qualitative analysis. They will be examined below in turn. Sub-section 6.2.2.1 will show the analysis of suggestory formulae with a developing deceptive trust focus, sub-section 6.2.2.2, more specifically sub-section 6.2.2.2.1, will feature the analysis of obligation statement requests with a focus on developing deceptive trust. A further sub-section, 6.2.2.2.2 will contain a brief analysis of four of the seven additionally discovered mood

derivable requests that were discovered during the other analyses and analysed within their own subsequently identified head act type. Want statement requests with a focus on assessing risks and isolating will be analysed in sub-section 6.2.2.3. The analysis of the head act types, and their support move functions have been integrated. The head act type being analysed is put in bold script in the extracts below.

6.2.2.1 Suggestory Formulae – Developing Deceptive Trust Focus

The suggestory formulae requests with a developing deceptive trust focus are framed using *what about* and *how about* constructions, as mentioned above. They were selected not only because of the reasons outlined in sub-section 6.2.1.1 but also because of their manipulation potential in online grooming contexts. The support move functions in suggestory formulae concordance lines featuring developing deceptive trust in G3 (0-299 minutes) and G4 (300-10,600 minutes) are shown in Tables 39 and 40. The support types have a similar distribution in G3 and G4.

Minimising is the most used support type in suggestory formulae with a DDT focus, followed by intensifying and giving grounds. The only difference in distribution is that the giving grounds support move function occurred once in G3 but not in G4. Minimising occurs most frequently in G4 and occurs once in G3. This was done through hedging and discourse markers in G3. One further support move that is used in G4, in addition to hedging and discourse markers, is vague language. The following are some examples of this in G3 (extract 6.1) and G4 (extracts 6.2 and 6.3).

Table 39: Support move function in suggestory formulae concordance lines in G3

support move	#	%
minimising	1	33.3
intensifying	1	33.3
enhancing interpersonal relations	1	33.3
giving grounds	0	0.0
total	3	100.0

Table 40: Support move function in suggestory formulae concordance lines in G4

support move	#	%
minimising	3	42.9
intensifying	2	28.6
enhancing interpersonal relations	2	28.6
giving grounds	0	0.0
total	7	100.0

In G3 minimising support moves are used to minimise the imposition of a face-threatening act such as asking about the target’s relationship status. In G4, minimising is used to mirror the target’s language and vague language is used to obscure language to the point where only the groomer and target understand what is meant. Vague language almost acts as a code between the groomer and target and deepens the bond between them.

Extract 6.1

[Context: This interaction appears roughly 130 turns into the conversation, groomer and target are making small talk and discuss relationships.]

Turn	Speaker	Message
1	t	what about that girl in the pic with u ? she looks hot §_HAPPY-SMILEY_§
2	g	be right back
3	g	ok
4	g	she was just a friend from a while back
5	t	oh
6	t	u look pretty friendly §_HAPPY-SMILEY_§
7	g	i've been single for 6 months or so
8	g	oh and what about you §_HAPPY-SMILEY_§
9	t	i broke up with troy couple weeks ago
10	t	single now §_SAD-SMILEY_§

As can be seen in extract 6.1, the target uses an emoticon in the first turn. They ask the groomer about his profile picture and the other person in it. The groomer does not answer right away, excusing himself (*be right back*). When he does answer, he uses the minimiser *just* (*she was just a friend from a while back*). The target then pays the groomer a compliment, again using an emoticon *u look pretty friendly §_HAPPY-SMILEY_§*. In turn seven, the groomer further explains that the other person is a friend by stating he has *been single for 6 months or so*. Turn eight is the request concordance line, which is introduced by the discourse marker *oh*. It consists of a request for information about the target’s relationship status and an emoticon to minimise the imposition of the request. The groomer asks the target a personal question, which is face threatening but is also signalling his interest in the target, using positive politeness, and building trust, because they are discussing their personal relationships. The groomer seems to be reciprocating the target’s language by echoing a similar

question to the one in turn one. He might also be using emoticons to assimilate the target's language emphasising the deceptive trust building.

In G4, the minimiser used is *lol* rather than an emoticon in turn 13, which contains the request head act in extract 6.2. The context is also different. The groomer tries to convince the target to engage in sexual activities, in this case taking indecent pictures (turn two) and masturbation (turn six) and then indirectly asks the target whether he can be their boyfriend (turn 13).

Extract 6.2

[Context: This extract is taken around 450 messages into the conversation. The groomer and target are discussing a potential offline meeting and the groomer has used implicit language to introduce sexual content, which is getting more explicit in the extract.]

Turn	Speaker	Message
1	t	wait a bit because mom still up
2	g	orig: ahhh ³⁶ haha afraid your mommie is going to see naughty pictures ?
3	t	shell b in bed 930
4	t	yeah
5	t	she be pissed
6	g	or are you afraid she 's going to see you rubbing your little pussy
7	t	: ">
8	t	i don't wan na get n trouble
9	g	HAPPY-SMILEY_§ but you should rub your little pussy
10	t	taht don't rely feel rite
11	t	i want a real bf
12	t	not me lol
13	g	lol what about me
14	t	i dunno if u my brother r bf r waht lol
15	t	rite now we just chat lol
16	g	mm if I was your brother , then my thoughts on you would be just wrong §_HAPPY-SMILEY_§

These instructions to engage in sexual activities are framed as challenges and are also connected to the target's mother catching the target. In turn one, the target expresses concern that their mother is still awake and wants to postpone the activities. In turn two the groomer challenges the target and uses negative impoliteness to ridicule the

³⁶ Generally the original spelling of words that have been standardised using VARD (indicated by *orig*;) are deleted, except in cases where the normalisation is deemed incorrect and the original spelling is presented alongside the attempted normalisation.

target. He does this again in turn six when asking the target if they are afraid that they will be caught masturbating and dismisses their fears of getting in trouble in turn eight (*i don't wan na get n trouble*) by saying they should do it anyway (*HAPPY-SMILEY_§ but you should rub your little pussy*) and using an emoticon as a minimiser to mitigate the direct request for masturbation. The conversation then moves on to their relationship when the decoy says they want a real boyfriend, in turn 11 and 12. The groomer asks *what about me*, introduced by the minimiser *lol*. This is likely due to the face threatening act of asking the target to be their partner after establishing the role this partner plays in sexual activities just a moment before. *Lol* mitigates this face threatening act, and the target echoes the minimiser in the next two turns *i dunno if u my brother r bf r waht lol*. They also express confusion about their relationship and say *rite now we just chat lol* (turn 15). In the final turn, the groomer defines their relationship by stating his thoughts about the target would be inappropriate if they were siblings, using another minimising emoticon. This exchange also shows that the groomer uses morals to defend some of his behaviours, namely having inappropriate thoughts about an underage target by using incest as an example of something morally *just wrong* while not directly addressing the inappropriateness of the conversation considering the target is believed to be a minor.

The other minimising example in G4 is vague language use, which occurs in turn seven in extract 6.3.:

Extract 6.3

[Context: The interaction below concerns an offline meeting that the target is supposed to hide from their parents by saying they are staying with a friend. It occurs roughly 7,000 turns into the contact between groomer and target.]

Turn	Speaker	Message
1	t	how was work ?
2	g	sucks as usual
3	t	busy ?
4	g	so what do we do now
5	g	nope
6	g	am home now
7	g	what about the friend thingy
8	g	you there?
9	t	yea i guess we can do that

10	t	better not get sick to quick lol
11	g	well did you have any other way in mind ?
12	g	yeah
13	t	thatd probably work
14	g	when can you ?
15	t	hmm
16	t	u busy chattin ?
17	g	nope
18	g	just to you
19	g	so when can you be with a friend ?
20	g	do you want to do it day or night ?
21	t	wats better ?
22	g	i can do it friday night , or sat day , or sat night , or sund day

The groomer and target have talked about this ploy earlier in the interaction and only refer to it in turn seven as the *friend thingy*. The interaction starts with some small talk (turns one to six). The groomer attempts to steer the conversation towards the agreed ploy in turn four (*so what do we do now*). The target picks up the hint and agrees that they can do *that*, being vague towards the plan. It does not become clear they are arranging to meet until turns 19-22 where the groomer asks when the target is free and when they can *be with a friend*. He then refers to the plan as *it* in turn 20 being vague once again. Both the target and groomer seem quite secretive about their agreed plan using vague language to obscure their meaning. The interaction is difficult to understand without background knowledge, which has the desired effect of creating a bond between groomer and target and emphasising the (deceptive) trust between them. Vague language is used as a code between the two.

Intensifying is the second most used support move in G4 and it occurs once in G3. This is done by request reinforcement either in the same turn or in several turns throughout the interaction. G3 features one example of this, while suggestory formulae requests with a focus on developing deceptive trust in G4 contain two. Request reinforcement in suggestory formulae requests is used for clarification purposes in G4 (see extract 6.4) and to harass the target in G3 (see extract 6.5). Extract 6.4 shows an interaction that features a request reinforcement in turn three. In turn one, the groomer encourages the target to ask him anything they want to know, seeking to establish trust and make the target more dependent on him, replacing the target's support network to make them more compliant and willing to go along with suggested actions. Once the target confirms and asks a question in turn two (*yea ok y does my age bug ya*), the

groomer asks the target a question instead (turn three). He does so by asking a general question *how about desert* and immediately repeating the request for information *whats your favourite desert*, which shows use of positive politeness, as it attends to the target's interests. It is a request for information from the target, which is aligned with the developing deceptive trust grooming intention. The following turns show that a misunderstanding has happened.

Extract 6.4

[Context: The conversation takes place around 200 turns into the contact between groomer and target. They are making small talk, exchanging information about their general likes, dislikes and the groomer also throws in questions about the target's relationship status.]

Turn	Speaker	Message
1	g	well if there is anything yah like to ask..just ask I shall answer you
2	t	yea ok y does my age bug ya
3	g	how about desert , whats your favorite desert ?
4	t	oh my god cheesecake
5	g	why what you mean ?
6	t	cheesecake u no ?
7	g	never mind dumb ques lol
8	g	haha strawberry cheesecake
9	g	or cherry
10	g	no what I meant when I asked why what you mean when you asked y does ur age bug me

While the target is answering the groomer's dessert question in turn four (*oh my god cheesecake*), the groomer attempts to clarify what the target meant by their question about their age in turn two (*y does my age bug ya*). However, the target mistakenly tries to explain cheesecake instead (*cheesecake u no ?*). It is only when the groomer explicitly asks the target in turn 10 (*no what I meant when I asked why what you mean when you asked y does ur age bug me*) that the conversation is back on track. All of the questions the groomer asks, and what he proposes in turn one, are oriented towards the online grooming intention of developing deceptive trust by exchanging general information about likes and dislikes featuring positive politeness strategies to mitigate the face threat of the age-related question from the target, which points out the inappropriateness of groomer and target having this conversation in the first place.

The groomer also backtracks in turn seven after the misunderstanding occurred to minimise the imposition on the target (*never mind dumb ques lol*).

The request reinforcement instance in G3 is different in that it does not occur in the same turn but is instead repeated three times over the course of 18 turns (turns two, 14, 17 in extract 6.5).

Extract 6.5

[Context: The interaction happens in the middle of the interaction. The groomer and target are arranging to meet, and the groomer is asking about the target's house, narrowing down the exact location and a description of the house.]

Turn	Speaker	Message
1	g	how many rooms fence or no trees ?
2	g	is the number of the house visible from the street
3	g	?
4	g	mailbox ?
5	t	3 rooms n two trees out front
6	g	cars ?
7	t	u can park on the street
8	t	in front
9	g	ok sure it 's not for permitted cars
10	t	no cars at house no
11	g	what does your neirbors house look like
12	t	no its a free street its cool
13	g	i want to make sure i can find it
14	g	what about the house number is it visibtle from the street baby
15	t	lol lem me go look i never pain attention
16	g	?
17	g	see if the house number is visible
18	g	ok
19	t	yah u can see it from the street

The groomer starts asking the target questions about the house (*how many rooms fence or no trees ?*). He first asks whether the house number is visible from the street in turn two. This is emphasised by the question mark in turn three. He asks the target another question in turn four (*mailbox ?*). At this point the target answers the question the groomer asked in turn one (*3 rooms n two trees out front*). Before waiting for an answer to his other questions, the groomer proceeds to ask another question in turn six (*cars ?*). The target answers this question, which is discussed in more detail in turns seven to ten. After not answering the other questions, the groomer repeats the question

about the house number in turn 14 (*what about the house number is it visible from the street baby*). This time, he uses a term of endearment *baby* to recall closeness between him and the target and persuade them to answer the question. The target says they have to check because they never pay attention. Turn 16 consists of another question mark showing how impatient the groomer is. He then prompts the target to again check if the house number is visible from the street in turn 17, repeating the request for information once again. Throughout the exchange, the groomer is impolite and attacking the target's negative face of being free from imposition, as he harasses the target until they give him an answer to his question. The questions are delivered direct to the point of only consisting of one word and a question mark without any mitigation of this negative face attack, except for the one term of endearment, which addresses the target's positive face needs of being accepted by others. This is reminiscent of the findings in chapter five (section 5.2.1.1) in which the sub-corpus of short duration grooming (G3) is very focused on exact location of the target and how one groomer harasses a target by asking for their address nine times within 300 lines of conversation using threatening language.

Enhancing interpersonal relations by using terms of endearment is present in both G3 and G4 in suggestory formulae requests. These terms of endearment are used to address the target's positive face needs and show a closeness between groomer and target, directly contributing towards the grooming intention of developing deceptive trust, which is used as a trust foundation to proceed with other grooming intentions. We have already seen an example of a term of endearment in G3 extract 6.5 above (turn 14) where it is used to remind the target of the closeness between them and the groomer to get the information the groomer wants. The following (extract 6.6 and extract 6.7) are the two instances in G4 where it is also used to emphasise the familiarity of groomer and target, evoking trust:

Extract 6.6

[Context: The interaction happens around 2,600 turns into the contact between groomer and target. They have just discussed a person in the groomer's class and their absent sense of humour, and the target is now attempting to make small talk.]

Turn	Speaker	Message
------	---------	---------

1	t	watcha doing ?
2	g	be right back
3	g	back
4	g	sorry
5	g	nothing
6	g	just filling up some rebate forms
7	t	k
8	g	how about u kiddo

In the first example (extract 6.6), the groomer uses a term of endearment (turn eight) after briefly excusing himself and delaying answering the target's question (*watcha doing ?*) in turn one. He apologises for his brief absence from the interaction in turn four, which addresses the target's positive face needs. The use of *kiddo* in turn eight might be in response to the short answer from the target in turn seven (*k*), which could indicate the target is upset by the groomer's leaving. *Kiddo* is thus used to show the target they can trust the groomer and a way to mitigate the positive face threat of not giving them his full attention.

Extract 6.7

[Context: The interaction starts 14 lines into the contact between groomer and target. They are getting to know each other and have just exchanged basic information, such as their age, sex and location]

Turn	Speaker	Message
1	t	im chelsea
2	g	bobby here nice name
3	d	ty ! ntmu bobby
4	t	what u do 4 fun bobby
5	g	lots of things sweetie
6	g	what about u ?

After each stating their names in turns one and two, the groomer immediately pays the target a compliment (*nice name*), which attends to the target's positive face needs. The target thanks the groomer and says it is nice to meet him (turn three). They then ask what the groomer likes to do (*what u do 4 fun bobby*). In turn five the groomer noncommittally says *lots of things sweetie*. The term of endearment is here used to show the target they can trust the groomer. This is done by using an affect term that might be used by parents and friends of the target. Using *sweetie* also addresses the

target's positive face needs. Terms of endearment are inherently part of the developing deceptive trust grooming intention and positive politeness strategies. As seen in chapter five, for instance, terms of endearment are used to build trust and develop the relationship in the long grooming duration sub-corpus (G4), which seems to have resurfaced here to support requests for information. The sole term of endearment in G3 seems to be used to speed up the process of receiving the information the groomer wants.

To summarise, minimising (lol, vague language) in requests by long duration groomers (G4) is used to challenge the target and mitigate sexual content, while vague language is also used to create a mutual language or code. Specifically, the expression *friend thingy* is used to hide a ploy to meet up with a target. This code is used to strengthen the bond between the target and groomer and to promote secrecy. In short duration groomer language (G3) minimisers (emoticons) seem to be used to assimilate the target's language use, mirror the target's language, and minimise the imposition of face threats to the target's negative face. Intensifying by long duration groomers in G4 is used for deceptive trust purposes, asking about general likes/dislikes, questions, and small talk. It is used very generally for developing deceptive trust purposes. In G3, the one request reinforcement is used by short duration groomers to find out the exact location of the target and make sure the groomer can find them. This is reminiscent of the finding in chapter five that G3 groomers are more focused on knowing the target's precise location and that G4 groomers take more time to build trust with the target.

The enhancing interpersonal relations support move in G4 is used generally for trust building, evoking trust, and using a term used by people close to the target to appear trustworthy while in G3 there is only one occurrence, which is used primarily to speed up a request for information, which also happens to be in the form of a request reinforcement. The function in G3 is more specific than the function in G4. Negative Impoliteness (ridicule, harassment) is used in both G3 and G4 to convince the target to do something (giving the groomer the exact location and masturbation/taking indecent photos, respectively). This suggests impoliteness also plays a role in negotiating power dynamics and requesting actions/information from the target.

6.2.2.2 Obligation Statements – Developing Deceptive Trust Focus

6.2.2.2.1 Obligation Statement Requests

The analysis will now move on to the second request head act type that was analysed qualitatively, namely obligation statements with a focus on developing deceptive trust. This head act type consists of the following structures: you have to, u have to, you will have (to), u will have (to). In terms of politeness, these structures are understood as inherently impolite, as they are commands, which threaten the target’s negative face.

Tables 41 and 42 show the support move functions in obligation statement concordance lines with a focus on developing deceptive trust in G3 (0-299 minutes) and G4 (300-10,600 minutes). The support move functions in these requests look quite different in G3 and G4. Minimising is the most used support move function in G3, while intensifying is second most used in G4. Enhancing interpersonal relations is the second most used support type in G3, while being the least frequently used one in G4. Giving grounds and intensifying are only used once each in G3 while they are the most and second most used support moves in G4.

Table 41: Support move function in obligation statement concordance lines in G3

support move	#	%
minimising	6	60.0
enhancing interpersonal relations	3	30.0
giving grounds	1	10.0
intensifying	0	0.0
total	10	100.0

Table 42: Support move function in obligation statement concordance lines in G4

support move	#	%
giving grounds	8	44.4
intensifying	7	38.9
minimising	2	11.1
enhancing interpersonal relations	1	5.6
total	18	100.0

Minimising is the most frequently used support move function in G3 in obligation statement requests with a focus on developing deceptive trust, appearing in over half of the requests (60%) while they are the third most frequently used type in G4, supporting only 11.2% of the request. Minimising is done by using discourse markers,

hedges, and vague language in G3. In G4, minimising is done by using hedges. A few such hedges (*I guess, huh*) and a discourse marker (*well*) in G3 appear in turn four in the extract 6.8 below:

Extract 6.8

[Context: This interaction appears just under 60 turns into the contact between groomer and target. They have introduced themselves, exchanged their age, sex and location and are now getting to know each other.]

Turn	Speaker	Message
1	g	im surprised your on here and not out with your friends
2	t	i just moved here so i don't really have any friends
3	g	oh
4	g	well i guess you have to make some huh
5	g	lol
6	t	i was supposed to see my friends near my grandma's in milwaukee but my mom went to the outlet malls in illinois instead
7	t	i do
8	t	probably will when i start school on monday
9	g	ya
10	g	she didn't take you with her
11	t	no she said it was a girls weekend away

The groomer expresses surprise that the target is chatting online instead of spending time with their friends (turn one) to which the target replies they just moved there and do not have friends (turn two). The groomer uses a discourse marker *oh* in turn three and then says the target has to make some friends, adding the hedge *I guess* acting as a minimiser here. The request is also framed by another discourse marker *well*. The minimiser makes his request sound noncommittal and less like a command. He also adds *huh* to get confirmation from the target. The two then continue to discuss friends. The target suggests they will make friends when they start school in turn eight (*probably will when i start school on monday*). The pair also briefly discuss the target's mother not taking them with her (turns six and 10), touching on the assessing risks and isolating grooming intention. Overall, this interaction is part of the developing deceptive trust grooming intention, as relationships are discussed, and the groomer tentatively encourages the target to make friends. The obligation statement request, which is usually a direct request is considerably minimised and becomes an indirect suggestion or encouragement, rather than a command.

The same hedge minimisers (*I guess*) in obligation statement requests in G4 is quite similar (turn 13 in extract 6.9). The interaction itself is a good example of asynchronous computer mediated discourse where the interlocutors do not provide straightforward answers to messages but instead bounce from topic to topic because the messages sent have not yet been read or received when a new message has been sent. Thus, the answer to the groomer's question in turn four, whether the target lives north or south of Santa Rosa, is not answered until turn eight and the groomer's question in turn 12 *so your going to school ?* receives an answer in turn 17.

Extract 6.9

[Context: The interaction starts roughly 500 turns into the contact between groomer and target. The groomer has just asked the target when they would like to meet up and they are starting to discuss details of an offline meeting.]

Turn	Speaker	Message
1	g	where in the north bay doyou live ?
2	t	i 'm near santa rosa
3	g	far north bay
4	g	north or south of santa rosa
5	g	is your mom gone yet ?
6	t	nah
7	t	i guess she decided to leave in the morning
8	t	south
9	g	how many other guys are you chatting with tonight ?
10	t	none
11	t	i was trying to figure out what to wear tomorrow
12	g	so your going to school ?
13	g	I guess you have to if you 're mom 's still around
14	g	what are you going to wear if I come over ?
15	t	whatever u want me to
16	g	lol
17	t	i 'm only going to school if she 's still here when i get up in the morning
18	g	well since I do n't know what your wardrobe is , you orig: 'll i'll have to pick something out before I get there

Turn 13 shows the obligation statement (*you have to*) modified by the hedge *I guess* and the explanation move *if you 're mom 's still around*. The interaction is mainly led by questions from the groomer and answers by the target. The groomer asks if the target is going to school in turn 12 and answers his own question in the next turn, saying *I guess you have to if you 're mom 's still around*. The support moves make the

direct obligation statement request indirect, and the groomer attempts to sympathise with the target.

Vague language as a minimising support move is only used in G3 obligation statement requests with a focus on developing deceptive trust. Extract 6.10 features an example of vague language (turn 20) to mitigate the introduction of sexual content and mitigating the face threat to the target's negative face.

Extract 6.10

[Context: The interaction happens around 260 conversational turns into the overall contact between groomer and target. They are discussing the target's sexual history and relationships.]

Turn	Speaker	Message
1	t	kinda almost was gon na w my 1 bf
2	g	what made y'all nt go all the way ?
3	t	i don't know lol
4	g	did he have a condom ?
5	t	said so
6	g	did u make the choice to stop or did he ?
7	t	came close a few times
8	g	orig: ahh haha
9	t	just kinda didn't lol
10	g	do u wish yal had had sex ?
11	t	sometimes
12	t	might
13	g	huh ?
14	t	lol i guess i think i might next time
15	g	what the next time u are in a relationship ?
16	t	depends
17	g	what does it depend on ?
18	t	lol if its a nice guy
19	g	oh
20	g	would u have to be dating the guy to mess around with him and or possible have sex with him ?
21	t	idk i don't know lol kinda did some stuff w cpl guys that weren't really my bf
22	g	what all did u do with them ?
23	t	just tuchin n making out

The entire interaction is a discussion about the target's past relationships, sexual experience levels and thus contributes towards both developing deceptive trust and seeking sexual gratification. The vague language support move appears in turn 20. Here, the groomer makes a distinction between *messing around* and *having sex*, which

he uses in the same turn. He is explicit when saying *having sex* while also maintaining a certain vagueness and not specifying what *messing around* entails. Early in the interaction, the groomer uses vague language to not be too explicit. Examples of this are *go all the way* in turn two and avoiding the word *sex* in turn six (*did u make the choice to stop or did he ?*). He then gets more explicit in turn 10 when he outright asks the target if they wish they had had sex with their previous boyfriend. After that, the groomer asks follow-up questions, which do not involve explicit language before being both vague and explicit in turn 20 while making a request for information.

Another minimising support move function in G3 is the discourse marker (*well*) in turn 14 in extract 6.11. The obligation statement request head act is *you have to listen to MOI*, which is minimised by *well* and *I'll be the judge of that*.

Extract 6.11

[The extract appears approximately 280 turns into the contact between groomer and target. They are discussing weight after the target said their mother was home making dinner and they do not want to eat. The target compares themselves to a cow, which prompts the groomer to ask whether they chew with their mouth open.]

Turn	Speaker	Message
1	t	lol no i just look like a cow
2	t	im orig: soooooo so fat
3	g	i doubt that
4	t	i am
5	g	well , its my opinion that counts
6	t	i need to lose like 15 pounds
7	g	OH MY GOD !
8	g	if u got ta lose only that much , your not fat
9	t	i am im like 100 pounds now
10	g	your a orig: tard retard
11	g	thats not fat
12	t	ya it is
13	g	thats good
14	g	well , i i'll be the judge , your not fat and you have to listen to MOI
15	t	lol
16	g	im the boss of u and your the boss of me ! : p
17	t	cool
18	t	go bring me something to drink lol
19	g	lol
20	g	u got it

The minimising in question *well* (turn 14) minimises the direct obligation statement request *you have to listen to MOI*. This command is also minimised by *i'll be the judge*. The groomer tells the target they are not fat, and they have to listen to him (turn 14), claiming to be the boss of them and the target is the boss of him (turn 16). This is further supported by his claim that his opinion is the one that counts in turn five. Earlier on in the interaction (turn 10), the groomer calls the target a retard, attacking their positive face.

The second most used support move function in G3 is the enhancing interpersonal relations support move, which is done by the speech act of promising and using terms of endearment. This also appears in G4 but to a lesser extent (only one promise appears in obligation statement requests in G4). One example from G3 is the following exchange in extract 6.12 (turn eight contains the obligation statement request).

Extract 6.12

[Context: This interaction occurs relatively early on (around 170 turns) in the overall contact between groomer and target. They have just exchanged names after already discussing the sexual history of the target.]

Turn	Speaker	Message
1	t	so how old ru
2	g	24
3	g	you ?
4	t	13
5	t	remember
6	g	wow very young
7	t	so u not ocming
8	g	but you have to keep the secret cutie
9	t	of course
10	t	i ain't stupid
11	t	lol
12	g	bicaue is jail time for me
13	t	its our little secret

Extract 12 shows the use of a term of endearment *cutie* (turn eight) to support the obligation statement request *you have to keep the secret*. The groomer knows how young the target is (established in turn one) and even comments on this in turn six. This prompts him to command the target to keep the conversation a secret (turn eight),

mitigated by the term of endearment, which addresses the target's positive face and attempts to establish trust, as terms of endearment are usually reserved for people who know the target well. The stakes are laid out in turn 12 (because is jail time for me), and the target echoes the groomers request for discretion in turn 13 (*its our little secret*). This interaction is simultaneously aligned with the grooming goals of assessing risks and isolating and developing deceptive trust.

The enhancing interpersonal relations support move in G4 shown in extract 6.13, which takes the form of a promise (turn five), is a little different in that the groomer prompts the target to ask them anything they like. The interaction starts with the groomer proclaiming that he can teach the target anything in turn one, implying teaching them about sex. The target questions this (*anything ?*) but the groomer reassures them that they can by saying *any thing u want* (turn four) and adds the promise *all u have to do is ask* (turn five).

Extract 6.13

[Context: The extract starts roughly 100 conversational turns into the overall contact. The groomer and target have just discussed (sexual) relationships. Teaching is used as an implicit mention of sexual activity. The target asks if the groomer has taught a lot and whether they should be jealous.]

Turn	Speaker	Message
1	g	i can teach u thing about life any question u have
2	g	if that make any senses
3	t	anything ?
4	g	any thing u want
5	g	all u have to do is ask
6	t	like what
7	g	that is up to u
8	g	any subject from a to z
9	t	u know everything in the world ? §_HAPPY-SMILEY_§
10	g	no
11	g	just the basics
12	g	may be a little more
13	g	street smart yes book smart so so
14	t	u seem smart
15	g	thanks

The imposition of requesting the target to ask the groomer anything about sex is mitigated by *all* (turn five). The groomer also supposedly puts the target in control of

the conversation by saying *any questions u have* (turn one), *any thing u want* (turn four) and *that is up to u* (turn seven) using positive politeness to attend to the target's positive face needs. It also shows that the groomer thinks very highly of himself and thinks he can teach the target, establishing a student-teacher relationship and shifting the power dynamics. The target questions his knowledge in turn nine but adds an emoticon and pays him a compliment in turn 14, which the groomer accepts (*thanks*).

A similar exchange to 6.13 appears in extract 6.14, taken from G3. Here, the obligation statement request is in turn seven (*all you have to do is trust in me*) and it is supported by *and the rest will come natural*. It is also supported by the promise in turn five (*I'd show you everything*).

Extract 6.14

[Context: The interaction starts around 300 conversational turns into the contact between groomer and target. They are discussing sexual activities, implicitly)]

Turn	Speaker	Message
1	g	well you say you trust mewell trust me when I say it would feel so good
2	t	i do trust u
3	g	we 'd be lovers
4	t	wow i never been that before lol
5	g	well that 's okI 'd show you everything
6	t	thats sweet of u
7	g	all you have to do is trust in meand the rest will come natural

The groomer and target are talking about having sex, which the groomer does by using vague expressions such as *it would feel so good* (turn one), *lovers* (turn three), *everything* (turn five) and the *rest* (turn seven). He emphasises trust and the romantic side of things rather than being explicit and enhances this bond between them by promising to show the target (turn five).

Giving grounds is the most used support move in G4 in obligation statement requests with a developing deceptive trust focus. It also occurs in G3 but only once, which is

shown in extract 6.15. The explanation *u have to give me direction* (turn eight) is the giving grounds for the requests support move.

Extract 6.15

[Context: The interaction starts approximately 150 conversational turns into the overall contact. They have just discussed a phone call and are now discussing a planned offline meeting]

Turn	Speaker	Message
1	g	I ca n't go tomorrow
2	g	I have to be here
3	t	y >
4	g	Work
5	t	o i see
6	t	til when ?
7	g	Ill just come tonight
8	g	But u have to give me direction
9	t	r we just gon na smoke out ?
10	g	Its on u
11	g	Tell me so I can be prepare
12	t	would we do more ?
13	g	Like what do u wan na do ?

The groomer and target are discussing an offline meeting, but the groomer cannot make it the next day (turns one and two) because he has to be somewhere for work (turn four). In turn seven, the groomer instead suggests visiting the target *tonight* and adds that they have to give him directions. He is thus asking the target for personal information so they can meet. In a way, the groomer is offering a visit in exchange for receiving the target's exact location. The rest of the interaction concerns planned activities (turns nine to 13). The groomer repeatedly says the target has to make the decision by saying *its on u* (turn 10) and *what do u wan na do* (turn 13). This focus on what the target wants to do supposedly puts them in control and is a negotiating power dynamics strategy as defined in the original OGDM. It could also be a mitigation strategy for asking the target for their address to make up for this intrusion and re-establish the (deceptive) trust.

One of the giving grounds support moves, which takes the form of a threat in G4 can be seen in the extract 6.16 below (turn eight).

Extract 6.16

[Context: The extract appears approximately 1,130 conversational turns into the overall contact between groomer and target. The groomer wants to buy the target some underwear to wear when they meet offline.]

Turn	Speaker	Message
1	g	i will buy you something pretty and sexy though
2	g	i like that about you though , it 's sexy on you , works for you
3	t	aww ty
4	g	i like black or red , you have a favorite ?
5	g	you like lace at all ?
6	t	either is cool n yea lace is cool
7	g	i'll look for that then
8	g	you have to promise to wear it though or i'll feel really bad
9	t	ok
10	g	nice

The groomer promises to buy the target underwear to wear when they meet each other (turn one). They discuss the details (turns four to seven) before the groomer makes the threat in turn eight: *you have to promise to wear it though or i'll feel really bad*. He tries to get the target to promise they will wear the underwear and makes them feel guilty if they refuse by saying he will feel *really* bad (intensifying the outcome).

Another giving grounds support function, which includes a condition in G4 in the obligation statement request with a focus on developing deceptive trust can be seen in example 6.17. The condition is that if the target wants the groomer to trust them, they have to trust the groomer (turn nine). As can be seen, the groomer dominates this interaction with only one turn out of 11 being made by the target, which hinders the target linguistically as they are unable to answer and is a negative politeness strategy:

Extract 6.17

[Context: The conversation below happens roughly 820 turns into the overall contact. The groomer has just talked about sexual activities they would like to do with the target at an offline meeting and the conversation turns to getting in contact via phone.]

Turn	Speaker	Message
1	g	and how can I call you when your mom is gone ?
2	g	so I can call when I get there
3	g	you have a home ph
4	g	I mean , I need some way of calling you once I hit Jackson
5	g	are you worried or scared to give that info to me

6	g	scared I your mom might find out I have it
7	g	if so do n't be
8	g	I 'd never call or come over there unless ya told me to call or cum ok
9	g	if you want me to trust you you have to trust me right
10	g	did you pass out ?
11	t	no im thinking

In the first turn, the groomer asks how they can call the target once their mother is gone. He wants to let the target know when he reaches their location (turn two). In turn three, the groomer asks *you have a home ph*, repeating the earlier request for a phone number. In the following turn, the groomer provides an explanation for the repeated request (*I mean, I need some way of calling you once I hit Jackson*). The target does not respond to any of these repeated requests, prompting the groomer to ask if they are worried about disclosing this information. In turn six, he assumes the target is scared their mother will find out (*scared I your mom might find out I have it*). As the target still does not respond, the groomer tries to reassure them it is okay for them to share their phone number (turns seven and eight). He promises to never call the target or visit them unless the target consents to a call or visit, which is another positive politeness strategy. The groomer also uses *cum* instead of *come* in the same turn. It is unclear whether this is meant as a sexual innuendo or is simply a typo. The groomer asks the target to trust him but adds a condition in turn nine, after the target still has not responded: *if you want me to trust you you have to trust me right*. He wants the target to trust him enough to disclose personal information. However, he flips that statement on its head to imply he cannot trust the target if they do not give him their phone number. He also adds a tag question *right* at the end to ask the target for confirmation that he does not receive as the target still does not respond. This is another negative impoliteness strategy, as the groomer puts the target's indebtedness to him on record.

In his last turn, the groomer asks the target whether they passed out as they have not replied. Finally, the target replies saying they are thinking. It is interesting that the groomer does not let up after the target repeatedly ignores requests and withholds a reply and their phone number. The groomer gets more and more insistent and finds possible reasons why the target might not want to share their phone number. In the end, he implies he cannot trust the target without this information, showing trust as a

kind of currency. Essentially, he says that the target would not have a problem with giving their phone number to the groomer if they really trusted him.

Intensifying is the second most frequent support move in G4, which is done using boosting adverbials and tag questions in G4. No such intensifying appears in G3. This suggests G4 groomers intensify the already direct obligation statement requests, while G3 groomers make them less direct and minimise them as seen in extracts 6.8, 6.10 and 6.11 earlier. Extracts 6.18 and 6.19 are examples of an intensifying support move in G4. Turn nine in 6.18 contains the obligation statement request.

Extract 6.18

[Context: The extract below starts roughly 3,300 turns into the overall contact between groomer and target. They are discussing what the groomer wants to buy the target and have already talked about movies and video games, which the groomer refers to as *them* in turn one.]

Turn	Speaker	Message
1	g	yes i know you like them but i would want to get you something else
2	t	like what ?
3	g	clothes or jewelry
4	t	yeah I like both lol
5	g	lol
6	g	well if its clothes you have to go so you can try them on because I do n't know your size
7	g	what size do you wear babe
8	t	oh i need shoes !
9	g	ok then for sure you have to go for shoes
10	g	because you know what kind you like

The groomer wants to buy something for the target, which is not a movie or video game (turn one) but rather clothes or jewellery (turn three). He says the target has to come along because the groomer does not know the target's size (turn six). The target ignores this question likely due to the asynchronous nature of chat log communication but says they need shoes (turn eight). The obligation statement request *you have to go for shoes* in turn nine is intensified by *for sure*. He also explains this further in turn 10 (*because you know what kind you like*). The groomer is actively trying to convince the target that they need to join him on the shopping trip.

Extract 6.19 shows the groomer emphasising romantic feelings in an interaction that is focused on sexual activities. The obligation statement request appears in turn six (*you have to always know that i love you*) and is intensified by the boosting adverbial *always*.

Extract 6.19

[Context: The interaction starts approximately 2,240 turns into the overall contact between groomer and target. They are discussing sexual activities such as masturbating and also a planned offline meeting. The groomer says he has to teach the target a lot.]

Turn	Speaker	Message
1	g	im just trying to teach you about how to make loveand what guys like to do..and all that
2	t	ok cool
3	t	cause ur my bf ?
4	g	yes
5	t	ok cool
6	g	and you have to always know that i love you
7	t	orig: lovs loves u 2
8	t	: x
9	g	thanksnice §_HAPPY-SMILEY_§

The groomer expresses that he wants to teach the target about sex, but he is using vague language (*make love*) to mitigate the imposition and refers to sex as *and all that* (turn one). The target asks if he wants to do this because he is their boyfriend (turn three), which the groomer agrees with. He then says the target has to know he loves them (turn six), which he intensifies by using the adverbial *always*.

Overall, minimising to make the direct obligation statement request less direct seemed to be the preferred support move function in short duration groomer language (G3), while long duration groomers (G4) intensified and gave grounds for the obligation statement request. The minimising by G3 groomers was done to the point that an obligation statement request became a suggestion. Vague language was also used to minimise the imposition on the target's negative face and mitigate discussion about sexual activities. Long duration groomers (G4) used threats and conditions while giving grounds for making an obligation statement request. Trust became a currency

between groomer and target. They also used positive politeness to increase this trust. Negative impoliteness also occurred in G4; Groomers hindered their targets linguistically, threatened them, tried to make them feel guilty and implied the target could not be trusted if they were not willing to share personal information with the groomers. In the G3 example, the groomer offered to visit the target if they disclosed their exact address. Intensifying, which did not occur in obligation statement requests in short duration groomer language (G3), was used by long duration groomers (G4) to emphasise the romantic side rather than the sexual side of their relationship with the target. Enhancing personal relations was used in both G3 and G4 extracts to establish a student-teacher relationship connected to teaching the target about sex. It also co-occurred with vague language use in G3, where sexual content was mitigated. Groomers in both sub-corpora also used positive politeness to address the target's positive face.

6.2.2.2 Mood Derivable Requests

As mentioned above, seven mood derivable requests were identified during the qualitative analysis, four of which belong to G3 (short grooming duration) and two of which are taken from G4 (long grooming duration). Most of these were identified during the qualitative analysis of obligation statement requests. Four examples (three from G3 and one from G4) will be analysed qualitatively below but these cannot be taken as representative of all mood derivable requests in the groomer corpus. The identified mood derivable requests feature all support move functions in both sub-corpora.

One mood derivable request from G3 that is supported by three support move functions is shown in extract y.1. The mood derivable (*don't turn me into the cops*) in turn four is supported by a minimising *lol* in the same turn and a promise (*u will have fun*) in turn 10.

Extract y.1

[Context: The interaction starts roughly 210 turns into the overall contact between groomer and target. They are discussing a planned face-to-face meeting, including sexual activities.]

Turn	Speaker	Message
1	t	im so excited
2	t	this is going to be so much fun
3	g	hell yea it is
4	g	as long as u don't turn me into the cops lol
5	t	i wouldn't do that
6	t	§_KISS-SMILEY_§
7	g	ok good
8	t	§_HAPPY-SMILEY_§
9	t	i want to have fun
10	g	o u will have lots of fun
11	t	§_HAPPY-SMILEY_§
12	g	§_KISS-SMILEY_§
13	g	i want u bad

The promise is also supported by a minimising discourse marker (*o*) and an intensifying support move (*lots*). The minimising *lol* in the request (turn four) shows the groomer is joking, which serves as a positive politeness strategy and strengthens the trust groomer and target have established. The target reassures the groomer they will not turn him in and underlines this with a minimising emoticon in turn six. The groomer accepts that the target will not get him caught (*ok good*), re-establishing the previously established (deceptive) trust. The target then says *i want to have fun*, referring to having sex with the groomer during the face-to-face meeting. The groomer reiterates that the target *will have lots of fun*, boosted by *lots* and introduced by the discourse marker *o*. This is done to mitigate the negative face threat to the target and promising fun in exchange for keeping the meeting a secret. The groomer also uses emoticons to mitigate sexual content and to address the target's positive face needs.

Another example of a mood derivable request supported by a discourse marker and an explanation to give grounds for the request also from G3 can be seen in extract y.2 (turn five).

Extract y.2

[Context: The interaction happens 100 turns into the conversation and the groomer and target are discussing wanting to party, which turns into plans for an offline meeting.]

Turn	Speaker	Message
1	g	ill pick u up if u want

2	t	hmm
3	t	how do i know ur not a serial killer ? lol
4	t	oo look i stumped u on that one
5	g	okay how about this here my phone number call me to see i i am liget edit ³⁷
6	t	lol okay one second lem me go find the phone
7	g	cool
8	g	if u don't call i know u r a wimp

The groomer offers to pick the target up in the first turn. The target does not seem convinced (*hmm*) and asks the groomer how they know he is not a serial killer in turn three, adding a *lol* at the end to show they are not serious. The target then revels in the fact the groomer does not reply immediately (*oo look i stumped u on that one*). In turn four the groomer gives the target his phone number, instructs them to call him using a mood derivable request *call me*. This request is introduced by the discourse marker *okay* and the explanation (*to see i i am liget*). The target agrees and says they have to find a phone. The last turn shows the groomer challenging the target. While this exchange and the sharing of personal information, in this case the groomer's phone number, serves to establish trust, it also has an element of negotiating power dynamics: the groomer and target are playfully competing and challenging each other. This is particularly obvious in turns four and eight (*if u don't call i know u r a wimp*) where the groomer calls the target names, attacking their positive face.

Another mood derivable request (*name it*) appears in turn 13 of extract y.3 also taken from G3 and is supported by a promise (*you will have it*).

Extract y.3

[Context: The interaction starts approximately 140 turns into the overall contact. The groomer and target are discussing details of a face-to-face meeting, including activities such as having sex, taking drugs, and drinking alcohol.]

Turn	Speaker	Message
1	g	you don't get sore after a short period of sex do you ?
2	t	nah u gon na have pot rite
3	g	most likely

³⁷ This (*edit*) means the phone number was removed from the original chat log by a volunteer from the PJ Foundation

4	g	you asked if I am going to hurt you
5	t	pot relaxs me tho
6	g	no , I am not just going to bend you over and ram it in
7	g	I will be fingering you while I am licking and nibbling on your clit
8	t	what kinda booze u bringing
9	g	what do you like/
10	g	?
11	t	all kinds
12	t	whatever gets me drunk qwick
13	g	what do you want name it and you will have it
14	t	malibu rum
15	g	and coke ?
16	t	hell yea
17	t	vanilla coke tho
18	g	no problem

The groomer and target discuss having sex and smoking marijuana to relax (turns one to seven). They then move on to alcohol and the target asks what type the groomer is bringing in turn eight. The groomer asks what type the target prefers (turn nine), to which the target replies *all kinds* and *whatever gets me drunk qwick* (turn 12). The groomer then states *name it* again (turn 13) and offers a promise in support of the request in turn 13 (*and you will have it*). This is the promise turn, which is also part of a condition: the target has to name it and the groomer will supply it. This supporting promise is a positive politeness strategy of promising a ‘gift’, which addresses the target’s positive face needs of being liked. Additionally, the groomer does not seem to have any qualms with supplying alcohol and drugs to a person he believes to be underage. In fact, he seems to go out of his way to make this happen and does not react to the target’s statement *whatever gets me drunk qwick*. He instead ignores these questionable statements and focuses on gathering information from the target, which suggests he is fixated on making this offline meeting happen and having sex with the target, which would already be classed as statutory rape and with being under the influence of alcohol and drugs would be classed as rape even if the target were not underage. Despite the interaction relating to potentially criminal behaviour, the request stays direct, and the face threat is mitigated by positive politeness (the implicit promise of a gift).

An example of a mood derivable request taken from G4 can be seen in turn five in extract y.4. The mood derivable request takes the form of the imperative *tell me*, which

is minimised by *just*. The groomer and target talk about what are presumably sexual activities, as the groomer praises the target on their behaviour when they discussed *some stuff* in turn one using vague language and positive politeness. The target asks for confirmation (*yea?*) in turn two, which the groomer provides (*yes u did*), reiterating the praise.

Extract y.4

[Context: The extract starts roughly 350 lines into the overall contact. The groomer refers to a previous conversation in which the groomer and target have presumably discussed sexual activities.]

Turn	Speaker	Message
1	g	u did good the other night in asking about some stuff ok
2	t	yea ?
3	g	yes u did
4	g	i know it was probably hard for u
5	g	if it make u uncomfortable just tell me
6	g	i like to here what u have to say
7	t	u don't make me uncomfortable
8	t	we talked along time i know u good
9	g	k
10	g	tell me what u know about me then

The groomer assumes what the target felt (uncomfortable), saying it was *probably hard* (turn four) to talk about sex. This is a positive politeness strategy, which contributes to the online grooming intention of deceptive trust, since the groomer expresses this presupposition in knowledge terms, rather than assumptions conveying they know the target well. The mood derivable request in turn five (*tell me*) is minimised by *just* and is additionally supported by two explanations *if it make u uncomfortable* and *i like to here what u have to say*. The latter explanation emphasises the groomer's interest in the target's opinion and seems to be a way of giving the target the floor. The groomer seemingly puts the target in control and says they can stop the conversation at any time if they feel uncomfortable. In a way, the groomer says the target has control and the conversation does not have to be about sexual content while also encouraging the target to speak about sexual activities. This seems rather manipulative, as the target will likely want to impress the groomer and will continue discussing things with which they may not be comfortable.

It is evident that the direct mood derivable requests that are made using the imperatives (*don't turn me in, call me, name it and tell me*) need a significant amount of support functions to mitigate the face threat. All examples show more than one support move function to make the request less direct. These support moves sometimes occur in the same turn or subsequent turns. The groomers also use politeness strategies to address the target's face needs. One groomer attacks his target's negative face by calling them names and challenging them. However, the support moves around this suggest he is joking, which can be a way of bonding. As mentioned above, no conclusions can be made about mood derivable requests in G3 or G4 in general as these are only a few examples, and they are not the focus of the analysis.

6.2.2.3 Want Statements – Assessing Risks and Isolating Focus

Finally, the last head act type that was analysed qualitatively was want statements with a focus on assessing risks and isolating. This head act type is made up of *want you to* and *want u to* constructions. The support move functions in want statement concordance lines, with a focus on assessing risks and isolating in G3 (0-299 minutes) and G4 (300-10,600 minutes), can be seen in Tables 43 and 44.

Table 43: Support move function in want statement concordance lines in G3

support move	#	%
giving grounds	5	50.0
enhancing interpersonal relations	2	20.0
intensifying	2	20.0
minimising	1	10.0
total	10	100.0

Table 44: Support move function in want statement concordance lines in G4

support move	#	66.7
intensifying	2	33.3
giving grounds	1	0.0
minimising	0	0.0
enhancing interpersonal relations	0	0.0
total	3	66.7

Overall, support moves in want statement requests with a focus on assessing risks and isolating are more varied in short duration groomer language (G3) than in long duration groomer language (G4), which only has two support moves (intensifying and giving grounds for request). Giving grounds is the most frequently used support move function in want statement requests in G3 with a focus on assessing risks and isolating.

Enhancing interpersonal relations and intensifying are the second most used support moves in G3 in want statement assessing risks and isolating requests. Minimising only appears once in G3 in want statement assessing risks and isolating requests. Giving grounds as a support move function is used to give reasons for making the request (explanation) and giving grounds for a condition in G3.

The one instance in G4 is also giving reasons for making the request. Extract 6.20 below shows an explanation move in G3 (turn seven).

Extract 6.20

[Context: The interaction below starts around 430 conversational turns into the overall contact between groomer and target. They are discussing the details of a face-to-face meeting including how they will get in touch and where they will meet.]

Turn	Speaker	Message
1	g	oh ok is it too quite near ur house ?
2	g	like at this time
3	t	what do you mean
4	g	i mean at this is it too quiet around ur house , not much traffic or something ?
5	t	oh we are in a like group of houses but there 's a lot of them.. we get cars through here though
6	t	is that what you mean
7	g	yes and i was asking that will people watch u come out of the house , i don't want u to get in trouble
8	t	ohhhh no.. the neighbors do n't really talk to us or anything ,

The groomer and target are discussing a face-to-face meeting and the groomer tries to find out more details about the target's surroundings. He asks a question in the first turn (*oh ok is it too quite near ur house ?*) and qualifies this by adding *like at this time*. The target seeks clarification in turn three (*what do you mean*). The groomer attempts to explain (*i mean at this is it too quiet around ur house , not much traffic or something ?*) in turn four. The target answers saying they live in a group of houses but do have cars go through (turn five) and asks *is that what you mean* in the next turn. The groomer finally provides an explanation for why he is making this request for information in turn seven by saying *yes and i was asking that will people watch u come out of the house i don't want u to get in trouble*. This is simultaneously oriented towards the grooming intention of assessing risks and isolating and arranging further

contact, as the groomer is specifically seeking reassurance no one will be there, and to arranging an offline meeting.

In G4, the groomer also does not want the target to get into trouble (turn 13, 6.21). However, the situation is different. Here, the groomer would like to talk to the target on the phone to arrange a meeting (*we mite have to talk on the phone to make plans*). The target responds that they do not have access to a phone right now.

Extract 6.21

[Context: This interaction starts approximately 2,500 conversational turns into the overall contact between groomer and target. The groomer has just asked about the target's mother's schedule and is proposing that they meet.]

Turn	Speaker	Message
1	g	it would be eayier in the daybut anytime is goodwe mite have to talk on the phone to make plans
2	t	ok cool
3	t	i don't got the phone right now
4	g	that s okwho has it ?
5	t	my mom
6	t	can i call u when she gets home ?
7	g	yeah..but i mite not be home
8	t	ok
9	g	it is it cell phone ?
10	t	i got a calling card
11	g	cool
12	t	i think i got some minutes on it
13	g	i don't want you to get in trouble ... will she find out ?
14	t	can it b deleted off the cell ? right ?
15	g	i think sobut it will show up on the bill
16	t	k
17	g	i want t you to call ... but we have to work this out

This prompts the groomer to ask more details in turns four (*who has it ?*) and nine (*it is it cell phone ?*), contributing to the online grooming intention of assessing risks and isolating. When the target reveals they have a calling card and would be borrowing their mother's phone, the groomer becomes hesitant and says he does not want the

target to get into trouble, asking if their mother will find out. The groomer's worry increases as the interaction continues, and when the target asks if the call can be deleted off the cell phone, in turn 14, the groomer says *i think sobut it will show up on the bill*. In his last turn, the groomer says he wants the target to call him but says *we have to work this out*, including both the target and groomer in the activity, which is a positive politeness strategy. Earlier on in the interaction, the groomer says he might not be home if the target were to call him later (turn seven). This could be an attempt to convince the target to call right away, rather than delaying and putting a bit of pressure on the target.

A different example of giving grounds for a condition is extract 6.22 taken from G3. The request occurs in line one (*i want to see you tonight*), when the groomer conveys the wish to see the target that night. He adds that he will not be around the next night (turn two). In turn three, the target expresses their dislike of this.

Extract 6.22

[Context: The interaction appears roughly 450 conversational turns into the overall contact. The groomer and target discuss setting up a face-to-face meeting that night.]

Turn	Speaker	Message
1	g	i want to see you tonight
2	g	i doubt ill make it back out here tomorrow night
3	t	o that sucks
4	g	no chance at all tonight ?
5	t	ya but ur gon na have 2 wait until my sister goes to sleep
6	g	how long you think that will be
7	t	shes in the bathroom right now
8	t	hello
9	g	im here
10	g	just knock her out
11	t	lol
12	t	sure
13	g	no don't be that mean
14	t	u told me 2 lol
15	g	i don't want you to hurt her but i want to see you tonight
16	t	i do 2
17	g	our mom and dad are in bed now tho right
18	t	they just got in there they aren't asleep yet
19	g	hopefully they are tired
20	t	ya

Turn four consists of the groomer asking to see the target again that night but the target replies with *ya but ur gon na have 2 wait until my sister goes to sleep*. The groomer's impatience can be seen when he asks how long that will be, in turn six. Unprovoked, he suggests knocking the sister out, in turn 10, again displaying his impatience. He does not add any minimisers to this request for violence and the target awkwardly answers *lol* and a sarcastic *sure* (turns 11-12). The groomer then says *no don't be that mean* to which the target replies *u told me 2 lol* in turn 14, again adding a minimising *lol* to diffuse the situation. Turn 15 is the condition turn: the groomer says he does not want the target to hurt their sister but that he does want to see them, repeating the earlier head act type. The implication is that the only way to see each other is for the target to hurt their sister, which is quite disturbing. The target says they also want to see the groomer, ignoring the suggestion of violence. The rest of the interaction concerns whether the target's parents are asleep yet, which also contributes towards assessing risks and isolating. The sudden introduction of violence as a means to see the groomer and a very extreme form of making sure the target and groomer are alone (physical isolation), is very unexpected and jarring, especially as there is nothing to indicate the groomer is not serious about this proposal.

Intensifying is the other support move found in both G3 and G4 in want statement requests with a focus on assessing risks and isolating. This appears in the form of tag questions in G3 to ask for confirmation from the target and by using boosting adverbials in G4. Extract 6.23 is taken from G3 and shows a tag question in turn eight, which is also the request turn (*i want you to check something for me first*).

Extract 6.23

[Context: The following is an example in which the groomer and target discuss a phone call. But before the target can call the groomer, he wants to make sure the chat log is not archived. It appears roughly 560 conversational turns into the overall contact.]

Turn	Speaker	Message
1	g	will you have to get offline to call ?
2	t	no
3	t	but ill hacve to go to the other room
4	t	because the cordless doesn't reach in here
5	t	is taht ok ?
6	g	yes

7	g	but
8	g	i want you to check something for me first - ok ?
9	t	okay
10	g	above this box is the word File - see it ?
11	t	uhhuh
12	g	click on it and go to Privacy Options ...
13	t	okay after i do that i cant type to you
14	g	ok - go down til you see Archive and select it
15	t	okay tell me what do to n i wont tpe back for a minute
16	g	is there a check mark in the Enable Archiving box ?
17	t	no
18	g	whew !
		!
		!

The interaction begins with a discussion about the target calling the groomer (turns one to six). The target has to go to a different room as their cordless phone does not reach their current position. The target asks if that is okay in turn five, to which the groomer replies *yes* and immediately follows it up with *but* (turn seven). He then asks the target to *check something for me first* and adds a tag question, asking the target for confirmation before he has outlined the request. Continuing without waiting for the target's response, the groomer describes how the target can find the archive chat settings (*above this box is the word File - see it ?*). He seems to be ignoring the target as he describes the next steps (*click on it and go to Privacy Options ...*), (*is there a check mark in the Enable Archiving box ?*). Ignoring them is a positive impoliteness strategy. After the target says *no* in turn 17, the groomer reacts relieved (*whew*). The fact he uses a tag question at the end of turn eight but then continues without waiting for the target's response is interesting, as it is certainly impolite. He does wait for confirmation after adding another tag question in turn 10. Yet, he is still ignoring the target who has to explain twice that they will not be able to reply if they carry out the steps outlined by the groomer (turns 13 and 15). The whole interaction is aligned with the assessing risks and isolating intention, as the groomer tries to establish whether there is a record of their conversations.

Extract 6.24 shows one example of intensifying in the G4 want statement assessing risks and isolating requests. The request itself appears in turn five (*i want u to feel like u can tell me anything*) and is intensified by the boosting adverbials *ever*, *anything*

and the tag question at the end of the turn. The use of the first name of the target conveys a certain closeness between groomer and target.

Extract 6.24

[Context: The interaction occurs approximately 270 conversational turns into the overall contact between groomer and target. They are discussing their relationship.]

Turn	Speaker	Message
1	t	you mean a lot to me to
2	g	ok do u feel like u can tell me anything ?
3	t	yea i guess
4	t	like what ?
5	g	it don't matter i want u to feel like u can tell me anything if there is ever anything wrong or ur upset about or just anything cory u can tell me ok
6	t	ok
7	g	ok good ! !

The interaction centres on trust and the groomer and decoy discuss their relationship (turn one). The groomer asks if the target feels like they can tell him anything (turn two) to which the target replies noncommittally *yea i guess* and then asks for examples. The groomer says *it don't matter i want u to feel like u can tell me anything if there is ever anything wrong or ur upset about or just anything cory u can tell me ok*. The boosters *anything* and *ever* are used to emphasise that the groomer is always there for the target, intensifying his hold on the target, which is oriented towards the assessing risks and isolating grooming intention. He tries to be the target's support network by becoming close to them, which is also closely linked to developing deceptive trust. The target agrees (turn six), which the groomer echoes enthusiastically *ok good !!*, using exclamation marks to underline his eagerness.

The two supporting moves the two sub-corpora have in common are intensifying and giving grounds. With the latter, the theme of not wanting the target to get into trouble is also shared. In both cases, the groomer explains their caution to the target. In the long grooming duration sub-corpus (G4) example, the groomer uses pressure to get the target to do what he wants. He also uses positive politeness to counter-act this impoliteness. The supporting moves in G3 also show how short duration groomers use positive impoliteness (ignoring the target), specific and open-ended threats and inciting violence to get their way. Groomers seem to be very goal focused and will do

anything to achieve their objective (e.g., checking if a chat archive exists or arranging an offline meeting). These threats and use of positive impoliteness seem very extreme. They are also very different from the other head act types where impoliteness was not quite as severe. This use of positive impoliteness does support the evidence for different duration approaches to grooming the fifth chapter (section 5.5) suggested.

In contrast, intensifying in G4 shows how long duration groomers attempt to persuade their target to perform an action by intensifying their interest in them, using positive politeness to remind them they are loved and to trust the groomer in want statement requests. This seems like a more subtle approach to manipulate the target, increasing their dependence on them and intertwining developing deceptive trust and assessing risks and isolating.

6.3 Conclusions on Manipulative Requesting Behaviour and Duration

This chapter has examined request structures in two different duration sub-corpora. It has done so by first outlining what constitutes a request, outlining request head act types and support move functions and exploring these in the two sub-corpora. It then analysed the support move functions in three request structures in G3 (0-299 minutes) and G4 (300-10,600 minutes) qualitatively. Requests as a speech act can be used to manipulate a target, which this chapter has shown. Groomers use request structures to advance their communicative intentions and ultimately manipulate their targets into complying. This ranges from manipulating targets into giving up personal information to engaging in discussions about sexual content to agreeing to meet face-to-face. The chapter also found that duration makes a difference concerning how groomers manipulate and negotiate power dynamics through requesting and the type of support move functions that groomers use strategically and differently in both duration sub-corpora.

The analysis of suggestory formulae showed that G3 and G4 groomers use the same support moves and prefer minimising over the other three support moves. Their use of these support moves differs. Minimising is used by short duration groomers (G3) to minimise the imposition of the potential face threat while long duration groomers (G4) add another component, vague language to create a coded language between groomer

and target and strengthen the bond while minimising the imposition. Minimisers are also used to echo and assimilate the target's language. Overall, the type of minimising is more diverse in G4, using more minimising types and more general in G3, with a focus on a few specific minimising types. Intensifying, more specifically request reinforcement in suggestory formulae requests, is used by short duration (G3) groomers to harass the target and gain information over several conversation turns while long duration (G4) groomers use it to clarify in the same turn. The enhancing personal relations support move is used by both G3 and G4 groomers to evoke trust by using terms normally reserved for family, friends, and significant others. The difference is that short duration groomers (G3) use it to get information they wanted from the target, which is a more specific function than long duration groomers (G4) who use it to generally build deceptive trust and emphasise familiarity.

Obligation statement requests with a focus on developing deceptive trust are supported differently in G3 and G4. Short duration groomers (G3) prefer minimising support moves, which suggests they minimise the otherwise direct obligation statement requests considerably turning the command into a suggestion. Long duration groomers (G4) use minimising less frequently than other support moves. Their most used support move is giving grounds followed by intensifying suggesting they intensify rather than minimise the already direct obligation statement requests. In G3 (short duration), vague language is used to minimise the imposition on the target's negative face and mitigate discussion about sexual activities. Intensifying is used by G4 (long duration) to achieve a similar effect and to emphasise the romantic side of the relationship, rather than focusing on sexual activities. In G4, groomers threaten targets while justifying the obligation statement request they make. Trust becomes a currency between groomer and target and is supported by positive politeness strategies. G4 groomers also use negative impoliteness: hindering targets linguistically, threatening, invoking guilt, and implying targets cannot be trusted if they do not willingly share personal information. Both groomers in G3 and G4 use enhancing personal relations to establish a student-teacher relationship connected to teaching the target about sex. It also co-occurs with vague language use in G3, where sexual content is mitigated.

The seven mood derivable requests seem to take a significant amount of support moves and use all support move functions. They are all made using the imperative

forms, which are most direct explaining this need for mitigation and multiple support move functions for one request. Groomers use politeness strategies alongside support moves and one groomer also uses negative politeness in conjunction with minimising support moves, suggesting he is joking. No conclusions can be made about mood derivable requests in online grooming in general. However, these findings suggest they take several support move functions to mitigate their directness.

Want statement requests with a focus on assessing risks and isolating in G3 and G4 show that short duration groomers (G3) use all support move functions for these requests while long duration groomers (G4) prefer intensifying and giving grounds. Both groups share not wanting to get the target into trouble. One difference is that the G4 groomer uses slight pressure on the target to get them to do what he wants. Positive politeness is used to counteract this impoliteness.

The varied use of supporting moves in G3 shows how short duration groomers use impoliteness (ignoring the target), specific and open-ended threats and inciting violence to get their way. Groomers seem to be very goal focused and would do anything to achieve their objective. G4 groomers on the other hand use intensifying to persuade the target to do something by increasing their interest in them, reassuring them they are loved by using boosting adverbials and generally making them more dependent on the groomer. This seems like a more subtle approach to manipulate the target compared to G3 groomers.

This chapter has examined manipulative requesting behaviour of online groomers by first defining requests in the context of online grooming and defining their component parts, head act types and support move functions. The specific methodological steps of extracting formulaic request structures were outlined before the chapter examined three head act types with different orientation (developing deceptive trust, assessing risks and isolating) in more detail. The analysis found that requests can be used effectively and strategically to manipulate targets into compliance and duration also makes a difference to the support move functions that online groomers use. Overall, short duration groomers (0-299 minutes) are very determined and harassed, elicit information, and use vague language to mitigate sexual content. They also use threats to manipulate their targets. Long duration groomers (300-10,600 minutes) emphasise

the trust base with their targets, use coded language and threats to manipulate them. They also use trust as a currency and make use of both positive and negative politeness to counteract the impoliteness. The thesis will now move on to discussing these findings, and those of chapters four and five, their implications and significance before drawing conclusions.

7. Chapter 7: Conclusions on Online Groomers' Discourse

7.1 Introduction

This chapter will summarise the main findings from the previous empirical chapters (chapters four-six) and will explain why they are significant, what they mean in real world terms and what the implications for online grooming language research are. The first section is organised into three main parts related to the three empirical chapters. As such, chapter four will be discussed in sub-section 7.2, chapter five in sub-section 7.3 and chapter six will be discussed in sub-section 7.4. The sub-sections will first answer the respective research questions of the chapter and will summarise the main findings. They will then move on to discussing the implications of these findings in the context of the literature reviewed in chapter two of this thesis and the wider field of online grooming language research. Lastly, they will address applications of these findings to real-world contexts, such as law enforcement and educational settings. Section 7.5 will draw conclusions by summarising the main findings compared to the thesis aims and reflecting on the thesis as a whole. Limitations will be discussed in section 7.5.1 and future research proposed in section 7.5.2. Conclusions follow in section 7.6.

7.2 RQ1: Are the Features of Online Groomer Language Distinct?

The first analytical chapter, chapter four compared a corpus of online groomer language to a digital chat language reference corpus (PAN2012) to establish the features of online groomer language and to find out whether these features are distinct. This study sought to fill a gap in existing knowledge about online grooming *language*, in which ironically most work has been conducted within non-Linguistics academic fields. The main research question in Chapter four thus asked what the features of online grooming language are and whether they are distinct compared to a general digital language chat corpus, PAN2012. The selection of PAN2012 as the reference corpus in chapter four was justified because it has been used in Machine Learning studies and was created for a competition about detecting online grooming language. It consists of Internet Relay Chat (IRC) language taken from technical support chats, general chats, and consensual discussions about sex between adults. The chapter also

asked how online grooming intentions are realised linguistically by online groomers. As part of the analysis, keywords were derived, analysed, and sorted into the online grooming intentions from the OGDM, originally developed by Lorenzo-Dus et al. (2016).

7.2.1 Online Groomer Language Features are Distinct

The chapter found that online groomer language is distinct, when compared to PAN2012. The top keywords were highly frequent, most were eight to 32 times more frequent in the online grooming corpus (GN) than they were in PAN2012 with a small minority being up to 265 times more frequent. Online grooming intentions appeared in these top keywords, further supporting the finding that online groomer language is distinct. When examining all keywords under consideration (n=178), all grooming intentions from the OGDM could be linked to the keywords, which also validates the model because a different methodology and larger dataset identified the same intentions. The KWIC analysis also found some specific linguistic realisations of online grooming intentions and strategies, which related to studies into online grooming communication (see chapter two, section 2.5.3). These individual intentions and strategies are reviewed next by referring back to the studies reviewed in section 2.5.3 of the literature review. These studies and their methodology are outlined in Table 3 (section 3.1.1).

Developing deceptive trust emerged as the most prevalent online grooming intention in the keyword analysis. It had four strategies associated with it: relationship terms, small talk, bonding, and exact location. In the literature, the most prevalent strategies within deceptive trust development are eliciting personal information and strengthening the groomer-target relationship. In the original OGDM, a different focus emerged: sexual and non-sexual compliments, aligned with different “speed” of grooming, which were further investigated in Lorenzo-Dus and Izura (2017). Compliments likely did not emerge as salient in chapter four of this study because of the use of a corpus assisted methodology. Specifically, compliments often present as strict multi-word strings (Jucker et al., 2008), rather than one-word units, which was the focus of the analysis in chapter four. Instead, relationship terms (*girlfriend*) and terms of endearment (*sweetie*) emerged as salient to negotiate the groomer-target relationship. These terms were also often used in conjunction with other developing

deceptive trust strategies to maximise their effect and success of gaining the target's trust. Two keywords, *precious* and *adventure*, were counterparts to these explicit relationship terms, by avoiding explicit reference to relationships. They were implicit relationship terms and were also used by groomers to address the targets' perceived negative face threat.

Bonding as a relationship strategy was done by groomers using the keywords *smile*, *smiling*, *mad dreaming* and *promise*, which showed their linguistic construction of affection towards their targets, addressing and fulfilling the targets' positive face needs. Relationship terms and bonding seemed to be an explicit linguistic realisation of how groomers sought to build and strengthen a relationship with their targets. Trust development is a recurring theme in the online grooming literature with studies showing that groomers regularly use strategies such as 'using/eliciting statements/promises of trust' (Chiang & Grant, 2017: 129) and 'trying building mutual trust' (Pranoto et al., 2015). Other studies report that groomers often state "I love you" or "I like you", more frequently with female targets than with male targets, (van Gijn Grosvenor & Lamb, 2016: 586). This is also called 'romantic expressions' (p. 11) and 'statements of longing contact' (Kloess et al., 2017a: 11) and 'feeling of love and exclusiveness' (Gupta et al., 2012). The bonding strategy identified in chapter four is closely related to these strategies, as the groomer expressed and negotiated how he felt about the target. Promises were part of this strategy, and they were also connected to the online grooming intentions of seeking sexual gratification and negotiating power dynamics.

Small talk was realised linguistically by using keywords like *Laundry*, *nada*, *bills* and *chattin* to maintain the conversation. Other keywords (*snack*, *pee*, *shower*, *youngest*, *dressed*) that were primarily used for small talk purposes were also used by groomers to gain sexual gratification in a third to 36% of the analysed concordance lines showing the interconnectedness of developing deceptive trust and seeking sexual gratification. One more collocation unit (*thinking + about*), which the groomer used to fantasise about the target, also showed evidence of this close link. The last broad strategy within developing deceptive trust related to the target's exact location, which is an aspect of the exchange of personal information strategy within the original OGDM. The groomers used *detroit*, *beach*, *addy* and *house* to ask the target for their

exact location. This focus on exact location is also aligned with one of the most prevalent strategies within developing deceptive trust in the literature: Eliciting personal information is also referred to as ‘exchange of personal information’ (Lorenzo-Dus et al., 2016), ‘stating/eliciting asl (age, sex, location)’ (p. 129) and ‘requesting email address’ (Chiang & Grant, 2017: 129) ‘exchanging email address / picture / web-cam information’ (Gupta et al., 2012: 5) and ‘geographical location’ (Williams et al., 2013).

Seeking sexual gratification was the second most prominent online grooming intention in the keyword analysis. Four strategies were identified: sex acts, physical contact, clothes/appearance, and contraception. The strategies identified in chapter four do not appear as such in the reviewed literature, likely because they are more specific than the strategies previously identified due to using a CADS methodology.

Groomers used both implicit and explicit desensitisation keywords strategically to gain sexual gratification. *Tingly, excited, massage* and *gentle* are examples of the implicit words, while *foreplay, doggie, orgasm, oral, swallow* and *sexually* illustrate keywords that are sexually explicit. In using some of these sexual gratification keywords, the groomers used a push-pull rhetorical structure (Montgomery, 2006; Lorenzo-Dus & Kinzel, 2021): They used a sexually explicit phrase or term in close proximity with a sexually implicit phrase or term. Sometimes the implicit side of the push-pull structure was used to advance the developing deceptive trust grooming intention, which provides further evidence of the close link between the two intentions.

The keywords *pyjamas, thongs, bras, cloths,* and *panties* were used to refer to the target’s appearance, fantasise about them sexually and request photos of them in minimal clothing. This strategy of clothes/appearance has been identified in Chiang and Grant (2017), labelled as ‘inquiring about target's clothing’ (Chiang & Grant, 2017: 132). In their work, this strategy is part of their immediate sexual gratification move, showing that the groomer gains sexual gratification from these questions and discussions about the target’s clothing. The strategy is also closely related to fantasy enactment and asking for pictures of the target, which has been identified by previous studies, likely due to Briggs et al.’s (2011) distinction between fantasy-driven and contact-driven groomers and most studies using PJ data, which is by definition

contact-driven. Asking for and sharing sexual pictures is also called ‘exchanging pictures of sexual nature’ (Gupta et al., 2012: 5) and ‘providing/requesting sexual photographs’, (Chiang & Grant, 2017: 132).

In chapter four, groomers also used the strategy known as reframing to present sex as beneficial and assigned discourse roles: Teacher (groomers) – pupils (targets). *Snuggle, kiss, hug, cuddle, and tickle* were for instance used to normalise the groomers’ desire to touch the target inappropriately. One last strategy within sexual gratification identified in chapter four was related to contraception, with the following keywords: *pill, condoms, and wear*. These keywords were used to discuss future meetings involving sex and introduce contraception into the conversation, which is a specific realisation of the explicit desensitisation strategy.

Four keywords (*privates, blanket, tummy, embarrassed*) that were not aligned with the strategies within sexual gratification were used to discuss offline meetings involving sex with the target, describe sex acts and instruct targets to masturbate. They were also linguistic realisations of the explicit desensitisation strategy. Contraception is not generally referenced in the reviewed literature, except for (1) one mention in Chiang and Grant (2017) in which the groomer promising to bring a condom is labelled ‘suggesting ways to improve target’s sexual enjoyment/lessen target’s anxieties towards sex’ (Chiang & Grant, 2017: 129). (2) Condom is also part of the LIWC biology category within the sexual stage in Black et al. (2015). (3) Egan et al. (2011) identified condom in an exchange in which a particular groomer was disinterested in the target’s concerns. They found that being disinterested and unconcerned was linked to fixated discourse. Apart from these rare mentions of specifically condoms, the identification of the strategy of discussing contraception as part of the seeking sexual gratification grooming intention in chapter four is therefore important. It reveals how groomers plan meetings that include sexual encounters with the targets, educate targets about contraception and ask about their experience levels and previous sexual encounters.

Another important finding in my keyword analysis that is not aligned to the above four strategies but fits within sexual gratification was the use of infantilising language, such as *tummy* instead of stomach, to instruct the target to masturbate or describe sex acts.

This is very similar to one of the variables outlined in Pranoto et al. (2015), ‘using child related vocabulary (X9): Sometimes to mention the sexual organs, the predators use the words used by the children to name their sexual organs’ (Pranoto et al., 2015: 360). Even though *tummy* is not a sexual organ, it was used in highly sexualised contexts (describing sex acts and feelings of arousal and instructing the target to masturbate). This example also shows a power imbalance, as the groomer assigns discourse roles to himself (teacher) and the target (pupil). The usage of the word *tummy* is linked to the strategy of reframing. According to O’Connell (2003) ‘the usual rationale for this approach is that the adult is somehow perceived as a mentor who will guide the child to a greater understanding of his or her own sexuality’ (O’Connell, 2003: 10). Child-like language was also used to make the target feel understood, to assimilate their language and present sex as beneficial, in other words reframing the sexual content and topics, which was also mentioned in other studies. O’Connell (2003) calls this strategy ‘promise that by engaging in these activities the child will grow to become a wonderful lover.’ (O’Connell, 2003: 10). It is called reframing as ‘learning experiences, games, or skills’ in Lorenzo-Dus et al. (2016: 49) and ‘portraying sex as fun/pleasurable/beneficial’ in Chiang and Grant (2017: 129). In van Gijn Grosvenor and Lamb (2016), groomers specifically told male targets that they ‘could learn and they would be better prepared when they later engaged in sexual activities with other boys’ (van Gijn Grosvenor & Lamb, 2016: 587). Interestingly, Pranoto et al. (2015) had a specific variable called reframing, which was statistically significant (p value: 0.000) and part of their logistic model development, unlike the child language variable (p value: 0.986).

The analysis in chapter four also showed a particularly strong relationship between developing deceptive trust and seeking sexual gratification, which appeared frequently in most keywords. The mentioned push-pull structure is a good example of this connection, which is also present in the original OGDM (Lorenzo-Dus et al., 2016). It is also a finding by Lorenzo-Dus and Kinzel (2021) who observe that groomers use explicit terms in close proximity to vague language terms to ‘advance the process of Sexual Gratification through explicit sexual desensitisation. It also served to advance the Deceptive Trust Development process by framing sexual activity in non-sexual terms’ (Lorenzo-Dus & Kinzel, 2021: 204). They concluded that groomers use vague language to embed sexual intent ‘covertly, couching it in terms that make

interpretation of friendship and romance likely, which in turn helps to develop trust of their deceiving goals.’ (Lorenzo-Dus & Kinzel, 2021: 206).

Arranging further contact, the third most prominent online grooming intention in the keyword analysis, had three strategies: phone, day/time, online. Arranging further contact is also identified in the reviewed studies. Groomers arranged to talk to the target on the phone (*313, 714, cal, voice, anytime*), asked the target to switch platforms (*yahoo.com*) and arranged logistics of further offline and online contact (*Tues, 9pm, sunday, Saturday*). Groomers gave their targets options to choose from to supposedly give them control of the situation, showing a connection between arranging further contact and negotiating power dynamics. Switching to other platforms might be an aspect of assessing risks and isolating, specifically physical isolation was connected to arranging further contact.

Offline further contact was also discussed by focusing on the target’s exact location (*Mapquest, Daytona*), specific landmarks near them (*bowling, cedar, gas*) and planned activities (*motel, bowling, theater, mall*). Arranging further contact offline has previously been identified in the literature. This is unsurprising given the nature of how the PJ Foundation operated and created their archive, which was by having groomers arrange face-to-face meetings at a specific address to get them arrested and convicted for engaging in online grooming (see chapter one, section 1.1 and chapter three, sub-section 3.1.3 for more detail on the PJ). This strategy is called ‘approach’ in Lorenzo-Dus et al. (2016), ‘talking about meeting’ (p. 596) in public places, houses, cars in van Gijn Grosvenor and Lamb (2016), ‘arrange further contact and meeting’ in Pranoto et al., (2015: 361) and ‘preparing to meet offline’ in Egan et al. (2011: 14). Arranging further contact by phone was the second most prevalent strategy in chapter four (sub-section 4.3.3.2). The literature also mentions further contact by phone, called ‘requesting photos, videos, phone calls, voice messages to verify target identification’ (Chiang & Grant, 2017: 133), ‘requesting photos, videos, phone calls for ID verification’ (Chiang & Grant, 2017: 131), ‘used an online chat website that allows users to communicate via instant messaging, voice, and video chat’ (Kloess et al., 2017a: 7), ‘use of a mobile phone to communicate’ (Kloess et al., 2017a: 8) and ‘cell phone’ (Quayle & Newman, 2017: 4).

Arranging further contact online only had one keyword associated with it (chapter four, sub-section 4.3.3.4), which was likely due to the data source. PJ data did not include transcripts of phone conversation, although they did occur between groomers and PJ volunteers. Two previous studies refer to arranging further contact online. Quayle and Newman (2017), in one of the few studies using groomer-child data, describe it as ‘evidence of movement from Internet related activity to the use of a mobile phone to exchange texts and images’ (p.8). This occurred in 28 of the cybertip.ca reports they analysed. Gauz (2016) proposes a strategy simply called ‘arranging further contact’ which entailed ‘attempts by men to communicate with ‘children’ (i.e., decoys) via other means, which generally involved more immediate and/or personal mediums, such as telephone, email, and social networks (Gauz, 2016: 72). Moving to different online platforms could also be used as a risk mitigation technique by the groomer so as not to get caught. Arranging further contact is an important part of the online grooming process and this thesis therefore proposes that arranging further contact should incorporate online, offline and phone contact.

The assessing risks and isolating grooming intention also had four strategies attached to it: Secrecy, supervision, replacing parental figure and criticising parental figure. The latter two were part of the mental isolation strategy and the former ones were part of the physical isolation strategy. In chapter four, *sneak* and *neighbors* were used to assess the risks of the neighbours detecting and observing the offline meeting and to convince the target to sneak out to meet the groomer. *Grandma* and *aunt* were the keywords that groomers used to ensure no supervision while the groomer and target were chatting and to ensure the target’s confidentiality. *Tuck* and *moms* were also used by groomers to realise the mental isolation strategy. *Tuck* referred to groomers wanting to tuck their target in at night, replacing their parental figure, which they might also use to gain sexual gratification, suggesting a connection between mental isolation and seeking sexual gratification. *Moms* was used to ask about the target’s mother’s schedule, talk about the groomer’s own mother and criticise the target’s mother, both mental and physical isolation strategies. Assessing risks and isolating seemed to be closely connected to all other grooming intentions.

Negotiating power dynamics behaved differently to the other grooming intentions and to the identified strategies in the OGDM. Only role reversal appeared in the

concordance lines. In terms of keywords, the three keywords (*jail, cops, cop*) were connected to the strategy of legality. This grooming strategy is mentioned by van Gijn Grosvenor and Lamb (2016) who refer to it as ‘asking if they’re undercover decoys’ (p. 596). It is called ‘directly confronting to ensure that child is not a cop/police-agent’ in Gupta et al. (2012: 5), and ‘asking the target whether he/she was an undercover police officer’ in Black et al. (2015: 143). Groomers used these keywords to talk about the potential repercussions of talking to the target. They also accused targets of being undercover cops in undercover sting operations. Groomers adopted low-risk behaviour, an example of role reversal. This suggests that negotiating power dynamics might be realised differently and not identifiable on the one-word level.

7.2.2 Implications for Online Grooming Research: A Revised Terminology

The discussion in section 7.2.1 shows that keyword analysis adds nuance to the proposed terminology of online grooming communicative intentions derived from the literature in chapter two (section 2.5.3). This terminology is generated from online grooming models that were developed mostly from outside Linguistics and that thus provide a good basis for understanding how the online grooming process functions communicatively (see chapter two section 2.2.1). The model/terminology proposed in this thesis (see Figure 32) is the first one that is derived by extracting terms used in the key online grooming language literature, which was subsequently tested in an empirical study of online groomer language compared to general digital chat language, building on the recent work within Linguistics using a CADS approach (Lorenzo-Dus et al., 2016; Lorenzo-Dus et al., 2020) and solidifying this knowledge.

To recap, the five communicative intentions were:

- (i) Developing deceptive trust, which referred to the groomers’ attempts at befriending their targets and building a relationship with them, which is deceptive based on the ulterior motive of the groomer to sexually abuse the target.
- (ii) Seeking sexual gratification, which referred to online groomers’ attempts to desensitise their targets by introducing sexual language, discussions of the targets’ sexual history and experience levels and gaining sexual gratifications from these instances.

- (iii) Assessing risks and isolating, which referred to the groomers' efforts to keep the interaction secret to mitigate detection and arrest, isolating the target from their support systems and ultimately replacing this support network.
- (iv) Arranging further contact was defined as the groomers arranging contact with the target, which was primarily face-to-face contact, but also incorporated other forms of contact.
- (v) Negotiating power dynamics was defined as the groomers manipulating or influencing the targets by adopting low-risk behaviours, challenging them, threatening them, or blackmailing them.

In light of the findings and discussions above, Figure 32 shows a slightly revised terminology of online grooming communicative intentions and strategies. The developing deceptive trust grooming intention includes bonding, praise, and sociability, which are all summarised under a “developing and strengthening groomer-target bond” umbrella term. The seeking sexual gratification intention has been adapted to include further detail about the groomer educating the target by discussing contraception. This is used to desensitise the target and to discuss potential face-to-face meetings, which include sex. The communicative intention now also includes using child-language such as *tummy* in highly sexualised contexts to assimilate the target's language and frame themselves as a mentor. This is related to the reframing strategy but distinguished because it emerged in its own right in chapter four.

More information is also added to the arranging further contact communicative intention. The analysis confirmed that groomers arrange further contact in various ways (offline, online, by phone). Switching platforms has implications for the assessing risks and isolating strategy, as it could be used to switch to a platform that is more encrypted and on which the interaction is less likely to be discovered by law enforcement or parents of the target.

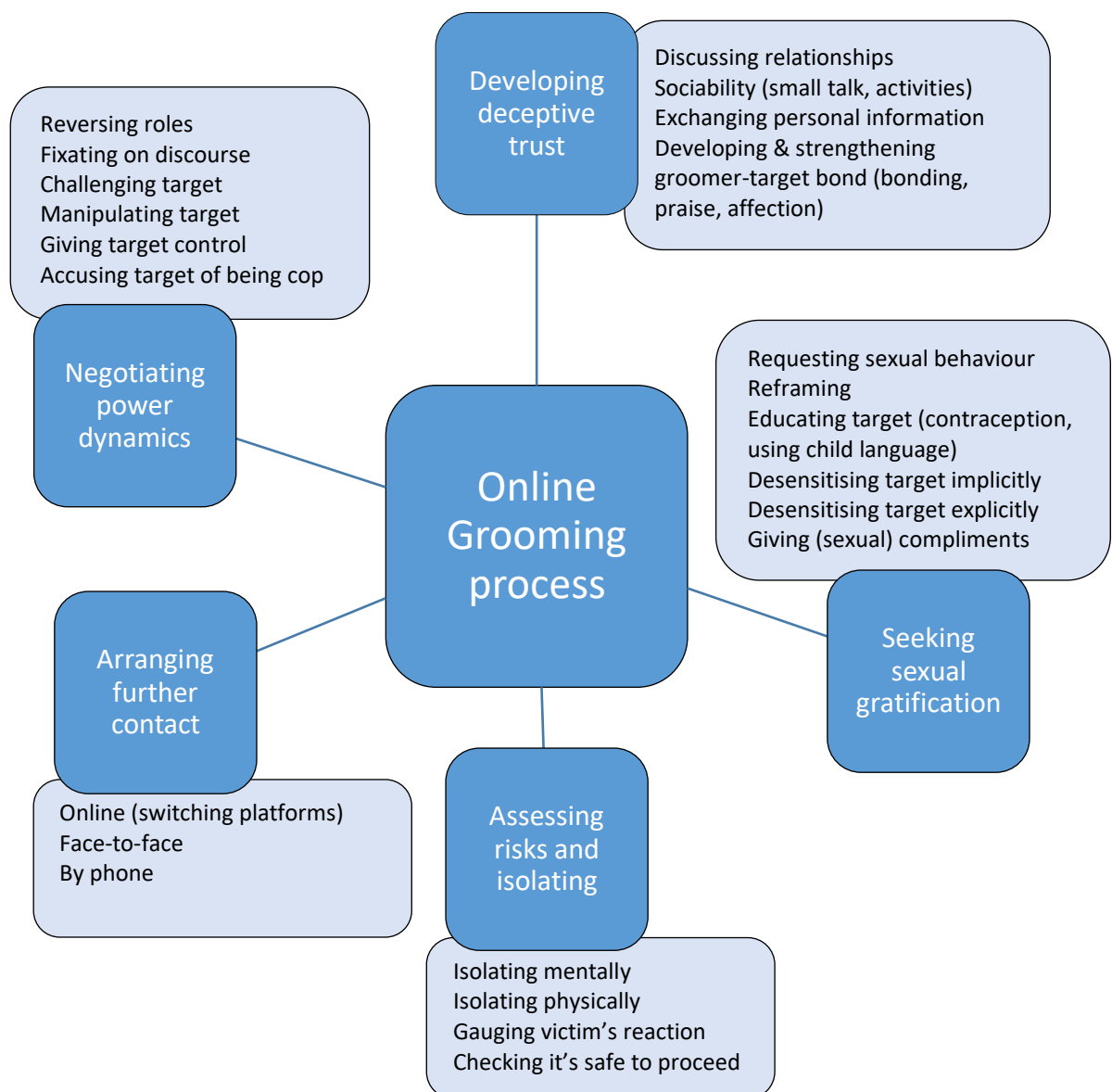


Figure 32: Revised terminology for online grooming communicative intentions

A final amendment is made in the negotiating power dynamics online grooming communicative intention: adding fixating on discourse, which emerged in chapters five and six when the groomer repeatedly and insistently asked the target for their address and whether their house number was visible from the street. Another aspect added to the negotiating power dynamics intention is accusing the target of being a cop, which emerged as a finding in chapter four and is also suggested in the literature.

Manipulating the target has also been added to this intention, which operates at a slightly different level than the other grooming intentions and somewhat transcends these communicative intentions.

Previous attempts at developing online groomer typologies in Psychology focused on groomer behaviour to understand their motivation, rather than groomer language (see chapter two, section 2.3.1.1). Briggs et al.'s (2011) distinction of contact-driven and fantasy-driven offenders has been called into question by Broome et al. (2018). The findings from this study further support that the distinction does not hold up. The online groomers in the PJ archive would all be considered contact-driven, as they all arranged face-to-face meetings with their targets and that is how they were caught and arrested. They do also engage in fantasy constructions about the targets and form deceptive relationships with them, which were strong themes throughout the analysis. All groomers, regardless of how long they interacted with their targets sought and gained sexual gratification from their interactions with their targets, which goes against Briggs et al.'s (2011) distinction (see section 7.3.2.).

7.2.3 New Resources for Online Grooming Detection Algorithms

As outlined in the introduction chapter (chapter one, section 1.3), the knowledge gained from chapter four could be used to inform Machine Learning algorithms and online grooming detection software. The findings also add further detail to the OGDM, which can in turn be used to inform Machine Learning algorithms that track or detect online grooming in conversations. Thus, 116 keywords within their respective online grooming intentions were identified (see Table 45 for an extract) in chapter four (for a full list see the appendix in section 9).

Table 45: Extract of table of 116 keywords identified within grooming intentions

#	OG intention	Strategy	Use	Keyword
1	DDT	Relationship	Term of endearment	sweetie
2	DDT	Relationship	Term of endearment	hun
3	DDT	Exchange pers. info	Exact location	addy
4	DDT	Exchange pers. info	Exact location	beach
5	SSG	Implicit	Vague language	excited
6	SSG	Explicit	Push-pull	foreplay

7	AFC	offline	logistics	park
8	AFC	phone	Phone call	cal
9	AFC	All	Day/time	saturday
10	ARISO	physical	secrecy	sneak
11	ARISO	mental	Replace figure	parental tuck
12	NPD	-	legality	jail
13	NPD	-	legality	cop

These terms as well as the more qualitative analysis about their context can be used to train a Machine Learning algorithm to recognise patterns commonly used by online groomers. More research into these specific keywords in context might be needed to derive longer strings of words, like those in Lorenzo-Dus et al. (2020), in which two and three-word collocational units were analysed and OGDM processes mapped onto them. Some of the keywords (n=52) identified in chapter four contributed to more than one grooming intention (see Table 46 for an extract), which can be helpful in identifying overlap between strategies and intentions (for a full list see the Appendix in section 9).

Table 46: Extract of table of 52 keywords with overlapping grooming intentions

#	Grooming intentions	Primary intention	Keyword
1	DDT + SSG + ARISO	DDT	adventure
2	DDT + SSG + NPD	DDT	youngest
3	DDT + AFC + SSG	DDT	tonight
4	DDT + SSG + ARISO + AFC + NPD	DDT + SSG + NPD	promise
5	DDT + AFC	DDT	addy
6	SSG + DDT	SSG	snuggle
7	SSG + AFC	SSG	condom
8	SSG + AFC	SSG	pill
9	SSG + ARISO + DDT	SSG	embarrassed
10	DDT + SSG + NPD + ARISO	DDT	nosy

7.3 RQ2: Does Duration Affect the Online Grooming Process?

Chapter five examined duration of online grooming and whether it has an impact concerning the types of online grooming intentions and strategies groomers use. There is some limited evidence from the literature that online grooming duration varies and

is both influenced by and influences the online grooming process. Lorenzo-Dus and Izura (2017) found evidence that “speed” of duration directly influenced the type and number of compliments groomers gave their targets. Considering this, chapter five asked whether the usage of specific words/grooming intentions is associated with different duration of grooming and whether different duration-based groomer profiles can be established and what the duration cut-off point of these are. The chapter drew up a duration overview in 100 minutes steps according to metadata gathered and decided on three cut-off points for analyses (300 minutes, 100 minutes, and 800 minutes) and six duration-based sub-corpora.

7.3.1 Duration Affects the Grooming Process: Two Different Approaches

The findings of the keyword analyses suggest that duration does influence which and how many grooming intentions groomers use. Shorter duration grooming interactions, which ranged from an engagement time of just 17 minutes to 299 minutes with their targets, used four grooming intentions (Deceptive trust development, arranging further contact, sexual gratification and assessing risks and isolating). Finding all grooming intentions in the shortest sub-corpus (0-99 minutes) further supports the hypothesis that online grooming is non-sequential. The more time groomers spent with their targets the fewer grooming intentions were identifiable in the keywords. Groomers spending 17-299 minutes with their targets used four grooming intentions (only missing negotiating power dynamics), while groomers spending 800-10,600 minutes interacting with their targets used two grooming intentions (developing deceptive trust and seeking sexual gratification). The seeking sexual gratification keyword (*sexy*) was also used for developing deceptive trust purposes, similar to a term of endearment. This difference in language was most clearly in the keyword analysis that compared sub-corpus G3 (0-299 minutes) with sub-corpus G4 (300-10,600 minutes). Groomers spending more time with their targets (300-10,600 minutes) seemed to focus on building and strengthening the deceptive trust relationship. They used terms of endearment (*princess, baby*), expressed feelings and emotions (*miss, missed, happy, love*), used emoticons (*X*), discussed activities (*busy, went, doing*) and used sociability (*sweet, dream, nite, sleep, bed*) to convince the target they were in a loving and caring relationship. In contrast, groomers spending less time with their targets (17-299 minutes) used a variety of grooming intentions, only missing negotiating power

dynamics. Developing deceptive trust had a different focus from the long duration groomers. Short duration groomers elicited personal information (*asl, address, zip, house, live, where, give*), sought out visual information (*profile, pics, cam, pic*), mitigated and used negative politeness (*prob, thanx*) and only used one term of endearment (*honey*). Groomer spending less time with their targets also arranged further contact face-to-face (*hang, come, meet, chill*), and by phone (*number, cell, call, phone*). They also seemed to be more focused on seeking sexual gratification by asking about the target's sexual history (*virgin*), experience levels (*condom, sex*) and instructing them to masturbate (*finger*).

The results of the second analysis, comparing the language of groomers who spent 0-99 minutes and those spending 100-199 minutes to each other, indicated that the language of those two groups was homogenous. They were focused on developing deceptive trust (*babe, m, k, quest*), arranging further contact (*meet, call*) and negotiating power dynamics (*please, want*). The findings therefore suggested that grooming duration does not make a difference below the cut-off point of 300 minutes. Negotiating power dynamics by clustering otherwise positive politeness strategies to the point they became impolite, begging, and evoking guilt emerge as unique to groomers who spent less than 300 minutes interacting with their targets.

The third keyword analysis, comparing sub-corpus G5 (0-799 minutes) with sub-corpus G6 (800-10,600 minutes), confirmed patterns in the first analysis showing that groomers who spent 300-10,600 minutes interacting with their targets seemed to be split into two groups: Those who spent 300-799 minutes grooming and those that spent over 800 minutes grooming. Groomers who spent less than 800 minutes chatting with their targets used mainly developing deceptive trust and seeking sexual gratification. They also used the grooming intentions of assessing risks and isolating and arranging further contact, which were the same four grooming intentions identified for the under 300-minute group in analysis one. Their focus on deceptive trust development strategies was slightly different, though, with a focus on sociability (*bye, later, yeah, ok, oh*) and two terms related to relationships (*sweetie, girl*). This group (under 800 minutes) had the highest number of sexually explicit keywords (*fuck, dick, horny, suck, sex, lick, pussy, cum*) that were used to introduce sexual language, describe sex acts, and ask if the target wants to have sex with the groomer. The findings suggested

the language of this group (under 800 minutes) was more similar to that of groomers who spent less than 300 minutes interacting with their targets and was part of the short duration approach. The focus of the over 800-minute group was on developing deceptive trust above all other grooming intentions. The groomers complimented the target (*special*), used terms of endearment (*baby*), expressed their emotions (*KISS-SMILEY_§*, *x*, *aww*, *miss*) and used relationship terms (*girlfriend*) to negotiate and strengthen the foundation of the deceptive trust. The same sexually explicit word (*sexy*) that was identified in the above 300-minute group in the first analysis appeared again in the over 800-minute group. Similar to the first analysis, this keyword was not overtly sexual but was more implicit.

Overall, the three keyword analyses showed that there are two distinct grooming approaches with different grooming intentions and that the cut-off point seems to be within the sixth sub-corpus, so around 800 minutes. There was a clear difference to the language used by groomers spending less than 800 minutes grooming. A specific cut-off point could not be defined, as duration lies on a continuum.

7.3.2 Implications of Two Online Grooming Duration-Based Approaches

Although some studies in the literature report differences in the average duration of online grooming, only one study to date (Lorenzo-Dus & Izura, 2017) has analysed it systematically focusing on one aspect of online grooming: compliments (see chapter two, section 2.2.2). The findings of chapter five confirm that duration makes a difference not just to compliments but also other online grooming communicative intentions. This supports the finding of different duration language profiles.

The findings from chapter five show that all online grooming intentions were found in the short online grooming duration approach corpora. Four of them (developing deceptive trust, arranging further contact, seeking sexual gratification, assessing risks and isolating) in G3 (0-299 minutes) and three of them (developing deceptive trust, arranging further contact, negotiating power dynamics) in the 0–99-minute sub-corpus (G1). There was also evidence of seeking sexual gratification in the shortest sub-corpus, although it was difficult to pinpoint at the one-word level (chapter five, section 5.3.3). These findings support the claim (see Lorenzo-Dus et al., 2016) that the online

grooming is comprised of inter-linked rather than linear processes. Black et al. (2015) suggest that sexual topics are introduced within the first 20% of the interaction and Winters et al. (2017) state it happened within 33 minutes, which is supported by the findings in chapter five, too.

Furthermore, there was clear evidence for two different online grooming approaches based on grooming duration. This had been hinted at in earlier research, with Kloess et al. (2017a) describing two different approaches, one gentle and the other one direct and highly sexualised. O’Connell (2003) states that different conversational patterns occurred in the sexual stage, with groomers wanting to uphold a relationship with their target entering the sexual stage ‘gently and the relational framing orchestrated by the adult is for the child to perceive the adult as a mentor or possible future lover’ (O’Connell, 2003: 10). The goal was ‘forming a loving/lasting relationship/friendship’ (p. 10) but groomers might also attempt to organise face-to-face meetings with their targets. Sexual content also ranged from ‘mild suggestions to explicit descriptions’ (p. 10). Kloess et al. (2017a) did not have the linguistic tools and large dataset to further elaborate on their observation and generalise their findings, as they only looked at 29 groomer-child chat logs. O’Connell’s (2003) interpretations were specifically about the sexual stage and did not look at the overall online grooming process. The studies were also not able to validate these claims using a larger dataset. The findings from chapter five provide these tools and linguistic knowledge based on a large corpus of online grooming that confirm there are two different grooming approaches based on duration. The findings also suggest that this does not just apply to the online grooming intention of seeking sexual gratification but changes the way groomers use strategies within deceptive trust development and other grooming intentions. This will be discussed further in section 7.5.2.

7.3.3 Tailoring Law Enforcement Efforts to Tackle Online Grooming

The knowledge gained from the duration-based keyword analyses can be used to specialise and tailor efforts by law enforcement to these two specific duration approaches identified, as the introduction chapter mentioned. As part of the analysis, 79 unique keywords have been identified as part of the short duration grooming

approach (n=54) and the long duration grooming approach (n=24). An extract of this list can be seen in Table 47 (See the Appendix for a full list in section 9).

Table 47: Extract of keywords associated with the two different duration approaches

#	Grooming approach	Intention	Use	Keyword
1	short	DDT	exact location	asl
2	short	DDT	exact location	address
3	long	DDT	relationship term	girlfriend
4	long	DDT	term of endearment	princess
5	short	SSG	explicit desensitisation	condom
6	short	AFC	offline	hang
7	short	AFC	offline	Sunday
8	short	AFC	online	here
9	short	ARISO	assessing risks	older
10	short	NPD	role reversal	want

Counter-grooming profiles of these two grooming approaches could be created by law enforcement and training for adults with child-safeguarding responsibilities (counsellors, teachers, therapists), parents and children could be created outlining what these two different grooming approaches entail and what to watch out for. This prevention material would be more nuanced than previous material, highlighting that there is not just one distinct online groomer approach, but several duration-based ones. These approaches are highly complex and adaptable, and change based on how much time the groomer has to invest in the interaction.

7.4 RQ3: Does Duration Influence Groomers' Requesting Behaviour?

Chapter six examined power dynamics and manipulation in groomer language through analysing request sequences based on two different duration groups. This chapter was informed by findings from the previous two empirical chapters and was primarily qualitative (Discourse Analysis) in its approach. Previous chapters showed that duration influenced groomers' level of directness and use of sexually explicit language. There is an inherent power imbalance between the groomer and the target, which leaves the target vulnerable to being manipulated. Overall, the grooming process is action oriented: the groomer wants the target to do something (e.g., give him their phone number/address, engage in sexual discussions, masturbate, meet them

offline). Requests are quite likely to be used by groomers to achieve their communicative grooming intentions. Directive speech acts, which requests are part of, have been studied in terms of their illocutionary requestive force. Moreover, all such force is always manipulative: the groomer's ulterior motive is to sexually abuse their target. This motive is kept from their targets, which is one of the defining features of manipulation (van Dijk, 2006). Chapter six thus asked how requests are realised by groomers and whether grooming duration influences realisation. Request realisation (aligned to grooming duration) was examined using established methods in Speech Act Theory, which focus on both the request head act and its internal / external modification.

7.4.1 Duration Influences Online Groomers' Requesting Behaviour

The chapter's findings show that duration does influence groomers' requesting behaviour. Groomers from the under-300-minute (G3) and over-300-minute (G4) group used support move functions differently and strategically. As regards realisation of head acts in the two sub-corpora, there were notable differences between the two sub-corpora. Suggestory formulae in G4 (81%) were more often focused on developing deceptive trust than in G3 (44.4%). Obligation statements also more often contributed to developing deceptive trust in G4 (77.7%) compared with G3 (39.7%). Assessing risks and isolating in want statements received high levels of support in both G3 (57%) and G4 (56%). These three head act types were subsequently focused on in more detail, concentrating on the support move functions.

In terms of suggestory formulae, the same support move functions were used by both groups but their use of them differed. Minimising was used by short duration groomers (G3) to minimise the imposition of the potential face threat. Long duration groomers used more diverse forms of minimising, such as vague language and created a coded language with their target and assimilated their language, which strengthened the deceptive bond while minimising the imposition. Intensifying in suggestory formulae requests was used by long duration groomers to clarify in the same turn, while it was used to harass and elicit information by short duration groomers over several turns. Enhancing interpersonal relations was used by short duration groomers (G3) to elicit

information, while long duration groomers (G4) generally used it to build trust and emphasise familiarity.

Obligation statement requests had different support move functions in the two groups. Short duration groomers preferred minimising support moves, while long duration groomers used giving grounds and intensifying. This suggested that short duration groomers minimised the direct obligation statements considerably and turned them into suggestions. In contrast, long duration groomers intensified these direct obligation statements. Short duration groomers also used vague language to minimise the imposition on the target's negative face and mitigate sexual content. Long duration groomers (G4) used intensifying to emphasise the romantic nature of the relationship, rather than focusing on sexual activities. Threats were used by long duration groomers in obligation statements and trust was used as a currency, supported by use of positive politeness strategies. Negative impoliteness strategies were also used: long duration groomers hindered targets linguistically, threatened them, evoked guilt, and implied targets could not be trusted if they did not volunteer personal information. Both groups established a student-teacher relationship connected to educating the target about sex, by using enhancing personal relations. In addition, short duration groomers used vague language to mitigate this sexual content.

Short duration groomers (G3) used all support move functions in want statement requests. Long duration groomers only used intensifying and giving grounds. Neither group wanted the target to get into trouble, but long duration groomers used slight pressure on the target to achieve their goal. They used positive politeness strategies to counteract the impoliteness.

Overall, short duration groomers were very goal-oriented and used impoliteness, specific and open-ended threats, and incited violence, in short, did anything to achieve their goals. Long duration groomers used intensifying to persuade the target to do something. They also increased their interest in the targets, told them they were loved and tried making targets more dependent on them, an overall more subtle approach to manipulation.

7.4.2 Towards a Description of Two Groomer Duration Profiles

The findings from chapter five and six show that there is evidence for at least two grooming approaches based on different online grooming duration: A short grooming approach (under 300 minutes) and a long grooming approach (over 300 minutes). These two grooming approaches have different online grooming intentions and strategies associated with them (see section 7.3.2) and groomers use requests differently. An overview of the approaches can be seen in Figure 33. Short duration grooming conversations are very goal-oriented and seek instant gratification, which often takes the form of asking for visual material (photos, webcam), arranging a face-to-face meeting with their target to abuse them sexually or gaining sexual gratification via fantasy enactment. They frequently use highly sexualised language with their targets, which is somewhat mitigated with vague language when making requests.

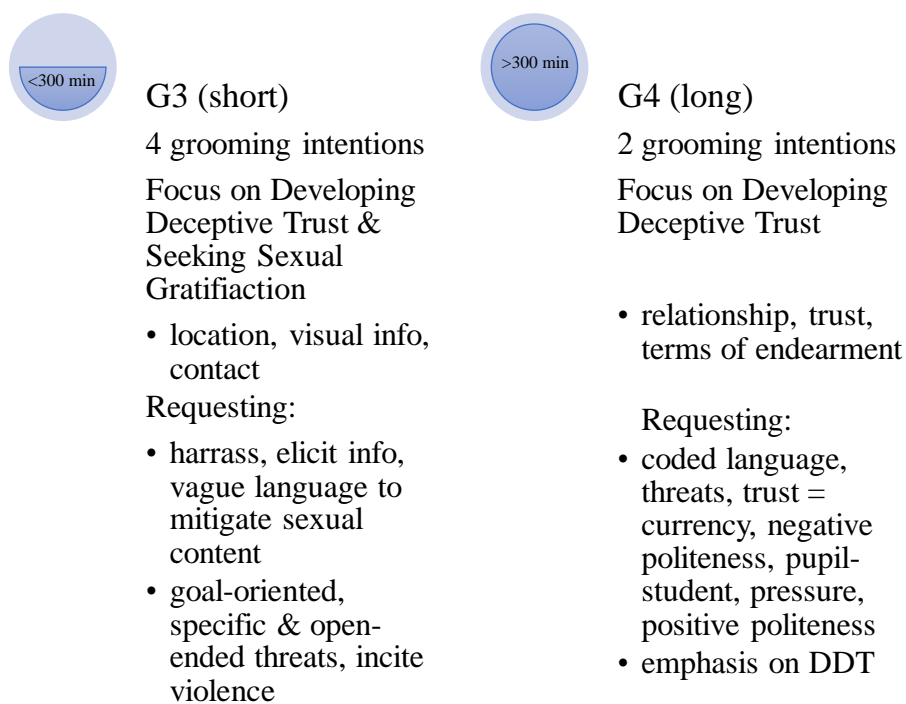


Figure 33: The two duration-based grooming approaches

Short duration grooming conversations elicit personal information (exact location, visual) and groomers do not shy away from making threats, harassing their targets, or inciting violence. This resonates with Webster et al.'s (2012) category of the hypersexualised offender group, whose approach was described as follows:

Their contacts with young people were highly sexualised and escalated very quickly. Their offence supportive beliefs involved ‘dehumanising’ young people. They tended not to personalise contact and so did not seem to be using the phone or other personal media like the other groups of offenders. (Webster et al., 2012: 85)

The difference is that the short duration grooming in this study did engage in developing deceptive trust, even though seeking sexual gratification was foregrounded. Groomers within this group also sought to schedule face-to-face meetings, but this is likely because all groomers in the PJ archive did so. Webster et al. (2012) report that some groomers in this group did seek out meetings and that they ‘seemed particularly susceptible to undercover police operations’ (Webster et al., 2012: 86). The short duration groomers in this study also arranged further contact by phone, unlike Webster et al.’s (2012) categorisation. Other aspects of their language such as the focus on sexual content, goal orientation and being very determined does seem more congruent with the hyper-sexualised classification.

Long duration groomers in this study, on the other hand, seemed to be very focused on building a relationship with their targets, making them feel loved, and using terms of endearment. Sexual content was implicit, and trust was emphasised. In their requesting behaviour, this trust became a currency, which could be earned or taken away. Long duration groomers created a coded language with their targets, which only they would understand to deepen the deceptive trust. They did also use threats and put pressure on their targets, but this was counter-balanced by using positive politeness strategies. Webster et al.’s (2012) intimacy-seeking group was described very similarly. They ‘seemed to spend a significant amount of time online talking to the young person before they met the victim. All men in this group went on to meet the victim to develop or further the ‘intimate relationship’ (Webster et al., 2012: 82). This was also the case in long duration groomers. Their approach to sexual content was described as ‘[p]rolonged and frequent, sexual conversations not introduced early, slow build up, as if getting to know a friend.’ (Webster et al., 2012: 82), which seems very similar to the long grooming duration approach, which used trust and implicit sexual content to emphasise the relationship with the target. Webster et al. (2012) also described this group as having ‘feelings of a ‘consenting relationship’ and feeling

‘love struck’. Idealised romantic fantasy.’ (Webster et al., 2012: 83), which is similar to how long duration groomers expressed their feelings to their targets, used terms of endearments and generally focused on bonding.

Webster et al.’s (2012) made an important point about the adaptable group, which is that ‘the speed at which contact developed could be fast and/or slow according to the how the young person responded to contact’ (p. 83). This is not something that could be analysed in the present study. It is possible that some of these groomers take different approaches with different targets and adapt their style accordingly. They might use a short duration approach with one target and a long duration approach with another one. This is something that would be useful to analyse further (see section 7.5.2).

These two online grooming duration approaches identified in this study require different methods of detection and prevention. Training for young people that focuses on specific aspects from these duration approaches (e.g., rapid introduction to sexual content in online conversations, strangers appearing to want to be friends) and safeguarding action plans to stop both grooming approaches in their tracks (e.g., questioning people’s intentions if they start asking inappropriate questions, not revealing information about location or parent’s schedule) could be developed and given to young people, schools, and parents. This could be similar to the Stop TIME online resources (Stop TIME online, 2017), which is an anti-grooming activity pack explaining Lorenzo-Dus et al.’s (2016) OGDM in simple and accessible terms. It was developed by the NSPCC and Swansea University for people with safeguarding responsibility and young people. The acronym TIME stands for Trust, Isolate, Measure and Enjoy, which are equivalent to deceptive trust development, isolation, compliance testing and sexual gratification from the original OGMD, respectively. The activity pack can be used in small groups or one-to-one activities and has been developed in collaboration with young people, social workers, law enforcement and youth services (Stop TIME online, 2017). A similar activity pack could be developed for the findings of this thesis, specifically focused on differences in the two grooming approaches. Resources and training for spotting manipulative requesting behaviour could also be developed and given to young people, teachers, and parents. Further

research is needed with other online grooming databases analysing different aspects of online grooming according to these grooming duration profiles (see section 7.5.2).

7.5 Drawing Conclusions on a Study of Online Groomer Language

This thesis aimed to analyse the features of online groomer discourse by providing the first detailed quantitative and qualitative linguistic analysis of a sizeable corpus of online grooming language. Based on the keyword analysis in chapter four, it can be concluded that online groomer language is distinct and the OGDM developed by Lorenzo-Dus et al. (2016) is robust. The analysis identified more detailed linguistic realisations of some of the strategies within online grooming intentions of the OGDM and a list of keywords used, some of which are used for multiple online grooming intentions (see sub-section 7.2.2). The analysis also adds knowledge to the OGDM and online grooming research as a whole and developed a more nuanced overview of online grooming communicative intentions and strategies (see Figure 32, section 7.2.2), which can be further developed and used to prevent online grooming and safeguard children.

The thesis also examined the impact duration has on online groomers' discourse, which has not been analysed systematically. The findings show that there are two different online grooming approaches based on duration of online grooming. Different online grooming intentions and strategies are used for these two approaches. This finding makes a significant contribution to online grooming language research, as it challenges the view of online grooming language as homogenous. Previous research, especially in Machine Learning attempted to describe the characteristics of online grooming language as a whole, distinguishing it from other language (e.g., Pendar, 2007; Kontostathis et al., 2009; Bogdanova et al., 2014; Pranoto et al., 2015), rather than creating different online groomer language profiles to analyse in detail and compare and contrast to each other.

The findings of this thesis also open up a new area for duration-based research into different aspects of online grooming. Duration-based language profiles could be used by law enforcement to tailor their efforts to detect online grooming and arrest online groomers. The last aim of the thesis was to examine how groomers use manipulative

requests and how they negotiate power dynamics and whether duration influences their requesting behaviour. The findings further established that there are two duration-based grooming approaches. The findings suggest short duration grooming is much more goal-oriented and highly sexualised compared to long duration groomers. Short duration groomers also look for instant gratification and they will stop at nothing to achieve their goal, using threats and inciting violence to achieve their aims. This information could be used to develop specialised training to schools to highlight the differences in language of online groomers, to raise awareness of the different approaches and ultimately to aid in online grooming prevention. Law enforcement could also use the findings to create different counter-grooming profiles and tailor their response of online grooming detection and arrests.

The academic community's response to the social problem of online child sexual grooming has been characterised by a comparative scarcity of language-based studies that use linguistic tools and combine quantitative and qualitative methods to analyse online grooming. Linguistic input using mixed method analysis of online grooming *language* has been largely missing. Research in Psychology and Criminology has focused on groomer characteristics, motivations (Briggs et al., 2011; Merdian et al., 2013) and typologies (Kettleborough & Merdian, 2017; DeHart et al., 2017), children's perspectives, characteristics, and risk factors (Katz, 2013; Whittle et al., 2014a; Whittle et al., 2014b; Kloess et al., 2017b). Studies on the language of the grooming process have mostly used content-based categories with the goal of developing online grooming models, identifying grooming processes and strategies. These models have seen the grooming process as being either sequential (Egan et al., 2011; Gupta, et al., 2012; Black et al., 2015; Kloess et al., 2017a) or non-sequential (Williams et al., 2013; Lorenzo-Dus et al., 2016; van Gijn-Grosvenor & Lamb, 2016; Quayle and Newman, 2017, Chiang and Grant, 2017, 2018). A few recent studies have used CADS to analyse aspects of the online grooming process (Lorenzo-Dus & Kinzel, 2019; 2021, Lorenzo-Dus et al., 2020), which lends itself well to online grooming language research and provides much needed linguistic scrutiny. This thesis contributes new insights to this emerging research direction. One particular aspect, online grooming duration, has not received much attention in the literature and has been identified as an overlooked but potentially important aspect. This thesis

addresses this gap in knowledge through a systematic examination of the duration of online grooming using a corpus of online grooming language.

In the context of an area of research (online grooming and more broadly online child sexual abuse/exploitation) characterised by significant terminological and conceptual proliferation – with the consequent risk of confusion – this thesis proposes a new working terminology. This comprises five online grooming communicative intentions with strategies. The empirical chapters of this thesis (chapter four-six) also address gaps in literature, such as providing a keyword analysis of online grooming language, compared with a general digital chat language corpus, and a detailed linguistic analysis of online grooming intentions. Keywords belonging to these grooming intentions were identified and examined in detail, which provides further insights into the online grooming communicative intentions stemming from the literature. Duration of online grooming as a second gap in the literature was addressed in chapter five by creating six duration-based sub-corpora of online groomer language and comparing them to each other using keyword analyses. The keywords were also analysed qualitatively. The findings demonstrate that there are two duration-based online grooming language approaches with a likely cut-off point around 800 minutes. Groomers interacting less than 800 minutes with their targets are using four grooming intentions but focusing on visuals, exact location and arranging further contact with the target while groomers interacting more than 800 minutes with their targets put more focus on establishing a trusting and loving bond with their targets. Groomer duration does not make an impact below 300 minutes of interaction, as the language is homogenous and exhibits the short duration grooming approach. Manipulation was the third focus of this thesis, specifically how groomers negotiate power dynamics in manipulative request sequences and whether duration influences the requesting behaviour, which was analysed in chapter six. Three specific head act types were chosen for closer analysis, and it was found that duration does impact this manipulative requesting behaviour, providing more detail to the emerging two duration-based approaches to online grooming.

The methodology of the thesis, CADS, proved to be successful in providing more linguistic detail to online grooming language research. It allowed me to examine a 3.7-million-word corpus of online groomers in its entirety but also to carry out fine-

grained close readings of statistically significant patterns therein to address the thesis aims and research questions. At times, this was challenging, especially when it came to combining a content-based framework (OGDM) and applying it to lexicogrammatical categories (keywords, collocations). This challenge was addressed in chapters four and five by first identifying the keywords and working on the one-word level and then adding the content-based qualitative discourse elements of the OGDM when the keywords were examined in context and not restricted to the one-word level. Politeness and impoliteness strategies were also identified on this level, rather than trying to address them using raw keywords. CADS was especially useful in this, as it requires the researcher to take a close look at the keywords in context, rather than relying on numbers and percentages alone.

This dichotomy of lexico-grammatical categories and content-based frameworks became challenging again in chapter six while using a corpus to analyse a pragmatic concept – requests. As mentioned in chapter three, like most speech acts, requests do not operate on the one-word level, which can be used as search terms in a corpus. They are made up of an almost endless possibility of multiple-word units that make it impossible to identify all requests in a given corpus using search terms/strings only. Instead, I reviewed request literature from Pragmatics and found a framework of head act types I could operationalise as search queries in the corpus to extract a subset of requests that were formulaic and analysed further using fine-grained qualitative analysis. This gave an insight into how online groomers use manipulative requests to convince their targets to perform actions and also shows that these two very different methodological aspects can be combined successfully using a CADS approach.

7.5.1 Limitations

One limitation of this thesis, as mentioned in the introduction chapter (chapter one, section 1.1) is the reliance on PJ data. This is due to the difficulty in securing groomer-child data from police forces, which involves questions about data storage, handling, analysis and potentially asking the subjects for consent. Groomer-target data has been used in the majority of studies and there are still as yet unexplored aspects of it that need answers. Duration and requesting behaviour were two such aspects, which this thesis has analysed successfully. The analysis in this thesis also focused on the

groomer contributions to the conversations and the groomer language. The target language was only kept for context purposes and to have a full picture of the requesting sequences in chapter six. It was not analysed beyond that, and no claim is made about how successful the decoys are at pretending to be underage children. Additionally, having worked with this data closely for over three years, I did not get the impression that groomers adapted their language and behaviour to accommodate chatting to decoys. In fact, the analysis suggests that they were unaware this was the case and thought they were talking to underage individuals until they showed up to a face-to-face meeting and were arrested. The analysis showed an array of different behaviours groomers used and adapted based on the time they invested in the interaction. This behaviour is complex and rich and adds knowledge to online grooming language research, which may be validated using groomer-child data. While the focus of the analysis was on the groomer language, it is possible that the decoys used language and made linguistic choices that elicited particular responses from the groomers.

Use of groomer-child data had in fact been initially planned for – and shared with the research team I belong to for use in this thesis. Work on anonymising this data had taken place for nine months in a secure building with strict access rules and stringent security measures when the first covid-19 lockdown in the UK made it impossible for me to access the secure location in order to complete data anonymisation and to progress to data analysis. At that point, I was getting ready to design my third empirical chapter and the initial plan was to use the groomer-child data to validate the findings from the previous empirical chapters. However, when I did not know when I would be able to return to campus, I made the decision to continue using PJ data for chapter six and focus on requesting behaviour. So, one limitation of this thesis is that I was unable to validate some of the findings from the first two empirical chapters using groomer-child data. This is something that needs further research (see subsection 7.5.2). Nevertheless, the alternative that I pursued (examining manipulative requesting behaviour) is novel regardless of data source in online grooming research. Therefore, this thesis still makes an original contribution to the research field. Future research could also compare decoy language to children language to find out whether the decoys' linguistic choice leads to some of the groomers responses.

7.5.2 Future Direction in Online Grooming Language Research

Faced with a social challenge that is exponentially growing, future research into online groomer language is still needed. One area that this future research should focus on is the meaning making potential of emoticons (see chapter five, section 5.3) and other onomatopoeic/textual representations of paralinguistic features such as laughter (*haha, lol*), which were not considered in this thesis. This could be expanded to include GIFs, emoji and reactions to messages that have been introduced to social media platforms recently. Digital language keeps changing as technology changes and it is important to adjust tools used in Discourse Analysis and other disciplines. As an example, Susan Herring's model of Computer-Mediated Discourse Analysis, which was originally developed in the first wave of Digital Discourse research initially encompassed four language domains: structure, meaning, interaction and social behaviour (Herring, 2003, 2004). It was adapted according to technological developments and multimodality was added as a fifth domain (Herring & Androutsopoulos, 2015).

Another area could be focused on other communicative features of online grooming concerning the duration-based language profiles. These aspects could be specific speech acts (questions, demands) or focus on one online grooming intention in more detail to find out how it is used by groomers spending different lengths of time with their targets. Validation from other datasets (groomer-child) is also needed to further define the duration-based online groomer language profiles. Another aspect mentioned in section 7.4.2 is to further analyse the grooming duration language profiles and find out if groomers switch between them. The language of groomers that interact with many different targets could be analysed to find out if they switch between a long and short duration grooming approach depending on how the target responds. This is similar to Chiang and Grant's (2018) approach of authorship profiling in Forensic Linguistics but focused on duration-based language approaches. The different conversations would first have to be classified according to the duration-based language profiles.

The specific keyword lists (section 7.2.3) could also be analysed further by calculating their collocates and developing two- or three-word units, similar to Lorenzo-Dus et

al.'s (2020) study. This could then be used to inform Machine Learning algorithms to detect online grooming language more accurately. Overall, further research into online grooming language should make use of findings from other fields (Psychology, Machine Learning, Criminology) and use interdisciplinary research to tackle online grooming. This thesis has shown that Linguistics can contribute to research into this *communicative* phenomenon and that other fields have also made valuable contributions and laid the groundwork. Machine Learning in particular could gain a lot of knowledge from Linguistics to fine-tune their modules and more accurately identify online groomer language.

7.6 Conclusion

This chapter has discussed the findings from chapters four, five and six and drawn conclusion. It was divided into three sub-sections according to the three different chapters and first summarised the main findings and answering the research questions in sub-sections 7.2.1, 7.3.1 and 7.4.1. The significance and potential applications of the findings from chapter four, including for Machine Learning algorithms trained to detect online grooming were outlined in sub-section 7.2.3. The applications of chapter five, which suggest tailoring law enforcements' efforts to detect and prevent online grooming, could be found in sub-section 7.3.3. Lastly, the incorporation of duration-based online grooming language profiles into training and safeguarding resources for young people, teachers, and parents, were discussed in sub-section 7.4.2. The theoretical implications for the field of online grooming language research were also addressed in sub-sections 7.2.2 and 7.3.2. The chapter then drew conclusions, outlining the aims of the overall thesis first and stated how the thesis addressed them. A reflective summary of the thesis and methodology followed. Limitations were outlined in sub-section 7.5.1. Future directions of and recommendations for online grooming language research were made in sub-section 7.5.2.

8. References

- ACCCE. (2020). *Statistics 2020*. ACCCE. (2020). Retrieved from: <https://www.accce.gov.au/resources/research-and-statistics/2020statistics> [last accessed November 2021]
- Aijmer, K. (2015). “Will you fuck off please”. The use of please by London teenagers. *Pragmática Sociocultural / Sociocultural Pragmatics*, 3(2), 127–149. <https://doi.org/10.1515/soprag-2014-0028>
- Aijmer, K. (2019). ‘Ooh whoops I’m sorry! Teenagers’ use of English apology expressions. *Journal of Pragmatics*, 142, 258-269. <https://doi.org/10.1016/j.pragma.2018.10.017>
- Alipour, M., & Nooreddinmoosa, M. (2018). Informality in Applied Linguistics Research Articles: Comparing Native and Non-Native Writings. *Applied Linguistics*, 4(2), 349-373. <https://doi.org/10.32601/ejal.464196>
- Almaged, S. (2021). Disseminating knowledge: A discourse analysis of terrorism in TED talks. *Heliyon*, 7(2), 1-8. <https://doi.org/10.1016/j.heliyon.2021.e06312>
- Anthony, L. (2020). AntConc (Version 3.5.9) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from <https://www.laurenceanthony.net/software>
- Austin, J. L. (1975). *How to Do Things with Words: The William James Lectures Delivered at Harvard in 1955*. Oxford Scholarship Online. <https://doi.org/10.1093/acprof:oso/9780198245537.001.0001>
- Baker, P. (2006). *Using Corpora in Discourse Analysis*. Continuum.
- Baker, P., Gabrielatos, C., Krzyzanowski, M., McEnery, T., & Wodak, R. (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. <https://doi.org/10.1177/0957926508088962>
- Baker, P. (2009). The BE06 Corpus of British English and recent language change. *International Journal of Corpus Linguistics*, 14(3), 312–337. <https://doi.org/10.1075/ijcl.14.3.02bak>
- Baker, P. (2014). *Using Corpora to Analyse Gender*. Bloomsbury.
- Baker, P. (2015). Introduction to the Special Issue. *Discourse & Communication*, 9(2), 143–147. <https://doi.org/10.1177/1750481314568549>
- Baker, P., & Levon, E. (2015). Picking the right cherries? A comparison of corpus-based & qualitative analyses of news articles about masculinity. *Discourse & Communication*, 9(2), 221-236. <https://doi.org/10.1177/1750481314568542>

- Baker, P., & Vassey, R. (2018). A corpus-driven comparison of English and French Islamist extremist texts. *International Journal of Corpus Linguistics*, 23(3), 255-278. <https://doi.org/10.1075/ijcl.17108.bak>
- Baron, A., & Rayson, P. (2009). Automatic standardization of texts containing spelling variation, how much training data do you need? In M. Mahlberg, V. González-Díaz, & C. Smith (Eds.), *Proceedings of the Corpus Linguistics Conference (CL2009)*. University of Liverpool, Liverpool. 20-23 July 2009. Retrieved from: http://eprints.lancs.ac.uk/42529/1/314_FullPaper.pdf [last accessed June 2021]
- BBC. (2015, June 17). *Reports of child abuse rising sharply, NSPCC says*. BBC News. Retrieved from: <https://www.bbc.com/news/uk-33160361> [last accessed June 2021]
- Bednarek, M. (2018). *Language and Television Series: A Linguistic Approach to TV Dialogue*. Cambridge University Press.
- Bennett, N., & O'Donohue, W. (2014). The Construct of Grooming in Child Sexual Abuse: Conceptual and Measurement Issues. *Journal of Child Sexual Abuse*, 23, 957-976. <https://doi.org/10.1080/10538712.2014.960632>
- Berger, C., Friginal, E., & Roberts, J. (2017). Representations of immigrants and refugees in US K-12 school-to-home correspondence: An exploratory corpus-assisted discourse study. *Corpora*, 12(2), 153-179. <https://doi.org/10.3366/cor.2017.0115>
- Berson, I. R. (2003). Grooming cybervictims: The psychosocial effects of online exploitation for youth. *Journal of School Violence*, 2(1), 5-18. https://doi.org/10.1300/J202v02n01_02
- Black, P., Wollis, M., Woodworth, M., & Hancock, J.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 & Neglect*, 44, 140-149. <https://doi.org/10.1016/j.chiabu.2014.12.004>
- Blommaert, J., Smits, L., & Yacoubi, N. (2020). Context and Its Complications. In A. De Fina & A. Georgakopoulou (Eds.), *The Cambridge Handbook of Discourse Studies* (1st ed., pp. 52-69). Cambridge University Press. <https://doi.org/10.1017/9781108348195.004>
- Blum-Kulka, S. & Ohlstein, E. (1984). Requests and Apologies: A Cross-Cultural Study of Speech Act Realization Patterns (CCSARP). *Applied Linguistics*, 5(3), 196-213. <https://doi.org/10.1093/applin/5.3.196>
- Blum-Kulka, S., House, J., & Kasper, G. (2011). Cross-Cultural Pragmatics: Requests and apologies. In Archer, D. & Grundy, P. (Eds.). *The Pragmatics Reader* (pp. 344-253). Routledge.

- Bogdanova, D., Rosso, P., & Solorio, T. (2014). Exploring high-level features for detecting cyberpedophilia. *Computer Speech and Language*, 28, 108-120. <https://doi.org/10.1016/j.csl.2013.04.007>
- Bolander, B., & Locher, M. A. (2020). Beyond the online offline distinction: Entry points to digital discourse. *Discourse, Context & Media*, 35, Article 100383. <https://doi.org/10.1016/j.dcm.2020.100383>
- Bou-Franch, P., Lorenzo-Dus, N., & Blitvich, P. G.-C. (2012). Social Interaction in YouTube Text-Based Polylogues: A Study of Coherence. *Journal of Computer-Mediated Communication*, 17(4), 501–521. <https://doi.org/10.1111/j.1083-6101.2012.01579.x>
- Brackenridge, C. H. (2001). *Spoilsports: Understanding and preventing sexual exploitation in sport*. Routledge. <https://doi.org/10.1136/bjism.36.4.311>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77-101. <https://doi.org/10.1191/1478088706qp063oa>
- Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge University Press. <https://doi.org/10.1017/9781316410899>
- Briggs, P., Simon, W.T., & Simonsen, S. (2011). An Exploratory Study of Internet-Initiated Sexual Offenses and the Chat Room Sex Offender: Has the Internet Enabled a New Typology of Sex Offender? *Sexual Abuse: A Journal of Research and Treatment*, 23(1), 72-91. <https://doi.org/10.1177/1079063210384275>
- Brown, P., & Levinson, S.C. (1987). *Politeness: Some Universals in Language Usage*. Cambridge University Press.
- Brown, P., & Levinson, S. C. (1978). Universals in language usage: politeness phenomena. In E.N. Goody, (Ed.) *Questions and Politeness Strategies in Social Interaction* (pp. 56–289). Cambridge University Press.
- Broome, L. J., Izura, C., & Lorenzo-Dus, N. (2018). A systematic review of fantasy driven vs. contact driven internet-initiated sexual offences: Discrete or overlapping typologies? *Child Abuse & Neglect*, 79, 434–444. <https://doi.org/10.1016/j.chiabu.2018.02.021>
- Buchanan, L. (2016). *The Dilemmatic Nature of Luring Communication: An Action-Implicative Discourse Analysis of Online Predator and P-J Member Interaction*. [Doctoral dissertation, The University of Iowa]. Iowa Research Online. Retrieved from: <https://ir.uiowa.edu/cgi/viewcontent.cgi?article=6403&context=etd> [last accessed June 2021]

- Cardei, C., & Rebedea, T. (2017). Detecting sexual predators in chats using behavioral features and imbalanced learning. *Natural Language Engineering*, 23(4), 589–616. <https://doi.org/10.1017/S1351324916000395>
- Chiang, E., & Grant, T. (2017). Online grooming: moves and strategies. *Language and Law / Linguagem e Direito*. 4(1), 103-141.
- Chiang, E., & Grant, T. (2018). Deceptive Identity Performance: Offender Moves and Multiple Identities in Online Child Abuse Conversations. *Applied Linguistics*. Vol. 40(4), 675–698. <https://doi.org/10.1093/applin/amy007>
- Cheong, Y.-G., Jensen, A. K., Guðnadóttir, E. R., Bae, B.-C., & Togelius, J. (2013). Detecting Predatory Behaviour in Game Chats. *IEEE Transactions on Computational Intelligence and AI in Games*. 7(3), 220 - 232. <https://doi.org/10.1109/TCIAIG.2015.2424932>
- Craven, S., Brown, S., & Gilchrist, E. (2006). Sexual grooming of children: Review of literature and theoretical considerations. *Journal of Sexual Aggression*. 12, 287-299. <https://doi.org/10.1080/13552600601069414>
- Culpeper, J. (1996). Towards an anatomy of impoliteness. *Journal of Pragmatics*. 25(3), 349–367. [https://doi.org/10.1016/0378-2166\(95\)00014-3](https://doi.org/10.1016/0378-2166(95)00014-3)
- Culpeper, J. (2005). Impoliteness and entertainment in the television quiz show: the weakest link. *Journal of Politeness Research*. 1(1), 35–72. <https://doi.org/10.1515/jplr.2005.1.1.35>
- Culpeper, J. (2011). *Impoliteness: using language to cause offence*. (Studies in interactional sociolinguistics). Cambridge University Press.
- Culpeper, J., & Tantucci, V. (2021). The Principle of (Im)politeness Reciprocity. *Journal of Pragmatics*, 175, 146–164. <https://doi.org/10.1016/j.pragma.2021.01.008>
- Curl, T.S., & Drew, P. (2008). Contingency and Action: A Comparison of Two Forms of Requesting. *Research on Language and Social Interaction*. 41(2), 129–153. <https://doi.org/10.1080/08351810802028613>
- Davis, N. (2016, September 7). *Online grooming of children often 'alarmingly fast', researchers find*. The Guardian. Retrieved from: <http://www.theguardian.com/society/2016/sep/08/online-grooming-of-children-often-alarmingly-fast-researchers-find> [last accessed June 2021]
- DeHart, D., Dwyer, G., Seto, M.S., Moran, R., Letorneau, E., & Schwatz-Watts, D. (2017). Internet sexual solicitation of children: a proposed typology of offenders based on their chats, emails, and social network posts. *Journal of Sexual Aggression*. 23(10), 77-89. <https://doi.org/10.1080/13552600.2016.1241309>

- del-Teso-Craviotto, M. (2006). Language and sexuality in Spanish and English dating chats. *Journal of Sociolinguistics*, 10(4), 460–480. <https://doi.org/10.1111/j.1467-9841.2006.00288.x>
- del-Teso-Craviotto, M. (2008). Gender and sexual identity authentication in language use: The case of chat rooms. *Discourse Studies*, 10(2), 251–270. <https://doi.org/10.1177/1461445607087011>
- Demjén, Z. (2016). Laughing at cancer: Humour, empowerment, solidarity and coping online. *Journal of Pragmatics*, 101, 18–30. <https://doi.org/10.1016/j.pragma.2016.05.010>
- Dobs, A. M., & Garcés-Conejos Blitvich, P. (2013). Impoliteness in polylogal interaction: Accounting for face-threat witnesses' responses. *Journal of Pragmatics*, 53, 112–130. <https://doi.org/10.1016/j.pragma.2013.05.002>
- Dresner, E., & Herring, S. C. (2010). Functions of the Nonverbal in CMC: Emoticons and Illocutionary Force. *Communication Theory*, 20(4), 471–471. <https://doi.org/10.1111/j.1468-2885.2010.01362.x>
- Drouin, M., Boyd, R., Hancock, J.T., & James, A. (2017). Linguistic analysis of chat transcripts from child predator undercover sex stings. *The Journal of Forensic Psychiatry & Psychology*, 28(4), 437–457. <https://doi.org/10.1080/14789949.2017.1291707>
- Duguid, A. (2010). Investigating anti and Some Reflections on Modern Diachronic Corpus-Assisted Discourse Studies (MD-CADS). *Corpora*, 5(2), 191–220. <https://doi.org/10.3366/cor.2010.0105>
- Dunning, T. (1993). Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*. 19(1), 61–74.
- Durán-Muñoz, I. (2019). Adjectives and their Keyness: a corpus-based analysis of tourism discourse in English. *Corpora*, 14 (3), 351–378. <https://doi.org/10.3366/cor.2019.0178>
- Dynel, M. (2014). Participation framework underlying YouTube interaction. *Journal of Pragmatics*, 73, 37–52. <https://doi.org/10.1016/j.pragma.2014.04.001>
- Eelen, G. (2002). *A critique of Politeness Theories*. Routledge. <https://doi.org/10.4324/9781315760179>
- Egan, V., Hoskinson, J., & Shewan, D. (2011). Perverted Justice: A Content Analysis of the Language used by Offenders Detected Attempting to Solicit Children for Sex. In R. M. Clark (Ed.), *Antisocial Behaviour: Causes, Correlations and Treatments* (pp. 119–133). Nova Science Publishers.
- Elliott, I., & Beech, A. (2009). Understanding online child pornography use: Applying sexual offense theory to internet offenders. *Aggression and Violent Behavior*, 14, 180–193. <https://doi.org/10.1016/j.avb.2009.03.002>

- Everitt, B.S. (2002). *Cluster Analysis* (3rd edition). Edward Arnold.
- Ferrara, K., Brunner, H., & Whittmore, G. (1991). Interactive Written Discourse as an Emergent Register. *Written Communication*, 8(1), 8–34. <https://doi.org/10.1177/0741088391008001002>
- Flöck, I. (2011). Suggestions in British and American English: A Corpus-Linguistic Study. *Bochumer Linguistische Arbeitsberichte*, 3, 67-81.
- Flöck, I., & Geluykens, R. (2015). Speech Acts in Corpus Pragmatics: A Quantitative Contrastive Study of Directives in Spontaneous and Elicited Discourse. In J. Romero-Trillo (Ed.), *Yearbook of Corpus Linguistics and Pragmatics 2015: Current Approaches to Discourse and Translation Studies* (pp. 7–37). Springer International Publishing. https://doi.org/10.1007/978-3-319-17948-3_2
- Franklin, E. (2015). Some Theoretical Considerations in Off-the-Shelf Text Analysis Software. *Proceedings of the Student Research Workshop associated with RANLP 2015*. 8-15. Retrieved from: <http://aclweb.org/anthology/R15-2002>
- Frediani, C. (2019). A pragmatic reversal: Italian per favore ‘please’ and its variants between politeness and impoliteness. *Journal of Pragmatics*, 142, 233-244. <https://doi.org/10.1016/j.pragma.2018.09.008>
- Gabrielatos, C., & Baker, P. (2008). Fleeing, Sneaking, Flooding A Corpus Analysis of Discursive Constructions of Refugees and Asylum Seekers in the UK Press, 1996-2005. *Journal of English Linguistics*, 36(1), 5-38. <https://doi.org/10.1177/0075424207311247>
- Gabrielatos, C. & Marchi, A. (2011, 5 November). *Keyness: Matching metrics to definitions* [Conference presentation]. Corpus Linguistics in the South, University of Portsmouth, United Kingdom. Retrieved from: <http://eprints.lancs.ac.uk/51449> [last accessed June 2021]
- Gabrielatos, C. (2018). Keyness analysis. In C. Taylor & Marchi, A. (Eds.). *Corpus Approaches to Discourse: A critical review* (pp. 225-258). London: Routledge.
- Gagne, C. (2018). Indirectness and entitlement in product requests in British service encounters. *Journal of Pragmatics*, 133, 1-14. <https://doi.org/10.1016/j.pragma.2018.05.015>
- Garcés-Conejos Blitvich, P., & Bou-Franch, P. (2019). *Introduction to Introduction to Analyzing Digital Discourse: New Insights and Future Directions*. In P. Garcés-Conejos Blitvich, & P. Bou-Franch (Eds.), *Analyzing Digital Discourse: New Insights and Future Directions* (pp. 3–22). https://doi.org/10.1007/978-3-319-92663-6_1
- Garcia McAllister, P. (2014). Speech acts: A synchronic perspective. In K. Aijmer & C. Rühlemann (Eds.), *Corpus Pragmatics: A Handbook* (pp. 29-51). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139057493.003>

- Garrido Ardila, J. A. (2019). Impoliteness as a rhetorical strategy in Spain's politics. *Journal of Pragmatics*, 140, 160–170. <https://doi.org/10.1016/j.pragma.2018.11.017>
- Garrison, A., Remley, D., Thomas, P., & Wierszewski, E. (2011). Conventional Faces: Emoticons in Instant Messaging Discourse. *Computers and Composition*, 28(2), 112–125. <https://doi.org/10.1016/j.compcom.2011.04.001>
- Gauz, M.J. (2014). 'This Would Be Much Funner in Person' A Qualitative Grounded Theory Analysis of Cybergrooming. [Master Thesis, Portland State University]. PDX Scholar. Retrieved from: http://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?article=2803&context=open_access_etds [last accessed June 2021]
- Georgakopoulou, A. (2017). 'Whose context collapse?': Ethical clashes in the study of language and social media in context. *Applied Linguistics Review*, 8(2–3), 169–189. <https://doi.org/10.1515/applirev-2016-1034>
- Gillespie, A. (2004). "Grooming": Definitions and the law. *New Law Journal*, 154, Article 71240, 586–587.
- Goffman, E. (1967). *Interactional Ritual: Essays on Face-to-Face Behaviour*. Doubleday.
- Goffman, E. (1974). *Frame Analysis: An Essay on the Organisation of Experience*. Harvard University Press.
- Grant, T., & McLeod, N. (2020). *Language and Online Identities: The Undercover Policing of Internet Sexual Crime*. Cambridge University Press.
- Guice, K. (2016). *Predators, decoys, and teens: A corpus analysis of online language* [Unpublished Master thesis]. Hofstra University.
- Gupta, A., Kumaraguru, P., & Sureka, A. (2012). Characterizing Pedophile Conversations on the Internet using Online Grooming. ArXiv e-prints. Retrieved from: <https://arxiv.org/pdf/1208.4324.pdf> [last accessed June 2021]
- Gonçalves, P., Araújo, M. Benevenuto, F. & Cha, M. (2013). Comparing and Combining Sentiment Analysis Methods. *Proceedings of COSN, the first ACM conference on Online social networks*, 27-38. Retrieved from: <https://dl.acm.org/citation.cfm?id=2512951> (Last accessed June 2021) <https://doi.org/10.1145/2512938.2512951>
- Hall, R.C.W., Hall, R.C.W. (2007). A profile of pedophilia: Definition, characteristics of offenders, recidivism, treatment outcomes, and forensic issues. *Mayo Clinic Proceedings*, 82(4), 457-471. <https://doi.org/10.4065/82.4.457>
- Hamilton-Giachritsis, C., Hanson, E., Whittle, H., & Beech, A. (2017). "Everyone deserves to be happy and safe": A mixed methods study exploring how online

- and offline child sexual abuse impact young people and how professionals respond to it. *Safer Children York*. Retrieved from: <https://www.saferchildrenyork.org.uk/Downloads/impact-online-offline-child-sexual-abuse.pdf> [last accessed June 2021]
- Hardie, A. (2012). CQPweb - combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17(3), 380–409. <https://doi.org/10.1075/ijcl.17.3.04har>
- Hardie, A. (2014). Log ratio – An informal introduction. *Centre for Corpus Approaches to Social Sciences (CASS)*. Retrieved from: <http://cass.lancs.ac.uk/?p=1133> [last accessed June 2021]
- Hardt-Mautner, G. (1995). ‘Only Connect.’ Critical Discourse Analysis and Corpus Linguistics. *UCREL Technical Papers*. Retrieved from: <http://ucrel.lancaster.ac.uk/papers/techpaper/vol6.pdf> [last accessed June 2021]
- Haugh, M., & Watanabe, Y. (2017). (Im)politeness Theory. In B. Vine (Ed.), *The Routledge Handbook of Language in the Workplace* (pp. 65-76). Routledge.
- Haugh, M., & Culpeper, J. (2018). Integrative pragmatics and (im)politeness theory. In C. Ilie, & N. R. Norrick (Eds.), *Pragmatics and its Interfaces* (pp. 213–239). John Benjamins. <https://doi.org/10.1075/pbns.294.10hau>
- Herring, S. (2003). Computer-mediated discourse. In D. Schiffrin, D. Tannen, & H.E. Hamilton. (Eds.), *The Handbook of Discourse Analysis* (pp. 612-634). John Wiley & Sons, Incorporated.
- Herring, S. C. (2004). Computer-Mediated Discourse Analysis: An Approach to Researching Online Behavior. In S. Barab, R. Kling, & J. H. Gray (Eds.), *Designing for Virtual Communities in the Service of Learning* (1st ed., pp. 338–376). Cambridge University Press. <https://doi.org/10.1017/CBO9780511805080.016>
- Herring, S. & Androutsopoulos, J. (2015). Computer-Mediated Discourse 2.0. In D. Tannen, H. E. Hamilton, & D. Schiffrin (Eds.), *The Handbook of Discourse Analysis* (2nd ed., pp. 127-151). John Wiley & Sons, Incorporated.
- Hofland, K., & Johansson, S. (1982). *Word frequencies in British and American English*. Norwegian Computing Centre for the Humanities.
- Home Office. (2018, September 2). Tackling child sexual exploitation online. *GOV.UK*. Retrieved from: <https://www.gov.uk/government/news/tackling-child-sexual-exploitation-online> [last accessed June 2021]
- Howitt, D. (1995). *Paedophiles and sexual offences against children*. John Wiley and Sons.

- Inches, G., & Crestani, F. (2012). Overview of the International Sexual Predator Identification Competition at PAN-2012. *CLEF 2012*. Retrieved from: https://pan.webis.de/downloads/publications/papers/inches_2012.pdf [last accessed June 2021]
- ITV. (2017, April 3). *NSPCC: 325% rise in 'child grooming' in North East*. ITV News. Retrieved from: <https://www.itv.com/news/tyne-tees/2017-04-03/charity-reports-massive-rise-in-child-grooming-in-the-north-east> [last accessed June 2021]
- Jaworska, S., & Nanda, A. (2016). Doing Well by Talking Good: A Topic Modelling-Assisted Discourse Study of Corporate Social Responsibility. *Applied Linguistics*, 39(3), 373–399. <https://doi.org/10.1093/applin/amw014>
- John, P., Brooks, B., & Schriever, U. (2019). Speech acts in professional maritime discourse: A pragmatic risk analysis of bridge team communication directives and commissives in full-mission simulation. *Journal of Pragmatics*, 140, 12–21. <https://doi.org/10.1016/j.pragma.2018.11.013>
- Jones, R. H., Chik, A., & Hafner, C. A. (2015). *Discourse and Digital Practices: Doing discourse analysis in the digital age*. Routledge.
- Jucker, A. H., Schneider, G., Taavitsainen, I., & Breustedt, B. (2008). Fishing for compliments: Precision and recall in corpus-linguistic compliment research. In A. H. Jucker & I. Taavitsainen (Eds.), *Pragmatics & Beyond New Series* (Vol. 176, pp. 273–294). John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.176.15juc>
- Katz, C. (2013). Internet-related child sexual abuse: What children tell us in their testimonies. *Children and Youth Services Review*, 35, 1536–1542. <https://doi.org/10.1016/j.childyouth.2013.06.006>
- Khazraie, M., & Talebzadeh, H. (2020). “Wikipedia does NOT tolerate your babbling!”: Impoliteness-induced conflict (resolution) in a polylogal collaborative online community of practice. *Journal of Pragmatics*, 163, 46–65. <https://doi.org/10.1016/j.pragma.2020.03.009>
- Kettleborough, D. G., & Merdian, H. L. (2017). Gateway to offending behaviour: permission-giving thoughts of online users of child sexual exploitation material. *Journal of Sexual Aggression*, 23(1), 19–32. <https://doi.org/10.1080/13552600.2016.1231852>
- Kilgarriff, A. (2009). Simple match for keywords. In M. Mahlberg, V. González-Díaz, & C. Smith (Eds.), *Proceedings of the Corpus Linguistics Conference, CL2009*. Liverpool: University of Liverpool. Retrieved from: http://ucrel.lancs.ac.uk/publications/CL2009/171_FullPaper.doc [last accessed June 2021]

- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovár, V., Michelfeit, J., Rychlý, P., Suchomel, V., (2014). The Sketch Engine: ten years on. *Lexicography*, 1(1), 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Kim, J., Kim, J. Behzadi, M., & Harris. I. G. (2020, 11-16 May). Analysis of Online Conversations to Detect Cyberpredators Using Recurrent Neural Networks. *Proceedings of the Workshop on Social Threats in Online Conversations: Understanding and Management (STOC-2020)*, 15–20. Retrieved from: <https://www.aclweb.org/anthology/2020.stoc-1.3.pdf> [last accessed June 2021]
- Kinzel, A. (2016, June 6-10). From Keywords to Discursive Legitimation: Representing 'kuffar' in the Jihadist Propaganda Magazines. *Proceedings of Corpus Linguistics Fest 2016*, 1607, 26-33. Retrieved from: <http://ceur-ws.org/Vol-1607/kinzel.pdf> [last accessed June 2021]
- Kloess, J. A., Beech, A. R., & Harkins, L. (2014). Online Child Sexual Exploitation: Prevalence, Process, and Offender Characteristics. *Trauma, Violence, & Abuse*, 15(2), 126–139. <https://doi.org/10.1177/1524838013511543>
- Kloess, J.A., Hamilton-Giachritsis, C.E. & Beech, A.R. (2017a). Offense Processes of Online Sexual Grooming and Abuse of Children Via Internet Communication Platforms. *Sexual Abuse*, 1-24. <https://doi.org/10.1177/1079063217720927>
- Kloess, J.A., Hamilton-Giachritsis, C.E. & Beech, A. R. (2017b). A descriptive account of victims' behaviour and responses in sexually exploitative interactions with offenders. *Psychology, Crime & Law*, 23(7), 621-632. <https://doi.org/10.1080/1068316X.2017.1293052>
- Knoll, J. (2010). Teacher sexual misconduct: Grooming patterns and female offenders. *Journal of Child Sexual Abuse*, 19, 371–386. <https://doi.org/10.1080/10538712.2010.495047>
- Komar, S. (2015). Linguistic Features of Persuasive Communication: The Case of DRTV Short Form Spots. *ELOPE: English Language Overseas Perspectives and Enquiries*, 12(2), 29–52. <https://doi.org/10.4312/elope.12.2.29-52>
- Kontostathis, A., Edwards, L., & Leatherman, A. (2009, April 30-May 2). *ChatCoder: Toward the Tracking and Categorization of Internet Predators* [Conference workshop]. Ninth SIAM International Conference on Data Mining, Nevada, United States. Retrieved from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.151.6501&rep=rep1&type=pdf> [last accessed June 2021]
- Kozłowska, H. (2019, July 24). *There's a global movement of Facebook vigilantes who hunt pedophiles*. Quartz. Retrieved from: <https://qz.com/1671916/the-global-movement-of-facebook-vigilantes-who-hunt-pedophiles/> [last accessed June 2021]

- Krone, T. (2004). A typology of online child pornography offending. *Trends & Issues in Crime and Criminal Justice*, 279, 1–6.
- Leberg, E. (1997). *Understanding child molesters: Taking charge*. Sage.
- Lanning, K. (2001). Child molesters: A behavioral analysis for professionals investigating the sexual exploitation of children. *National Center for Missing & Exploited Children*. Retrieved from: <https://www.icmec.org/wp-content/uploads/2015/10/US-NCMEC-Child-Molesters-A-Behavioral-Analysis-Lanning-2010.pdf> [last accessed June 2021]
- Locher, M., & Watts, R. (2005). Politeness theory and relational work. *Journal of Politeness Research*, 1(1), 9–33. <https://doi.org/10.1515/jplr.2005.1.1.9>
- Locher, M. A. (2008). Relational work, politeness and identity construction. In G. Antos, E. Ventola, & T. Weber (Eds.), *Handbooks of Applied Linguistics. Volume 2: Interpersonal Communication* (pp. 509–540). Mouton de Gruyter. <https://doi.org/10.1515/9783110211399.4.509>
- Lorenzo-Dus, N., & Bou-Franch, P. (2013). A cross-cultural investigation of email communication in Peninsular Spanish and British English: The role of (in)formality and (in)directness. *Pragmatics and Society*, 4(1), 1–25. <https://doi.org/10.1075/ps.4.1.01lor>
- Lorenzo-Dus, N., Izura, C., & Pérez-Tattam, R. (2016). Understanding grooming discourse in computer-mediated environments. *Discourse, Context and Media*, 12, 40-50. <https://doi.org/10.1016/j.dcm.2016.02.004>
- Lorenzo-Dus, N., & Di Cristofaro, M. (2016). # Living/minimum wage: Influential citizen talk in twitter. *Discourse, Context & Media*, 13, 40–50. <https://doi.org/10.1016/j.dcm.2016.02.003>
- Lorenzo-Dus, N., & Izura, C. (2017). “cause ur special”: Understanding trust and complimenting behaviour in online grooming discourse. *Journal of Pragmatics*, 112, 68-82. <https://doi.org/10.1016/j.pragma.2017.01.004>
- Lorenzo-Dus, N., Walker, L., & Kinzel, A. (2017). The Role of Discourse Analysis in Terrorism Studies. In M. Conway, L. Jarvis, O. Lehane, S. Macdonald, & L. Nouri. (Eds). *Terrorists’ Use of the Internet: Assessment and Response* (pp. 158-169). IOS Press. <https://doi.org/10.3233/978-1-61499-765-8-158>
- Lorenzo-Dus, N., Kinzel, A., & Walker, L. (2018). Representing the West and “non-believers” in the online jihadist magazines Dabiq and Inspire. *Critical Studies on Terrorism*, 11(3), 521-536, <https://doi.org/10.1080/17539153.2018.1471081>
- Lorenzo-Dus, N., & Kinzel, A. (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(1), 1-30. <https://doi.org/10.18573/jcads.31>

- Lorenzo-Dus, N., Kinzel, A. & Di Cristofaro, M. (2020). The Communicative Modus Operandi of Online Child Sexual Groomers: Recurring Patterns in their Language Use. *Journal of Pragmatics*, 155, 15-27. <https://doi.org/10.1016/j.pragma.2019.09.010>
- 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. In M. Fuster-Márquez, J. Santaemilia, C. Gregori-Signes, & P. Rodríguez-Abruñeiras (Eds), *Exploring Discourse and Ideology through Corpora* (pp. 189-210). Peter Lang.
- Ly, T., Murphy, L., & Fedoroff, J.P. (2016). Understanding Online Child Sexual Exploitation Offenses. *Current Psychiatry Reports*, 18(8), 74-75. <https://doi.org/10.1007/s11920-016-0707-0>
- Maillat, D., & Oswald, S. (2009). Defining Manipulative Discourse: The Pragmatics of Cognitive Illusions. *International Review of Pragmatics*, 1(2), 348–370. <https://doi.org/10.1163/187730909X12535267111651>
- Mao, R.L. (1993). Beyond politeness theory: ‘Face’ revisited and renewed. *Journal of Pragmatics*, 21, 451-486. [https://doi.org/10.1016/0378-2166\(94\)90025-6](https://doi.org/10.1016/0378-2166(94)90025-6)
- Marchenko, S. (2017, November 1). *Web of Darkness: Groomed, Manipulated, Coerced, and Abused In Minutes*. Biometrica. Retrieved from: <https://www.biometrica.com/icmec-online-grooming/> [last accessed June 2021]
- Marwick, A. E., & boyd., d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1), 114–133. <https://doi.org/10.1177/1461444810365313>
- Mautner, G. (2019). A research note on corpora and discourse: Points to ponder in research design. *Journal of Corpora and Discourse Studies*, 2, 1-13. <https://doi.org/10.18573/jcads.32>
- McEnery, T., McGlashan, M., & Love, R. (2015). Press and social media reaction to ideologically inspired murder: The case of Lee Rigby. *Discourse & Communication*, 9(2), 237–259. <https://doi.org/10.1177/1750481314568545>
- McGhee, I., Bayzick, J., Kontostathis, A., Edwards, L., McBride, A., & Jakubowski, E. (2011). Learning to identify internet sexual predation. *International Journal on Electronic Commerce*, 15(3), 103-123. <https://doi.org/10.2753/JEC1086-4415150305>
- McLeod, N., & Grant T. (2017). “go on cam but dnt be dirty”: linguistic levels of identity assumption in undercover online operations against child sex abusers. *Language and Law/Linguagem e Direito*, 4(2), 157-175.

- Merdian, H., Curtis, C., Thakker, J., Wilson & Douglas, N., & Boer, P. (2013). The three dimensions of online child pornography offending. *Journal of Sexual Aggression*, 19(1), 121-132. <https://doi.org/10.1080/13552600.2011.611898>
- Merdian, H., Moghaddam, N., Boer, D., Wilson, N., Thakker, J., Curtis, C. & Dawson, D. (2018). Fantasy-Driven Versus Contact-Driven Users of Child Sexual Exploitation Material: Offender Classification and Implications for Their Risk Assessment. *Sexual Abuse: Journal of Research and Treatment*, 30(3), 230–253. <https://doi.org/10.1177/1079063216641109>
- Merrison, A. J., Wilson, J.J., Davies, B. L., & Haugh, M. (2012). Getting stuff done: Comparing e-mail requests from students in higher education in Britain and Australia. *Journal of Pragmatics*, 44, 1077-1098. <https://doi.org/10.1016/j.pragma.2012.05.009>
- Montgomery, M. (2006). Broadcast news, the live ‘two-way’ and the case of Andrew Gilligan. *Media, Culture & Society*, 28(2), 233-259. <https://doi.org/10.1177/0163443706061683>
- Mortensen, K. K. (2017). Flirting in online dating: Giving empirical grounds to flirtatious implicitness. *Discourse Studies*, 19(5), 581–597. <https://doi.org/10.1177/1461445617715179>
- Nelson, P. (2021, June 2). Deep Dive: FBI estimates 500,000 online predators are a daily threat to kids going online. (2021, June 2). KOAA. Retrieved from: <https://www.koaa.com/news/deep-dive/fbi-estimates-500-000-online-predators-are-a-daily-threat-to-kids-going-online> [last accessed November 2021]
- NSPCC. (2018, November 30). *Protecting children from grooming*. NSPCC Learning. Retrieved from: <https://learning.nspcc.org.uk/research-resources/briefings/grooming/> [last accessed June 2021].
- NSPCC. (2020a, September 3). *Police record over 10,000 online child sex crimes in a year for the first time*. NSPCC. Retrieved from: <https://www.nspcc.org.uk/about-us/news-opinion/2020/2020-09-03-cybercrimes-during-lockdown/> [last accessed June 2021]
- NSPCC (2020b, November 13). *Instagram most recorded platform used in child grooming crimes during lockdown*. NSPCC. Retrieved from: <https://www.nspcc.org.uk/about-us/news-opinion/2020/instagram-grooming-crimes-children-lockdown/> [last accessed June 2021]
- O'Connell, R. (2003). A typology of child cyberexploitation 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> [last accessed June 2021]

- O'Connor, C., & Joffe, H. (2020). Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *International Journal of Qualitative Methods*, 19, 1-13. <https://doi.org/10.1177/1609406919899220>
- Olson, L. N., Daggs, J. L., Ellevold, B. L., & Rogers, T. K. K. (2007). Entrapping the Innocent: Toward a Theory of Child Sexual Predators? Luring Communication. *Communication Theory*, 17(3), 231–251. <https://doi.org/10.1111/j.1468-2885.2007.00294.x>
- Partington A. (2004). Introduction: Corpora and discourse, a most congruous beast. In: A. Partington, J. Morley, & L. Haarman (Eds.), *Corpora and Discourse* (pp. 11-20). Peter Lang.
- Partington, A. (2008). The armchair and the machine: Corpus-Assisted Discourse Studies. In C.T. Torsello, K. Ackerley, & E. Castello (Eds.), *Corpora for university language teachers*. Peter Lang.
- Partington, A. (2010). Modern Diachronic Corpus-Assisted Discourse Studies (MDCADS) on UK newspapers: An overview of the project. *Corpora*, 5(2), 83-108. <https://doi.org/10.3366/cor.2010.0101>
- Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and Meanings in Discourse: Theory and Practice in Corpus-assisted Discourse Studies (CADS)*. John Benjamins.
- Pendar, N. (2007, 17-19 September). Toward Spotting the Pedophile Telling victim from predator in text chats. *International Conference on Semantic Computing (ICSC 2007)*, 235 – 241. <https://doi.org/10.1109/ICSC.2007.32>
- Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015a). The development and psychometric properties of LIWC2015. Austin: University of Texas at Austin, Texas. Retrieved from: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf [last accessed June 2021]
- Pennebaker, J.W., Booth, R.J., Boyd, R.L., & Francis, M. E. (2015b). Linguistic Inquiry and Word Count: LIWC2015: Operator's Manual. Austin: University of Texas at Austin, Texas. Retrieved from: https://s3-us-west-2.amazonaws.com/downloads.liwc.net/LIWC2015_OperatorManual.pdf [last accessed June 2021]
- Perelmutter, R. (2018). Globalization, conflict discourse, and Jewish identity in an Israeli Russian-speaking online community. *Journal of Pragmatics*, 134, 134–148. <https://doi.org/10.1016/j.pragma.2018.03.019>
- Perverted Justice. (2019). *Ceasing decoy operations, Deck convicted again*. perverted-justice.com. Retrieved from: <https://web.archive.org/web/20190102172915/http://www.perverted-justice.com/> [last accessed June 2021]

- Pranoto, H. (2015). Logistic Models for Classifying Online Grooming Conversation. *Procedia Computer Science*, 59, 357-365. <https://doi.org/10.1016/j.procs.2015.07.536>
- Quayle, E., & Newman, E. (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. <https://doi.org/10.1186/s40163-016-0050-0>
- Rayson, P., D. Archer, & N. Smith. (2005). "Vard versus word: A comparison of the ucrel variant detector and modern spell checkers on english historical corpora". *Proceedings from the Corpus Linguistics Conference Series*. 1(1), ISSN 1747-939
- Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4), 519-549. <https://doi.org/10.1075/ijcl.13.4.06ray>
- Rayson, P. (2015). Computational tools and methods for corpus compilation and analysis. In D. Biber & R. Reppen (Eds.), *The Cambridge Handbook of English Corpus Linguistics* (pp. 32–49). Cambridge University Press. <https://doi.org/10.1017/CBO9781139764377.003>
- Rideout, V., Foehr, U.G., & Roberts, D.F. (2010). Generation M²: Media in the Lives of 8- to 18-Year-Olds. *KFF - Henry Kaiser Family Foundation*. Retrieved from: <https://www.kff.org/other/event/generation-m2-media-in-the-lives-of/> [last accessed June 2021]
- Rychlý, P. (2008). A lexicographer-friendly association score. In P. Sojka, & A. Horák (Eds), *Proceedings of Recent Advances in Slavonic Natural Language Processing RASLAN* (pp. 6-9). Masaryk University.
- Salkin, A. (2006, December 13). *Web Site Hunts Pedophiles, and TV Goes Along*. New York Times. Retrieved from: <https://www.nytimes.com/2006/12/13/technology/13justice.html> [last accessed June 2021]
- Salter, A. (1995). *Transforming trauma: A guide to understanding and treating adult survivors of child sexual abuse*. Sage. <https://doi.org/10.4135/9781452232102>
- Savić, M. (2018). Lecturer perceptions of im/politeness and in/appropriateness in student e-mail requests: A Norwegian perspective. *Journal of Pragmatics*, 124, 52-72. <https://doi.org/10.1016/j.pragma.2017.12.005>
- Schiffrin, D., Tannen, D., & Hamilton, H. E. (2015). Introduction to the First Edition. In D. Tannen, H. E. Hamilton, & D. Schiffrin (Eds.), *The Handbook of Discourse Analysis* (pp. 1-7). John Wiley & Sons, Incorporated.
- 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 & Grant (2017, 2018). *Language and Law / Linguagem e Direito*, 5(1), 97-102.

- Scott, M. (1997). PC analysis of key words – and key key words. *System*, 25(2), 233-245. [https://doi.org/10.1016/S0346-251X\(97\)00011-0](https://doi.org/10.1016/S0346-251X(97)00011-0)
- Scott, M. (2020). WordSmith Tools version 8 [Computer Software], Stroud: Lexical Analysis Software.
- Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139173438>
- Sgroi, S. M. (1982). *Handbook of clinical intervention in child sexual abuse*. Lexington Books.
- Smith, C., Adolphs, S., Harvey, K., & Mullany, L. (2014). Spelling errors and keywords in born-digital data: A case study using the Teenage Health Freak Corpus. *Corpora*, 9(2), 137–154. <https://doi.org/10.3366/cor.2014.0055>
- Spencer-Oatey, H. (2005). (Im)Politeness, Face and Perceptions of Rapport: Unpackaging their Bases and Interrelationships. *Journal of Politeness Research Language Behaviour Culture*, 1(1), 95-119. <https://doi.org/10.1515/jplr.2005.1.1.95>
- Spencer-Oatey, H. (2007) Theories of identity and the analysis of face. *Journal of Pragmatics*, 39(4), 639-656. <https://doi.org/10.1016/j.pragma.2006.12.004>
- Spencer-Oatey, H. (2008). *Culturally speaking: Culture, communication and politeness theory* (2nd edition). Continuum.
- Stop TIME Online. (2017). *An anti-online grooming activity pack*. NSPCC Learning. Retrieved from: <https://learning.nspcc.org.uk/research-resources/2017/stop-time-online/> [last accessed June 2021]
- Swales, J. 1981. *Aspects of Article Introductions: Aspects of Article Introductions: Issue 1 of Aston ESP research reports*. Language Studies Unit.
- Szabla, M., & Blommaert, J. (2020). Does context really collapse in social media interaction? *Applied Linguistics Review*, 11(2), 251–279. <https://doi.org/10.1515/applirev-2017-0119>
- Tagg, C., Baron, A., & Rayson, P. (2012). I didnt spel that wrong did i. Oops: Analysis and normalisation of SMS spelling variation. *Lingvisticae Investigationes*, 35(2), 367-388. <https://doi.org/10.1075/li.35.2.12tag>
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1), 24–54. <https://doi.org/10.1177/0261927X09351676>
- Taylor, C. (2015). Beyond sarcasm: The metalanguage and structures of mock politeness. *Journal of Pragmatics*, 87, 127-141. <https://doi.org/10.1016/j.pragma.2015.08.005>

- Thompson, D., & Filik, R. (2016). Sarcasm in Written Communication: Emoticons are Efficient Markers of Intention. *Journal of Computer-Mediated Communication*, 21(2), 105–120. <https://doi.org/10.1111/jcc4.12156>
- Thurlow, C. (2018). Digital Discourse: Locating Language in New/Social Media. In J. Burgess, A. Marwick, & T. Poell (Eds.), *The SAGE Handbook of Social Media* (pp. 135–145). SAGE Publications Ltd. <https://doi.org/10.4135/9781473984066.n8>
- UNICEF. (2012). Child Safety Online: Global challenges and strategies. *UNICEF*. Retrieved from: https://www.unicef-irc.org/publications/pdf/ict_techreport3_eng.pdf [last accessed June 2021]
- Vandergriff, I. (2013). Emotive communication online: A contextual analysis of computer-mediated communication (CMC) cues. *Journal of Pragmatics*, 51, 1–12. <https://doi.org/10.1016/j.pragma.2013.02.008>
- van Dijk, T. (2006). Discourse and manipulation. *Discourse & Society*, 17(3), 359–383. <https://doi.org/10.1177/0957926506060250>
- van Gijn-Grosvenor, E. L., & Lamb, M.E. (2015). Behavioural Differences Between Online Sexual Groomers Approaching Boys and Girls. *Journal of Child Sexual Abuse*, 25(5), 577–596. <https://doi.org/10.1080/10538712.2016.1189473>
- Van Olmen, D. (2019). A three-fold approach to the imperative's usage in English and Dutch. *Journal of Pragmatics*, 139, 146-162. <https://doi.org/10.1016/j.pragma.2018.11.006>
- Ventura, M. (2012). The Detection of Faking Through Word Use. In M. Ziegler, C. McCann, & R. Roberts (Eds.), *New Perspectives on Faking in Personality Assessment* (pp. 165-173). Oxford University Press.
- Wang, N., Kosinski, M., Stillwell, D. J., & Rust, J. (2014). Can Well-Being be Measured Using Facebook Status Updates? Validation of Facebook's Gross National Happiness Index. *Social Indicators Research*, 115(1), 483–491. <https://doi.org/10.1007/s11205-012-9996-9>
- Watts, R. J. (2010). Linguistic politeness theory and its aftermath: Recent research trails. In M. A. Locher, & S. L. Graham (Eds.), *Interpersonal Pragmatics* (pp. 43-70). De Gruyter Mouton.
- Webster, S., Davidson, J., Bifulc, A., Gottschalk, P., Caretti, V., Pham, T., Grove-Hills, C. T., Tompkins, C., Ciulla, S., Milazzo, V., Schimmenti, A., & Craparo, G. (2012). *European online grooming project final report*. NatCen. Retrieved from: <https://natcen.ac.uk/media/22514/european-online-grooming-projectfinalreport.pdf> [last accessed June 2021]
- Webman Shafran, R. (2019). Level of directness and the use of please in requests in English by native speakers of Arabic and Hebrew. *Journal of Pragmatics*, 148, 1–11. <https://doi.org/10.1016/j.pragma.2019.05.020>

- Whittle, H., Hamilton-Giachritsis, C.E., & Beech, A. (2014a). In Their Own Words: Young Peoples' Vulnerabilities to Being Groomed and Sexually Abused Online. *Psychology*, 5, 1185-1196. <https://doi.org/10.4236/psych.2014.510131>
- Whittle, H., Hamilton-Giachritsis, C.E., & Beech, A. (2014b). "Under His Spell": Victims' Perspectives of being Groomed Online. *Social Sciences*, 3, 404-426. <https://doi.org/10.3390/socsci3030404>
- Whittle, H., Hamilton-Giachritsis, C.E., & Beech, A. (2015). A comparison of victim and offender perspectives of grooming and sexual abuse. *Deviant Behavior*, 36(7), 539-564. <https://doi.org/10.1080/01639625.2014.944074>
- Williams, R., Elliott, I.A., & Beech, A. (2013). Identifying Sexual Grooming Themes Used by Internet Sex Offenders. *Deviant Behavior*, 34(2), 135-152. <https://doi.org/10.1080/01639625.2012.707550>
- Wilson, S.R., Kim, M., & Meischke, H. (1991). Evaluating Brown and Levinson's politeness theory: A revised analysis of directives and face. *Research on Language and Social Interaction*, 25, 215-252. <https://doi.org/10.1080/08351819109389363>
- Winters, G. M., Kaylor, L.E., & Jeglic, E.L. (2017). Sexual offenders contacting children online: an examination of transcripts of sexual grooming. *Journal of Sexual Aggression*, 23(1), 62-76. <https://doi.org/10.1080/13552600.2016.1271146>
- Wolak, J, Finkelhor, D., Mitchell, K.J., & Ybarra, M.L. (2008). Online "Predators" and Their Victims: Myths, Realities, and Implications for Prevention and Treatment. *American Psychologist*, 63(2), 111-128. <https://doi.org/10.1037/0003-066X.63.2.111>
- Wootton, A.J. (2005). Interactional and sequential features informing request format selection in children's speech. In A. Hakulinen & M. Selting (Eds.), *Syntax and lexis in conversation* (pp. 185–207). Benjamins.
- Xiao, R. (2015). Collocation. In D. Biber & R. Reppen (Eds.), *The Cambridge Handbook of English Corpus Linguistics* (pp. 106–124). Cambridge University Press. <https://doi.org/10.1017/CBO9781139764377.007>
- Zhang, G. (2013). The impact of touchy topics on vague language use. *Journal of Asian Pacific Communication*, 23(1), 87–118. <https://doi.org/10.1075/japc.23.1.06zha>
- Zinken, J. (2015). Contingent control over shared goods. 'Can I have x' requests in British English informal interaction. *Journal of Pragmatics*, 82, 23--38. <https://doi.org/10.1016/j.pragma.2015.03.005>

9. Appendix

Table 48: Positive and negative politeness strategies with examples from GN corpus

Positive Politeness	
Notice, attend to H (his interests, wants, needs, goods)	i can call you when i get gome .. no prob
Exaggerate (interest, approval, sympathy with H)	babe can i call u on the phone please ?
Intensify interest to H	is my princess busy
Use in-group identity markers	we have to work this out
Seek agreement	yeah, do you feel the same way
Avoid disagreement	yea i sup you are right , see i told u that u are so mature and concerned
Presuppose/raise/assert common ground	we both will be in lot of trouble if we talk so we wont
Joke	as long as u don't turn me into the cops lol
Assert or presuppose S's knowledge of and concern for H's wants	well that 's ok I 'd show you everything
Offer, promise	I wanna make your nite a special one too
Be optimistic	all you have to do is trust in me ... and the rest will come natural
Include both S and H in activity	we 'd be lovers
Give (or ask for) reason	why is that
Assume or assert reciprocity	cause we love each other
Give gifts to H (goods, sympathy, understanding, cooperation)	u are so sweet
Negative Politeness	
Be conventionally indirect	u did good the other night in asking about some stuff ok
Question, hedge	Well maybe if i come back i could bring u with me for like a week
Be pessimistic	(No examples found in corpus)
Minimise the imposition	never mind dumb ques lol
Give deference	(No examples found in corpus)
Apologize	sorry i havent been around lately i hadda go outta town for awhile
Impersonalise S and H	(No examples found in corpus)

State the FTA as a general rule	(No examples found in corpus)
Nominalise	actions speak louder than words
Go on record as incurring a debt, or as not indebting to H	how bout i owe you lunch and dinner tomorrow instead

Table 49: Impoliteness strategies with examples from GN corpus

On Record Impoliteness	
Positive Impoliteness	
Ignore/Snub the other	[groomer ignores target's message and instructs them to delete chat archive]
Exclude other from activity	(No examples found in corpus)
Disassociate from the other	i never came out and straight up said anything as hurtful as that
Be disinterested, unconcerned, unsympathetic	I mean, I need some way of calling you once I hit Jackson
Use inappropriate identity markers	(No examples found in corpus)
Use obscure secretive language	(No examples found in corpus)
Make the other feel uncomfortable	no , I am not just going to bend you over and ram it in I will be fingering you while I am licking and nibbling on your clit
Seek disagreement	I KNOW you are wrong
Use taboo words	or are you afraid she 's going to see you rubbing your little pussy
Call the other names	you are such a little slut now
Negative impoliteness	
Frighten	scared I your mom might find out I have it
Condescend, scorn, ridicule	orig: ahhh haha afraid your mommie is going to see naughty pictures ?
Invade the other's space	you know im gon na contact yahoo security and tell them your a minor and they will

	make sure you wont be able to get in there chatrooms
Explicitly associate the other with a negative aspect	i don't need shit like this in my life
Put the other's indebtedness on record	if you want me to trust you you have to trust me right
Hinder or block the other, either physically or linguistically	[groomer sends 12 messages without letting target send reply]
Off-record impoliteness	
Sarcasm	well enjoy it while you can
Withhold politeness (No examples found in corpus)	

Table 50: Keywords of GN compared with PAN2012

<p>Table 50 removed due to confidentiality reasons.</p>

Continued: Table 50 removed due to confidentiality reasons.

Continued: Table 50 removed due to confidentiality reasons.

Continued: Table 50 removed due to confidentiality reasons.

Continued: Table 50 removed due to confidentiality reasons.

Table 51: Keywords of G3 and G4

Table 51 removed due to confidentiality reasons

Continued: Table 51 removed due to confidentiality reasons

Continued: Table 51 removed due to confidentiality reasons

Continued: Table 51 removed due to confidentiality reasons

Table 52: All keywords of G5 and G6

Table 52 removed due to confidentiality reasons.

Continued: Table 52 removed due to confidentiality reasons.

Table 53: Full list of newly identified keywords in the first empirical chapter

Table 53 removed due to confidentiality reasons.

Continued: Table 53 removed due to confidentiality reasons.

Continued: Table 53 removed due to confidentiality reasons.

Continued: Table 53 removed due to confidentiality reasons.

Table 54: Keywords belonging to more than one online grooming intention

Table 54 removed due to confidentiality reasons.

Continued: Table 54 removed due to confidentiality reasons.

Continued: Table 54 removed due to confidentiality reasons.

Table 55: Unique keywords associated with the two grooming duration approaches

Table 55 removed due to confidentiality reasons.

Continued: Table 55 removed due to confidentiality reasons.

Continued: Table 55 removed due to confidentiality reasons.

Table 56: Keywords with more than one grooming intention

Table 56 removed due to confidentiality reasons

Continued: Table 56 removed due to confidentiality reasons.

10. Glossary

Corpus-Assisted Discourse Studies

A Linguistics approach or methodology combining quantitative and qualitative methods drawn from Corpus Linguistics and Discourse Analysis.

Corpus Linguistics

A Linguistics approach or methodology able to analyse large collections of language data (corpora) based on statistical measures.

Computer-Mediated Discourse Analysis

Discourse Analysis that analyses language through computer-mediated means (e.g. email, chat logs, social media posts). Also a framework by Herring (2003, 2004).

Discourse Analysis

A Linguistics approach informed by other academic disciplines analysing language in use.

Keyword in Context (KWIC) analysis

A term from Corpus Linguistics: Keyword in Context (KWIC) is a single instance of the word or string of words that is being investigated in its context, normally 10 to 20 words on either side of the word or string of words under investigation.

Log Likelihood

A term from Corpus Linguistics: A keyword measure that calculates the statistical significance of word frequencies compared to word frequencies in a reference corpus, measuring whether the difference is based on chance or a statistical difference is present.

Log ratio

A term from Corpus Linguistics: The binary logarithm of the ratio of normalised frequencies in a study corpus compared to a reference corpus. A value of 0 indicates a comparable frequency in both corpora, while a value of 1 indicates that the item is twice as frequent in one corpus. A negative value indicates the item is more frequent in the reference corpus than the study corpus.

Normalised Frequency

A term from Corpus Linguistics: The normalised frequency is a frequency of a word in a corpus per million words for easier comparison across corpora.

Online Grooming

A process by which an adult befriends a child using the internet with the ultimate goal of sexually abusing them online or “offline”.

Speech Act Theory

A term from Pragmatics, originally from Philosophy: Speech Act Theory is the theory that specific sentences perform actions. Speech acts come in three

forms: locutionary act, illocutionary act or perlocutionary act. More recently they have been described as context-dependent and in other categories, e.g. request, apology, command.