

ASTON UNIVERSITY

DOCTORAL THESIS

**Username Construction and Identity
Performance in Dark Web Child Sexual Abuse
Communications**

Daniela Schneevogt

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Applied Linguistics (Forensic Linguistics)

at the

Aston Institute for Forensic Linguistics

School of Social Sciences and Humanities

Aston University

April 2023

©Daniela Schneevogt, 2023

Daniela Schneevogt asserts her moral right to be identified as the author of this thesis.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright belongs to its author and that no quotation from the thesis and no information derived from it may be published without appropriate permission or acknowledgement.

Abstract

Child sex abusers use the dark web to exchange illicit imagery and share advice. As part of their identity work, users can make use of a number of resources – linguistic, social, cultural, situational – to form their online identity and create a unique username.

Employing corpus-assisted discourse analysis with a focus on social actor representation, this thesis investigates child sex abusers naming practices and uncovers the resources child sex abusers draw on when crafting their usernames and how those relate to users' discursively performed identities. The data analysed stems from a corpus of eight dark web discussion fora related to child sexual abuse consisting of 295,799 forum posts written by 32,410 unique users.

In study 1, the usernames are manually segmented and are then classified thematically into categories. 20 username categories emerged which are typical for users' naming choices in the dark web CSA context. The analysis shows that users draw on a number of resources to create their usernames and specific considerations regarding anonymity impact those choices. Interestingly, users in this high risk/low trust environment still opt for proper names as parts of their usernames.

For study 2, forum posts for the most frequent posters in the data (n=1837) are analysed through a corpus-assisted approach examining the usernames and forum contributions for a subset of 5 username categories. Findings suggest that users who identify in a certain way will also show linguistic behaviours related to that self-assigned identity and engage in discussions of topics which can be inferred from the username choice.

Study 3 dives deeper into the 5 selected subcorpora and applies a thematic as well as a detailed concordance analysis for top keywords identified against the full corpus of most frequent posters. The results show that while similarities between the subgroups are present, differences between the subgroups show clear distinctive features by type of user. Users engage in discussions of topics directly or indirectly linked to their self-assigned identities and different types of users are shown to be more or less concerned with topics such as age of consent, security measures or the exchange of CSAM.

In addition to adding to the linguistic literature on identity performance, child sexual abuse and offender communities, this research opens up possibilities for law enforcement policing such online environments: findings can aid agencies in identifying targets and aid undercover officers in choosing appropriate usernames when engaging in infiltration tasks.

Keywords: child sexual abuse, dark web, usernames, identity, corpus-assisted discourse analysis, social actors, online onomastics

Acknowledgements

It takes a village. This is not only true for raising children but also for completing your PhD research and dissertation. Some might even consider their final thesis to be their baby. My PhD village started out small when I joined my PhD programme in 2018, and consisted mainly of family and friends from 'before' and a handful of lovely humans – both fellow students and lecturers – whom I've had the pleasure of meeting during my Masters in Forensic Linguistics at Aston the year prior to my PhD start date. Over the last six years, this village has grown into a very special community of continuous support – both personally and academically.

I want to thank my main supervisor Tim Grant who has supported me throughout this journey, who went above and beyond to help me carry out the research I wanted to do, helping me jump many hurdles along the way. Thanks for your patience, Tim!

I want to thank my associate supervisors Nicci MacLeod and Gertrud Reershemius, as well as my prior associate supervisor Erika Darics, for all their valuable input over the years.

I want to thank Amanda Towler and Mark Haase of Hyperion Gray, LCC for their support with the data collection.

I want to thank my colleague Emily Chiang, who's work sparked my interest in Forensic Linguistics, for welcoming me so warmly to the community and supporting me throughout the years. You continue to be an inspiration, Emily!

I want to thank Tahmineh Tayebi, Ria Perkins, Geoff Morrison and Kzysztof Kredens for their invaluable input in the earlier stages of my project, from proposal development all the way through to my QR examination.

I want to thank my examiners Nuria Lorenzo-Dus and Felicity Deamer for their time, their comments and their genuine interest in my work, and our chair Graeme Hayes for making this a much less daunting experience than it could have been.

I want to thank my fellow Forensic Linguistics PhDs, past and present, at AIFL and beyond, for their company and support over those many years. Amy, Annie, Dana, Helena, Holly, Jordan, Julija, Lauren, Madi, Natascha, Neus, Solly – thank you for your friendship and for making this experience a much less lonely one!

I want to thank all my colleagues at AIFL, especially the FoLD team of Tim, Lucia, Sarah and Marton, for creating such a supportive work environment and sense of community which has always included PhD researchers.

I want to thank Aston's Postgraduate Research Society (APRS) for providing community across the university and introducing me to invaluable friends. Hassan, Rotimi, Monse – I couldn't imagine my PhD experience without your friendship and support.

I want to thank Eren Bilgen for working so tirelessly for and closely with PhD researchers at Aston. Your writing retreats are invaluable during key points of the PhD process, thank you!

I want to thank Amanda Beattie for her hard work to improve the PhD experience at Aston and for supporting me through my own challenges along the way.

I want to thank Michael for his immense personal and academic support in the early stages of my research.

I want to thank my dear friend Hannah for bringing so much love and joy to my early PhD experience. Your support was everything and I am sad you aren't here now for us to celebrate together.

I want to direct special thanks to my family for their support, understanding and patience throughout the years. Mama, thanks for always making sure I could follow my interests and supporting my choices. And thanks for never stopping to ask what I actually do. Tobias and Christian, thank you for always knowing how to make me laugh. Oma, thank you for all your kind words and handwritten letters. Doris, thank you for believing in me. Opa, thanks for all your trust and support, I'm sad you didn't get to see me complete this journey.

Some only become family later in life, by chance, by choice, maybe fate. They hold a special place in my heart and deserve a special mention here as well.

Jamie, thank you for all your love and support throughout the years, for truly understanding and for always knowing how to cheer me up when things were tough.

Ina, thank you for being my best friend, my confidante, the sister I never had. Your encouraging words and loving gestures have been key to me successfully completing this journey.

Barkley and Holly, my *greytest*, three- and four-legged friends, thank you for providing comfort, companionship and silliness when I needed it most.

And finally, Alex, thanks for being my rock for the past 4 years. Your love, support and understanding as well as your relentless encouragement and curiosity have been crucial to me completing my research. I couldn't have wished for a better partner to cheer me on. Thank you!

Contents

- Abstract** **i**

- Acknowledgements** **ii**

- 1 Introduction** **1**
 - 1.1 Child Sexual Abuse Communications Online 2
 - 1.2 Aims and Research Questions 3
 - 1.3 Originality and Contributions 5
 - 1.4 Thesis Structure 6

- 2 Theoretical Grounding and Orientation to Previous Research** **8**
 - 2.1 Identity Performance Online 8
 - 2.2 Group Identity Construction Online 13
 - 2.3 Online Discourse and Discussion Fora 14
 - 2.4 Online Onomastics and Username Research 16
 - 2.4.1 Usernames and Identity 17
 - 2.4.2 Usernames in Criminal Contexts and Communities 21
 - 2.5 Privacy, Anonymity and the Dark Web 22
 - 2.6 Criminal and Harmful Online Communities and Communications 23
 - 2.6.1 Extremist and Terrorist Discourse Online 24
 - 2.6.2 Profit-driven Crimes 27
 - 2.6.3 The Manosphere 28

2.6.4	Sex Work	30
2.6.5	Online Child Sexual Abuse Communications	31
2.7	Criminal Communities as Communities of Practice	40
2.8	The Role of Linguistic Analysis in Policing Online Crime	43
3	Methodology	46
3.1	Procedure	46
3.1.1	Data Sets: Collection & Preparation	47
3.1.1.1	The Full Corpus: DarkWebCSA	49
3.1.1.2	Analysing Most Frequent Posters: the DarkWebCSA-top corpus	51
3.1.1.3	Significant Username Subgroups: Subcopora for Studies 2 and 3	53
3.2	Methodological Approach	56
3.2.1	Critical Discourse Analysis	56
3.2.2	Corpus-Assisted Discourse Analysis	59
3.3	Linguistic Frameworks	60
3.3.1	Thematic Analysis	60
3.3.2	Social Actor Framework and Transitivity	62
3.3.2.1	Analysing Social Actors	63
3.3.2.2	Transitivity Analysis	66
3.3.2.3	Difficulties with Social Actor Analysis	68
3.4	Limitations	71
3.5	Ethical Considerations	74
4	Study 1: Username Categories in Child Sexual Abuse Fora	81
4.1	The Data Set: CSA Fora and Usernames	82
4.2	Analysis and Findings	83
4.2.1	Categorisation and Segmentation of Usernames	83

4.2.1.1	Data Exploration: Initial Observations	83
4.2.1.2	Data Preparation: Segmentation of Usernames	85
4.2.1.3	Segmentation Results	87
4.2.1.4	Annotation: Username Classification	93
4.2.2	Naming Themes and Patterns	93
4.2.2.1	Emerging Username Categories	95
	Age	95
	Animals	96
	Brands	97
	Childlove	98
	Colours	99
	Dark/Evil	100
	Dark Web	100
	Explicit Anonymity	101
	Family Relations	102
	Film, TV and Literature	103
	Food & Drinks	104
	Gender	104
	Names	105
	Nationality, Ethnicity and Location	108
	Occupation	108
	Paedophile	109
	Preferences	110
	Religion	110
	Sexual Orientation	111

Violence	112
4.2.2.2 Graphology, Typography and Multimodality	112
Punctuation and word boundaries	112
Numbers as placeholders	113
4.2.2.3 Other notable username features	114
Phrases	114
Random and keyboard sequences	115
Creativity and word play	116
4.2.3 Frequent vs Infrequent Posters	116
4.2.4 Reliability and Generalisability	118
4.3 Discussion of Findings	119
5 Study 2: Usernames and Identity Performance in Child Sexual Abuse Fora	125
5.1 The Data Set: Username Groups Among Frequent Posters	125
5.2 Analysis and Findings	126
5.2.1 Corpus tool and corpus analysis steps	126
5.2.2 Names	129
5.2.2.1 Age (of consent)	130
5.2.2.2 Criminality and imprisonment	132
5.2.2.3 Names and community	133
5.2.3 Family Relations	134
5.2.3.1 Family and incest	134
5.2.3.2 Pronouns	136
5.2.3.3 Proper names	137
5.2.3.4 Self-identification as paedophile	137
5.2.4 Childlove	138

5.2.4.1	Exchange of CSAM and security measures	138
5.2.4.2	The child-adult 'relationship'	138
5.2.4.3	Proper names	140
5.2.5	Violence	140
5.2.5.1	Violent acts and references to violence and body parts	141
5.2.5.2	Exchange of CSAM	142
5.2.5.3	Politeness features	142
5.2.6	Paedophile	144
5.2.6.1	Exchange of CSAM and security measures	144
5.2.6.2	Paedophilia and Paedophiles	145
5.2.6.3	Law Enforcement and the Law	146
5.3	Discussion of Findings	147
6	Study 3: Differences and Similarities between Types of CSA Fora Users	150
6.1	The Data Set: Selected Username Groups	150
6.2	Analysis and Findings	151
6.2.1	Top themes and keywords shared across subcorpora	152
6.2.2	Use of proper names across user groups	155
6.2.2.1	Child Victims and Their Families	158
6.2.2.2	Stories of Abuse: Fictional Characters	161
6.2.2.3	References to Community, Users and User's Families	162
6.2.3	Exchange of CSAM and Security	162
6.2.4	Forum Rules and Community Conventions	165
6.2.5	Family Dynamics	167
6.2.6	Suspected Abusers' Preferences and Victim Descriptions	174
6.2.7	Age	176

6.2.8	Pronouns	179
6.2.9	The Human Body	180
6.2.10	Explicit Mentions of Extreme Violence and Torture	182
6.2.11	The Paedophile Identity	183
6.2.12	The Child-Adult Relationship	186
6.3	Discussion of Findings	187
7	Discussion	193
7.1	Summary of Findings	194
7.2	Discussion: Usernames and Identity Performance in CSA-related Dark Web Commu- nities	196
7.3	Contributions	203
7.3.1	Contributions to Linguistic Research into CSA	203
7.3.2	Contributions to Onomastics Research on Usernames	204
7.3.3	Contributions to the Role of Linguistics in Policing Online Crime	204
7.4	Application of Findings	206
7.5	Future Research	207
7.6	Conclusion	208
7.7	Final remarks	209
A	Ethics Approval Aston University Ethics Committee	211
B	Study 2: Concordance Tables	214
C	Study 3: Thematic annotations and concordance samples	275
D	Full Data Set: DarkWebCSA	277
E	Full list of 32,410 usernames and respective post counts	278

F Python scripts for data extraction	279
G Full list of 10,222 usernames in reduced set and respective username segments	280
H Username Lists per Subcorpus	281
I Top keywords and bigrams by subcorpus	283
J CSA Subcorpora	284
Bibliography	286

List of Figures

- 3.1 Van Leeuwen’s (2013) network of discourse features for the representation of social actors. 64
- 3.2 Simplified list of Van Leeuwen’s (2013) discourse features by Babaii and Rashed (2017). 65

- 4.1 Holistic diagram showing initial observations regarding naming themes and patterns from the full username corpus. 84
- 4.2 Percentage of users in the DarkWebCSA corpus who posted once, 2-5 times, 6-10 times, 11-20 times and over 20 times, respectively. 117
- 4.3 Distribution of number of posts by number of users in the full DarkWeb-CSA corpus. . 118

List of Tables

2.1	Medium factors (Herring, 2007).	14
2.2	Situation factors (Herring, 2007).	15
3.1	Details of the full corpus and analysed subcorpora, including number of usernames and analysis stages.	48
3.2	Full list of fora in DarkWebCSA data set including information on languages present and main discussion topics.	50
3.3	Top segments (frequency \geq 4) in the DarkWebCSA-top username corpus (1837 core users).	52
3.4	Top segment bigrams (frequency \geq 2) in the DarkWebCSA-top username corpus (1837 core users).	53
3.5	List of all category annotations for segments of top 1837 users. Annotation of the data was carried out in NVivo.	54
3.6	Relevant categories for the selection of username groups to be analysed in Chapter 5 and 6 based on previous observations and findings. Those selected for detailed analysis in studies 2 and 3 are highlighted in bold.	55
3.7	Subcorpora by prominent username category (English posts only) used for keyword and concordance analysis with LancsBox. Details for the initial DarkWebCSA-top corpus are included for comparison.	56
3.8	Table introducing the process types available for transitivity analysis based on Halliday and Matthiessen (2013) including subtypes and participant types.	68

4.1	Examples of usernames and their respective username segments to illustrate the manual segmentation process.	87
4.2	Top segments (frequency \geq 10) in the 10,222 username corpus.	88
4.3	Top segment bigrams (frequency \geq 3) in the 10,222 username corpus.	89
4.4	Top segments and segment pairs in the reduced corpus referring to gender.	90
4.5	Top segments and segment pairs referring to the child victim's age.	91
4.6	Key segments and Segment pairs showing self-identification of the user as paedophile.	91
4.7	Key segments and segment pairs identifying the user as "childlover".	92
4.8	Key segments and segment pairs referring to sex and violence.	92
4.9	List of all username categories identified in the full corpus with examples for each category.	94
4.10	Examples of word boundary features and segments used by members in CSA-related dark web fora.	113
4.11	Functions of types of username segments by username category according to the inventory of social actor representation by Van Leeuwen (2013).	120
5.1	Description of subcorpora analysed in study 2.	126
5.2	Reference corpora for study 2.	126
5.3	Reference corpora used for keyword analysis with LancsBox.	129
5.4	Top 50 keywords for the Names subcorpus against all 5 reference corpora. (Cells stating <i>REDACTED</i> included keywords revealing specific usernames or forum names and have therefore been redacted from the public version of this thesis but can be found in the examiner appendices on the encrypted USB drive.)	131

5.5	Top 50 keywords for the Family Relations subcorpus against all 5 reference corpora. (Cells stating <i>REDACTED</i> included keywords revealing specific usernames or forum names and have therefore been redacted from the public version of this thesis but can be found in the examiner appendices on the encrypted USB drive.)	135
5.6	Top 50 keywords for the Childlove subcorpus against all 5 reference corpora. (Cells stating <i>REDACTED</i> included keywords revealing specific usernames or forum names and have therefore been redacted from the public version of this thesis but can be found in the examiner appendices on the encrypted USB drive.)	139
5.7	Top 50 keywords for the Violence subcorpus against all 5 reference corpora. (Cells stating <i>REDACTED</i> included keywords revealing specific usernames or forum names and have therefore been redacted from the public version of this thesis but can be found in the examiner appendices on the encrypted USB drive.)	141
5.8	Top 50 keywords for the Paedophile subcorpus against all 5 reference corpora. (Cells stating <i>REDACTED</i> included keywords revealing specific usernames or forum names and have therefore been redacted from the public version of this thesis but can be found in the examiner appendices on the encrypted USB drive.)	145
6.1	Description of subcorpora and DarkWebCSA-top corpus as reference corpus used in study 3.	150
6.2	Alphabetised lists of top 50 keywords for all five subcorpora with keywords highlighted that occur in two or more subcorpora among the top 50 list.	153
6.3	Comparison of themes derived from the top 50 keywords across all five subcorpora showing commonalities and differences between the corpora. The darker the yellow, the more subcorpora shared a theme. (*totals include overlapping thematic categories)	155

6.4	Keywords per subcorpus identified as proper names together with their absolute frequency and type of proper name. (Cells stating <i>REDACTED</i> included keywords revealing specific usernames or forum names and have therefore been redacted from the public version of this thesis but can be found in the examiner appendices on the encrypted USB drive.)	156
6.5	Number of top keywords per type of proper names by subcorpus	157
6.6	Top keywords annotated as CSAM theme by subcorpus	163
6.7	Top keywords annotated as Security theme by subcorpus	164
6.8	Top keywords annotated as LEA/crime theme by subcorpus	164
6.9	Top keywords annotated as Community theme by subcorpus.	165
6.10	Number of top keywords consisting of members' usernames per subcorpus	167
6.11	Top keywords annotated as Family & Incest theme by subcorpus.	168
6.12	Top keywords annotated as <i>Preference</i> theme by subcorpus.	175
6.13	Top keywords annotated as <i>Age</i> theme by subcorpus.	176
6.14	Top keywords annotated as <i>Pronouns</i> theme by subcorpus.	179
6.15	Top keywords annotated as <i>Body</i> theme by subcorpus.	180
6.16	Top keywords from the <i>Violence</i> subcorpus annotated as <i>Violence</i> theme.	182
6.17	Top keywords annotated as <i>Paedophile</i> theme by subcorpus.	183
6.18	Top keywords annotated as <i>Ontopic</i> theme by subcorpus.	185
6.19	Top keywords annotated as <i>Sex</i> theme by subcorpus.	185
6.20	Top keywords annotated as <i>Relationship</i> theme by subcorpus.	186
B.1	Random concordance lines for keyword <i>14</i> in Names subcorpus.	215
B.2	Random concordance lines for keyword <i>15</i> in Names subcorpus.	215
B.3	Random concordance lines for keyword <i>16</i> in Names subcorpus.	216
B.4	Random concordance lines for keyword <i>17</i> in Names subcorpus.	217

B.5	Random concordance lines for keyword <i>18</i> in Names subcorpus.	218
B.6	Random concordance lines for keyword <i>years</i> in Names subcorpus.	219
B.7	Random concordance lines for keyword <i>yr</i> in Names subcorpus.	220
B.8	Random concordance lines for keyword <i>age</i> in Names subcorpus.	221
B.9	Random concordance lines for keyword <i>consent</i> in Names subcorpus.	222
B.10	Random concordance lines for keyword <i>criminal</i> in Names subcorpus.	223
B.11	Random concordance lines for keyword <i>criminal</i> in Names subcorpus.	224
B.12	Random concordance lines for keyword <i>criminal</i> in Names subcorpus.	225
B.13	Random concordance lines for keyword referring to specific community name in Names subcorpus. The community name has been replaced with a placeholder here for security reasons.	226
B.14	Random concordance lines for keyword <i>daughter</i> in Family Relations subcorpus.	227
B.15	Random concordance lines for keyword <i>wife</i> in Family Relations subcorpus.	228
B.16	Random concordance lines for keyword <i>niece</i> in Family Relations subcorpus.	229
B.17	Random concordance lines for keyword <i>mom</i> in Family Relations subcorpus.	230
B.18	Random concordance lines for keyword <i>kids</i> in Family Relations subcorpus.	231
B.19	Random concordance lines for keyword <i>kids</i> in Family Relations subcorpus.	232
B.20	Random concordance lines for keyword <i>daughters</i> in Family Relations subcorpus.	233
B.21	Random concordance lines for keyword <i>uncle</i> in Family Relations subcorpus.	234
B.22	Random concordance lines for keyword <i>husband</i> in Family Relations subcorpus.	235
B.23	Random concordance lines for keyword <i>child</i> in Family Relations subcorpus.	236
B.24	Random concordance lines for keyword <i>parents</i> in Family Relations subcorpus.	237
B.25	Random concordance lines for keyword <i>family</i> in Family Relations subcorpus.	238
B.26	Random concordance lines for keyword <i>incest</i> in Family Relations subcorpus.	239
B.27	Random concordance lines for keyword <i>pedo</i> in Family Relations subcorpus.	240

B.28 Random concordance lines for keyword <i>pedos</i> in Family Relations subcorpus.	241
B.29 Random concordance lines for keyword <i>pedophile</i> in Family Relations subcorpus.	242
B.30 Random concordance lines for keyword <i>pedophile</i> in Family Relations subcorpus.	243
B.31 Random concordance lines for keyword <i>her</i> in Family Relations subcorpus.	244
B.32 Random concordance lines for keyword <i>my</i> in Family Relations subcorpus.	245
B.33 Random concordance lines for keyword <i>Ashley</i> in Family Relations subcorpus.	246
B.34 Random concordance lines for keyword <i>Chloe</i> in Family Relations subcorpus.	247
B.35 Random concordance lines for keyword <i>Liz</i> in Family Relations subcorpus.	248
B.36 Random concordance lines for keyword <i>Claire</i> in Family Relations subcorpus.	249
B.37 Random concordance lines for keyword <i>Kylie</i> in Family Relations subcorpus.	250
B.38 Random concordance lines for keyword <i>tor</i> in Childlove subcorpus.	251
B.39 Random concordance lines for keyword <i>data</i> in Childlove subcorpus.	251
B.40 Random concordance lines for keyword <i>encryption</i> in Childlove subcorpus.	252
B.41 Random concordance lines for keyword <i>security</i> in Childlove subcorpus.	252
B.42 Random concordance lines for keyword <i>relation</i> in Childlove subcorpus.	253
B.43 Random concordance lines for keyword <i>relation</i> in Childlove subcorpus.	254
B.44 Random concordance lines for keyword <i>hurtcore</i> in Violence subcorpus.	255
B.45 Random concordance lines for keyword <i>rape</i> in Violence subcorpus.	255
B.46 Random concordance lines for keyword <i>hurt</i> in Violence subcorpus.	256
B.47 Random concordance lines for keyword <i>blood</i> in Violence subcorpus.	256
B.48 Random concordance lines for keyword <i>cock</i> in Violence subcorpus.	257
B.49 Random concordance lines for keyword <i>mouth</i> in Violence subcorpus.	258
B.50 Random concordance lines for keyword <i>ass</i> in Violence subcorpus.	259
B.51 Random concordance lines for keyword <i>thanks</i> in Violence subcorpus.	259
B.52 Random concordance lines for keyword <i>please</i> in Violence subcorpus.	260

B.53 Random concordance lines for keyword <i>ontopic</i> in Paedophile subcorpus.	261
B.54 Random concordance lines for keyword <i>tor</i> in Paedophile subcorpus.	262
B.55 Random concordance lines for keyword <i>tor</i> in Paedophile subcorpus.	263
B.56 Random concordance lines for keyword <i>hidden</i> in Paedophile subcorpus.	264
B.57 Random concordance lines for keyword <i>anonymous</i> in Paedophile subcorpus.	265
B.58 Random concordance lines for keyword <i>pedo</i> in Paedophile subcorpus.	266
B.59 Random concordance lines for keyword <i>pedophile</i> in Paedophile subcorpus.	267
B.60 Random concordance lines for keyword <i>pedophiles</i> in Paedophile subcorpus.	268
B.61 Random concordance lines for keyword <i>pedos</i> in Paedophile subcorpus.	269
B.62 Random concordance lines for keyword <i>attracted</i> in Paedophile subcorpus.	270
B.63 Random concordance lines for keyword <i>LEA</i> in Paedophile subcorpus.	271
B.64 Random concordance lines for keyword <i>cops</i> in Paedophile subcorpus.	272
B.65 Random concordance lines for keyword <i>prison</i> in Paedophile subcorpus.	273
B.66 Random concordance lines for keyword <i>jail</i> in Paedophile subcorpus.	274

List of Abbreviations

AIFL	A ston I nstitute for F orensic L inguistics
DA	D iscourse A nalysis
CADS	C orpus- A ssisted D iscourse S tudies
CDA	C ritical D iscourse A nalysis
CL	C orpus L inguistics
CoP	C ommunity of P ractice
CMC	C omputer- M ediated C ommunication
CSA	C hild S exual A buse
CSAM	C hild S exual A buse M aterial
CSEA	C hild S exual E xploitation and A buse
IIOC	I ndecent I mages O f C hildren
KWIC	K ey W ord I n C ontext
MRA(s)	M en's R ights A ctivist(s)
NLP	N atural L anguage P rocessing
OCSA	O nline C hild S exual A buse
OCSEA	O nline C hild S exual E xploitation and A buse
POS	P art- O f- S peech
RQ	R esearch Q uestion
SA	S ocial A ctor

Chapter 1

Introduction

Recently, interest in analysing online criminal communities related to child sexual abuse has grown. The publicly unacceptable nature of paedophilic tendencies and the criminality of sexual relations with children leads to individuals seeking highly anonymous online settings to get together to receive support, give advice and share child abuse material with one another. For child sex abusers, the dark web is particularly attractive as it offers the highest level of anonymity and users often remain unidentifiable by law enforcement.

In recent years, few linguistic studies have attempted to describe the language of those individuals. Difficulties regarding access to these types of data sets are largely to blame. The author of this thesis has been fortunate enough to gain access to genuine offender-to-offender data from a number of dark web fora related to child sexual abuse and this work presents one of the first studies investigating language in these communities. More importantly, this work is the first of its kind examining suspected child sex abusers naming practices in dark web forum communications and provides insights into the relationship between online usernames and online identity performance.

Considering previous work on online criminal communities, identity studies as well as onomastic research into usernames, this work introduces three consecutive studies examining child sex abusers naming choices and identity performance on various levels, employing a corpus-assisted discourse analysis approach and making use of Van Leeuwen's well-established framework for the representation of social actors.

1.1 Child Sexual Abuse Communications Online

Child sexual abuse has been the focus of psychological and criminological research over the last decades. Recently, online child sexual abuse (OCSA) has received a lot of attention from the academic community, especially in psychology and more recently also from (computational) linguistics. The focus has thereby been on the study of online sexual grooming, with a number of studies being contributed by linguistics research (Chiang and Grant, 2017; Chiang and Grant, 2018; Chiang, 2019; Lorenzo-Dus, Izura, and Pérez-Tattam, 2016).

Yet, another important type of online interaction in this context which facilitates child sexual abuse, both online and offline, has so far largely been backgrounded: offender-to-offender communications. Online fora and groups where suspected offenders discuss child sexual abuse, give each other advice and support, and engage in the trading of illicit images of children (IIOC) are on the rise and are increasingly popular with offenders due to the high levels of privacy and anonymity these spaces provide.

The policing of such dark web communities comes with a number of new challenges. Undercover officers engaging in infiltration tasks have to perform with authenticity when portraying members of the community in-group. As communication and relationship building in these spaces occurs almost exclusively through language, linguistic analysis can offer useful insights to law enforcement.

Compared to online sexual grooming interactions, dark web communications between suspected offenders of child sexual abuse, however, have seen far less interest from the academic research community. This is largely due to data from this context not being readily available to researchers. The linguistic studies so far have employed a number of methods such as move analysis and Markov chains to investigate these online communities of offenders (Chiang, 2019; Chiang et al., 2021), and have shown first insights into how offenders build relationships in these online environments and how identities are performed.

Another aspect of identity are members' usernames. Usernames are a key element in constructing an online identity and building communities. Yet, so far, few studies exist which employ linguistic methods to study these phenomena. Studies investigating usernames in criminal online communities generally are still incredibly rare and none have investigated communities in the CSA context.

This work provides the first attempt at studying online naming practices of suspected child sexual abusers on the dark web through a corpus-assisted discourse analytic approach. The data for this work comprises 295,799 forum posts – including 32,410 usernames – from eight different dark web fora related to child sexual abuse (CSA) posted between 2014 and 2019. A selection of these usernames undergo a segmentation process to identify key segments and segment pairs present in the data, to then use a corpus-assisted approach to username and social actor analysis to investigate further the themes and patterns emerging from the data. Findings are linked to identity and onomastics literature and practical outcomes for undercover policing are considered.

1.2 Aims and Research Questions

Child sexual abuse is a persistent issue and a growing societal problem. Online child sexual abuse (OCSA) has increased drastically over the years, with a 762 per cent increase in OCSA reports to the Internet Watch Foundation over the last 10 years – from 41,877 in 2011 to 361,062 in 2021¹. This fact alone gives the first clear justification for continued research into OCSA, and this thesis aims to add to the literature by investigating yet another angle of online child sexual abuse interactions to widen understanding of this complex issue.

Another clear aim of this work is to progress linguistic research in this area. While there have been numerous studies investigating online sexual grooming from a linguistic perspective, studies analysing online communities of suspected child sexual abusers have only just started to appear

¹Source: <https://www.iwf.org.uk/news-media/news/iwf-of-prime-importance-in-fight-against-growing-threat-of-online-child-sexual-abuse/>

on the linguistic research landscape (see e.g. Chiang et al., 2021). And while others have researched usernames in online fora, and some even in criminal contexts (see e.g. Hämäläinen, 2015; Hämäläinen, Haasio, and Harviainen, 2021), to the best of my knowledge, this thesis constitutes the first linguistic study of usernames in the CSEA context. This research therefore aims to increase understanding of linguistic identity performance in CSA offender-to-offender interactions.

Many studies analysing group identity employing discourse analytic approaches focus on only a small number of texts since it is time- and labour-intensive. Many computational approaches to forensic text analysis, especially when moving away from purely lexical or grammatical structures, are – while covering large amounts of data – rarely grounded in linguistic theories and often fail to explain why and how the system got to the final output. Combining a discourse analytic approach with corpus linguistics, the aim is to analyse discursive structures through detailed manual analysis but for a much larger data set than would be possible with traditional manual discourse analysis.

By employing social actor analysis following an extensive segmentation and categorisation process (further outlined in chapter 3), this corpus-assisted study aims to achieve the following three research goals:

1. Increase the understanding of dark web child sexual abuse interactions and suspected offenders' identity performance in these online spaces
2. Gain insights into suspected child sex abusers' naming practices to increase understanding and to aid law enforcement in online policing of such communities
3. Explore the usefulness of a corpus-assisted approach and the application of Van Leeuwen's (2013) social actor framework for studying usernames

To reach these research goals, a detailed analysis of the extracted usernames as well as corresponding forum posts has been conducted to answer the following research questions:

1. What types of usernames are present in the corpus and what kinds of resources do child sex abusers draw on when crafting these usernames? (Study 1)
2. How do users represent themselves in terms of social actor representation through their choice of username? (Study 1 & 2)
3. How are usernames linked to users' posted content and their identity performance in the context of child abuse interactions on the dark web? (Study 2 & 3)

Rather than identifying all identity parts or aspects an individual user brings with them into the interaction, this work looks at which parts of child sexual abusers identities are significant to them and, more importantly, significant in gaining acceptance into the online community.

1.3 Originality and Contributions

Through a combined approach which employs both discourse analytic frameworks and corpus linguistics, this work offers a number of original contributions to research in the following areas:

First, it adds to existing OCSA research by being one of the first linguistic studies examining authentic data of offender-to-offender communications within the context of child sexual abuse fora on the dark web. Second, it adds to the existing literature on identity construction and performance, especially through the use of usernames in online discourse. This further contributes to the existing literature on online onomastics and username research, which so far has received little attention when concerned with criminal communities. And third, this work shows how an established framework like Van Leeuwen's (2013) inventory for Social Actor Representation can usefully be applied to online contexts and the study of usernames and online naming practices.

This work thereby aims to increase the understanding of linguistic identity performance in offender-to-offender communications in the CSEA context, especially in direct connection to users' naming choices in these online spaces. By doing so, this work hopes to benefit both future research and

education into OCSA and criminal online communities, as well as policing of online crime and dark web communities of suspected child sexual abusers, including the identification of individuals within those groups and the infiltration of such communities by undercover operatives.

1.4 Thesis Structure

Chapter 2 introduces the theoretical grounding for this work and discusses general findings regarding identity research, online communications, identity construction online, onomastics and username research followed by an introduction on issues regarding dark web research and dark web communities, including privacy and anonymity. It then presents findings from other research investigating harmful and criminal online communities and communications and concludes with a discussion of the role of linguistics in policing online crime.

Chapter 3 outlines the methodological approach for the work in the empirical chapters and discusses corpus-assisted discourse analysis as the ideal approach for the research context of this thesis. It then details the linguistic frameworks employed for the analysis with a strong focus on social actor analysis and its difficulties arising from applying it to online data.

Chapter 4 then presents the first part of the analysis which explores username categories in the studied child sexual abuse online fora. It briefly discusses the context of the data and describes the data set. This is followed by a description of the initial data exploration, data preparation and annotation process for segmenting and categorising the usernames followed by a detailed discussion of the findings.

Chapter 5 presents finding from the corpus-assisted analysis on identity performance. First, concepts and previous research on identity work and identity performance are discussed in the context of online communications. Then, the data set is presented and findings for different user groups compared to 5 reference corpora are presented and discussed followed by a detailed comparison of

the group-based findings. The chapter concludes with a discussion of the findings in relation to the criminal context.

Chapter 6 looks further into the 5 studied subgroups from chapter 5 and compares top keywords for the subcorpora against the CSA-top corpus. Then, results from a thematic analysis of the keywords are combined with a detailed concordance analysis of said keywords and findings are discussed regarding similarities and differences between the subgroups, also in terms of social actor representation.

Chapter 7 summarises and discusses the results presented in the previous chapters in reference to the posed research questions at the start of the thesis. It further discusses issues regarding reliability and generalisability of the presented findings. This is followed by a discussion of contributions of the research to linguistic and username research, identity theory, and the role of linguistics in policing online crime. Then, the application of findings and implications for future research in this area are discussed. The chapter concludes with a final conclusion on the presented work and some final remarks.

As a final note, due to the sensitivity of the data analysed in this thesis, all appendices except for Appendix A and B are stored on an encrypted, password-protected USB drive and are available to the thesis examiners only. They are not available for public consumption. Selected genuine data samples have been included as example usernames, segments and forum extracts throughout the thesis to underline findings and arguments made throughout the discussion sections.

Chapter 2

Theoretical Grounding and Orientation to Previous Research

2.1 Identity Performance Online

To study identity through naming, an understanding of identity more generally has to be established. This chapter introduces the relevant literature and research background on identity and discusses the linguistic approach to identity taken in this work.

Looking at names and naming, in this case usernames and online naming, as an essential part of identity construction requires the underlying assumption that there is such a thing as identity and that it is important to the human experience and to interpersonal interactions. Identity as a concept has widely been studied in various disciplines – most notably in psychology, sociology and linguistics – and it is therefore important to define what stance this work takes on the concept.

For the current research, a socio-cultural approach to identity is taken following the work of Bucholtz and Hall (2005). According to this approach, identity is understood as something people do through their actions, interactions and the representations they create of themselves and others. Identity can therefore be described as something flexible and context-dependant, rather than stable. Since it is constantly evolving and adapting to the interactional context, identity can never be studied in its completeness and can only ever be investigated in part (Aldrin, 2014).

If identity is performed through people's actions and linguistic choices, naming is evidently also part of such actions contributing to the construction of one's identity. In any given context, there exists a relationship between a name and the self, but also between the name and others – a social aspect of high importance in any community. This is why this work takes an interest in online naming in a particularly closed online community, as individuals' username choices need to be studied and understood in the context of the community.

In the Goffmanian sense, online identity construction can also be seen as a form of face-work acted out by means of a number of different linguistic practices, one of which is the practice of choosing a username (Kersten and Lotze, 2020). In their earlier work, Kersten and Lotze (2020) show that this practice is impacted by users' desire to connect with others in an online community while at the same time marked with a desire for privacy and anonymity. In the context of criminal communications online, this desire becomes a need which is of primary importance to the users in that context (Grant and MacLeod, 2020). This causes a dilemma to users in such spaces as they are trying to achieve opposing goals with their choice of username. Kersten and Lotze (2020) understand this as the problem of "privacy vs. authenticity" (p.125). Following the socio-cultural linguistic approach to identity inspired by Bucholtz and Hall (2005), online naming can therefore be understood as social positioning of the individual in relation to an online community of practice (CoP).

Following the work by Kersten and Lotze (2020), the work presented in this thesis adopts a post-modern view of identity to the analysis of online identity construction, where identity is made up of partial identities which are negotiated in relation to the online community in question. In addition, this work follows the basic principles of linguistic identity construction laid out by Bucholtz and Hall (2005): *emergence, positionality, indexicality, relationality, and partialness*.

Emergence is the first principle and suggests that identity is the result of an interactive negotiation process and something that is continually created, rather than static and internal as in the

traditional view of identity. *Positionality* refers to the principle whereby identity goes beyond just macro-level demographic categories such as age or gender, but further is made up of cultural and temporal positions as well as roles specific to the interaction; the individual uses language to position themselves within the particular discourse.

Indexicality refers to linguistic forms which index certain social meanings and relations. In this work, this is linked to how social actors are represented – including self-representation – where indexing can be achieved through, for example, categorisation (for instance through labelling oneself 'childlover'). Identity construction happens in relation to social groups (the dark web online CoP in this work) and indexical links are drawn from shared norms, values and associations within a community. *Relationality* relates to the fact that identities are not independent but their social meaning is always linked to other identity positions and to other social actors.

Finally, *partialness* refers to the principle of identities being intersubjectively constituted. Identities are formed intentionally or unintentionally, consciously or subconsciously, as a result of negotiation in interaction with other social actors or through perceptions of others and as a product of wider social and cultural structures or ideologies. Identity can thereby never be experienced (or studied) in full but only in part, and interpretation of identity is also only possible in part and as a collaborative effort between different interactants. Names can be seen as one such partial aspect of identity.

This partialness of identity is one of the main underlying assumptions necessary for the justification of the methodological approach taken in this work: the assumption of partialness of identity underlines the need for segmentation of the usernames into their respective parts and the need to annotate those parts individually.

Compared to offline discourse, online discourse such as that explored in the research presented in this thesis differs in that it offers the individual a high level of anonymity which they can use for identity play and deception in a way that would not be possible in offline interactions.

This also raises the question as to what resources are available to an individual to construct and

perform their (online) identity. Grant and MacLeod (2020) argue that individuals are not just able to draw from certain resources creatively and based on context freely, they are also constrained by social, cultural, linguistic and other factors as to which resources they can draw on in any given interactional context. They have developed a new model – the resource-constraint model – to explain the complex relationship between language and identity. The model aims to address certain contradictory observations regarding identity performance which remained from other attempts to explain the phenomenon.

Grant and MacLeod (2020) explain that, on one hand, individuals can perform a wide range of macro and micro-level identities across different contexts, suggesting a high degree of flexibility. However, forensic linguistic studies (e.g. Grant, 2012; Grant, 2022; Nini, 2023) have shown that some elements of an individual's linguistic identity performance may remain constant across situations, indicating that certain aspects of identity are difficult to separate from the individual, even when attempting to disguise their identity, as is the case in the CSA context of the present work.

Additionally, despite the potential for diverse identity performances, individuals may sometimes struggle to perform appropriate identities, especially in unfamiliar contexts. This suggests that identity performance is not entirely without constraints. The resource-constraint model proposes that individuals have a large range of potential identity performance elements at their disposal, which can be understood as resources. According to Grant and MacLeod (2020), these resources are derived from three main sources:

- **Physicality:** An individual's appearance and voice can influence their ability to successfully perform certain identities, such as gender identities, in line with societal norms.
- **Sociolinguistic history:** This includes details such as dialectal varieties of language that individuals are familiar with, as well as the social categories (e.g., gender, race, class) they are perceived as belonging to. While identities are socially constructed, some elements pre-exist specific interactions and serve as acquired resources for performing those identities.

- Interactional context: Individuals have expectations about appropriate performances based on previous experiences in similar contexts. They can also adjust their self-positioning throughout an interaction to ensure successful interactions.

It is important, however, that individuals have finite identity resources, which limits the identities they can perform. They cannot convincingly perform identities for which they lack the relevant resources, whether those are high-level social categories or more localised groups or communities. Individuals can acquire new identity resources over time through sustained effort and interaction within communities of practice – a notion introduced by Wenger (1998) and of high relevance for the current project (and discussed in section 2.7 below). However, this acquisition process is not immediate, but rather a gradual development which evolves through interpersonal interactions over time. Taking all the above details into account, the resource-constraint model accounts for the tension between the flexibility and stability of identity performances by positioning linguistic identity elements as resources that individuals can draw upon, acquire, and are constrained by – essential for understanding identity performances through username choices and forum contributions in online discussion fora, both on the open web and in the dark web context.

Grant and MacLeod's (2020) resource-constraint model is therefore one of the concepts guiding the analysis for the current study when considering the specific communicative context in which members of the studied dark web discussion fora find themselves in.

In this work, the interest lies in examining the identity work and performance of individuals based on the subgroups they belong to. Members of the community who choose to refer to themselves as, for example, paedophiles are regarded as similar and grouped together. This means that rather than looking at the individual alone, factors of group identity come into play. The following section therefore discusses related research on group identity in online contexts to bring together linguistic work on both individual identity performance as well as group identity construction.

2.2 Group Identity Construction Online

Social representations in discourse have been studied in a variety of ways. For example, Koller (2012) used Van Leeuwen's (2013) Social Actor framework as a starting point to study group identity in discourse. In her paper, she proposed a number of linguistic and semiotic features to ascertain which collective identities emerge from discourse, and to analyse the construction of those identities.

Further, in-group out-group relations in discourse were studied in Baxter and Wallace (2009) and Engström and Paradis (2015). The former introduced the concept of "linguistic exclusion" to refer to the methods by which the out-group is constructed linguistically.

Studying online discourses and online communities is generally still a new and under-researched area. A study on interpersonal relations in online discussion fora by Beuchot and Bullen (2005) demonstrates that the cultivation of interpersonality increases participation and expands the depth of discussion. In another study, Arguello et al. (2006) investigated how individuals gain acceptance in online groups. They showed that successful initiation of conversation is dependent on a large number of factors including the degree to which the initiator matches the in-group markers of the target group. The latter seems especially relevant when considering what different roles users might take on as members of a criminal online group and how they manage to identify and construct those in-group markers. Successfully identifying those markers of the target group, the criminal online community, is crucial to successful infiltration of such groups in the context of policing online crime.

A number of research projects have focused on studying member roles and usernames in different online contexts. Computer science and privacy research have addressed this topic and Wang et al. (2016) have used a Natural Language Processing study to identify linguistic descriptors of user roles in online communities, while Maloney-Krichmar and Preece (2002) have investigated membership roles, patterns of interaction, and group dynamics in an online health community. Further studies investigating usernames in online contexts are discussed in the following chapters.

2.3 Online Discourse and Discussion Fora

With computers and mobile phones being used for personal communication, new forms of discourse emerged. Computer-mediated communication (CMC) has been described as the "text-based human-human interaction mediated by networked computers or mobile telephony" (Herring, 2007, p.1). When studying such discourses, researchers faced the following problem: CMC could be studied as one of two basic modalities of language – speech or writing – appearing within one text, or the text could be classified as a new, third type being an intermediate between those two basic types. According to Herring (2007), early efforts by the research community to classify this type of language tended to over-generalise and described computer-mediated discourse (CMD) as if it were a homogeneous genre or a single type of communication. Furthermore, they were not only imprecise in describing such discourses, but lacked an understanding of social aspects and distinctions. For that reason, Herring (2007) introduced a more flexible classification scheme for CMD which classifies text samples in terms of clusters of features. By using such a classification scheme, aspects of both social and technical context influencing the language use in a specific discourse can be combined and adequately articulated. The classification system consists of two basic types of categories: medium factors and situation factors. For both types, a number of facets exist which can have several possible realisations (values). A detailed list of factors and their facets can be found in tables 2.1 and 2.2.

M1	Synchronicity
M2	Message transmission (1-way vs. 2-way)
M3	Persistence of transcript
M4	Size of message buffer
M5	Channels of communication
M6	Anonymous messaging
M7	Private messaging
M8	Filtering
M9	Quoting
M10	Message format

TABLE 2.1: Medium factors (Herring, 2007).

S1	Participation structure	<ul style="list-style-type: none"> • One-to-one, one-to-many, many-to-many • Public/private • Degree of anonymity/pseudonymity • Group size; number of active participants • Amount, rate, and balance of participation
S2	Participant characteristics	<ul style="list-style-type: none"> • Demographics: gender, age, occupation, etc. • Proficiency: with language/computers/CMC • Experience: with addressee/group/topic • Role/status: in "real life"; of online personae • Pre-existing sociocultural knowledge and interactional norms • Attitudes, beliefs, ideologies, and motivations
S3	Purpose	<ul style="list-style-type: none"> • Of group, e.g., professional, social, fantasy/role-playing, aesthetic, experimental • Goal of interaction, e.g., get information, negotiate consensus, develop relationships, impress/entertain others, have fun
S4	Topic or Theme	<ul style="list-style-type: none"> • Of group, e.g., politics, linguistics, feminism, soap operas, sex, science fiction, South Asian culture, medieval times, pub • Of exchanges, e.g., the war in Iraq, the project budget, gay sex, vacation plans, meta-discourse about CMC
S5	Tone	<ul style="list-style-type: none"> • Serious/playful • Formal/casual • Contentious/friendly • Cooperative/sarcastic, etc.
S6	Activity	<ul style="list-style-type: none"> • E.g., debate, job announcement, information exchange, problem solving, exchange of insults, game, flirtation, virtual sex
S7	Norms	<ul style="list-style-type: none"> • Of organization • Of social appropriateness • Of language
S8	Code	<ul style="list-style-type: none"> • Language, language variety • Font/writing system

TABLE 2.2: Situation factors (Herring, 2007).

The classification scheme was developed for research purposes specifically and can help bring all those different aspects of a specific discourse to the attention of the researcher when analysing a text (Herring, 2007). As the scheme allows for adaptation to any newly appearing types of online discourse without needing to rely on a single, pre-defined genre, it serves as a useful tool to provide contextual background for all kinds of discourse analysis focusing on CMC.

For the current work, the above classification scheme was used as a starting point and point of reference throughout the manual analysis stage to keep in mind the specific communication situation and context of the analysed text data.

The use of new media has also led to new social uses of language. Online communities through online discussion fora are a specific form of new media, and based on the type of community, language use can differ greatly and is dependant on the specific online community. Online discussion

fora provide platforms and safe spaces for a variety of communities. On top of general special interest fora, like those for gamers, hobby gardeners or animal lovers, they offer others a space to provide each other with help and support where it might be difficult to access the same level of support in the offline world. This is especially true for marginalised groups in society. Individuals discriminated against on the grounds of their sexual orientation or identity, their ethnicity or other parts of their identity find support in such communities. People suffering from rare diseases may find others in the same position they would not be able to interact with in real life. Special interests or political views might lead others to seek out online discussion groups and in many cases this is beneficial for both individuals and society.

However, these online spaces also provide a perfect environment for all sorts of crime and radicalisation. Criminals benefit from these spaces by gaining access to victims that they can groom online and by sharing advice as well as illicit material with other like-minded people: an ideal setup for scammers, terrorists and child sex offenders. Online discussion fora further provide criminals with a large pool of individuals to recruit for their purposes and to further spread their ideologies.

Dark web communities in particular offer criminals a number of benefits, such as increased anonymity and security, which makes this an attractive space for them to build their communities. A detailed discussion follows in Section 2.5.

2.4 Online Onomastics and Username Research

Usernames as a distinct naming category have been studied by many scholars over the past 30 years, with most works investigating usernames being published in the 2010s (Hämäläinen, 2022). Those studies have been conducted using a vast amount of different methodologies and within a number of different disciplines, ranging from onomastic and linguistic studies through behavioural science experiments to computational studies in data science (Hämäläinen, 2022). Several scholars over the years have provided detailed reviews of the literature (Raátz, 2011; Aleksiejuk, 2016;

Aleksiejuk, 2017; Hämäläinen, 2022), however, none of them offer a full review covering all works to date from different disciplines, theoretical backgrounds, methods and diverse research aims. The most extensive review of username studies in English is provided by Hämäläinen (2022) who has compared 103 studies.

2.4.1 Usernames and Identity

Usernames have been studied by a number of disciplines, including applied linguistics, onomastics, psychology, media studies and computer science. The area of socio-onomastics specifically, is still a very new area of research (see Aleksiejuk, 2016). When it comes to the question of online identity construction and self-naming, onomastics research recognises and highlights the close relationship between names and identity. It considers usernames as a form of face-work (Kersten and Lotze, 2020) where face is negotiated linguistically. Online naming can also be understood as a negotiated social process (Aldrin, 2014).

In their work, Kersten and Lotze, 2020 discuss the interconnectedness of naming, face-work, and stance in relation to individual identity and the relevance of a number of theoretical frameworks for identity construction across disciplines to onomastics, the study of names. In Western philosophy, the individual is seen as an indivisible entity, recognized by a unique name. More recent constructivist approaches shifted focus to interpersonal identity construction, viewing identity as a process rather than a fixed product. Post-modern theories describe identity as a 'patchwork' of partial identities which are all seen as relevant to different aspects of life. This relates to onomastics in that a person can have multiple names for various contexts. According to Kersten and Lotze (2020), Goffman's concept of face-work, where individuals present a polished version of themselves in social interactions, connects to online self-naming practices, which are seen as a form of face-work and identity construction. Bucholtz and Hall's (2005) principles of linguistic identity construction – emergence, positionality, indexicality, relationality, and partialness – are applied to onomastics, particularly in online contexts where usernames can indicate group membership and exclude outsiders.

Finally, Kersten and Lotze (2020) propose four main principles of onomastic identity construction:

- Using names to establish mono-referentiality to a unique referent
- Using names to model human consciousness
- Using names as a device to authenticate oneself as a rational agent
- Using names to position the individual in relation to social groups

They emphasise that name choice is a partial aspect of identity constitution and can be understood as an interactive negotiation process. It involves temporal and spatial positioning relative to groups, indexical references to social groups, and semantic interpretation in relation to certain contexts. They see name choice as a collaborative, partially controllable process and describe their framework as a useful tool for interpreting empirical data on online self-naming practices and understanding how individuals construct and present their identities through naming choices in digital spaces.

Aldrin (2014) argues that personal naming is a process of creating and negotiating social meaning and identity where each phase of naming involves choices that can be seen as social positionings. Consciously or not, individuals take up different social stances when selecting a name. Any name or action during the naming process can potentially be associated with certain social attributes, values, and groups. According to Aldrin (2014), this process of choosing a name involves three interconnected identity levels: micro, local, and macro. These levels influence each other in a flexible, circular process. Each choice made during naming influences subsequent choices and possibilities. In her study, she shows that parents, for example, use tangible resources when positioning themselves through naming. These may include sources of inspiration, traditions, input from others, name criteria, name types, individual names or spellings. The use of these resources potentially positions the parents socially, though the relationship between a resource and its social position is indirect and indexical. Aldrin (2014) further points out that identity construction through

naming is always embedded in context: the immediate local context, larger societal discourses, cultural traditions, and macro-societal structures all influence how identities are created and which social positions are available. This reliance on and restriction by certain resources ties in nicely with Grant and MacLeod's (2020) resource-constraint theory of identity (see detailed discussion in 2.1). Aldrin further argues that naming is never a completely free choice, even in the absence of regulations or laws, and is always embedded in and framed by other practices. Understanding the social meaning of specific naming choices requires rich information on how individuals describe, use, and relate to these choices. This complexity makes it challenging for researchers to interpret the social positioning created through particular naming resources – a challenge also encountered in this current work.

There are broadly two different approaches to analysing usernames and their meaning: the first is to conduct surveys or interviews with the users, to learn from users directly about intended meaning and motivation for their specific naming choices; the second one is to analyse the usernames without reaching out to the users. The latter approach is taken in the current work. This introduces the problem of ambiguity where there are several possible meanings for a username or segments of the name. This problem and countermeasures taken in the present study are discussed in detail in Section 4.2.4.

Most linguistic studies, especially those with a computational focus, take the second approach. To aid with the analysis, in addition to the usernames themselves some metadata is usually collected. This can help with both annotation and further analysis of the data. Hämäläinen, Haasio, and Harviainen (2021), for example, collected information on which of the users in their data set were either sellers or buyers of illegal drugs. In the current work, information such as number of forum posts, date and time stamps for each user's posts, as well as the language of the user's contributions has been collected.

Those studies that investigated a rather small number of usernames, employ a purely manual approach to analysing their data linguistically, like Stommel (2007), Aleksiejuk (2017), and Hämäläinen, Haasio, and Harviainen (2021). Studies that investigate patterns in a larger data set of usernames through a linguistic lens, such as Wood-Doughty et al. (2018), employ both corpus linguistic and other computational methods to study usernames.

In a study investigating usernames in an online *Buffy The Vampire Slayer* fan community, Gatson (2011) shows that some users chose usernames reflecting offline relationships between people, blurring the distinction between online and offline world and identities.

Kersten and Lotze (2020) have conducted an empirical study on usernames where they created a corpus of usernames which was analysed for lexis, syntax, morphology and orthographic features, and they combined this corpus approach with results from a qualitative questionnaire which examined users' motivation for their name choices. For the corpus part of their study, they collected 500 usernames from British online platforms.¹ 100 usernames each were collected from Twitter, Flickr, a TV forum, an IT forum and a newspaper forum and were then analysed using a predefined tag set. Kersten and Lotze (2020) understand self-naming online in two parts: First, as a "complex and dynamic socio-linguistic practice of authentication or anonymisation" (p.147), and second, in the Goffmanian sense, as a form of face-work.

Donlan (2020) conducted a linguistic analysis of authorial pseudonyms in online fanfiction communities and identified the most common formation methods across a set of 600 usernames as compounding, blending and variant spellings. They show that with these processes users create unique usernames that help them form memorable identities within the community to be recognised by others easily, thus showing that the creation of usernames in this way adds to building their online, and in this case, authorial identity.

In the context of eating disorders, Stommel (2007) studied usernames by 83 users from a

¹This study formed part of a larger international cooperation project where scholars conducted the same analysis on username corpora across 14 languages, see Schlobinski and Siever (2018).

German online forum and found that usernames were strongly linked to users' identity, especially through the common use of first names. Further, it is revealed that certain attributes prevalent in the usernames are linked to femininity and characteristics of individuals with eating disorders, such as smallness, weightlessness, depression or negative self-evaluation. This shows that the community context is of high importance for username creation and establishing an online persona or identity.

Further studies in cyberpsychology, identity and personality research as well as communication studies have all shown an interest in studying usernames (e.g. Perito et al., 2011; Cornetto and Nowak, 2006; Gross and Churchill, 2007; Graham and Gosling, 2012; Hassa, 2012), but none of these have considered any such contexts which involve illegal content or criminal activities.

2.4.2 Usernames in Criminal Contexts and Communities

Few studies so far have studied usernames in criminal contexts. Hämäläinen, Haasio, and Harviainen (2021) collected data on 1654 usernames from a Finnish online marketplace on the dark web which is frequented by sellers and buyers of illegal drugs. Users initiated contact through the online forum and then moved on to agree on meeting times and locations in a private messaging app called Wickr. Contacting another Wickr user requires knowing their username, which would need to be shared with buyers via the discussion forum. The study focuses on those shared Wickr usernames.

They find that many usernames in this community feature personal details such as the user's gender, age or location. In contrast to findings by Hämäläinen (2015) on AlphaBay users, members of this community do not appear to be using strategies in their usernames to imitate real-life brand names. Their analysis also shows that sellers and buyers in this Finnish online community only show minimal differences in features included in their usernames, which hints at an overlap of the two groups. Users' strategy in this high-risk context appears to be blending in rather than trying to stand out. Hämäläinen, Haasio, and Harviainen (2021) also show that users in this community show an unusually high amount of Finnish language names, even for an exclusively Finnish community,

which links to showing expertise by knowing a number of slang terms for illegal drugs but is also related to the commonly displayed racist attitudes in the community.

Even more so than in other anonymous discussion fora online, in criminal online communities and contexts users have a high need for anonymity as the practices they engage in come with great risks. At the same time, trading illegal substances on the dark web requires high levels of trust between the involved parties. Usernames are one way for users to establish themselves within a community as a trustworthy member.

2.5 Privacy, Anonymity and the Dark Web

In the early days of online communications, usernames were often understood as ad-hoc creations that were needed to set up accounts and facilitate communication but without much meaning behind those usernames (Bechar-Israeli, 1995) but many more recent studies show that usernames have developed to be carefully managed and highly meaningful constructions (e.g. Gatson, 2011; Aleksiejuk, 2016; Hämäläinen, 2015; Hämäläinen, Haasio, and Harviainen, 2021). Key factors such as authenticity and anonymity are highlighted by a number of studies into online names and online communities (see Kersten and Lotze, 2020, and Grant and Macleod, 2016).

To understand child sexual abusers' use of the dark web, it is necessary to first understand what the dark web is and how it relates to other parts of the internet. The internet can roughly be divided into the following three parts: the *surface web*, the *deep web* and the *dark web*. The *surface web*, also known as "clearnet" or the open web, is the searchable part of the internet most of us know and access regularly. It consists of publicly available websites which we can access through our favourite search engines and includes e-commerce, blogs or sites like Wikipedia or Youtube. The *deep web* differs from the surface web in that it is not searchable and it is encrypted. Most internet users, however, use the deep web regularly, for example when accessing email accounts or online banking. The deep web includes private fora and networks, netbanking and credit card accounts,

websites storing legal documents and medical records. The *dark web* is another part of the internet which is encrypted but as opposed to the deep web, the dark web cannot be accessed without specialist software, like the TOR browser. This makes the dark web the ideal place for criminal individuals and communities to communicate and to exchange goods and services. It includes online market places for illegal drugs and arms, online fora created by terrorist organisations or child sex abusers discussing and planning criminal activities as well as sharing advice and experiences, often on tor-encrypted websites. It should be noted, however, that using the dark web is not criminal behaviour in itself, and there are many legitimate uses of the dark web, such as private discussion and support fora for users who prefer staying anonymous, or safe spaces for individuals who live under authoritarian regimes and cannot communicate openly on other parts of the internet.

Child sexual abuse and the dark web are closely linked. While some forms of online CSA are carried out on the clearnet, such as online sexual grooming or various forms of blackmail, child sex abusers use the dark web to trade illicit imagery, exchange advice and share experiences. These spaces offer (potential) offenders a safe space to communicate with one another and affords them high levels of privacy and anonymity, which is especially important to individuals engaging in criminal behaviours and who therefore tend to operate in high-risk/low-trust environments (Grant and Macleod, 2016).

This shift of child sexual abuse and other criminal behaviours from offline to online spaces has also meant a shift in research activities into these crimes and communities. The following section discusses in detail the research that has been undertaken over the years, in various disciplines, to better understand such online communities and communications.

2.6 Criminal and Harmful Online Communities and Communications

In their paper on researchers monitoring online offenders, Aldridge and Decary-Hetu (2015) thoroughly analyse the shifting of crime from the physical world to the digital world. The authors not

only provide an analysis, but also include a set of data collection guidelines for researchers working in those areas and identify three main data sources for such research projects: Mirroring (making static copies of websites), monitoring (making periodic observations of websites) and leaks (capturing unwittingly disseminated information). The research presented in this thesis makes use of mirroring for the collection of data from child sexual abuse fora on the dark web. The fora as a whole have been captured and then the raw text data, together with some relevant meta data, has been scraped from this mirrored copy.

Choo and Smith (2008) identified online criminal communities as falling into one of three categories: Traditional organised criminals who operate and act offline but use the Internet for organisation purposes, offline groups conducting crime through the Internet, and criminal groups exclusively operating online. While individuals frequenting those fora might engage in criminal activities offline, the communities studied for this current research belong to the third category.

Further research has been conducted to study specific criminal online groups and communities. While many of the below focus on different types of crime than the ones studied in this work, many of the findings are still relevant for and applicable to the current research as they also focus on communities related to online illicit commodity trading and the discussion of criminal activities. While the works discussed in this section stem from a number of different academic disciplines, the main focus lies on linguistic research into criminal communications and how such research relates to the current work.

2.6.1 Extremist and Terrorist Discourse Online

Terrorism-related fora operate in similar ways as fora related to child sexual abuse: As Rogan (2006) shows, both are tightly controlled and password-protected spaces wherein the community plays the primary role. Bowman-Grieve (2013) studied those communities from a psychological perspective and concluded that involvement in terrorism fora can only be described as a group phenomenon and that it would not be possible to understand it when just looking at individual actors, suggesting the

high importance of looking at these online groups as close communities which need to be studied from a collective rather than individual perspective. Other researchers looked at further details of online terrorism groups: Janbek and Prado (2012) analysed the main goals of groups hosting such online sites, and Torres-Soriano (2013) took a mixed-methods approach studying the creation, evolution, and disappearance of Jihadi fora. Finally, a study by Saidi et al. (2017) investigated different approaches to studying such data. As online fora related to child sexual abuse share some features with those related to terrorism, such as the importance of ideology, the above findings are directly applicable to this research.

The first linguistic studies analysing harmful and criminal communities online have been concerned with extremist and terrorist groups. Brindle (2016) combines corpus linguistics with a CDA approach to study forum contributions referring to the subject of homosexuality on the white supremacist online forum *Stormfront*. His study of white supremacists' online discourse provides insights into the discursive construction of in- and out-groups and shows that homosexuality is used by *Stormfront* members to construct discourses of heterosexual masculinity, sexuality, gender and race.

Rothenberger, Müller, and Elmezeny (2018) investigate the discursive construction of group identity in six terrorist groups across 27 websites employing CDA. Their analysis shows that groups employ a number of common strategies to define their group identity, which includes demontage and destruction, justification and relativisation, as well as boundary work – the differentiation between “us” and “them” which is an essential element to identity construction. Members of such groups commonly construct shared group identities through narrative strategies which focus on rhetorical strategies such as shift of blame, displacement of responsibility and dehumanisation of victims. This type of identity work through manipulative language can be seen as part of the radicalisation process of such groups.

Following a corpus linguistic approach, Litvinova et al. (2018) study features of Islamic extremist rhetoric by comparing the language used by mainstream internet users and members of

a Russian-language Islamic extremist forum. The analysis focuses on forum entries discussing the 2010 Moscow Metro bombings. In line with Rothenberger, Müller, and Elmezeny's (2018) findings, they show that users focus on rhetorical strategies such as the us-versus-them dichotomy and war metaphors.

Recently, Esser (2021) conducted an analysis of language used by members of radicalised subcommunities on Parler (a US social networking platform known for its highly conservative user group and popularity with far-right extremists and conspiracy theorists). The study employed a combined approach of unsupervised clustering and corpus linguistics with a focus on conversations regarding the 2021 US Capitol riots and election fraud. Conversations on topics such as Trumpism, voter fraud, conspiracy theories, in-group favouritism and far-right extremist groups were the most prevalent in the corpus. Additionally, Esser was able to identify distinct language patterns indicating radicalisation: expressions of anger and violence, war terminology, the us-vs-them rhetoric and hate speech were all common parts of their violent and aggressive discourse legitimisation of a need for violence.

In her recent study investigating language in an online forum for white nationalists, Booth (2023) examines how identity boundaries are constructed and (re)negotiated by forum members and discusses the usefulness of the in-group concept. Not unlike the current work, Booth (2023) combines CDA and corpus linguistics to study the positioning of social actors in relation to others in the data and finds that, while an overwhelming amount of literature would suggest there is one coherent in-group in a community such as the one investigated, members of this white nationalist online forum often disagree on boundaries regarding their ideological as well as racial identities. This suggests that there is not one coherent identity that unites the so-called in-group but within this in-group there are multiple, partially overlapping identities that forum members subscribe to and consistently (re)negotiate during their forum discussions with other members. This directly relates to the research presented in this thesis, where communities of child sex abusers are investigated based on

their chosen usernames which divide the community (in-group) into further subgroups with differing and partially overlapping identity positions.

2.6.2 Profit-driven Crimes

Another type of criminal activity that has received some attention in the literature is profit-driven crimes. As with the terrorism groups, Chua, Wareham, and Robey (2007), who studied groups of auction fraudsters, showed that such crimes cannot be understood purely on an individual level and that the group dynamics need to be taken into account. The authors employed Social Disorganisation Theory to analyse how these groups interact with auction houses and authorities. Cracking- and malware-related communities have been studied by Benjamin and Chen (2012), which conducted a qualitative study of hierarchies in such fora. They thereby focused on how individual members climb to the top by providing valuable information and tools to the community. As users in child sexual abuse fora also participate in similar activities where they share information and provide advice to other community members, some features identified, like the ways hierarchies are created in such fora, are also applicable.

Another area which is of interest here is the illegal trade of drugs through online channels. From a linguistic perspective, few studies so far have investigated such communities but recently, Lorenzo-Dus and Di Cristofaro (2018) studied the discursive construction of trust in crypto-drug markets employing a corpus-assisted discourse study (CADS) approach. Focusing on buyers' and sellers' identities, they showed that users regularly discussed sellers' identities and that the construction of these identities was strongly related to members' values of expertise, integrity and benevolence.

Further, onomastics researchers have studied different online market places and communities involved with illegal drug trade (Hämäläinen, 2015; Hämäläinen, Haasio, and Harviainen, 2021). These studies focused on users' identities based on their usernames and are discussed in detail in Section 2.4 above.

2.6.3 The Manosphere

Another context that has received considerable attention in the linguistic literature in recent years is the manosphere. The manosphere is a collection of online fora and platforms which are dedicated to sharing anti-feminist and misogynistic ideologies (Marwick and Caplan, 2018). Much of this research has focused on studying 'incel' groups. Incels, or 'involuntary celibates', are known for their endorsement of violence towards women and sexually active members of society, leading to crimes such as the 2018 Toronto van attack, where a young man belonging to the incel community killed and injured a number of people when he drove his van into a group of pedestrians. To shine a light on the connection between online misogynist discourse and offline violent attacks, many scholars have since collected large sets of textual data to study these communities through a linguistic lens.

Jaki et al. (2019) studied hatred against women in incel discourse and showed that the use of misogynist and homophobic language in the community amounts to hate speech and often includes cases of direct incitement to violence against women in the offline world.

Heritage and Koller (2020) analysed social actors in the incel community on Reddit and were able to show that members have a hierarchical view of men in society, with the group of men seen as conventionally attractive on top of that hierarchy. Women, however, are not placed in such a hierarchy. The researchers analysed the most common male and female social actors in their corpus and, applying appraisal analysis, found not only that community members are highly concerned with physical attractiveness, but also that members see women as immoral, dishonest and capable of hurting men, while men are judged as incapacitated.

Further, Krendel (2020) analysed the discursive representation and construction of gender identity in the incel community by employing corpus-assisted social actor and appraisal analysis to the lemmas *woman*, *girl*, *man* and *guy* found in three incel subreddits. She finds that bare assertions about gendered behaviour are the focus of discussions in these groups with an especially well-defined feminine gender role. Women and girls are sexually objectified and described with

dehumanising language.

In a subsequent analysis of five manosphere subreddits, Krendel, McGlashan, and Koller (2022) find that while both men and women are referenced using relational terms, men are also referenced by kinship or in-group membership and women are often referenced using derogatory language. Complementing previous work, their analysis reveals that members of manosphere communities see women as violent towards men and as holding a privileged position over them in society.

In another study focusing on a community of Men's Rights Activists (MRAs), Lumsden (2019) analyses threads and comments in this group and in particular focuses on their discussions of trolling and gendered violence. Threads and comments in a subreddit have been analysed following a data-driven qualitative approach with a focus on thematic analysis. Findings show that members of the MRA community deny victimhood to women and feminists who have experienced online violence. Lumsden (2019) further shows that MRAs consider men to be the true victims of online violence and harassment and see feminism as the cause for the erasure of men's victimisation; their language of othering is evident throughout the discourse portraying women and feminism as the enemy. Lumsden (2019) also argues that the presence of eliminationist and genocidal rhetoric as well as dehumanising language in the data makes these discourses classifiable as hate speech.

Closed online communities of practice, like the incel community, often adopt a specific vocabulary and common language which includes metaphorical representations closely linked to the group's beliefs, values and underlying ideology. Conceptual metaphor theory sees metaphors as more than figures of speech, they are seen as linguistic elements which structure the way people think. Pražmo (2020) studies the incel community and investigates dehumanising metaphors used by community members to describe women. The findings suggest that incels' use of novel metaphorical terminology such as *femoid* or *foid* is rooted in the same motivation found in other types of hate speech: members of the target group are compared to non-human or cybernetic organisms to strip them of

their human-like qualities and thereby their human rights. The use of these metaphorical constructions allows incels to further objectify women and to justify their inhumane treatment of them. This study serves as another example of how CDA can improve understanding of how language is used to create in-groups and out-groups and how users use language to promote their ideology and justify their actions – both essential to the CSA communities studied in this work.

2.6.4 Sex Work

Sex work is prohibited or illegal in many jurisdictions. Online communities and platforms designed to support individuals in buying sexual services therefore makes for another interesting area for linguistic analysis. Discourses of sexual desire, partner preferences and address strategies in such an online sex work forum have been studied by Lawson and McGlashan (2017). Employing a corpus-assisted discourse approach, they analyse language used by both sex workers and sex buyers in the community. Keyword analysis hinted at specific uses of terms of address, politeness strategies and partner preferences and a subsequent qualitative analysis showed: sex buyers use terms such as *lady* to appear more polite and respectful towards the women on the site and the term *girl* is used to conjure images of youthfulness. Further, the abbreviation *wg* (*working girl*) – most commonly used by men on the site – is used to avoid other terms with negative connotations, to avoid having to use the full term which makes their desires seem less taboo but also to adhere with the communicative forms of the community. Terms such as *gent/gentlemen* are frequently used by sex workers to refer to sex buyers to show politeness, to acknowledge a potential hierarchy of age and to appear professional. The use of *client(s)*, *person* or *they* to refer to sex buyers on the site highlights the idea that they see their work as a business and could further serve as a distancing technique. Their analysis of language related to partner preference further reveals frequent use of objectifying language towards women, highly influenced by porn language, and sex buyers' language use regarding sexual desire reveals a strong wish for intimate relations with women of particular ages and body types.

A study by Carter et al. (2021) focuses on semantic themes and applies both manual and corpus-based thematic analysis to forum posts within a sex work forum. The posts analysed are contributions by both sex buyers and sex workers. Following an inductive approach, their manual analysis of 300,000 words revealed the following five themes of discussion in the community: preference sharing, personal narrative sharing, practical advice, philosophical issues and community maintenance. The first three themes were thereby often linked together. Additionally, they performed a corpus study of over 16,000,000 words, which included topic modelling and keyness analysis, and uncovered a number of additional themes including ethnicity, legality, desire and sexual hygiene.

2.6.5 Online Child Sexual Abuse Communications

Over the last three decades, interest in studying online communications related to child sexual exploitation and abuse has been growing. Most of this work has focused on online grooming, which is the process where "a person prepares a child, significant adults and the environment for the abuse of this child" (Craven, Brown, and Gilchrist, 2007, p. 279).

Most academic work on online grooming comes from psychology, criminology and computational research. Research has thereby focused to a large extent on victim vulnerability factors, offender typologies and the online grooming process itself. As online sexual grooming is a process mainly conducted through written language, it is perhaps surprising that linguistic research into this phenomenon is lacking.

Also, in linguistic research that has been conducted into CSEA, online sexual grooming has been at the centre of most studies. There are several reasons for this: first, online grooming has become a widely and publicly discussed societal problem which is growing and has received much media attention over the past years. Second, online grooming of children is an almost exclusively linguistic process which makes linguistic methods for analysis the obvious area and avenue for researching this phenomenon. And finally, volunteer groups of self-proclaimed "paedophile hunters" who pose as children online to identify child sexual abusers (see, e.g. www.perverted-justice.com)

have made available large amounts of linguistic data which can be used for academic research by CSEA researchers worldwide.

Linguistic research using such 'decoy' data has led to a number of valuable insights regarding offenders' language use and linguistic strategies. Chiang and Grant (2017) apply move analysis to the OCSA context by analysing moves and strategies found in chat log transcripts between child sexual abusers and adult decoys taken from the popular website *perverted-justice.com*. They identify 14 rhetorical moves that are commonly used in online grooming interactions. Further, they provide a new form of visualisations of these moves through colour-coded "move maps" which provide a promising resource to explore variation between grooming interactions.

Lorenzo-Dus, Izura, and Pérez-Tattam (2016) also make use of the *perverted-justice.com* corpus and analyse 24 chat logs in their study, employing Pragmatic Act Theory and Relational Work in their computer-mediated discourse approach. Their findings show that online grooming comprises of three phases – access, approach and entrapment – and identify online grooming as an "entrapment network" with a number of simultaneously occurring strategies. They further show that online groomers engage heavily in the development of trust and rapport with their victims and that they show a lot of relational work in their interactions. As a major result of this study, they developed the first empirical model of online grooming discourse. In another corpus-assisted study of decoy-offender chat logs, this time with a set of >600 transcripts, Lorenzo-Dus and Kinzel (2019) show the usefulness of corpus methods to gain a better understanding of online grooming strategies, especially as they allow for the investigation of connections between such strategic uses of language and other discourse features, such as sexual explicitness. Lorenzo-Dus, Kinzel, and Di Cristofaro (2020) then build on their earlier work and identify recurring patterns of language use in these chatroom interactions by identifying a number of linguistic structures (collocations) that are regularly used by online groomers. Those collocations are especially frequent where individuals attempt sexual gratification through the interaction. These findings can be directly related to the entrapment network

identified through the empirical model presented in Lorenzo-Dus, Izura, and Pérez-Tattam (2016) .

All these studies have furthered the understanding of online sexual grooming discourse and offenders' strategies when engaging in these acts. However, recently more researchers are also acknowledging the limitations of analysing such data sets and highlight the importance of analysing grooming conversations between offenders and genuine victims to be able to more accurately describe and understand the online grooming process. Work by Chiang and Grant (2017, 2018) which studied rhetorical moves of OCSA offenders suggested that interactions between offenders and decoys differ from those with genuine child victims in various ways. More specifically, they show that certain moves, such as Overt Persuasion and Extortion, are only present in data featuring genuine child victims but not when adults who pose as children interact with offenders online. Schneevogt, Chiang, and Grant (2018) later confirm this with an additional study which employs a corpus-linguistic approach. They analyse a corpus of 622 chat logs between offenders and decoys to investigate further whether such coercive and forceful language is present in these interactions. They find no extortion and only rare instances of overtly persuasive language which shows once again that while this sort of data is useful to some extent, CSA interactions between offenders and adult decoys are not truly representative of those with genuine child victims.

While researchers now acknowledge those limitations more, it has to be noted that genuine grooming data is still scarce and hard to obtain for many researchers due to the sensitivity of the data and the need to protect victims. This has meant that much of current research into grooming has been focused on offenders' rather than victims' discourse.

This primary focus on online sexual grooming and the widely available 'decoy' data sets of grooming conversations has had the further consequence of neglecting other forms of online communications crucial in the context of CSEA: one-to-one and group interactions between offenders. Those online communities of potential child sex abusers have so far only been addressed by a small number of studies.

Until recently, those online communities had only been studied by researchers in other academic disciplines, such as criminology (Holt, Blevins, and Burkert, 2010; Holt, 2010; Paquette and Cortoni, 2022; Steel et al., 2023), psychology (Malesky and Ennis, 2004; Young, 2008; Quayle and Taylor, 2011; Woodhams et al., 2021) and sociology (Nielsen, Aaskov, and Larsen, 2022) as well as through more technical approaches such as social network analysis (Frank, Westlake, and Bouchard, 2010; Westlake, Bouchard, and Frank, 2011; Westlake and Bouchard, 2016). Some of these studies directly relevant to this current research which examined child sex offender communities are discussed below and in subsection 2.8.

In criminology, Holt, Blevins, and Burkert (2010) have studied an online support community for and by paedophiles and focused on analysing how the use of the internet can enable the development of a subculture of paedophiles, including how these communities are used to justify members' sexual relationships with children. Using a qualitative approach studying five online fora, Holt, Blevins, and Burkert (2010) show that the values shared in these communities support and encourage the sexual abuse of children, both virtually and offline, clearly highlighting the importance of better understanding these communities as they directly impact offending. In a related publication, Holt (2010) further discusses the importance of qualitative criminological research into online sources to better understand – and ultimately police – communities of deviant and criminal individuals.

Recently, Paquette and Cortoni (2022) conducted a thematic analysis of police interviews with 20 men who committed online offences against children (consumption of CSAM, sexual solicitation of minors, or both) to study the cognitions supporting online sexual offending of children. The analysis identified eight cognitive themes which reflected four underlying themes relevant to these crimes. The themes covered topics such as interpersonal relationships, sexualisation of children, self perception and perceptions of the online environment. While the themes for the different types of offences against children are largely the same, the content uttered by the different types of offenders shows some significant differences: while all offenders agreed that the virtual world does not reflect

real life, those directly interacting with children talked about how their chats with their victims did not reflect their actual intentions, while those consuming CSAM talked about how, to them, CSAM does not represent real victims. Further, all participants talked about curiosity, but while solicitation offenders discussed their desire to find out how children talk about sex, CSAM offenders discussed their wish to find out how CSAM looked like. They further show that cognitions between online offenders and contact offenders largely overlap.

Steel et al. (2023) studied self-perceptions of CSAM offenders as well as their perceptions of CSAM, the legality of CSAM and their opinions regarding laws and sentencing guidelines associated with CSAM offences. The study used an online survey as the basis for a quantitative analysis. Findings include CSAM offenders minimisation of the severity of CSAM offences, offenders endorsement of child erotica and legality of virtual CSAM, offenders endorsement of treatment rather than prison time for online offences as well as offenders' strong views against sex offender registries.

The findings of both studies above highlight the relevance of studying cognitions of online offenders and offending communities to better understand these crimes, to improve prevention and policing of such crimes.

In another psychological study, Malesky and Ennis (2004) studied distorted cognitions in an online message board for paedophiles with an interest in teenage and/or pre-teen boys (self-identified 'boy lovers'). Their analysis showed that while users of these message boards used a range of different distorted cognitions, most could be classified as euphemistic labeling – where individuals labeled their relationships with children as 'romantic' and egalitarian – which was used by posters to justify their actions, highlighting that these types of online communities minimise the harm done to children and can potentially lead individuals with paedophilic interests to act upon their desires. Malesky and Ennis (2004) also show that members of this message board engage in a number of discussions on topics unrelated to paedophilia, children or child sexual abuse, showing these individuals' need for community. Sharing and discussing other, more normative interests outside their deviant interests

could serve to further normalise the latter. The study concludes that the use of these types of online spaces could increase the risk of (re)offending, in line with a number of past studies on this topic. Findings from this study show the importance of further studying these types of online communities to better understand how their use relates to offending behaviours, both online and offline. This is one of the aims of the research presented in this thesis which examines a number of different online fora with different foci regarding users' paedophilic interests and offending behaviour.

Young (2008) analysed 22 forensic interviews to better understand virtual sex offending behaviour from an addiction perspective. The offenders in this study were all first time offenders with no prior history of paedophilia or sex offences against children and were all classified as fantasy users. According to Young, these types of offenders all followed a pattern of five stages based on their internet addiction, with hopelessness or regret marking the final stage. While Young points out that many fantasy-driven users with no history of offending might enter these types of online spaces out of curiosity only, not being fully aware of the risks, users in the current study who join dark web fora to exchange illegal materials are very well aware of these risks and even openly discuss those risks and how to mitigate them with other users. While it is likely that some users initially started out as fantasy users as described in Young (2008) and remain addicted to the consumption of CSAM to some extent, they significantly differ from that studied population as they are fully aware of their deviant and criminal behaviours, even if they might not agree with the illegality of those behaviours.

More recently, Woodhams et al. (2021) have examined dark web CSA communities specifically and studied the behaviours of 53 anonymous suspects. This study is of particular interest for the work presented in this thesis as it studies a similar population and discusses a number of offender characteristics and behaviours also present and discussed in the empirical chapters 4, 5 and 6. Their findings suggest that almost all suspects self-reported being male and 34% of the individuals studied reported having access to children through family relations, either their own children or

children of extended family. A variety of different offending behaviour regarding CSA has been evident in their sample and most individuals were consuming and sharing CSAM. Most of the users also reported abusing or having previously abused children, both online and offline. In addition to the traditionally described dichotomy of contact- versus fantasy-driven offences against children, they also found that some users were abusing children via webcam or encouraging and enticing other users to abuse children sexually. Regarding their purposes for frequenting such online fora, users stated that they did it to access CSAM, to share CSAM and to interact with others like them. Another focus of conversation in the studied sample was users' discussion of risk management, security measures and maintaining anonymity. Particularly interesting and relevant for the present study, Woodhams et al. (2021) found that in their sample, the majority of individuals did not use usernames that were directly related to CSA or paedophilia. Further, some subjects disclosed their preferences for their victims' gender and age or the types of sexual and abusive acts they are interested in, especially when it came to accessing CSAM, with sadism and incest being the ones most commonly mentioned. Some also disclosed they additionally had a sexual interest in adults as well. Finally, Woodhams et al. (2021) also analysed the interactions between offenders. As suggested before in the literature, a sense of community was evident in the data and individuals did engage in conversation outside of their sexual interest in children. Users gave each other advice, shared experiences and thanked each other for contributions, much like in other types of online communities. Expressions of dominance and power were also observed. The textual data further revealed that suspects viewed children as sexual beings, minimising the harm caused by CSA and normalising their behaviours. Others objectified children and even stated they deserved the abuse, linking to the strong interest in sadistic behaviours by some of the studied subjects.

Another recent study by Nielsen, Aaskov, and Larsen (2022) investigates a vastly different kind of online community of paedophiles through a sociological lens: a web forum for and by 'virtuous paedophiles', individuals who identify as paedophiles but believe sexual contact with children to

be wrong and pride themselves in being non-offending, is examined regarding the forum's culture. Analysing the textual data in this particular forum, Nielsen, Aaskov, and Larsen (2022) find that the mainstream stigmatised understanding of paedophiles as child molesters limits not only the understanding of paedophilia but also limits individuals in their ability to remain non-offending. The authors believe that a supportive online community like this could actually be beneficial to individuals with paedophilic interests and help them remain non-offending paedophiles. This highlights that understanding the internet as generally harmful and increasing the risk of (re)offending for paedophiles – as was suggested by others like Young (2008) – is not to be generalised and individuals interacting with other paedophiles online can potentially prevent offending if those communities focus on non-offending and abide by the right beliefs and values, which see sexual acts between children and adults as wrong and something to be prevented.

In linguistic research, a small number of studies have recently contributed to the understanding of CSA-related online communities. Chiang (2024) investigated forum contributions by new members in one such online community and found that new members employ a number of typical rhetorical moves when attempting to gain acceptance into the community. Her research highlights the importance of understanding offender identities and being able to portray oneself 'authentically' in such a community. She identifies two features which are essential to gain membership to such a community: expressions of competence and demonstration of expertise. Chiang's findings can directly be applied to the policing of such online spaces, especially for infiltration tasks, as is discussed further in section 2.8.

Chiang also showed in her work that dark web communities of child sex abusers can be viewed as Communities of Practice, an insight directly applicable to the research presented in this thesis. This notion and its relevance for the current research is further discussed in section 2.7.

MacLeod and Grant (2021) have also analysed such dark web communications and were able to show that using linguistic analysis in this context to build linguistic profiles of forum members can

be beneficial for identity assumption tasks which undercover officers often have to engage in when investigating such criminal communities.

Chiang et al. (2021) analyse communications between suspected child sexual abusers in a dark web chatroom and offer the first linguistic description of authentic interactions between offenders in this context. Employing a combination of move analysis, which they use to uncover rhetorical strategies employed by group members, and Markov models, which enable the researchers to compare linguistic behaviours of different individuals, Chiang et al. (2021) identify a number of characteristic moves prevalent in this community's interactions: Greetings, General Rapport, Offering Indecent Images, Image Appreciation, and Image Discussion. They further show that there are significant differences between users' rhetorical strategies linked to levels of experience with dark web image-sharing and individual offending histories. Based on these findings, they provide recommendations for law enforcement as this research offers many practical applications for undercover agents working in the CSEA context, including law enforcements' understanding of functions and linguistic behaviours in this community as well as how officers can authentically perform offender identities in covert operations.

In a recent linguistic study analysing authority and power resources in online criminal interactions, Newsome-Chandler and Grant (2023) provide a practical framework for studying power and authority in such online communities. Their framework aims to assist law enforcement in policing such anonymous online spaces and further understanding of community roles and dynamics. They show how their framework can be used to study individuals' claims to power in such communities and, among others, provide the analysis of dark web CSA interactions as one of their examples. They find that the studied CSA forum exhibits hierarchies of power and exclusivity, and possession of novel and/or extreme CSAM presents an extralinguistic marker of power in this community. Newsome-Chandler and Grant (2023) identify nine commonly occurring power resources used by forum members. Those power resources occurring with the highest proportion in the CSA forum

they study are: Direct Personal Experience, Broad Topic Expertise, Direct Community Expertise, Community Specific Initialisms and Direct Technological Expertise. Their analysis further shows that individuals are not inherently more or less powerful but their use of power resources in context is what grants them power in a given interaction. Users can thereby often draw on a number of different resources and it appears that users who can draw on more resources in a given discussion than others are granted higher status and more power. This might, however, not translate to other topics discussed if the same individual does not have any relevant resources to draw on there. These findings allow for a better understanding of hierarchies and community roles and might thereby aid policing strategies such as infiltration of such online fora.

Holt, Bossler, and Seigfried-Spellar (2015) provide a comprehensive and multidisciplinary survey of the overall subject by connecting the literature on the various types of online criminal behaviours, by summarising the investigative strategies which are available to law enforcement agencies, and by investigating the role of technology as an enabler for criminal behaviour.

2.7 Criminal Communities as Communities of Practice

The research summarised in Section 2.6 shows that, overall, researchers seem to agree that we need to study these communities on a group or community level to fully understand the dynamics, hierarchies and ways in which members interact; further rationale to investigate not just single roles and social actors, but focus on collective identity and the notion of communities of practice.

The concept of Communities of Practice (CoP) was first introduced by Wenger (1998). They can generally be described as groups of people who share a certain passion for something they do, a mutual interest or a joint mission. Such groups also participate in learning to improve at whatever it is they do and they share by interacting with others in this group regularly, whereby this learning can either be intentional or incidental. Examples for a CoP could be a group of students from the same study programme or researchers within a specific field of academic research. In this work,

a group of child sex abusers sharing illicit materials and regularly interacting online on a specified dark web forum is understood as a CoP. According to Wenger (1998), the following three important characteristics need to be present for a group to be considered a community of practice, and the cultivation of such a CoP happens when group members develop all three parts in parallel:

- **Domain:** An identity defined by a shared domain of interest. Members display commitment to the domain, and shared competence often distinguishes members from outgroups although the expertise is not necessarily recognised or valued by society.
- **Community:** Information and advice is shared, and regular discussion takes place. Members engage in joint activities, and rely on each other for support. Relationships with regular – although not necessarily daily – contact are formed.
- **Practice:** Members are not just people with a shared interest in the topic, they also share expertise, resources, experiences, stories, tools, or tricks.

(Wenger, 1998)

The concept has since its original publication been extended for online communities (Zhang and Watts, 2008; Angouri, 2015) and proven useful to describe and categorise communicative behaviours in criminal dark web communities. Recently, Chiang (2019) studied contributions on CSA-related fora using Wenger's (1998) concept of Communities of Practice. Chiang (2024) showed that CSA offenders indeed operate as online communities of practice. She adapted Swales' (1981) framework of move analysis for this modern context and identified rhetorical moves employed by new forum members attempting to get accepted into these communities.

Since the data investigated in this work closely resembles the data analysed in Chiang (2019, 2024), the notion of CoP was also adopted for the present work and the groups of users interacting in the eight analysed fora are investigated as one large community of practice which allows for more specific subgroups also classified as distinct communities of practice.

Another potential framework to describe and study community structures commonly applied in other online contexts that has been considered for the current work is the concept of *affinity spaces* (Gee, 2005).

Affinity spaces have first been introduced by Gee (2005) who discussed it as an alternative to communities of practice, especially in the context of studying classroom teaching and online learning, including online gaming. While with communities of practice the focus lies on group membership, Gee's (2005) concept takes spaces as the starting point, rather than groups, as he believed the idea of membership was not always a useful concept to apply for the communities studied as membership might look very different from one community to the next. Gee found it difficult to start with the 'group' label as it immediately labels some individuals as 'in' or 'out' and found the boundaries too strict. In Gee's (2005) affinity spaces individuals can connect through networks working on common endeavours and pursuing shared interests "without the need to develop persistent, stable communities rooted in ideas of belonging and membership" (Pellicone and Ahn, 2014).

Affinity spaces as a way to describe online communities has found strong popularity among researchers investigating online learning and online gaming spaces (Gee and Hayes, 2012; Pellicone and Ahn, 2014; Jones et al., 2016; Barany and Foster, 2024). But also within these contexts, the original framework has received some criticisms: While it is true that the clear boundaries and sense of belonging essential to communities of practice are absent in many fluid online spaces to do with learning and gaming, and a combination of the two, and possibly other contexts as well, more recent work applying the framework has shown that some affinity spaces might be more exclusive and less welcoming than others, making them closer to the original idea of a community of practice which has stricter boundaries (Pellicone and Ahn, 2014; Gee and Hayes, 2012).

Affinity spaces and communities of practice, conceptually, do share some important features – such as a shared goal/common endeavour – but they also differ in crucial ways which are relevant to the context studied in this work: Unlike individuals in affinity spaces, users in the studied CSA-related

dark web fora have limited routes to participation in the fora and a limited number of options when it comes to gaining status and power within the community (see Newsome-Chandler and Grant, 2023). Further, newbies and masters are clearly distinguished in the dark web forum context and options for how to join the community (through signing up to the forum and sharing a specific kind of accepted content) are extremely limited.

While the author believes that the concept of an affinity space could potentially be applied when considering a wider context which explicitly includes, for example, paedophiles and (suspected) child sexual abusers which are not part of the studied online fora, in this particular work the given context is that of users in a selection of online fora where, by creating an account, individuals become members of the forum. Therefore, membership – something Gee saw as not always a given – is undisputed and clear for all studied individuals. Further, not only do individuals become members through signing up for the forum, some fora explicitly carry the title 'community' in their forum name and users call themselves and others 'members' and see this membership as an essential part of their online identity. Therefore the decision was made to follow the examples of Grant and MacLeod (2020) and Chiang (2024) and consider the communities investigated in this thesis as communities of practice.

2.8 The Role of Linguistic Analysis in Policing Online Crime

The effectiveness of online undercover police work for CSA-related communities has been the focus of several studies. In a study looking into CSA cases going to court, Mitchell, Wolak, and Finkelhor (2005) found that approximately 25 percent of all CSA-related arrests were conducted based on some kind of undercover work. Not only did these arrests produce high rates of guilty pleas, they further produced low rates of dismissed and dropped cases. In Martellozzo (2015) and Vendius (2015), the limitations of undercover policework were explored. Martellozzo (2015) concluded that the advancement of technology would decrease the effectiveness over time, while Vendius (2015)

concluded that borders and jurisdictions produced serious challenges for agents operating on the decentralised web. Finally, Krone (2005) analysed potential other strategies by the police, finding circumstantial evidence for the value of undercover work over e.g. extensive monitoring, but lamenting the lack of data on the subject. These results show the importance of developing tools, including linguistic analysis frameworks and methods, to advance undercover policing of online criminal activity.

Grant and MacLeod (2020) further demonstrated that analysing conversations on fora related to child sexual abuse (CSA) and building linguistic profiles of individual users can be useful for identity assumption tasks, thus showing the usefulness of linguistic analysis for online policing tasks. For infiltration, however, undercover officers need to analyse the overall language use and culture of a community when engaging in linguistic legend building, and more tools for such tasks which are informed by linguistic research have yet to be developed.

In computational research, fora related to CSA have recently been the subject of several studies using social network analysis (SNA) (Frank, Westlake, and Bouchard, 2010; Westlake, Bouchard, and Frank, 2011; Westlake and Bouchard, 2016). Those projects show that SNA is an excellent tool to locate key players and targets for infiltration, but it cannot directly contribute to the infiltration process. SNA studies can provide law enforcement – and other researchers – with an important starting point from where to further investigate a community and potentially dangerous individuals. While identifying key players is beneficial when trying to isolate the most prolific and dangerous abusers in a community to directly target those for investigation, SNA studies do not provide in-depth information on how the common forum user or the community as a whole operate and what makes a typical member in terms of their linguistic expression. As such, this approach forms a complement to linguistic strategies for social actor analysis and role identification, where linguistic analysis can provide further details on how different types of community members – based on their role in the community or their topical focus – interact with one another and which linguistic features

might help an undercover operative to convincingly portray themselves as a genuine member of this community.

Another challenge with covert investigations is that undercover operatives are often limited in what they are allowed to share or discuss when entering a criminal online community for infiltration purposes. Due to regulations around "agents provocateur" (UK) and "entrapment" (US), officers may not be allowed to share certain materials or suggest certain topics as to not 'entrap' the suspect by suggesting, offering or enticing them to commit a crime. This also includes the choice of username when infiltrating such a group as a new member. Deviations to options for what undercover operatives can choose as names or topics to introduce might change if research can provide significant evidence that introducing a certain topic or using a certain username is simply copying the behaviour of other genuine community members, rather than being something specifically chosen by law enforcement to lead a suspect to commit a crime, changes to the current rules and regulations might be possible.

If linguistic analysis, for example, could provide investigators with reasonable grounds for choosing a more sexually explicit or aggressive username based on the role the undercover officer is attempting to imitate when infiltrating an online community of child sex abusers, officers might be able to work around those limitations and back up their actions with findings from linguistic research. This offers another way in which the current work can contribute to the policing of OCSA, and online crime more generally, in a meaningful and practical way.

Chapter 3

Methodology

This chapter outlines the methodological approach taken in the three subsequent studies described in Chapters 4, 5 and 6. The first section outlines the overall procedure for data collection, preparation and analysis in this work, followed by an introduction to the general methodological approach taken in the three studies. Section 3.3 then introduces the different linguistic frameworks employed for the analysis of the forum data followed by a discussion of limitations of the presented studies. Finally, this chapter is concluded with a discussion of ethical considerations with regards to both data set and researcher well-being.

3.1 Procedure

The analysis of the data in this work is split into three parts, presented as Study 1, Study 2 and Study 3 in chapters 4, 5 and 6, respectively. This section gives an overview of the steps taken in order to collect, prepare and analyse the corpora in each of the studies.

The first study aims to provide an overview of the types of usernames present in the data and identifies common username categories as well as other noteworthy features detected during the analysis. First, usernames and post frequencies were extracted from the full data set (32,410 usernames). Afterwards, the author engaged in an initial exploration of the data and noted down first observations. Then, for the top 10,222 posters (post frequency \geq 3) the usernames were split into

their individual parts by means of a manual segmentation process. Those segments and segment pairs were then analysed and together with observations from the previous step, emerging themes and patterns were described and presented as a classification of 20 username categories. Each category was also analysed for social actor representation.

The second study then takes this analysis further by analysing usernames and respective forum posts for the top 1837 posters in the data set (post frequency ≥ 21). To this end, the top segments and segment pairs in the set were analysed and together with findings from the first study, five username categories were selected for further investigation. For these five categories, subcorpora of forum posts were created and analysed with a corpus assisted approach. These subcorpora – *Childlove*, *Family*, *Names*, *Paedophile* and *Violence* – were analysed for top keywords against a number of reference corpora and observed themes within the post text are discussed in section 5.2.

For the third and final study, the same five subcorpora as in the previous study were investigated. This time, top keywords against the full corpus of top users (DarkWebCSA-top) were analysed thematically and, following a detailed analysis of concordance lines from the corpus for top themes and keywords, including those which constitute social actors in the discourse, those themes are discussed in the context of the subgroups of users and the wider context, highlighting differences and similarities between the five user groups.

The three studies in this work were created following an inductive approach and specific steps taken in the analysis are best understood after having read about the previous study which influenced the design choices for the next study. Therefore, details regarding the specific analysis steps taken for each of the studies are discussed within the empirical chapters 4, 5 and 6.

3.1.1 Data Sets: Collection & Preparation

The data set for this work was kindly collected and provided by Hyperion Gray, a software company which, among other things, provides tools for law enforcement to analyse and investigate data on the dark web. The data set consists of dark web forum conversations from eight different fora related

to child sexual exploitation and abuse (CSEA). The eight fora were selected based on topic which had been assigned to the fora by Hyperion Gray, who identified these fora as involving discussions of child sexual exploitation and abuse and the exchange of child sexual abuse material (CSAM). The total number of forum posts in this data set is 295,799 and the posts were created between 2014 and 2019. The number of unique usernames in the full corpus is 32,410. An individual might post and participate in the forum under more than one username, so it is only tentatively assumed that the number of unique usernames matches the number of users. Individual users (based on unique usernames) contributed between 1 and 5557 posts within this time frame. The full list of usernames and their respective post counts can be found in Appendix E.

For different stages of the analysis in this thesis, different subsets of the data have been prepared, annotated and analysed. While the data exploration and initial categorisation has been carried out on the full data set of 32,410 usernames, only parts of the data have been analysed in the subsequent stages. For the username segmentation process, a reduced corpus of 10,222 users has been selected based on users' post frequencies (see details in 4.2.1.2). For studies 2 and 3, which involve corpus-assisted analyses of username groups and their respective forum posts, a subset of 1,837 users (the most frequent posters in the data set) was prepared for analysis. This data set is referred to as DarkWebCSA-top and subgroups of this set were analysed against a number of reference corpora in study 2 and 3. Table 3.1 below summarises the details for the different subsets of data. Details on the inductive process of narrowing down the data into smaller corpora for the different parts of the project are explained in detail in sections 3.1.1.1 and 3.1.1.2 below, and in the analysis sections for each of the empirical chapters.

	Full Corpus	Reduced Corpus	DarkWebCSA-top
Number of usernames	32,410	10,222	1,837
Used for the following analysis stages	Data exploration and categorisation	Segmentation and categorisation	Segmentation, categorisation and concordance analysis

TABLE 3.1: Details of the full corpus and analysed subcorpora, including number of usernames and analysis stages.

3.1.1.1 The Full Corpus: DarkWebCSA

The usernames analysed in this work are taken from eight dark web discussion fora with posts covering the time frame from 2014 to 2019. While the actual forum names have been omitted for security reasons, each forum has been given a descriptive label in order to highlight the main interest of the forum users: *Support Community 1 (SC1)*, *Support Community 2 (SC2)*, *Support Community 3 (SC3)*, *IIOC General*, *IIOC Girls*, *IIOC Boys*, *IIOC Babies* and *IIOC Hurtcore*, respectively.

While *SC1*, *SC2* and *SC3* all focus on support among members and contributions are often long and include narratives about fantasies and experiences, the other five fora mainly focus on the exchange of illicit imagery of children (IIOC). The IIOC fora comprise of posts that are generally shorter, often include links or references to external image or video sources, include IIOC within the posts, and contributions discussing those images and videos with one another, asking for recommendations and discussing preferences with other members. The sharing of IIOC is, however, not exclusive to those five fora and still happens in the support community fora as well.

Further, while the main language in all fora is English (see Table 3.2 for details), many usernames are in other languages, sometimes another alphabet or include individual words or phrases within the English username which are in a different language. Those usernames were still included in the analysis where possible and the author attempted to code those usernames and their segments based on translations for the individual segments. Usernames by users who exclusively posted in another language have also been included in the analysis for this chapter, and in cases where a username could be English or another language, it has been treated as an English language username and annotated accordingly.

A complete description of all fora can be found in Table 3.2.

Before any data could be extracted for analysis, the data set had to be prepared for safe usage and cleaned of a number of details. The raw data set included a number of potentially unsafe links to illegal websites and materials (such as child sexual abuse images and videos).

Forum	Content	Main language
Support Community 1 (SC1)	Support and advice; sharing stories; community focus	English
Support Community 2 (SC2)	support and advice; general paedophilia discussions; ideology; society	English
Support Community 3 (SC3)	Support and advice; sharing stories; community focus	English German ¹
IIOC General	Link, image and video exchange	English
IIOC Girls	Link, image and video exchange; preference for female victims	English
IIOC Boys	Discussions and exchange of images and videos; preference for male victims	English
IIOC Babies	Preference for very young children (=infants and toddlers; nepiophilia); discussions and exchange of images and videos	English
IIOC Hurtcore	Sadistic practices; violence, pain and death; exchange of videos and images	English

TABLE 3.2: Full list of fora in DarkWebCSA data set including information on languages present and main discussion topics.

Therefore, the first data cleaning exercise was to remove all links from the data and replace them with placeholders. Where the link referred to a dark web forum itself, the link was replaced with a specific placeholder identifying the forum so that this piece of information could be retained should it become important later in the analysis process. The links to each individual post (post link) made it possible to link each post to the forum it originated from. In an initial step, through a simple find and replace Python script, the following data cleaning steps have been carried out: the links to the forum posts have been replaced by placeholders referring to each of the eight fora, all other links in the data set have been replaced with a '[url_removed]' placeholder. All analysis presented in this dissertation has then been carried out on the cleaned-up version of the data set. The raw data was

¹It only transpired late into the analysis process that the main language for contributions in this forum is German. Since a significant amount of posts in this forum are in English, the decision has been made to include this forum in the analysis of the usernames, and to some extent for the analysis of forum posts.

presented in JSON format and both the raw and the cleaned version were stored and submitted as one csv file each. Both the raw data and the cleaned data set are available to the examiners in Appendix D.

The details available for each individual forum post in the complete data set are as follows: (0) thread ID, (1) post title, (2) post link, (3) date-time stamp, (4) post ID, (5) username, (6) post text (including quoted text), and (7) clean post text (no quoted text). For the purposes of this thesis, the two columns in the data set that were of highest importance are *username* and *clean post text*. The former column forms the basis of the full username corpus extracted for Study 1 as well as all other username subcorpora investigated, and the latter contains the forum posts extracted and analysed in studies 2 and 3. From the above eight dark web fora, the usernames have been extracted with a Python script written by the author. The script can be found in Appendix F.

3.1.1.2 Analysing Most Frequent Posters: the DarkWebCSA-top corpus

The data set analysed in chapter 5 comprises a subset of the original data set. Users among the original full corpus have been separated into two groups: frequent posters – those users who have posted more than 20 times within the given time frame – and infrequent posters – those individuals who have posted 20 times or less in the given time frame. For the following corpus-assisted analysis which looks into identity performances of users by user group (based on the identified categories detailed in Chapter 4), only usernames and post content by frequent posters (1837 users) have been considered.

To prepare the data for corpus analysis, first the usernames for the most frequent posters and the respective segments identified in the previous analysis stage in Chapter 4 have been extracted from the full data set. The extraction of the usernames from the data set was done with a simple Python script written by the author, with the usernames list as an input file. The script can be found in Appendix F. During study 1, the usernames of the top 1837 users which are the focus of this chapter have already been segmented. A subsequent corpus analysis identified the most

frequent segments and segment pairs in the corpus to gain a better understanding of what types of usernames are present in the data. The most frequently occurring segments and segment pairs within the frequent posters user group can be found in tables 3.3 and 3.4 below. Findings from these tables will be included and referred back to in analysis of the subgroups in Section 5.2.

Segment	Frequency	Segment	Frequency	Segment	Frequency
lover	50	one	8	l	4
boy	50	.	7	p	4
the	35	child	7	blue	4
pedo	30	anon	7	young	4
love	29	lvr	7	ped	4
2	23	mom	7	cum	4
-	21	fucker	7	smith	4
girl	20	bob	7	papa	4
man	19	life	7	yng	4
dad	17	b	6	luv	4
little	17	andy	6	hot	4
daddy	17	teen	6	radek	4
4	15	15	6	teddy	4
i	14	99	6	master	4
baby	14	jack	6	10	4
mr	14	lolita	6	13	4
1	13	king	6	22	4
loli	13	joe	6	25	4
fan	13	boys	6	wolf	4
a	12	bl	6	mann	4
s	12	in	6	44	4
666	11	is	6	66	4
bear	11	angel	6	wonder	4
69	11	candy	6	sweet	4
girls	11	curious	6	red	4
of	11	0	5	and	4
to	11	5	5	dick	4
guy	10	lil	5	diaper	4
me	10	star	5	roger	4
x	9	88	5	that	4
good	9	all	5	dream	4
toddler	9	kids	5	head	4
3	8	white	5	on	4
m	8	bad	5	so	4
12	8	john	5	stone	4
uncle	8	my	5	yo	4
new	8	kid	5	panda	4
not	8	dark	5	dave	4
123	8	cat	5		

TABLE 3.3: Top segments (frequency \geq 4) in the DarkWebCSA-top username corpus (1837 core users).

For the analyses in studies 2 and 3, the entirety of the 1837 usernames in the subset had to be grouped into subgroups. This was done based on both the frequency analysis carried out on

Bigram	Freq.	Bigram	Freq.	Bigram	Freq.	Bigram	Freq.
girl lover	5	active dad	2	- member	2	pedo bear	2
boy lover	4	astrid tolmo	2	so sweet	2	john doe	2
of the	4	retired -	2	magic dragon	2	joao 1206	2
little boy	3	wolf 69	2	loli lover	2	star dust	2
i m	3	4 life	2	diaper boy	2	nabokov is	2
lover boy	3	call me	2	to see	2	mr .	2
staff -	2	dom toretto	2	baby fucker	2	vicky fan	2
69 ridinghood	2	in the	2	baby boy	2	fun fare	2
white rabbit	2	- staff	2	radek radek	2	boy love	2
matthew 1914	2	ian owt	2	anon anon	2	teddy bear	2
little girl	2	4 me	2	a lot	2		
white bear	2	pedo papa	2	baby lover	2		

TABLE 3.4: Top segment bigrams (frequency \geq 2) in the DarkWebCSA-top username corpus (1837 core users).

segments and segment pairs as well as on the much more detailed segment annotations.

The segment annotation was carried out manually by the author for all segments of all 1837 usernames. To do so, the segmented usernames were imported into NVivo and there each segment has been annotated for one or more categories (see the full list of categories with detailed descriptions in section 4.2.1 in the previous chapter). Table 3.5 below shows the full list of categories found in the DarkWebCSA-top corpus and the respective segment and username count.

As usernames often comprise of several segments annotated for different categories, a single username can belong to more than one category. Therefore, some usernames were assigned to more than one category. An example is the username *pedodaddy* which was split into the segments *pedo* and *daddy*, assigning the username the categories *paedophile* and *family relations*, respectively. Therefore, the username is part of both subgroups. Additionally, a username might comprise of several segments belonging to the same category, which further explains differences between segment and username counts for some of the categories.

3.1.1.3 Significant Username Subgroups: Subcopora for Studies 2 and 3

Based on the above distribution and the top segments and segment pairs discovered in Chapter 4, the decision has been made to focus the analysis on a select number of subgroups. These include

Category	Segments	Usernames
Animals	67	66
Anonymity	13	11
Brand	3	3
Childlove	67	67
Colour	25	25
Community Role	9	7
Dark/Evil	13	13
Family Relations	72	68
Food	34	30
Gender: Female	5	5
Gender: Male	41	41
Location	14	14
IIOC	11	11
Names: First & Last	369	310
Names: Titles	39	39
New User	13	12
Occupation	10	10
Paedophile	46	46
Preference: Age	86	85
Preference: Appearance	5	5
Preference: Gender	83	83
Preference: Sexual Acts	30	29
Religious	21	18
Sexual Orientation	7	7
Violence	22	22

TABLE 3.5: List of all category annotations for segments of top 1837 users. Annotation of the data was carried out in NVivo.

the most common categories present among the username segments and segment pairs of the most frequent posters' usernames, as well as those categories with most usernames in that category among the DarkWebCSA-top corpus. Further, the author decided to also consider subgroups thought to be of high importance in the OCSA context and the policing of such online spaces. Categories with less than 20 usernames belonging to that category among the DarkWebCSA-top corpus have been considered too low frequency for further consideration and analysis at this point. Table 3.6 below shows the relevant observations used for the selection of the username subgroups analysed in this chapter. Six categories are key categories within the top categories, top segments and top segment pairs among the 1837 users. Additionally, the usernames belonging to the category *Violence* have been deemed relevant for LEA purposes and therefore been included for analysis in this chapter. Based on those observations and findings, the following 7 subgroups have been selected to be analysed in more detail: *Childlove*, *Family Relations*, *Names*, *Paedophile*, *Preference:Age*,

Preference:Gender, and *Violence*. After collection of the subcorpora, it transpired that – due to the significant overlap of both *Preference* categories with other username categories – naming themes within these subgroups were not distinctive enough to lead to valuable insights regarding those user groups. They have therefore been discarded from the analysis, leading to a set of 5 subgroups undergoing detailed analysis in chapters 5 and 6: *Childlove*, *Family Relations*, *Names*, *Paedophile*, and *Violence*.

Top segments (Top Users)	Top segment pairs (Top Users)	Top Categories (Top Users)	Additionally relevant for LEA
Names Preference: Age Childlove Preference: Gender Family Relations Gender Paedophile Animals	Preference: Gender Childlove Names Preference: Age Animals Paedophile Family Relations	Names (310 usernames) Preference: Age (85 usernames) Preference: Gender (83 usernames) Family Relations (68 Usernames) Childlove (67 Usernames) Animals (66 usernames) Gender (46 usernames) Paedophile (46 usernames)	Violence

TABLE 3.6: Relevant categories for the selection of username groups to be analysed in Chapter 5 and 6 based on previous observations and findings. Those selected for detailed analysis in studies 2 and 3 are highlighted in bold.

For each of these subgroups, all posts by all users in this subgroup have been collected into a subcorpus each for further corpus analysis. To extract the text data for each subcorpus, a Python script has been written which allows for the extraction of forum post text by username. For each subgroup, the script is fed the list of usernames belonging to that subgroup and the text data is then stored into a new corpus consisting of txt files which include the forum text, with one file per user including all forum entries of that particular user, where one line in the txt file equals one forum post. The full script for reference can be found in Appendix F. An initial exploration of one of the subcorpora showed the corpus results skewed due to non-English language data as the reference corpora are all English language corpora. The subcorpora for analysis have therefore undergone one further data preparation step and have been cleaned of all non-English forum contributions. This step has been carried out manually by the author for all selected subcorpora. Where a user has posted in

both English and (an)other language(s), the English posts remained in the subcorpus and only the non-English posts were removed. Where a user has only posted in other languages than English, the entire username file (username and posts) has been removed from the subcorpus. This led to the following subcorpora for analysis:

Corpus	Language	Files	Tokens	Types	Lemmas
DarkWebCSA-top	Various	1830	20818195	493856	496871
DarkWebCSA-en-violence	English	19	40689	5703	5108
DarkWebCSA-en-family	English	68	381097	19006	17304
DarkWebCSA-en-names	English	216	1712462	58228	53195
DarkWebCSA-en-childlove	English	64	1005121	35595	32824
DarkWebCSA-en-paedophile	English	45	337348	18057	16403

TABLE 3.7: Subcorpora by prominent username category (English posts only) used for keyword and concordance analysis with LancsBox. Details for the initial DarkWebCSA-top corpus are included for comparison.

For each subcorpus, one file corresponds to one username with one line per forum post. This means that the *DarkWebCSA-en-family*, for instance, comprises forum posts by 68 users. The full list of usernames per subcorpus can be found in Appendix H.

3.2 Methodological Approach

It should be noted that due to the large number of usernames, this work does not attempt to analyse particular individuals but focuses on broader patterns within the particular data set and research context.

3.2.1 Critical Discourse Analysis

The research in this thesis follows an approach drawing on Critical Discourse Analysis (CDA), a broad, problem-oriented interdisciplinary research movement wherein language is viewed as a form of social practice which is fundamentally interconnected with ideology, power relations, and social structures (Fairclough, Mulderrig, and Wodak, 2011; Fairclough, 2013; Wodak and Meyer, 2015).

CDA as a cross-disciplinary research movement employs a number of different approaches, all with a strong focus on identifying underlying issues related to inequality, ideology and unequal power distributions (Fairclough, Mulderrig, and Wodak, 2011; Flowerdew and Richardson, 2017). It is an interdisciplinary approach, interested in language in use, which aims to reveal what part those discourses play in shaping social processes and structures.

CDA aims to critically examine and expose how discourse functions in the production and reproduction of social inequalities and injustices by analysing the linguistic choices evident in discourses in relation to their wider socio-political context (Fairclough, 2013; Flowerdew and Richardson, 2017). Unlike critical linguistics, CDA does not just analyse the textual data in question but further considers social structures, power dynamics as well as creation and reception of the given text; interdiscursivity and intertextuality are also considered (Fairclough, 2013; Barker and Prentice, 2017). Within CDA, there are several influential yet distinct approaches, each operating with somewhat different theoretical models and analytical toolkits. These include Fairclough's dialectical-relational approach (Fairclough, 2013), Wodak's Discourse-Historical Approach (Riesigl and Wodak, 2001), van Dijk's socio-cognitive model (Van Dijk, 2006), and van Leeuwen's work on the representation of social actors (Van Leeuwen, 2013). There is no single prescriptive method used and researchers select methods and frameworks guided by their particular research questions (Flowerdew and Richardson, 2017). The key aim is thereby always to unpack the underlying ideologies, beliefs and power dynamics that manifest through the linguistic choices present in discourse (Wodak and Meyer, 2009). Despite differences in favoured definitions, theories and frameworks, CDA approaches are generally characterised by the following: an interest in naturally occurring language use; analysis extending isolated words and sentences, instead studying larger text units; examining language as social interaction and practice; incorporation of multimodal semiotic phenomena beyond just written text; a focus on the dynamic strategies and context involved in meaning-making; and analysis of diverse aspects of text and discourse structure like coherence, argumentation and rhetoric (Wodak and Meyer,

2009). The contexts that CDA scholars investigate are highly diverse, spanning print media, television, radio, internet and social media, as well as domains such as politics, law, business, medicine, institutional discourse, literature, advertising and many others (Wodak and Meyer, 2009). In addition to purely written text data, some CDA studies also analyse other semiotic modalities such as visuals, gestures and multimedia features. CDA's broader goal thereby is always promoting more equitable and just social practices.

Flowerdew and Richardson (2017) describe CDA as "an inter-disciplinary approach to language in use, which aims to advance our understanding of how discourse figures in social processes, social structures and social change". The primary unit of analysis is thereby always a text. CDA regards language as a form of social practice through which relations between social actors are expressed in discourse (Fairclough, Mulderrig, and Wodak, 2011). Through this social practice, ideologies in a community are reproduced and legitimised through the discourse itself (Wodak and Meyer, 2001). CDA, then, is a way to analyse textual data to uncover those underlying ideologies and beliefs, and aims to answer why a text has been written in a particular way, why certain linguistic choices have been made and what those mean in the wider, social context of that text. This includes not just ideologies but also other aspects such as social identities. This is the part that is the focus of this current work which aims to uncover intended meanings through close analysis of particular discourse features, in this case usernames and other, related social actors present in the forum discourse.

CDA is applied in contexts which are concerned with a social or societal problem. Child sexual abuse is one such problem researchers have a chance to tackle through their work, and online conversations between suspected offenders of child sexual abuse – interactions which predominantly occur through the means of written text – are therefore an area of research that clearly calls for a critical approach.

3.2.2 Corpus-Assisted Discourse Analysis

In its initial, more traditional form, CDA has always been concerned with the close analysis of a small number of texts which have been analysed by a single researcher and involved their own interpretations of linguistic features in a given social context. This has often been a point of CDA that has been criticised stating the potential researcher bias. One way to counter these effects is to employ more than one research method to increase validity of the conclusions made by the researcher or research team. One such way is to include a more quantitative approach, like corpus linguistics, to allow for the study of a larger text sample, making it possible to draw conclusions over a larger sample, and to uncover overall trends within the data (Riesigl and Wodak, 2001; Baker, 2006). Corpus linguistics covers a range of methods and techniques which aid researchers in analysing large collections of texts through quantitative observations regarding the studied text data and these quantitative descriptions of the data are extracted by means of specialised computer software (Brookes and McEnergy, 2020; Di Cristofaro, 2023). Unlike other approaches to analysing large sets of digital textual data, such as Natural Language Processing (NLP), which are often focused on numbers, descriptions and grounded in computer or information science, corpus linguistic tools are developed using theoretical knowledge from linguistic research to inform the quantitative analysis (Di Cristofaro, 2023). Corpus linguistics thereby follows the assumption that words and patterns of words can be analysed by examining their frequency of occurrence in a corpus as frequency offers a direct link to the meaning such words or phrases have in the language sample under scrutiny (Di Cristofaro, 2023). These patterns that can be uncovered using corpus tools might otherwise not be accessible for researchers, especially in very large data sets which would be impossible to annotate manually by a human researcher. Corpus linguistics tools can therefore aid the researcher in highlighting significant patterns of language use in the studied corpus, which can then be further analysed with more traditional, qualitative methods used in CDA research. Corpus-assisted CDA typically

follows an iterative process where the researcher moves between corpus methods and more traditional DA methods to allow the researcher to form new hypotheses about the data in question and test these through close reading of the data (Baker et al., 2008). Concordance tools are typically used to analyse qualitatively those significant linguistic patterns analysed by the statistical analysis. Corpus-assisted discourse studies (CADS) have been the choice of many CDA studies which focus on studying underlying beliefs and ideologies in large sets of digital text data, especially corpora collected from social media and online discussion boards (Baker, 2006; Baker et al., 2008; Booth, 2023; Koller, 2012; Krendel, 2020; Lorenzo-Dus and Kinzel, 2019; Wright, 2020).

It is to this end that the current work also employs such a combined approach to uncover insights into the language used by different types of CSA forum users – insights which would not be uncoverable through traditional, manual discourse analytic methods due to the vast amount of textual data available.

3.3 Linguistic Frameworks

The approaches discussed above offer researchers a variety of tools, concepts and frameworks to select for analysing corpora of different kinds of text data. This section introduces the linguistic frameworks chosen for the different steps of the analysis for the three studies presented in this work.

3.3.1 Thematic Analysis

Thematic analysis is a useful approach during the first stages of analysis to group observations thematically and get a first overview of what a text – or in this case a selection of texts and a list of usernames – is about. It can further be useful to highlight differences between different corpora, as is useful for study 2 and 3 in this work where themes across five subcorpora are investigated and compared. It is also a useful method to identify underlying meanings in textual data (Braun and Clarke, 2006) which is key in any enquiry employing a critical approach to analysing discourse.

In addition to the social constructionist view on language which the approach in this thesis takes, thematic analysis can also adopt a realist epistemology where language is treated as providing insights into individuals' cognition and motivation, enabling the researcher to theorise individuals' experiences and the meanings behind their utterances directly from their language use (Jaspal, 2020). Thus, thematic analysis is a flexible approach to language analysis which allows for the combination of different perspectives within the same study (Jaspal, 2020).

When applying thematic analysis, the researcher conducts a detailed manual review of the text data and codes the data for any themes emerging. According to Braun and Clarke (2006), thematic analysis aims to "identify, analyse and report patterns (themes) within data"(p.6). Usually, the researcher will thereby go through several iterations of data exploration and review, as has been done in this work by the author. Thematic analysis is an iterative, reflexive process (Mihás, 2023) and an entirely inductive and qualitative approach to textual analysis. Themes are defined as patterns of meaning which are observable across a corpus and they are formed through observations from the data in relation to the initially stated research questions of the project (Jaspal, 2020).

While themes are often described as "emerging" from the data, it is important to note that the researcher, in this case the author of this thesis, plays an important and active role in identifying those themes and assigning meaning to certain features within the data (Taylor and Ussher, 2001).

Furthermore, it is important to note that although thematic analysis has been used to qualitatively analyse textual data in many contexts, it originally is a framework developed for psychological research. Rather than being able to identify and make claims about specific linguistic features, thematic analysis provides a way to identify and describe more generally patterns of meaning within a text or across a set of texts (Braun and Clarke, 2006; Jaspal, 2020). It is therefore useful for the initial exploration of the data and a useful complement to the linguistic frameworks and methods applied to the textual data in this work.

3.3.2 Social Actor Framework and Transitivity

In his influential article on social actor representation, Van Leeuwen (2013) proposes a "sociosemantic" approach to Critical Discourse Analysis and introduces an inventory for analysing how participants in a given social practice are represented. It can be used as a tool which helps identify those representational choices and explore the beliefs and values underlying those choices. The approach is introduced and explained through an example; a text identified as racist discourse. By using the framework, the researcher reveals how the author allocated different roles to the different social actors in the discourse, how they have been either included or excluded, and how they have been categorised by the writer. In doing so, he uncovered the underlying ideology and beliefs of the author.

The social actor framework is commonly used in CDA studies to uncover hidden or underlying ideologies and power imbalances in discourses. It has widely been applied in a number of CDA studies to investigate social actor representation in, for example, news reports (Rasti and Sahragard, 2012) and translation studies (Babaii and Rashed, 2017). It is an approach in CDA which allows researchers to study the ways in which social actors are represented in discourse to uncover underlying beliefs and ideologies the author of the text holds. To this end, Van Leeuwen (2013) introduces a social actor (SA) network representing the linguistic choices an author can make to represent any social actor – be that groups, organisations, or individuals – in discourse.

The application of Van Leeuwen's (2013) SA framework to online discourses is still new but researchers have successfully shown its usefulness in a number of contexts. Krendel, McGlashan, and Koller (2022), for example, have analysed the representation of gendered social actors in the context of online misogyny in Reddit posts by members of the incel and Men's Right Activists communities, and were able to uncover underlying ideologies shared across those communities.

Most recently, Booth (2023) has used social actor analysis in her study on group identity and careers in a white nationalist online forum. Combining corpus linguistics with a CDA approach and

making use of both Van Leeuwen's (2013) framework for social actor representation and Halliday's (1985) transitivity analysis, Booth shows that forum members often disagree on the boundaries of their identities – both racial and ideological – and she raises the question whether the notion of the 'in-group' as it is often used and understood in the identity literature is still a useful concept when analysing such online groups. Booth argues that the 'in-group' in the white nationalist context should be regarded as a combination of overlapping, distinct identities – a view on group identity also highly valuable for the analysis of identity in this work.

In this work, the scope of the analysis of social actor representation is further extended to the study of individual identity, by investigating representations of the self in child sexual abuse interactions between suspected or potential offenders interacting in dark web discussion fora. The framework by Van Leeuwen (2013) for the classification and analysis of social actors is thereby applied to the linguistic choices made by members of those dark web communities, both for their usernames which clearly constitute a form of reference to themselves and within their contributed forum posts which form an extension of that self-identification through particular forms of reference to themselves and others.

3.3.2.1 Analysing Social Actors

To analyse social actor representation in any given text, Van Leeuwen (2013) developed an inventory of choices a writer can make when representing social actors in discourse. The inventory is used to show how social actors can be represented in discourse through a number of choices. It is based on Systemic Functional Linguistics, which is an approach which views language as a resource which allows authors to choose between different alternatives (Halliday and Matthiessen, 2013). This network is used as a guide to analyse social actors in the text or texts in question. The original system network of social actor representation is shown in Figure 3.1 below; Babaii and Rashed (2017) simplified the inventory later and provided a more accessible list of those discourse features seen in Figure 3.2 below.

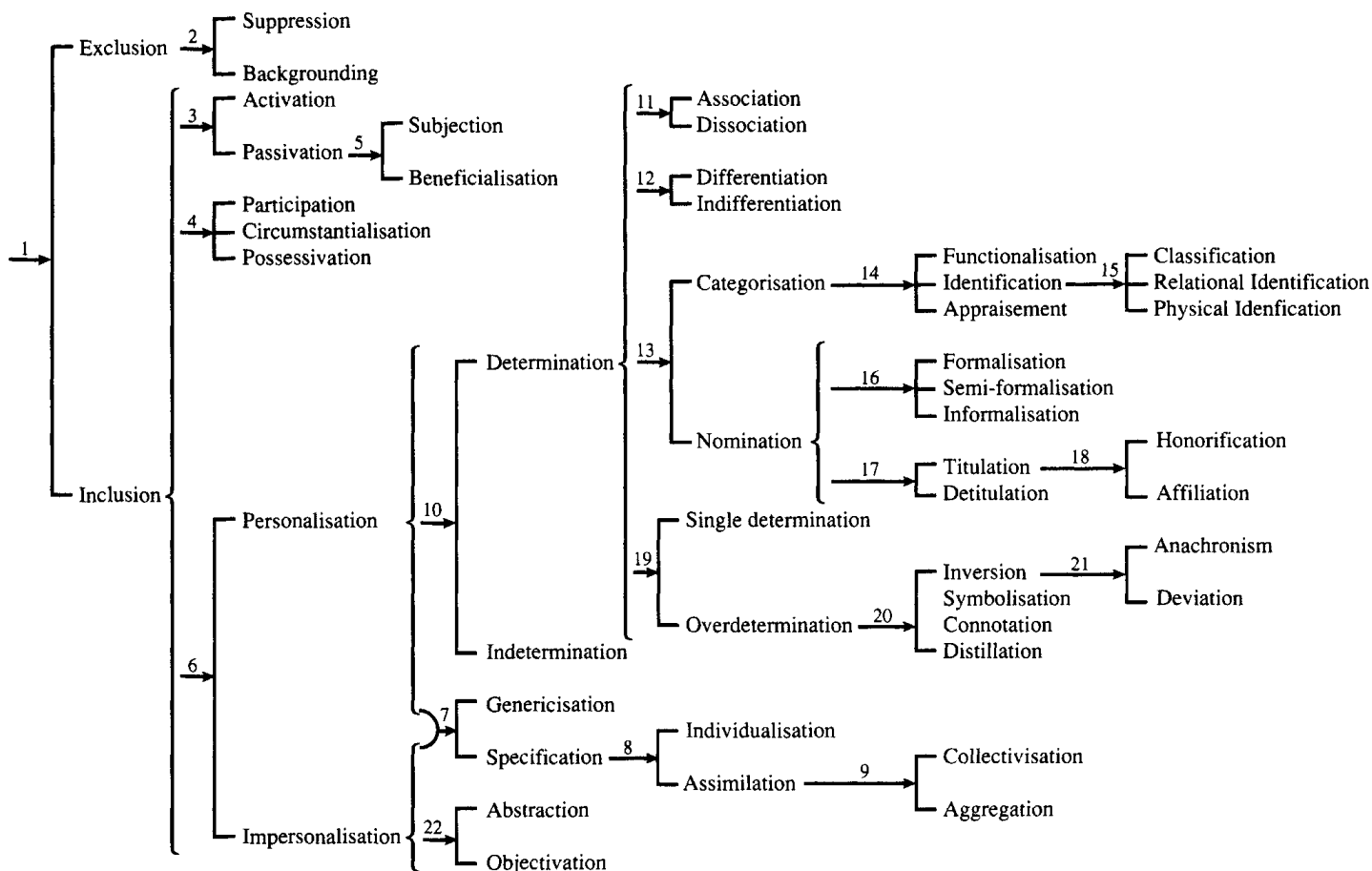


Figure 3.1 The representation of social actors in discourse: system network
 FIGURE 3.1. Van Leeuwen's (2013) network of discourse features for the representation of social actors.

In the Hallidayean view of language, every choice that an author makes carries a particular meaning which could have not been expressed in the same way if the author had made another choice. Applying this approach, researchers can analyse those choices systematically and uncover underlying beliefs, ideologies and intentions by the author of the text.

The first distinction made in the network is that of *inclusion* and *exclusion*. This is simply the distinction between the author's choice to include an explicit reference to a social actor, *inclusion*, or to completely exclude the social actor by not referencing them in the text at all. SAs excluded from the discourse can either be excluded with no trace through what Van Leeuwen (2013) calls *suppression*, or they can be excluded through *backgrounding*, where it is still possible to infer that an actor was involved but they are entirely deemphasised.

Table 1: The inventory of representation of social actors

Discourse Feature	Description
Inclusion & Exclusion	In any social practice, there are some social actors involved in the process. When representing the practice, some of these actors may be <i>included</i> (explicitly mentioned), some <i>excluded</i> . Exclusion with no trace is <i>suppression</i> (deletion). Exclusion with the possibility to infer the actor's involvement is <i>backgrounding</i> . In the latter case, the actor is de-emphasized, pushed into the background, but not completely deleted.
Role allocation	This feature concerns the roles that social actors are given to play in the representation. They may be <i>activated</i> (depicted as active, dynamic forces in an activity) or <i>passivated</i> (pictured as receiving end of the activity).
Genericization & Specification	Social actors can be represented through <i>generic</i> (as classes) or <i>specific</i> (as identifiable individuals) reference.
Association & Dissociation	This feature refers to forming and unforming the groups of social actors in a text as the texts proceeds.
Indetermination & Determination	<i>Indetermination</i> occurs when social actors are represented as unspecified, 'anonymous' individuals or groups, <i>determination</i> , when their identity is, one way or another, specified.
Nomination & Categorization	Social actors can be represented either in terms of their unique identity (<i>nomination</i>), or in terms of identities and functions they share with others (<i>categorization</i>).
Functionalization & Identification	<i>Functionalization</i> occurs when social actors are referred to in terms of an activity, in terms of something they do, for instance an occupation or role. <i>Identification</i> occurs when social actors are defined, not in terms of what they do, but in terms of what they, more or less permanently, or unavoidably, are.
Personalization & Impersonalization	Social actors can be <i>personalized</i> (represented as human beings) or <i>impersonalized</i> (through concrete or abstract nouns whose meaning does not include the semantic features of human).
Overdetermination	<i>Overdetermination</i> occurs when social actors are represented as participating, at the same time, in more than one social practice. Here the social actor comes to symbolize different social practices which may be related to one another, be opposite, etc.

Translation & Interpreting Vol.9 No.2 (2017)

FIGURE 3.2: Simplified list of Van Leeuwen's (2013) discourse features by Babaii and Rashed (2017).

156

For those SAs which are included, the original system network offer an array of choices for their representation. Authors can refer to SAs as individuals (*individualisation*) or refer to them in terms of groups they belong to. This can be achieved by either grouping individuals together and referring to them as a group of specific people (*collectivisation*), referring to such a group as a statistic

(*aggregation*) or by representing a group as an entire class without reference to specific individuals (*genericisation*). Individuals can be represented in form of *nominalisation* and Van Leeuwen (2013) here distinguishes between the following types of naming: *formalisation* (last name), *informalisation* (first name), *semi-formalisation* (both), *titulation* (Mr, Miss), *honorification* (Dr, Prof, Rev), or *affiliation* (reference through personal or kinship relation term). Nominalisations will become especially relevant in this work analysing usernames – which do not neatly fit into one of the original categories as can be seen in the analysis and discussion in chapters 4 and 5. Furthermore, social actors can be represented in conjunction with other individuals (*association*) or through *differentiation*, where they are explicitly distinguished from other SAs. Social actors may further be represented through *functionalisation* (what they do), *classification* (what they are) or through *relational* or *physical identification*, as well as by means of positive or negative *appraisement* or *impersonalisation*. As can be seen from the original network, there are a number of additional, very specific categories of SA representation which can be included in the analysis. Some which are of particular importance for the analysis carried out in this work are, for example, *relational identification* and *objectivation*. In addition, it can be seen that many of the above mentioned categories appear to be exclusive of one another through the visual representation in the network, whereas in reality a single social actor might be represented in ways which include different categories of opposite ends of the network – a key flaw in the original framework.

3.3.2.2 Transitivity Analysis

While the social actor framework's function is to investigate the lexical choices made by the community members to represent certain social actors, including representation of the self through their usernames, Halliday's (1985) transitivity model allows for a systematic analysis of verb processes according to the following six categories: behavioural, existential, material, mental, relational and verbal. Categorising the verb processes attached to the social actors identified before could help

reveal what kinds of agency and character were given to those social actors by the author of a specific post, further revealing a more detailed view of that person's underlying ideology and the groups' values. This is why Halliday's model is often used alongside the SA framework to further analyse authors' choices in the representation of self and others. Van Leeuwen (2013) himself includes it in his analysis in much of his work, including his initial influential article analysing "Our Race Odyssey" and introducing the original SA framework.

So while the SA framework is useful to study representations and language choices made regarding how to represent a certain individual or group, transitivity within systemic functional linguistics offers a way to investigate actions and processes in relation to such actors. Transitivity is divided into participants and the processes or actions they are involved in and allows to answer the question: who does what to whom (Halliday and Matthiessen, 2013).

The six process types to carry out the transitivity analysis together with their subtypes and participants have been summarised for ease of access and application, and are found in Table 3.8 below.

Transitivity analysis which includes both processes and participants allows the researcher to analyse representation of social actors beyond just the noun phrase and thereby adds analytical value to this corpus-assisted study of usernames and social actors. It allows, for example, to distinguish between *passivation* and *activation* of social actors in the discourse. This approach differs from much of the traditional works employing transitivity analysis where the author of the text is typically just one person and participants can easily be assumed and remain the same throughout the text. In the analysis provided in this thesis, texts are written by a large number of different authors and it could therefore lead to additional, valuable insights to further investigate which SAs fulfill which participant role with regards to a certain verb process.

Process Type	Process	Participants
Material	Doing	actor - carries out the action goal - affected by the process
Mental	Sensing/Experiencing Sub-types: perceptive (seeing, noticing, hearing) cognitive (thinking, remembering, understanding) desiderative (wanting, agreeing, intending) emotive (liking, fearing, regretting)	senser - the one who experiences phenomenon - the thing being experienced
Relational	Identifying/Classifying Sub-types: intensive (A is B) possessive (A has B) circumstantial (A is at B)	Neither entity effects the other
Verbal	Saying	sayer - the one conveying the information target receiver (indirect) - the one being spoken to verbiage - the one being spoken about
Behavioural	Behaving (sleeping, breathing, laughing, crying)	behavior
Existential	Existing/Happening	existent

TABLE 3.8: Table introducing the process types available for transitivity analysis based on Halliday and Matthiessen (2013) including subtypes and participant types.

3.3.2.3 Difficulties with Social Actor Analysis

As can be seen from the section above, the original system network introduced by Van Leeuwen (2013) is very complex and not easily applicable and not always intuitive through the many available categories. Furthermore, categories might overlap, and those overlapping categories might lie on different node groups of the network, indicating that they should be exclusive when they might indeed co-occur. This is a point of particular importance for the current study, as usernames are made up of different segments which function differently from proper names or other representations typically found in other text types. The original network was developed analysing larger written texts, such as fiction books and news articles. Applying it to newer types of discourse, especially online discourses, means that it might need to be more flexible and perhaps extended or further simplified.

Darics and Koller (2019) recently developed a simplified way to study agency using social actor representation by introducing a new toolkit to explore questions of agency and SA representation

in discourse. Their approach guides the researcher along a number of questions to examine those representations. Following their model, the following questions need to be answered to reveal underlying ideologies and beliefs:

- Who is absent or implicitly or explicitly present in the text? (If social actors are absent from a text, have they been strategically excluded)?
- How are social actors represented: as active or passive, as more or less agentive, and in personal or impersonal ways? The next question is how can actors be represented in a text? Are they included as active participants in or passive recipients of an action? In addition, what degrees of semantic agency are allocated to them?
- What is the relationship between communicator, audience, and third parties? Having examined the language used in the text and established how it is employed to include or exclude social actors, activate or passivate them, bestow different degrees of semantic agency on them, and refer to them in personal or impersonal ways, we might want to zoom in on the relationship between those social actors and the author and audience of the text: Who is the creator of the text? Who is the text intended for? Is the audience explicitly mentioned in the text? What third parties are mentioned?
- Who is explicitly present in the text, what actors are implicitly present, and who is absent? Are the social actors depicted as active or passive, what degree of agency is allocated to them, and are they referred to in a personal or an impersonal manner? What kind of relationship is created in the text between the communicator, the audience, and third parties?

The first three sets of questions are concerned with identifying social actors, the final set helps the researcher recognise and reveal the possible intentions of the author. Following these sets of questions, any text can be analysed for social actor representations, agency and ideology without having to rely on the original, much less intuitive systems network developed by Van Leeuwen

(2013). While developed as a toolkit for business communication, Darics and Koller's (2019) toolkit also offers a great resource for researchers studying other discourses to find their way into the data and focusing on those linguistic choices regarding social actor representation.

Another issue arising with social actor analysis is its original intention for studying long texts written by one individual, such as media or newspaper articles. Applying this approach to those types of texts helps uncover underlying ideologies held by either an individual author or a certain group, such as in the cases of news media from a specific political background. When applying the framework to newer forms of text, such as social media or forum data, the texts are usually much shorter and even if we have a lot of text data available for one author, this is often not from one single text and might therefore have different intended audiences which makes it harder to interpret certain lexical or grammatical choices. The few studies to this date that have attempted to study social actor representation within larger corpora have taken the approach of extracting keywords from the corpora with the help of corpus software to then focus on the analysis of those SAs which occur frequently in the text and are thereby assumed to be representative of the text of group of texts (Alzahrani, 2013; Booth, 2023; Sánchez Moya, 2018). This same approach is taken in the current work, where subgroups of users are analysed (see Chapters 5 and 6) collectively through keyword analysis.

In the case of dark web communications like the ones analysed in this work, texts are also typically short and instead of having large amounts of text data for each individual, a large collection of textual data by different authors is available. This is why, for the purpose of this study, users and their naming choices as well as representations of self and others are examined not just on an individual but rather on a group level. While specific choices of self-representation within a username might be unique to the individual, choices of self- and other-representation in both usernames and forum contributions which match across users hint at subgroups of users with the same set of values, beliefs and ideological views; their intentions behind certain representational choices are assumed

to be the same or at the very least similar to one another.

While there are certain issues with the original framework introduced by Van Leeuwen (2013) and newer, adapted or simplified versions like the ones by Babaii and Rashed (2017) and Darics and Koller (2019) are valuable additions to the literature and help students and researchers new to the framework in understanding it better and carrying out a number of useful analyses, there is also value in applying the original framework with all its detailed categories it has to offer. Even though not all categories have been relevant for the current work, following the original framework has been beneficial in identifying details and minor distinctions between users' representations of themselves and others, and the author has therefore made the choice to apply the original framework and analyse social actor representation in subgroups of the studied dark web fora by closely comparing representational choices to the nodes in the original systems network.

3.4 Limitations

The research presented in this thesis does come with a number of limitations both due to restrictions and constraints for which parts part of the data have been analysed in detail but also due to limitations to do with the analytical approach and the analysis frameworks chosen.

Due to time constraints, certain avenues of investigation have not been further explored but could lead to important insights about users' naming choices and identity work within OCSA contexts.

As introduced in the methodology section, as a common accompanying analysis tool, transitivity analysis offers a valuable addition to social actor analysis as it allows to consider participants (social actors) and the processes they engage in together when investigating issues regarding social actor representation but also identity. Especially in the context of online forum discussions, where there are many different authors of many different shorter texts, including an additional analysis step investigating transitivity for common types of username – and thereby identity – parts would add

another layer to the results presented in this study and are a clear avenue for future research on this and similar data sets.

Examining and annotating only roughly a third of the usernames available in the original corpus certainly led to some avenues not being explored. The initial exploration of the full data set revealed a number of potential username categories which have not been identified (or only in very low frequency) among the reduced corpus. While the justification for the reduction still holds (see section 4.2.1.2) and users who only contributed one or two posts over six years certainly are less relevant and not suitable for the analysis of deeper identity work issues in the current context, studying those infrequent posters' naming choices could still lead to useful insights. Examining those username choices further could, for example, allow researchers to compare them to high frequency posters and give insight into what makes a successful vs an unsuccessful name choice when trying to gain acceptance in this community. For computational work on username construction and naming choices, a larger annotated corpus would also be beneficial for training and testing purposes of such systems. A fully annotated set of usernames or username segments would also facilitate comparative research to better understand how similar or different naming choices in the present context really are compared to other online fora, for example in gaming or self-help communities.

In the current study, usernames with less than 20 occurrences per category among the most frequent posters have been automatically excluded from further analysis due to low frequency. Some of those user groups, however, could nonetheless be of high relevance to the OCSA context and especially the policing of such online spaces. For example, studying the linguistic behaviours of users choosing to self-identify through their role within the (online) community or as new members of the community could be important for policing tasks when trying to gain a better understanding of the roles and hierarchies within a community of child sex abusers. In her work investigating the moves and strategies of new users attempting to gain access and acceptance in an online CSA community, Chiang (2024) goes a first step into that direction and this could be build on by analysing

the contributions by users in the current data set who self-identify as new users. Another such group could be those users choosing to identify through their ethnicity, nationality or geographic location as part of their name choice. If an analysis of their forum contributions would hint at those username segments actually reflecting their true – offline – locations or nationalities, this could be a potential lead in investigations when trying to locate and identify offenders through their online presence and behaviours.

The corpus analysis of username segments and segments pairs showed that some of the most commonly used segments within members' usernames are alternative spellings or misspellings of the same type of segment and carry the same meaning, such as the segments *lover* and *lvr*, *young* and *yng*, *ped* and *pedo*. When looking deeper into those alternative spellings for different kinds of segments in the full corpus, it is revealed that for some segments a much larger set of alternative spellings is present, such as *girrrl*, *g1rl*, *gurl*, *gril*, all meaning *girl*, or *pedo*, *pedro*, *p3do*, *P3D0*, *ped*, *pd*, all meaning *paedophile*. To facilitate a much more precise and accurate analysis of username segments using corpus tools, and to be able to generalise over similarly spelled items with the same meaning, possibly even over several languages, a second version of the username segment corpus with normalised spellings would be beneficial and the author hopes to conclude this step before publication of any further research on the present data set. Having such a normalised list of username segments also allows for better visualisation of the data when presenting findings to non-expert or non-academic audiences.

Finally, the choice of reference corpora for study 2 should briefly be discussed here as well. While the author provided justifications for the choice of reference corpora used for the keyword analysis in the second study, she is aware that other corpora exist which might seem more suitable to the given research context. Reference corpora specific to online text data, such as the TenTen corpora, would grant an even better insight into the studied online communities and how their contributions compare to more general online language use. While such a large reference corpus with more

comparable language data is obviously desirable for a project like the current one, at the time that decisions had to be made on which corpora to choose, the TenTen corpus was only available through SketchEngine, an online corpus tool which a) the researcher could not obtain a license for through their institution for the entire length of the project, and b) as an online tool was not suitable since the data investigated in this work was not to be uploaded anywhere and only stored on an encrypted drive and analysed on a local machine.

3.5 Ethical Considerations

When undertaking a research project like this, certain ethical considerations need to be made. This section discusses ethical issues regarding data collection and storage, data access and distribution, publication of findings, and researcher wellbeing.

Although operated on the dark web, the discussion fora analysed in this work can be considered public data as it is possible to access the fora with the Tor browser and by creating a new user account. It is important to mention here that except for those users who have shared IIOC in the fora, it is unknown to the researcher if the users analysed in this work have committed any crimes. It does, however, show their willingness to commit certain offences or their interest in discussing potentially criminal acts with others online. It is unknown to the researcher if any of the described criminal acts actually took place or if any of the analysed individuals have been charged with or prosecuted for any child sex abuse, trafficking or distribution offences. Therefore, the choice was made to refer to individuals in these fora as *suspected offenders* rather than *offenders*.

Investigating unlikeable or criminal subjects or analysing (textual) data by (suspected) offenders and other types of distressing and sensitive data, as is often the case within the field of forensic linguistics, comes with a number of challenges. Barber (2021) discusses a number of such challenges and offers a list of questions researchers should ask when approaching research which examines (potentially) vulnerable subjects. One question researchers must ask is if their research subjects are

vulnerable or susceptible to harm. It is unlikely that the users who are the subjects of my research will experience mental or physical harm through the research but they might be legally vulnerable within or outside of the context in which they have published their content online. Publishing results could make them vulnerable but since usernames are not published together with actual posts, and usernames are not linked to actual offline identities, this harm is highly unlikely.

Another aspect to take into account is if publishing research with subjects subscribing to extreme ideologies will directly or indirectly be disseminating or promoting such ideologies. Barber (2021) raises the question of whether acknowledging and reproducing such ideologies can be harmful in itself. She argues that it is key to balance researcher bias and subjectivity with scientific objectivity. This is also true for the current study: while there is a minimal risk that some findings detailing views by the studied subjects could indirectly promote such views, the benefits of this study for policing and prevention of child sexual abuse offences, online and offline, clearly outweigh such risk.

Barber (2021) further highlights that researchers always need to consider if victims might be revictimised through conducting or publishing their research. Grant and MacLeod (2020) also discuss the need for protection of victims and in their research solve this by not approaching victims for consent and by anonymising the data. Since in the current work the victims' identities are unknown, this has not been considered but the risk of revictimising the children who are being discussed by the suspected offenders in the dark web fora is very low; proper names given by forum members are likely not real as they would make detection of the studied individuals offline easier. Therefore there is low to no risk of direct revictimisation, though there is of course always the slightest possibility a victim might read a publication on this data and recognise a situation described. However, since in this thesis no full accounts given by users are reproduced and only individual sentences or partial sentences are reproduced for the purpose of illustrating findings, this again is highly unlikely to occur with the presented work in this thesis.

Another important point that Barber (2021) raises is that of researcher desensitisation and the

risk of harm to the researcher. For the studies in this work, there were no images or videos analysed which minimised the risk to an extent; the text data itself, however, contained some harrowing details and it was impossible to know beforehand what topics exactly would be discussed by the forum members. One way to minimise the risk or mitigate the impact of analysing such data discussed by Barber (2021) which have also been applied for the current work is the direct and regular exchange with colleagues who work on similar types of data and are familiar with the subject. Further, the author's supervisory team and a psychologist appointed by the Aston Institute for Forensic Linguistics (AIFL) have been consulted as well. Barber (2021) further discusses the risk of being targeted by the individuals or communities studied. For this work there currently are no publications completed and the author has so far refrained from sharing details on social media about the details of the research they are conducting due to this risk factor. While this has potentially protected the researcher from harm, it had a negative effect on career progression and networking as the author has not been able to share her research findings with a wider audience outside a very narrow audience working on similar data.

Finally, Barber (2021) describes the problem of being an outsider of the community that is studied, something also highly relevant for the current work. As the researcher is not part of the community studied, holds opposing views and is not fluent in the community's language, it can even be challenging to decipher 'familiar terms' (Barber, 2021) or know the meaning behind commonly used acronyms; the author of this work has made use of online dictionaries and consulted literature as well as with individuals policing such online spaces to mitigate this factor.

Kredens, Haworth, and Atkins (2022) addressed issues and challenges regarding ethics of forensic language data in their talk at the 2022 IAFLL conference in Porto. One of the key ethical questions discussed is that of anonymisation. While there are a number of ethical and legal requirements to adhere to when embarking on a forensic linguistics research project, how to match these requirements is not always straight forward and often rules or recommendations regarding anonymisation

are vague or ambiguous even, especially when it comes to the collection and preparation of forensic language data. Research councils and university ethics frameworks dictate that researchers have to respect participants rights to privacy, confidentiality and anonymity but in many cases where forensic language data is concerned this is only partially possible, or not possible at all. With forensic speech data, for example, this can never fully be anonymised at the analysis stage but researchers can anonymise when disseminating findings. In other cases, it might be easier to remove identifying information but what exactly constitutes identifying information is often left to be decided by individual ethics committees or researchers. In the case of the present study, the forum members' usernames themselves are the focus of the study, therefore a lack in anonymisation is to be expected as the researcher needs to provide examples to the reader to be able to effectively present and discuss findings. By conducting the initial segmentation process of the usernames, however, the author has found a way to study the members' usernames without having to provide all of them as full usernames. The list of full usernames is only available to the examiners of this thesis and will not be published anywhere. Examples of real usernames to illustrate findings have been carefully selected to avoid sharing any identifying information. Further, since the researcher does not have any information regarding the users' offline identities and users in the studied context are highly concerned with anonymity themselves, there is hardly any identifying information to share in the first place. Kredens, Haworth, and Atkins (2022) further highlight the challenge of defining what constitutes 'identifying' information and show that there's potentially a much larger pool of options considering that authorship analysis as a forensic linguistic discipline itself is concerned with finding ways to identify individuals without details present which are commonly understood as identifying information (names, addresses, date of birth, place of work, etc.). Other important considerations Kredens, Haworth, and Atkins (2022) discuss are the representativeness of examples chosen to illustrate findings when the researchers are highly concerned with anonymity as well as the importance of audience. They highlight that features which are of linguistic interest are often identifying in

nature, especially seen from an authorship analysis perspective, therefore some researchers decide to not present such findings to general audiences or to swap real examples with constructed ones. This, however, poses a risk of falsifying data and skewing findings and could even cause career damage by not sharing valuable findings – both issues relevant for the project presented here. Further, they argue, it weakens the strength of the research and prevents useful practical outcomes, e.g for policing. As the current work is concerned with findings being applicable to the policing context, this is highly relevant and an argument for discussing genuine examples from the data – here and in future publications.

To ensure the wellbeing of the researcher, other researchers in the AIFL and others potentially affected by exposure to the data in this project, the following precautions have been taken: All data is stored on an encrypted data medium that is kept locked away in the secure data lab at AIFL to prevent third-party access, as has been agreed per Aston University's ethical approval process. The full ethics approval form can be found in Appendix A. Further, to prevent onlookers from exposure to any kind of sensitive material, all research and analysis on the data sets has been carried out at the university office only. Since the data could reveal information about unresolved criminal cases, the researcher also had the obligation to report any identifying information found throughout data analysis to the relevant authorities.

Finally, the potential harm to the individual researcher needs to be considered thoroughly. Literature on the topic shows that there are a number of risks to be considered and an array of mitigating factors and coping strategies are available.

Burns et al. (2008) studied the impact of working on CSEA materials. In their study, practitioners reported a number of issues that impacted them. This includes awareness (or a lack thereof) of the scope of their work which can lead to being not properly prepared or overwhelmed with the task. Practitioners also reported physical and emotional impact such as fatigue, headaches or the inability to fulfil responsibilities in the home. Further, intrusive thoughts or images were common but

dependent on the type of content and the length of exposure to it. Practitioners further reported they found it difficult to not be able to talk about the materials they were exposed to but did not want to traumatise others. They also commonly felt they had gotten overly protective or even paranoid regarding children and their safety. Strategies to better cope with this type of work used by participants of Burns et al.'s (2008) study included gradual introduction to material, mental preparation, dissociation and compartmentalisation, monitoring self, taking breaks, remaining analytical (just data), as well as planning where, when and how to engage with the data. Burns et al. (2008) further found a number of mitigating factors which included personal characteristics, supervision, humour, sense of control, organisational support, social support and psychological support.

Elstein and Kredens (2020) interviewed a number of forensic linguistic practitioners to better understand occupational stress in this line of work. They find that forensic linguists make use of a number of coping strategies to protect themselves against harm such as burnout and vicarious trauma. These coping strategies are similar and partially overlapping with the ones found in Burns et al. (2008) and include desensitisation, exchange with others, clearly separating work and personal life, preparing oneself for reading the material and, importantly, recognising their work as a positive contribution to society. A number of these coping strategies have also been used by the author and are discussed in the following sections, with a strong focus on separating work and personal life as well as social and psychological support.

Since working with sensitive and potentially distressing data could cause psychological harm to the researcher, appropriate access to psychological support has been provided to the author of this work. For researchers working at Aston's Institute for Forensic Linguistics (AIFL) this is standard practice and support was ensured through regular meetings with a visiting psychologist provided through AIFL. As an additional precaution, the data was kept on campus and no analysis work requiring close reading of the data has been carried out from home.

Even when a number of precautions are taken and everyone involved in the research does their

best to protect the researchers' wellbeing, there is always the potential for unexpected harms. For this author it meant that even when not working on the data from home, accessing the psychological support provided and working in a supportive environment at AIFL which fosters exchange and collaboration with colleagues working on similar distressing data sets, there have been times where the author had to step away from the data for a while to protect themselves. This has happened either when topics discussed by the suspected offenders in the data set have been very graphic – such as in the *Violence* subcorpus – or when topics discussed by forum members unexpectedly related to the author's own life. This is something one can never be fully prepared for, the importance is to recognise when researchers need to step away and take a break from analysing such distressing data and to know where to turn for support; in this case this was colleagues, friends, professional psychological support provided by the AIFL and the supervisory team. It was further important to always be prepared to slightly change the direction of the project should the researcher want distance to part of the data set due to personal reasons. This has always been an option for this project, has frequently been discussed with the supervisory team at various points throughout the PhD journey, and it has impacted some changes made for the empirical chapters.

Chapter 4

Study 1: Username Categories in Child Sexual Abuse Fora

Usernames form an important part of any forum user's online identity and can even determine the person's success within an online community. When deciding on a username to join an online community, individuals have a number of choices available to them. These linguistic choices are in part based on restrictions set out by the platform – like a maximum number of characters allowed in a given forum – but are predominantly based on the individual's own resources and constraints as well as their knowledge of and familiarity with the online community they are attempting to join. In the context of CSA, other considerations such as that of entering a low-trust/high-risk environment can further impact the choices users make when creating their usernames. At the same time, users still have a need to be recognised within the community and might reveal certain parts of their identity through their usernames. Because of that, others are able to make certain conclusions about the individual based on their username.

To find out what kinds of usernames are present and prevalent in an online community of suspected child sexual abuse offenders, it is necessary to find out which parts usernames are commonly made of and what sorts of themes these parts cover. To this end, usernames in this study are first segmented to separate the individual meaning-bearing segments of a username and then categorised thematically.

The chapter is structured as follows. The first section covers a brief recap of the research context including related literature into dark web communications and username research and links the analysis to the policing of online crimes and spaces. This is followed by an introduction of the data set of usernames. Following this, this chapter presents the analysis process including initial observations, the username segmentation process as well as the annotation and categorisation of username segments, and a detailed review of the findings which covers all identified categories and patterns that emerged from the data through a combination of quantitative results and detailed corpus examples. This is followed by a discussion of frequent and infrequent forum posters as well as issues regarding reliability and generalisability of the findings. The chapter concludes with a discussion of the findings.

4.1 The Data Set: CSA Fora and Usernames

In this chapter, the analysis is based on two different (sub)corpora. In the first step, initial observations regarding the types of usernames present in the data have been collected and described based on a manual reading of all 32,410 usernames in the full DarkWebCSA data set. The DarkWebCSA corpus includes data from eight dark web fora frequented by individuals discussing and committing child sexual abuse. Those fora have different foci, such as general exchange of IIOC, discussion of CSA of boys/girls/babies, or the distribution of particularly violent and sadistic CSAM. A table detailing the characteristics of each forum can be found in 3.2 in the previous chapter.

In a second step, usernames for the top 10,222 posters (post frequency \geq 3) from the full DarkWebCSA data set have been analysed. This set of 10,222 usernames is referred to as 'reduced corpus' in the remainder of the thesis. The 10,222 usernames analysed in this chapter were first split into their individual parts by means of a manual segmentation process (details in section 4.2.1 below), then those segments (and segment pairs) were analysed and – together with observations

from the previous step – emerging themes and patterns were described and presented as a classification of 20 username categories. Each category was also analysed for social actor representation.

4.2 Analysis and Findings

In the first part of this section, the procedure for data preparation as well as segmentation, categorisation and annotation of the usernames is presented. This is followed by the presentation of overall themes emerging from the data, a description of the categories identified and an analysis of most common segments and segment pairs across all usernames. Then, findings regarding the frequency of users' contributions to the discussion fora are discussed and linked to the findings from the segmentation and categorisation task. Finally, issues regarding reliability and generalisability of the findings are discussed.

4.2.1 Categorisation and Segmentation of Usernames

This section first introduces the steps taken for preparation of the data set and the manual segmentation process including initial data exploration, all illustrated by examples. It then describes the annotation process for assigning categories to the usernames and their individual segments followed by a detailed description of the different categories together with example usernames for each of the described categories.

4.2.1.1 Data Exploration: Initial Observations

Before attempting any structured annotation or classification of the data, the author first examined the full set of 32,410 usernames manually to extract some initial observations and trends within the data to further structure the analysis process and to identify some initial types of usernames which form the basis for the developed username categories in the later chapters. After the initial data

exploration, initial observations have been collected in a holistic diagram presented in figure 4.1 below.



FIGURE 4.1: Holistic diagram showing initial observations regarding naming themes and patterns from the full username corpus.

The diagram includes a rough classification of the type of username categories present in the data but also additional naming features such as graphology, typography, multimodality and the use of numbers. Themes observed by the author cover a range of topics that users include in their usernames to craft their online identities: users’ self-identification through features such as gender, age, appearance, ethnicity, nationality or occupation; the use of proper names (own or not, real

or fictional); community roles; explicit anonymity; sexual preferences and sexual orientation; self-identification as paedophile; references to hobbies, interests, music, film, TV, literature or gaming; references to mythology or religion; animals; colours; or brand names. Further, users tend to use numbers within their names for a number of reasons: to indicate age, to refer to sex or sexual acts, to refer to other dates or even bible verses. Numbers are also used in a creative way to replace phonemes, letters or even entire words within the name with a number. Numbers and letters are also commonly used seemingly randomly in form of sequences or as keyword sequences covering a number of letters or numbers which occur in the same line on the computer keyboard. Other creative decisions include alternative spellings, repetitions (of words, numbers, letters), the use of punctuation, as well as the indication of word boundaries through hyphens, underscores, capital letters or spaces. Some users also choose entire phrases or sentences as their usernames, others combine a number of the mentioned features to engage in word play. The size of the bubbles in the diagram do not indicate salience, bigger bubble sizes are used to indicate top-level themes as identified by the author. It should also be noted here that the colour coding in the diagram is used mainly as a visual aid to differentiate between different kinds of features identified, to aid readability of the diagram but also serves as the basis for developing the classification scheme.

From these first observations, a number of common themes or username types have already become apparent, such as names, age, preferences, family relations or self-identification as paedophile. The more detailed analysis of username segments in this chapter and further analysis of usernames of the top 1837 users in Chapter 5 confirms their importance in the OCSA context.

Next, the usernames had to be prepared for the segmentation process and further analysis. A complete description of those steps is given in the following sections.

4.2.1.2 Data Preparation: Segmentation of Usernames

Identifying common parts and patterns in usernames used by child sex abusers online can aid understanding of what kinds of resources individuals draw on when creating their usernames which

can in turn aid law enforcement in performing more convincingly during identity assumption and infiltration tasks as well as increase understanding of how their usernames relate to individuals' relevance within the online community and therefore for the investigation. To identify any themes and patterns, the usernames were split into segments to annotate each of those segments and to carry out any further analysis.

This segmentation of the usernames has been carried out manually by the author for 10,222 of all 32,410 usernames. The remaining usernames have been omitted from the detailed analysis for two reasons: first, the remaining usernames belong to users who have posted only once or twice over the entire time span analysed and have therefore deemed less relevant as they don't offer a lot of text per user to be analysed in terms of identity performance and linguistic expression and therefore are less relevant to the law enforcement and policing context which is one of the main motivations for this project; and second, due to time constraints and restrictions put in place during the COVID-19 pandemic¹ which meant the decision was made to exclude the usernames for the least frequent posters from the analysis. More details regarding post frequencies are discussed in section 4.2.3. The 10,222 usernames have been split into meaningful parts which in most cases translates to segmentation into several words or splitting compound words into individual ones where appropriate. Other segments that have been separated from the rest of the name are acronyms and additional standalone letters, as well as standalone numbers, strings of numbers and strings of letters. Any punctuation within the usernames has been recorded as individual segments as well. Table 4.1 below illustrates this on a number of examples from the data set. This then created a full table of usernames and their individual username segments. The segmentation has been carried out using an Excel spreadsheet where each segment of the username translates to one cell, with one row of the spreadsheet per username. A full list of the 10,222 usernames and their segments is provided in Appendix G.

¹In accordance with the ethics guidelines of Aston University's Research Ethics Committee, the data had to be kept within the Aston Institute for Forensic Linguistics on an encrypted drive and locked in the secure data lab, and the author was not able to work from the university office for a number of months due to government lockdowns and then only on select days for a number of months during 2020 and 2021.

Username	S1	S2	S3	S4	S5	S6	S7
Plum	Plum						
pedodaddy	pedo	daddy					
secretmum	secret	mum					
turkishmen999	turkish	men	999				
Torpedo1965	Tor	pedo	1965				
luvptpuss46	luv	pt	puss	46			
themanwhoknocks	the	man	who	knocks			
xxxlolilolixxx	xxx	loli	loli	xxx			
lovetohurtem69	love	to	hurt	em	69		
TheBigBadWolf69Ridinghood	The	Big	Bad	Wolf	69	Ridinghood	
I_Love_Green_Olives	I	_	Love	_	Green	_	Olives

TABLE 4.1: Examples of usernames and their respective username segments to illustrate the manual segmentation process.

4.2.1.3 Segmentation Results

The segmentation results for the full set of 10,222 usernames can be seen in Tables 4.2 and 4.3 on the following pages. The first table provides the list of top segments (based on frequency) drawn from this reduced set. It includes all key segments (extracted through analysis in the corpus tool LancsBox) with a frequency of 10 or above. The top segment pairs are presented in the second table which includes key segment bigrams extracted with the corpus tool. All top segments pairs with a frequency of 3 or higher were included in the analysis. A closer look at those tables shows a number of prominent themes:

Numbers A total of 1413 number segments are present within the top segments. They refer to years, ages, or age preferences, but also include a number of sequences such as *12345*. The high frequency of low numbers such as 1, 2, 4, and 3 suggest that these are numbers added to a name in cases where the username was already taken and the number is added to distinguish that member from another with an almost identical name.

Names Among the key segments in the reduced corpus, 290 segments are proper names. These are names referring to either the member themselves, by real name or most likely a pseudonym, or

Segment	Freq.	Segment	Freq.	Segment	Freq.	Segment	Freq.	Segment	Freq.
lover	238	young	30	12345	20	smith	15	nepi	11
boy	204	ass	30	star	20	lolita	15	oh	11
pedo	171	13	29	1234	20	sex	15	up	11
the	162	to	29	dr	20	angel	15	onion	11
man	158	toddler	29	for	20	111	15	kinder	11
love	149	bad	29	anal	19	l	14	time	11
1	146	p	28	am	19	day	14	tiny	11
2	144	master	28	bi	19	der	14	you	11
girl	97	new	28	on	19	66	14	jimmy	11
i	96	cat	28	6	18	jb	14	captain	11
123	95	7	27	wolf	18	them	14	world	11
loli	90	d	27	rape	18	mister	14	max	11
baby	79	not	27	her	18	7	14	johnny	11
4	73	kids	27	fucker	18	horny	14	2016	11
mr	73	5	26	jack	18	c	13	deep	11
69	72	hard	26	2018	18	g	13	bum	11
12	67	hurt	26	dog	17	n	13	james	11
x	66	cp	26	1	17	cute	13	h	10
little	65	no	26	21	17	white	13	r	10
daddy	56	joe	26	green	17	fox	13	t	10
dad	55	red	26	bl	17	pt	13	pink	10
-	54	and	26	el	17	sexy	13	andy	10
a	51	big	26	so	17	nice	13	doe	10
me	51	teen	25	lvr	17	devil	13	0	10
dark	50	is	25	like	17	papa	13	19	10
666	49	fun	25	0	16	feet	13	32	10
3	48	mom	25	cunt	16	2015	13	mark	10
guy	48	child	25	it	16	101	13	play	10
girls	48	pussy	25	xxx	16	u	12	hc	10
fan	47	king	25	licker	16	tod	12	that	10
john	42	j	24	billy	16	tom	12	hebe	10
luv	42	15	24	luvr	16	18	12	curious	10
one	40	23	24	lol	16	25	12	here	10
s	39	dick	24	cock	16	55	12	kiddy	10
of	39	my	24	ped	16	lord	12	lust	10
11	38	9	23	perv	16	em	12	crazy	10
uncle	38	b	23	dude	16	dream	12	mad	10
boys	38	10	23	all	16	yng	12	pretty	10
m	37	in	23	k	15	hidden	12	lick	10
lil	37	yo	23	14	15	dreamer	12	magic	10
bear	36	kid	23	33	15	cool	12	loving	10
bob	35	good	22	77	15	peter	12	evil	10
.	34	tor	22	xx	15	2000	12	pedro	10
anon	33	just	22	hot	15	2017	12	shadow	10
22	33	cum	21	real	15	test	11	heart	10
99	33	17	21	candy	15	2	11	wonder	10
88	31	who	21	super	15	20	11	sam	10
black	31	777	21	old	15	34	11	100	10
fuck	31	sweet	21	life	15	44	11		
user	31	8	20	blue	15	raper	11		

TABLE 4.2: Top segments (frequency \geq 10) in the 10,222 username corpus.

Bigram	Frequency	Bigram	Frequency	Bigram	Frequency
boy lover	19	love little	4	am i	3
i love	19	to fuck	4	bum bum	3
girl lover	19	i need	4	cum on	3
i m	14	pedo girl	4	star dust	3
i am	14	a boy	4	boy raper	3
loli lover	14	lover 123	4	12 yo	3
baby lover	9	bla bla	4	baby girl	3
little boy	8	billy bob	3	bi pd	3
tor pedo	8	girl 2	3	4 u	3
i like	8	man 123	3	boi lover	3
john doe	8	cookie monster	3	joe joe	3
mr .	7	uncle bob	3	the prof	3
pedo bear	6	lover of	3	lover 2	3
4 me	6	yo girl	3	- man	3
lover boy	6	love 2	3	cherry popper	3
the one	6	h .	3	boo boo	3
in the	5	pedo gril	3	i luv	3
pt hc	5	why not	3	the librarian	3
lil girl	5	a lot	3	little anna	3
lil boy	5	the guy	3	who s	3
baby boy	5	cp luv	3	old man	3
of the	5	papa bear	3	the bear	3
just me	5	like em	3	bi ped	3
it s	5	g lover	3	mic mic	3
boy love	5	pure evil	3	active dad	3
little girl	4	dark side	3	loli pop	3
pedo mom	4	the real	3	yo love	3
on the	4	wang rong	3	baby love	3
bi pedo	4	m a	3	love them	3
man 69	4	lover 69	3	diaper boy	3
feet lover	4	teddy bear	3	teen boy	3
love girls	4	lover 50	3	4 kids	3
child lover	4	dude dude	3	little girls	3
hurt meh	4	one and	3	loli luvr	3
not a	4	love is	3	4 boys	3
4 life	4	rong jun	3	big daddy	3
ass lover	4	dad of	3	pedo pedo	3
lover 1	4	anon anon	3		

TABLE 4.3: Top segment bigrams (frequency \geq 3) in the 10,222 username corpus.

to the name of a child victim. Except for *doe* and *smith*, these all refer to first names. In the top segment pairs, 26 such pairs refer to a proper name.

In addition, titles or honours such as *mr*, *master*, *king*, *dr*, *mister*, *lord* and *captain* all also appear among the top segments, with *mr* being the most common one with a frequency of 73.

Gender 669 segments among the top segments refer to gendered terms. These might refer to the member’s own gender or to their preference in the victim’s gender. The analysis of top bigrams also reveals a number of segment pairs referring to gender. They are displayed together in Table 4.4 below.

Gender	Freq.	Preference:gender	Freq.	Gender	Freq.
boy	204	boy lover	19	man 69	4
girl	97	girl lover	19	pedo girl	4
man	158	lil girl	5	girl 2	3
guy	48	lil boy	5	man 123	3
girls	48	baby boy	5	pedo gril	3
boys	38	boy love	5	the guy	3
dude	16	little girl	4	dude dude	3
m	37	love girls	4	- man	3
b	23	g lover	3	old man	3
		boy raper	3	diaper boy	3
		baby girl	3		
		boi lover	3		
		diaper boy	3		
		teen boy	3		
		little girls	3		
	669		87		32

TABLE 4.4: Top segments and segment pairs in the reduced corpus referring to gender.

Preferred victim age A large number of users in the reduced corpus also refer to their preference in the child victim’s age through their usernames, with 551 individual segments and 86 segment pairs in the key lists (see Table 4.5).

Self-identification as paedophile Another common theme among the top segments and segment pairs is users referring to themselves through their identity as a paedophile – with 218 segments and 38 segment pairs among the top most frequent keywords and bigrams. The details are shown in Table 4.6 below.

Preference:age	Freq.	Preference:age	Freq.
loli	90	loli lover	14
baby	79	baby lover	9
little	65	little boy	8
lil	37	pt hc	5
young	30	lil girl	5
toddler	29	lil boy	5
kids	27	baby boy	5
teen	25	little girl	4
child	25	love little	4
kid	23	12 yo	3
yo	23	baby girl	3
lolita	15	little anna	3
l	14	yo love	3
pt	13	baby love	3
tod	12	diaper boy	3
yng	12	teen boy	3
kinder	11	little girls	3
tiny	11	loli luvr	3
kiddy	10		
	551		86

TABLE 4.5: Top segments and segment pairs referring to the child victim's age.

Paedophile	Freq.	Paedophile	Freq.
pedo	171	tor pedo	8
ped	16	pedo bear	6
nepi	11	pedo mom	4
pedro	10	bi pedo	4
hebe	10	pedo girl	4
		pedo gril	3
		bi pd	3
		bi ped	3
		pedo pedo	3
	218		38

TABLE 4.6: Key segments and Segment pairs showing self-identification of the user as paedophile.

Childlovers Another apparent theme visible in the key segments and segment pairs is users identifying as "childlovers". 543 key segments and 47 key segment pairs can be grouped to belong to this category of usernames, making this one of the most prevalent themes together with those segments and segment pairs referring to the victim's age. Table 4.7 below shows the details:

Childlove	Freq.	Childlove	Freq.
lover	238	boy love	5
love	149	child lover	4
fan	47	lover 1	4
luv	42	lover 123	4
lvr	17	lover of	3
luvr	16	g lover	3
bl	17	lover 69	3
like	17	lover 50	3
		boi lover	3
		lover 2	3
		yo love	3
		baby love	3
		love them	3
		like em	3
	543		47

TABLE 4.7: Key segments and segment pairs identifying the user as "childlover".

Sex and Violence The final significant theme emerging that will be discussed in this section is that of sex and violence, represented with 268 key segments and 29 key segment pairs in this selection. It is a concerning finding, showing that a large number of users in the reduced corpus feel a need to include these terms in their usernames as part of their online identity. The table below shows the details:

Sex and violence	Freq.	Sex	Freq.	Violence	Freq.
hard	26	man 69	4	pt hc	5
hurt	26	to fuck	4	hurt meh	4
anal	19	lover 69	3	boy raper	3
rape	18	cherry popper	3		
cum	21	cum on	3		
deep	11				
fuck	31				
fucker	18				
licker	16				
raper	11				
lick	10				
hc	10				
sex	15				
lust	10				
cp	26				
	268		17		12

TABLE 4.8: Key segments and segment pairs referring to sex and violence.

4.2.1.4 Annotation: Username Classification

Once all usernames had been split into their parts, usernames had to be annotated with the categories they belonged to. The approach for developing the relevant categories has been predominantly inductive. While some categories had been selected in advance based on previous research on child sex abusers as well as literature on computer-mediated discourse, and online discussion fora and online naming in particular, the majority of categories evolved from the data itself and new categories were added or existing categories refined as the exploration, segmentation and annotation processes progressed. The final classification system consists of 20 categories, which includes sub-categories, and can be seen in Figure 4.9 below.

As different parts of a username can belong to different categories, most usernames have been annotated as belonging to more than one category. This further allowed the author to investigate categories that commonly co-occur and therefore reveal frequent patterns in the data. The classification was carried out manually by the author using an Excel spreadsheet with one row per username.

In addition to the above categories, usernames have also been annotated for multimodality, typography and graphological features. This includes features like capitalisations, repetitions, punctuation, alternative spellings and numbers standing in for letters, words or phonemes. A detailed discussion of such features and their prevalence in the data set follows in Section 4.2.2.2.

4.2.2 Naming Themes and Patterns

With a data set of this size, many observations can be made and discussed to gain insights into the community studied. For example, among the entire data set, usernames consist of a minimum of one and a maximum of 7 username segments, with the shortest usernames just being one character long, such as *X*. These types of findings, however, give little to no insight into users' identities. To this end, the username segments have been studied for their (intended and/or assumed) meanings and grouped by category to get the big picture of how members of child sexual abuse communication

Category Name	Sub-categories	Example Usernames
Age		Torpedo1965, kids1977
Animals		Badwolf, blackbird, frog32, TheCat
Brands		DrPepper1, Jackdaniels80
Childlove		BabyBoyLove, lilgrllvr, CLover, MikeTheBL, kidlove, childlover25,
Colours		redmike, white_bear, blackcat
Dark/Evil		dark666666, evildevil1412, MrDevil
Dark Web		darkweb183, torfan2, toruser44
Explicit Anonymity		AnonFrance, anon59, AnonymousFox
Family Relations	Parents Grandparents Aunts/Uncles Siblings Family Spouses	AbusiveDad, sweetdaddy, pedodaddy, Mom2Katelyn, secretmum GranddadUK, nepigrandad niecelover, UncleGeorge, Uncle Bob FrenchSister, luvmysis pedofamily, familylover03 PedoCouple, curiouscouple
Film, TV & Literature		tylerdurden, anakin, wakanda, 1gandalf1, puffmagicdragon
Food & Drinks		Candy, Tomato, pizzaman, Plum
Gender	Female Male	secretmum, PedowomanGER pizzaman, pedodaddy, themanwhoknocks
Names	Own Victim Popular Culture Known Paedophiles Titles	Dimitri, HappyHarry, MikeTheBL Mom2Katelyn, Krissfan, Sammylover1 tylerdurden, RonWeasley nabokovisreturned, JosefFritzl DrPedo, mrcomputer23, MrNO
Nationality, ethnicity, location		englishdaddy, turkishmen999, DutchMike, Swedenped, AnonFrance, GranddadUK, PedowomanGER, ArizonaBob
Occupation	Law Enforcement Other	totallynotFBI, FuckAllLEA PedLibrarian, Researcher
Paedophile	Paedophilia Hebephilia Nepiophilia	newped1, pedodaddy, Pedobear, Pedo4Life hebelover2, unclehebe nepigrandad, shynepi2016
Preference: age	Baby Toddler Preteen Teen	babylover9004, BabiesRFun, Ben4Babies ToddlerAddict, tods4me luv10yogirls, luvtpuss46 lolilove210, youngloli, xxxlolilolixxx
Preference: gender	Male Female No preference	Ygboylover, BoyzInDiapers, mad4boys girlzlover2014, ilikegirls12, younggirlhunter BiPed, biped515
Preference: other	Body Parts Sexual Acts Appearance	buttllove, cuntlover, ilovepussies cumongirls3, child69, suckmeofflg darkgirlsllove, Blondlover
Religion		DarkMessiah, Matthew1914, crazymonk
Sexual Orientation	Gay Bisexual	gayMatt, gaybeast BiPed, biped515
Violence		sadist1, ChildRapist, hurtgirls22234

TABLE 4.9: List of all username categories identified in the full corpus with examples for each category.

fora identify themselves through their usernames. The following sections discuss in detail the 20 categories that emerged, together with some other distinct features identified in the initial stages of analysis.

The reader will find that at times the findings presented in the following sections present some vagueness regarding their quantification. The reason for this is two-fold: as the findings in this chapter are based partially on the initial observation gathered through manual reading of the full DarkWebCSA user list, which was a purely qualitative analysis in form of close reading of the data, there are no numbers to accompany this part of the analysis; in addition, an overlap between the categories annotated after segmentation of the usernames further prevented quantification of the results.

4.2.2.1 Emerging Username Categories

This section explores and describes emerging themes and patterns in the data which are drawn from the initial exploration and review of the username data from the full corpus (all 32,410 usernames) and the above described analysis of the most common segments and segment pairs within the reduced set of 10,222 usernames which has undergone the manual segmentation process.

Following the segmentation and classification procedure, a number of themes emerged from the analysis of the usernames and have been summarised as categories below. The 20 identified categories are described and discussed in alphabetical order in the following pages.

Age As in many other online communities users in the examined dark web CSA fora also often indicate their own age as part of their username. In most cases, this is done explicitly through including their birth year as one of the username segments, such as in the the usernames *whiterabbit72* or *penny1989*, or through mentioning their exact age in years as part of their usernames, as done by user *Mutti26*.

It is important to note here that while many users include this information as part of their username, it is unclear to what extent users are truthful here and include their actual age or birth year. Given the high-risk environment in which they interact, it is likely that some may choose to pick a year or age which is similar to their own but not exact. In many cases it is also unclear if users include their birth year at the end of the username or if they simply include a random number as the username without numbers was already taken.

A number of users also decide to just hint at their own age by describing themselves as *young* or *older*, or by indicating they are a grandparent through their username. While this doesn't give the reader an exact age, it does allow the reader to form a first impression regarding the individual's age based on those descriptors.

This form of self-identification through age is clearly a form of classification in terms of social actor representation, that is, social actors which are referred to based on major classes society uses to differentiate, which includes age. Which ever way users choose to include hints about their own age within their usernames, they do so to classify themselves as belonging to a certain group or class of people within society and it is apparent that this forms an important part of users' online identities, perhaps also due to conventions already established in other online communities.

Animals Animals as part of members' usernames appear to be a popular choice by members of these fora. Examples include *Moose*, *whitebear* and *blackcat*. 66 of the most frequent posters include references to animals in their usernames. The most common choices include *bear*, *cat*, *rabbit* and *wolf*. While some of these usernames might be references to popular characters in children's TV and literature – such as in *whiterabbit72* (Alice in Wonderland) or *TheBigBadWolf69Ridinghood* – many appear random. Some could, however, be linked to the person's offline identity through associations between the animal and person characteristics and identity features – such as cleverness for the fox, being old and wise like the grey wolf, or being big, calm and lovable like a panda.

In addition to this, connections between naming conventions in CSA communities and other communities may need to be considered here. With *bear* being the most popular choice, one might wonder if it connects to the use of the term in male gay culture, where the term has been used to describe bigger (often obese) and hairier men. Within bear culture, men of different age, appearance or ethnicity might have different names within the community: younger men are often described as cubs, older men as polar bears, and the terms Black Bear, Brown Bear and Panda Bear often refer to gay men of Black, Hispanic and Asian ethnicity (Moskowitz et al., 2013). Since the emergence of bear culture in the 1980s, the term has gained in popularity and signifies values of self-acceptance and maturity through members consciously adopting a masculine aesthetic (Manley, Levitt, and Mosher, 2007). As it is such an established term in gay culture where it has developed into its own subculture, it is likely that individuals in the CSA community have adopted some of this language to hint at their offline identity, including sexual orientation, through usernames such as *whitebear*, *Pedobear* or *SadPanda*.

Connecting the above to social actor representation, one can see how users' self-referral by means of an animal can be understood as either identification – by relating the physical appearance of a certain animal to the person – but also as a form of impersonalisation, namely abstraction, where a social actor is not represented as being human but instead by means of a quality assigned to them by the representation used – in this case, the qualities of a particular animal.

Brands A number of brands are commonly included in forum users' usernames, with example names such as *2CocaCola* or *heineken*. The most commonly referenced brands are those of drinks, both alcoholic and non-alcoholic, and could simply be a reference to the user's favourite beverage or alcoholic drink. Other popular choices are car brands or tech companies with these most likely again showing the user's preference for a particular type of car, device or service. These are username choices also found in other online contexts such as general interest discussion fora.

There are also some username choices more specific to the CSA context. References to sweets

commonly known to be brands enjoyed by children are one of these types of naming choices. These just imply the reference to children or paedophilia without explicitly mentioning this connection.

Another brand that is commonly used as part of users' names is *Pampers*, often used synonymously for diaper or nappy. In those cases the brand stands in for the actual item and the user, through the brand name in their username, shows their sexual preference for children young enough to still wear diapers. These are individuals known as nepiophiles who have a sexual interest in infants or toddlers. This is an example for one way in which users hint at their victim preference through their choice of username. Other examples of this practice are discussed further in paragraph 4.2.2.1.

While there might be certain associations connected to the brand chosen by a user, in cases where the brand is the only meaning-carrying segment in the username, it does not refer to an individual directly. In terms of social actor representation, this clearly constitutes exclusion in form of backgrounding: the individual who is referred to with the username is not directly mentioned (included) in the name choice but the reader can reasonably assume who the person is. Without context this is, however, hard to identify. But even if unclear to the reader if the username might refer to the forum member themselves or perhaps refers to their victims, in both cases the social actor is backgrounded.

Childlove Usernames that fall into this category are those that explicitly self-identify as childlover, boylover or girllover as part of their usernames. This is either done through including the full term as a username segment, or alternative spellings or shortforms of the same terms. Example usernames in this category are *Childlover2016*, *LennyCL* or *girlluva*.

Within CSA communities, individuals often distinguish between those who willingly hurt children, both physically and emotionally, and are interested in inflicting pain to the children, and those individuals who believe themselves to be in loving relationships – often called pseudo-intimate relationships in the psychology literature – with the children and believe that what they call "childlove" is something that not only does not harm the child but actually is beneficial to the children.

With 67 users among the top 1837 users in the data set self-identifying as childlovers of some sort, this is also a significant subgroup of individuals in this community. While few include other details along them being childlovers – such as their preferred age or gender of the child, or themselves being a parent – most users in this subgroup choose to solely identify through their childlover identity when choosing their username for the community. This suggests that members belonging to this subgroup see this as a significant part of their own identity and as the single most significant identity feature to share with others in the community, highlighting their opposition to certain sexual acts performed by child sex abusers who engage in and enjoy explicit violence towards their child victims.

In terms of SA, this type of representation constitutes differentiation, a form of determination where individuals or groups are referred to by means of how they differ from others, drawing a clear distinction between the 'self' and the 'other', between in-group and out-group. Furthermore, this choice of username can be seen as a form of positive appraisal, as users evaluate themselves as being the good ones through this lexical choice. By self-identifying this way by choice of username, members not only self-identify as part of one particular group, they also clearly distance themselves from another.

Colours Colours form part of a number of members' chosen usernames in the CSA community. Among the top 1837 users, 25 have included a colour as a segment of their username. Most of these are used to further describe another part of the username, such as in *white_bear*, *gingerbeard222*, or *blackninja*. A number of these usernames (7 across the top users) use it to further describe an animal. As we have seen already in the discussion of the Animals category, this can often be an attempt to disclose a certain part of the user's offline identity, such as their physical appearance, or could also indicate their age (*white*, *grey*) or their gender (*blue*, *pink*). All those uses of colours to refer to themselves constitute some form of categorisation in social actor representation, either by means of classification or physical identification.

The use of colours as parts of users' chosen online names has also been reported in other onomastic studies and therefore appears to be a common naming practice across different types of online special interest communities.

Dark/Evil Users falling into this category openly self-identity as bad, dark or evil in some way, indicating their awareness of sexual relations with children being wrong – not necessarily in their own view but society's view of them and their actions. Segments such as *dark*, *evil*, *devil* and *demon* are common among usernames in this category. Self-identification in this way clearly constitutes a kind of (negative or positive) appraisal, a form of categorisation, within SA representation. It is thereby unclear if users evaluate themselves and their social practice as something negative or positive. As a community that opposes mainstream views, much like other countercultures, it can be assumed that in many cases members of a pro-CSA community would positively appraise their being evil and possibly even hold pride with regards to their abuse practices. A keyword-in-context (KWIC) analysis with LancsBox of the terms annotated for this category was conducted to explore this further but this has been inconclusive: no data using such terms to describe themselves through negative or positive appraisal has been found and users did not engage in any discussions of their chosen usernames for this category either.

In many cases, username segments indicating dark or evil co-occur with segments related to religious beliefs, suggesting that not only do these users acknowledge society's non-acceptance of their sexual preferences and acts but they also are aware of the views shared among religious communities (in most cases Christianity) regarding child sexual abuse. This can be seen in examples such as *DadgoingtoHell*, *Antichrist* or *SinSeeker*.

Dark Web This constitutes one of the categories which appear to be more specific to the communicative and social context of the corpus. Within this category, usernames explicitly referring to dark web terminology are found. Examples in this category are username segments such as *tor*,

onion (both referring to the software "The Onion Router") or *darkweb*. By referring specifically to the software used to stay anonymous online, these users highlight both their wish for anonymity but perhaps also their expertise with using the software. Social actors represented by means of a social practice they regularly engage in, something they do, are represented by functionalisation but in this case belonging to a group of people with expertise in a certain area with which they identify strongly could be seen as a form of classification. In addition, as clear from usernames such as *toruser123456*, it is also used to directly highlight their anonymity as key part of their online identity in this community.

Users choosing to classify themselves as experienced users of encryption software and dark web environments might do so to highlight their expertise and position themselves on top of the hierarchy common for these types of online communities. In their work investigating power resources in, among others, online CSEA interactions on the dark web, Newsome-Chandler and Grant (2023) found nine power categories in their data used by users to make claims to power and assert authority in their interactions. Usernames of users claiming to be experienced users of dark web technologies could be seen as using one such power resource, namely 'technological expertise', something particularly valued in this high-risk context.

Explicit Anonymity A large number of users in the data set have chosen a username which explicitly mentions their (wish for) anonymity. This type of representation can be understood as either abstraction, a form of impersonalisation in the SA framework where individuals are represented by means of a quality assigned to them by the representation, in this case 'anonymity', or as a form of categorisation by which the social actor is personalised in some way but characterised entirely through their anonymity. Usernames including the segments *anon* and *anonymous* are most common. While explicit anonymity as part of the username is a significant feature over the entire data set, only few of the frequent posters chose such a username, indicating that frequent posters wish to be recognised and identified in some way, such as through their sexual preferences,

their demographic information or their interests outside of the forum topic's scope.

Explicit anonymity being extremely popular with those users who rarely post in the forum suggests that such users have no interest in building any sort of relationship with the rest of the forum community and therefore have no need to share any identifying information as part of their usernames.

Interestingly, a handful of users explicitly stating their status of anonymity do include another descriptor which hints at their offline identity, such as *France* in *AnonFrance*. This suggests that some users' need to identify through their gender, location or sexual preferences might be greater than their need for complete anonymity, or that offering such a small identifying detail within their community raises their status within the community, both in terms of authenticity and trust as well as recognisability.

It should be noted here that other users might indicate their (wish for) anonymity through other choices within their usernames, such as forum member *toruser123456* does through their use of the segments *user* or the keyboard number sequence *12345*. These segments, however, have not been annotated as explicit anonymity but are discussed under other categories.

Family Relations Identified as usernames which include references to family relations are all those usernames where the user chose to either self-identify as being closely related to the child or identify themselves through their relationship to a spouse or their family as a whole. This representational type of social actor constitutes relational identification in Van Leeuwen's systems network: SAs are represented by their kinship relations.

Those identifying themselves through their relation to the child most commonly identify as either mothers, fathers, or uncles. But other relations, such as *sis* in *pedosisluv*, do occur, although rarely. These types of usernames could also be seen as classification in SA terms, where the users classify themselves as being a parent, grandparent, etc. and could be used as a display of power and

authority by claiming 'topic expertise through personal experience' (Newsome-Chandler and Grant, 2023).

A number of users choose to identify through their relation to another family member, most commonly their spouse, such as *wife* in *NonPedoWife*.

Finally, there are usernames within this category indicating more than one person is involved in either the abuse of a child or running the user account. Such usernames include segments such as *family* in *pedofamily*, or *couple* in *curiouscouple*.

Film, TV and Literature Including popular characters from film and TV is already a well-known choice for usernames in other contexts. Users in dark web fora discussing child sexual abuse seem to be no exception. Popular choices by members are characters from popular book/film series such as Harry Potter and Lord of The Rings, as seen in examples such as *RonWeasley* or *1gandalf1*. These choices could either hint at the author's offline character or appearance, or simply at those being their favourite film or book series, all equating to some form of identification through classification or possibly physical identification.

More relevant for the particular context, however, might be choices relating to children's literature or TV shows, as well as those names referencing works which directly relate to the topic of paedophilia or child abuse. Examples with references to children's titles are usernames such as *timmyturner* (The Fairly OddParents) or *gru* (Despicable Me). These might be referencing titles that the adult uses in interactions with children in their own life, or they might simply be chosen for being know characters and therefore being more easily remembered by others in the community. In addition to the SA categories already discussed, these choices could also be understood as forms of backgrounding as the social actor, the individual suspected of engaging in child sexual abuse, is not referenced directly but understood to be the actor by the reader.

Those names in this category which directly reference paedophilia or the sexual abuse of children clearly are chosen for that exact reference. They might indicate furthermore what victim preference an individual has and can be seen as a form of classification in SA terms. An example of this is the username *hardcandy*, referencing the 2005 film of the same name which features the character of a teenage girl who communicates with a paedophile online and tricks him into meeting her and attempts to get a confession from him admitting his criminal acts.

Food & Drinks Users choosing references to food and drink, most commonly fruit or vegetables, as parts of their usernames are common in this data set. This ranges from simple one-segment usernames such as *Plum*, *Tomato* or *Cookie* to more complex names such as *Peaches Everywhere*.

The author can only speculate as to the meanings behind those choices but some explanations appear to be more likely than others. Users might choose such a username for 1 of 3 reasons: 1) their personal preference for the actual food, 2) the food represents a nickname of some sort, such as *Cookie* or *Candy*, referring to a real person, either themselves or a victim, or 3) to refer to a particular body part and thereby a possible sexual preference by the individual. The latter is likely for names such as *Peaches Everywhere* – with peaches referring to bums – and *cherrypoppy* – referring to the act of taking someone's virginity as "cherry popping".

The above possible interpretations mean that in terms of social actor representation, these usernames could be understood as either 1) backgrounding or classification, 2) nomination, or 3) backgrounding of the adult and objectivation (somatisation) of the child.

Gender Users in the investigated fora often include information about their own gender as part of their usernames; identification through classification in terms of SA representation. Common segments of usernames in the data include *man*, *guy*, *girl*, and *boy*. Others indicate their gender by including the segment *f* or *m* – for female or male, respectively – into their usernames or by revealing

their gender through a first name which is commonly understood to be female or male. Examples within this category are *PolishGuy*, *anon-man*, *islandwoman* or *Pedogirl1992*.

Another common way a user might indicate their gender identity with their username is through proper names. Many users do include a first name of some sort, often a name that is commonly understood to be used for only one gender. If the name used is the person's actual first name in real life or not is thereby irrelevant; the choice is made purely to indicate the individual's gender to other users. This hypothesis is further supported by the common use of the segment pair *john doe* which is commonly understood as reference to a male deceased person of unknown identity. Proper names as popular username segments are further discussed as their own category in the next section.

Names Proper names are the most commonly used username segments in this corpus. Given the high-risk discourse context and users' assumed wish for anonymity to avoid detection by law enforcement, this appears to be a key finding that requires further analysis. Users thereby choose to either include just a first name or just a last name, to combine a first and a last name, or to include a title as well. The types of proper names introduced in the username choice do, however, not always refer to the member themselves, at least not in a direct sense. In the paragraphs below, the different types of names used as parts of people's usernames are introduced and discussed alongside examples for each type.

A member's own name The most obvious choice of a username would appear to be someone's own name – either in full, in part or through a nickname or pseudonym. In this high-risk context, however, it is highly unlikely that a member would, for example, choose to reveal their entire name. When usernames just include a first name, like in *UncleGeorge* this could, however, reflect their actual first name offline – a form of nomination in SA terms. Since it is a common enough first name, the individual likely does not risk detection through the name but other users can easily

remember the name and refer to them by that name, just like we are used to in offline environments. Users who included both a first and last name in their chosen username, likely are using either a made-up name or the name of a well-known person or fictional character (both discussed below).

Celebrities Many users chose a celebrity's name as part of their username. This can be politicians, actors, musicians, etc. Examples for this are *donaldtrump4561*, *stanlee*, or *GreenDay*.

Others choose names referring to well-know CSA-related individuals, those that are either known for or suspected of having committed child sexual abuse. These include usernames such as *nabokov-isright* or *JosefFritzl*.

Users selecting usernames of the second kind, clearly choose such a name to identify themselves as a paedophile or child sex abuser to the rest of the community without having to reveal anything more about their own offline identity. It might also show others their expertise in the subject by knowing of that particular person, their story and possible allegations or convictions against them. This choice constitutes a form of classification within the SA network and may further include positive appraisal of the individual being referred to.

Fictional characters Another type of username which includes proper names is that of fictional characters. Users hereby choose names of popular characters from film, TV or literature as their usernames. Examples are the usernames *Alex DeLarge* (A Clockwork Orange), *RonWeasley* (Harry Potter) or *1gandalf1* (LOTR). Reasons for choosing such a name may be either that they themselves enjoy that particular work of fiction and the character referenced is simply their favourite character in the story, or users might choose the character based on specific person characteristics which are meant to invoke those associations in the reader to perceive the user the same way. The latter might be true, for example, for a user such as *FreddyKrueger*. All those cases constitute a form of classification through the choice of someone else's name but can also be perceived as a form of backgrounding of the individual who chose the username.

Victims' names Another type of usernames is that of users who name their victims in their usernames. This is done partially to show admiration for a particular child, but also to show appreciation for abusers who produced a certain set of IIOC. It can also show a user's expertise as the username choice shows they know about a particular set of images or videos. Alternatively, it could be used by the forum member as a strategy to let others know what kinds of IIOC they are sharing. Example usernames include *VickyFan4Life* or *CuteLittleAnna*.

In terms of SAs, these usernames can be understood as a combination of classification (individuals who enjoy the same kind of content), backgrounding (of the author, in some cases) or even as a form of positive appraisal (of the child and/or the creator or distributor of the IIOC content).

Titles In addition to proper names, a number of users also decide to include titles or forms of address as parts of their username. This includes username segments such as *Mr*, *Miss*, *Lady*, or *Dr*. While in some cases this might constitute part of the name of a well-known person or fictional character, in other cases it can serve the purpose of revealing a detail about the member's offline identity.

Connecting this to the social actor framework by Van Leeuwen (2013), users choosing to include some form of title in their usernames perform self-representation through titulation. Titulation is usually a form of nomination and is typically used to represent a social actor through their unique identity. Nomination is typically realised through proper nouns, especially proper names. This seems true, to some extent, for usernames such as *LadyLucy* or *DrJones*. However, in the dark web CSA context studied here, titulation appears to also be used as a form of categorisation instead. Categorisation is present when social actors are represented through functions or identities they share with others – their social or identity categories. Examples from the corpus such as *mrX* and *mrpedo* show that in these cases it is much harder to know this username to reference a specific person. As the most commonly used titles in the corpus are Mr, Miss and Dr, and they often are paired with other segments which suggest categorisation of some sort – either through relational or

physical identification – titulation in this context can be considered as a form of categorisation itself, typically gender. This suggests that titles are used by users in online CSA fora when they want to stay anonymous but at the same time deem it important to share their own gender already at the time of creating their username. In this sense, titles in this context have the same function as other gender markers, such as the segments *man* or *girl*.

Nationality, Ethnicity and Location While users on dark web child sexual abuse fora are generally concerned with anonymity to protect themselves against protection and prosecution by law enforcement, their need for community and being recognised within that community means that users commonly choose to reveal some sort of identifying information regarding their own person within their usernames. When crafting their usernames, many users include details regarding their racial, ethnic or national identity by including non-English language words, locations, and national or ethnic descriptors, as well as slurs or derogatory terms referring to someone's race, ethnicity or national background. Example usernames among this category include *englishdaddy*, *GranddadUK* and *PolishGuy*.

These types of details to refer to oneself clearly constitute a form of identification through classification according to Van Leeuwen's social actor categories. Users describe and represent themselves as belonging to a certain group defined as a kind of class by society.

Occupation A small number of users in the reduced corpus include references to their occupation as parts of their usernames. Within the reduced set of usernames, three usernames each feature the segment pairs *the librarian* and *the prof*. Other segments in the reduced set referring to someone's occupation are *teacher* or *cop*.

One group of users should receive particular mention within this category and this is users referring to themselves as (not) being a member of a law enforcement body. Example segments for this category are *LEA*, *cop* and *agent*.

In SA terms, self-identifying by means of occupation clearly constitutes functionalisation, a form of categorisation, that is, social actors represented through something they do. This suggests that these users also see themselves predominantly in this view and highlight parts of their identity that have to do with actions and activities, like their occupation, rather than through identification – even though it could be argued that some perceive their occupation as something they permanently are rather than something they do. Whichever the case, clearly to those users the fact that they belong to a particular occupation is key enough to make this the first thing communicated to other community members through the choice of username. When being negated though, as in the case of the username *totallynotFBI*, the user represents themselves by means of differentiation, by precisely what they are not.

Paedophile Many users choose to openly declare themselves as paedophiles through their usernames. The segments *ped* and *pedo* are among the most commonly used ones across the entire data set. Both in the reduced corpus and in the subcorpus of the 1837 most frequent posters, a large part of the user base included such a segment in their usernames – 218 and 34 respectively.

It is important to note, however, that while in many cases self-identification as a paedophile seems to merely serve as a descriptive feature – a case of identification by means of classification in SA representation – , in some cases this description is accompanied by an evaluation by the user: Examples such as *ProudPedoPapa* suggest the user views their own paedophilia as a positive aspect of their personality, while other members, such as *pedointhecloset*, exhibit a more negative view of themselves and paedophiles as a whole. Others appear to view paedophilia itself as neutral but assign positive or negative value to specific acts. The user *responsiblepedophile* clearly is an example of one such forum member who believe they are different from other paedophiles as they behave "responsibly".

Preferences Many usernames explicitly mention the users' preferences regarding their victims. This includes preferences regarding age, gender or ethnicity of the child, as well as preferences for certain body parts or explicit sexual acts. In doing so, users refer to different social actors in different ways: while the author themselves identifies by means of classification in terms of their sexual preferences, the child victims are represented by either classification through major demographic identifiers, such as age or gender, or through physical identification and simultaneously objectivation by being reduced to certain body parts.

The most common preference included explicitly as part of a member's username in the present data is the gender preference. For both the reduced corpus and top users corpus, the segment pairs *girl lover* and *boy lover* feature in the top three segment pairs. With such a large portion of users in the reduced corpus mentioning their preference for either male or female children explicitly through a segment (or several) in their usernames, this type of identifying feature appears to be one of the most important and successful within this Community of Practice (CoP).

The second most important preference by potential child sex abusers is the mention of age preference. This is done either explicitly through the choice of segments or segment pairs such as *loli lover*, *baby lover* or *12 yo* – or implicitly. The implicit mention of preference is done through specifying another preference further, such as age, as in the segment pair example *teen boy*, or through self-identification by the user as paedophile, hebephile or nepiophile, but these cases have already been discussed separately in the previous section titled *Paedophile*.

Religion While only two such usernames have been identified among the top segment pairs within the core user group, it is still worth mentioning here as a number of low frequency users have also opted for a religiously inspired username.

The username segment pair occurring twice within the top users corpus is *matthew 1914*, referring to the following bible verse:

Jesus said, 'Let the little children come to me, and do not hinder them, for the kingdom of heaven belongs to such as these.' (Matthew 19:14, NIV)

The fact that the same verse with such clear reference to the context of the forum occurs twice within a corpus of 1837 users suggests this verse and the religious context to be of high importance to those users and also of some relevance to the community. At the very least, choosing a username so explicitly religious has not hindered these two users from becoming an integral part of the community, having provided 685 and 27 posts contributions, respectively.

So while not one of the most prominent categories, there clearly is a subgroup of users who do identify strongly with their religious beliefs and deem it necessary and important to share this directly through their usernames by means of representing themselves through classification – belonging to a group of people of a certain faith community.

Sexual Orientation As discussed in the Paedophile section above, many users in the CSA context identify themselves through their paedophilia (or hebephilia, nepiophilia). There is evidence that users do see this as a form of sexual orientation and wish for it to be recognised as such by society.

There are, however, also other users who do specify their sexual orientation – such as gay or bisexual – explicitly within their usernames. Within the reduced corpus, there are, for example, 7 users who have chosen *gay* as one segment of their username. Self-identifying by means of sexual orientation constitutes classification in the SA network by Van Leeuwen (2013).

Interestingly, out of those forum members in the reduced set whose username includes the segment *bi*, two combine this with the segment *pedo*, *ped* or *pd* as part of their usernames, indicating that this might also be a way within the community to express an individual's sexual interest in children of both (or all) genders. It can thereby be understood as another way of expressing their victim preferences – in this case as 'no preference regarding gender'.

Violence Users who have chosen username segments that fall into this category are expressing their acceptance of and willingness or wish to commit violent acts against children. Commonly used segments or segment pairs to express violence in the reduced corpus are: *hurt* (26), *hc* (10) and *boy raper* (3). Users choosing to refer to themselves in such a way represent themselves by two means: 1) by means of identification through classification as belonging to the a group of people enjoying or engaging in such acts, and 2) by means of differentiation as members of the community who do not share the believes of those referring to themselves as 'childlovers'.

Among the group of top users, however, references to violence are not found among the top segments or segment pairs, indicating that users who choose such usernames are less involved in the fora generally, meaning they contribute fewer posts overall.

4.2.2.2 Graphology, Typography and Multimodality

In addition to the above thematic categories, a number of graphological, typographical and multimodality features have been identified in the username data when analysing the username segments. Some of the most frequent ones are discussed in this section.

Punctuation and word boundaries Users in the dark web CSA context chose a number of different punctuation as segments of their usernames. In the reduced corpus, 54 hyphens and 34 periods were identified. Among the top users, 21 hyphens and 7 periods were identified as username segments. When analysing such usernames which include these punctuation marks as parts of their usernames, it becomes clear that these are used predominantly to distinguish between different words in the username and to thereby establish word boundaries. When examining the corpus for other punctuation marks, it becomes apparent that there is others used for the same purpose, such as the underscore. In addition to punctuation used to visually separate words within a username, some users choose capitalisations or spaces to achieve the same goal. Others combine some of

these features. Examples of all those different types of achieving a visual boundary between different words within a username can be seen in table 4.10 below.

Feature	Segment	Username examples
Capitalisation		TheCat, UncleGeorge, HappyHarry, DutchMike
Hyphen	-	Lover-of-Childs, hat-man, candy-gurl, I-Am-Lost
Period	.	Sunny.Meyer, Pedo.bear, Mr.Strangelove, Dr.Who
Underscore	_	Batman_V_Superman, I_Love_Green_Olives, white_bear
Space		The Librarian, Le Baron, First Love, proud uncle

TABLE 4.10: Examples of word boundary features and segments used by members in CSA-related dark web fora.

While the author cannot know for sure what the reason behind these choices is, it is assumed that choosing visual cues to establish word boundaries within usernames not only increases readability. Choosing punctuation marks over spaces might also be a choice made by many users based on past experiences with other online platforms which often do not allow for spaces within usernames. Aleksiejuk (2017) and Kersten and Lotze (2020) also found many such typographical features indicating word boundaries in their research on Russian clearnet message boards and UK online communities, respectively, showing that these features are common across contexts and languages. Further, establishing word boundaries through the above mentioned features ensures a greater chance of other members in the community remembering and recognising the username in the future – a key factor in becoming an accepted and established member of this online CoP.

Numbers as placeholders A number of usernames showed that members use numbers as placeholders for letters, phonemes or even entire words that have been replaced. Examples of this are: *pl34sur3*, *b4rn3y* and *p3d0*, *BabyPedo4Ever*, *boy4me* or *mad4boys*. Members might make these choices for a number of reasons: 1) to be able to choose a number that has otherwise already been taken by simply replacing some of the letters or phonemes with numbers, 2) to shorten an otherwise long username, or 3) to stand out by making the username more creative or visually interesting or

appealing to others.

Before the emergence of CMC, linguistic research has typically focused on content and structure of language but as with CMC users combine aspects of traditional written communication with other, more visual aspects of communication, such aspects have become part of studying CMC and therefore usernames (Aleksiejuk, 2017). Non-linguistic symbols such as numbers have also been found to represent and replace words or parts of words in other contexts, and typographic creativity has been observed globally Danet and Herring, 2007; Naruszewicz-Duchlińska, 2003. Studies examining the communicative role of typographic features in usernames suggest that such features could be used to, for example, demonstrate originality (Scheidt, 2005); increase aesthetic value and attractiveness of the username (Van Langendonck, 2007); express the user's emotional state (Bechar-Israeli, 1995); or "visually extend the semantic potential of a text" (Aleksiejuk, 2017, p.175).

4.2.2.3 Other notable username features

In addition to the categories and features already discussed above, this section introduces a small number of additional features found during the initial exploratory stage of this research.

Phrases Another common feature found in the corpus is usernames consisting of entire phrases or sentences, such as *helloiamjoe32*, *CallmeDaddy26*, *imkindanewtothis*, *iamback*, *whyamihere69*, or *whocare!?*.

Phrases which constitute statements about the individual's likes or preferences, often regarding their choice of victim or sexual practice, such as the usernames *iliketights* or *ilovelittleboys* are frequently found, but more generic and seemingly out of context ones are present in the data as well, such as *IhateFrogs*.

Random and keyboard sequences Especially among the most infrequent posters in the data set, users tend to choose random or keyboard sequences either as parts or as their entire username. Examples for seemingly random sequences of numbers and/or letters are usernames such as *z0x9c8v7b6*, *uheuwfihbrvu* or *8189500725*. It is, however, entirely possible that those sequences do represent some sort of meaning to the particular user and it is unknown to the author and perhaps also unknown to the rest of the online community. As users who have chosen such usernames are also users who have only posted once or twice and do not interact with others on the forum, this suggests that these types of usernames, likely also because they would be hard to remember, are unsuccessful username choices in this particular online environment if the goal is to become an active member and part of the CoP – something that might be of interest to LE agents attempting to infiltrate such communities. It can, however, be assumed that many such users never created their account with the goal of becoming a permanent part of the community and have never intended to post more frequently. Many such users only ever joined to share a particular link to IIOC and it can be assumed that this might be users who create a new account every time they access the forum.

Commonly used sequences are number sequences such as *123* and *54321* or sequences based on keyboard layout such as *qwertz*, *qwerty* or *asdf*. Interestingly, while choosing such sequences might seem to be one of the safest ways to stay entirely anonymous online, choosing a sequence based on the layout of one's keyboard might reveal at least some detail about this person's identity. As not all languages use the same keyboard layout, some sequences might reveal what the user's dominant or everyday language is. A username such as *qwertzuiopü*, which represents the entire top row of letters on a German computer keyboard, might indicate that the user who created this username might be a native speaker of German or live in a location where German computer keyboards are readily available or commonly used in home and workplace settings.

It should be noted that, while these random sequences might seem unusual as a name choice as they would be hard to pronounce, usernames are as visual as they are linguistic in nature and

therefore any sequence of symbols (as allowed by the specific website) is a valid username. As something not usually possible in offline names, it can be seen as another tool of identity performance in online spaces (Aleksiejuk, 2017).

Creativity and word play The final type of usernames found in the data which will be discussed in this chapter are usernames which display a high amount of creativity and/or some form of word play. Examples for these types of usernames are *WaltDickney*, *thorpedo*, *walterego* and *CopyPasta*. While it is hard to group these usernames according to their actual meaning, they all do have in common that users have put at least some amount of effort into creating them which suggests these individuals have an interest in being accepted in the community more long-term and also being recognised and remembered through that particular name choice. Many names that can be included here do however have something in common: they often include some form of reference to sex or paedophilia into their names, showing that users make use of their creativity to perform a specific identity not just based on their own interests or preferences but also directly related to the forum content.

4.2.3 Frequent vs Infrequent Posters

Based on unique usernames alone, this data set offers over 32,000 usernames and their corresponding forum post contributions to be analysed. When looking at the data more closely, however, it can be seen that over half of the users (53%) posted only a single time. Another 30% of users posted only 2-5 times during the given period and only 6% of the users (1837 posters) posted more than 20 posts. This final group is considered to be *frequent posters* for the purpose of this study. See the the detailed chart of distribution of posts per users in Figure 4.2 below.

It is important to note that within the group of frequent posters there is still large variation between how often the users have posted within the 6-year time frame: while over half of the users in the group of frequent posters (1032 users) have posted between 21-50 times, there is other users who

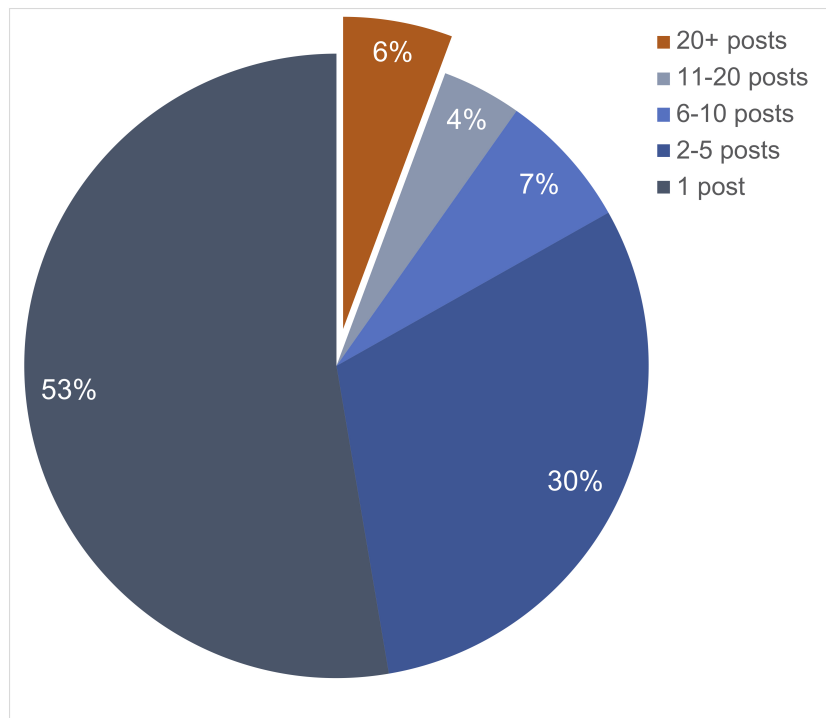


FIGURE 4.2: Percentage of users in the DarkWebCSA corpus who posted once, 2-5 times, 6-10 times, 11-20 times and over 20 times, respectively.

have posted several hundred or even several thousand times. The user who has posted most frequently has contributed to the forum discussions with a total of 5557 posts. See the full distribution of posts per users in Figure 4.3 below. A full list of number of posts per individual user can be found in Appendix E.

The distinction between frequent and infrequent posters is especially important in the context of analysing and understanding identity and its performance in these fora. While users who post infrequently still make some claims about their identity through their usernames, as has been discussed in the previous section, a detailed manual analysis into how users perform these claimed identities within the discourse can only be achieved if enough relevant text data for the users is available. Therefore, the corpus-assisted analysis in Chapter 5 has been carried out solely on data on the identified 1837 frequent posters.

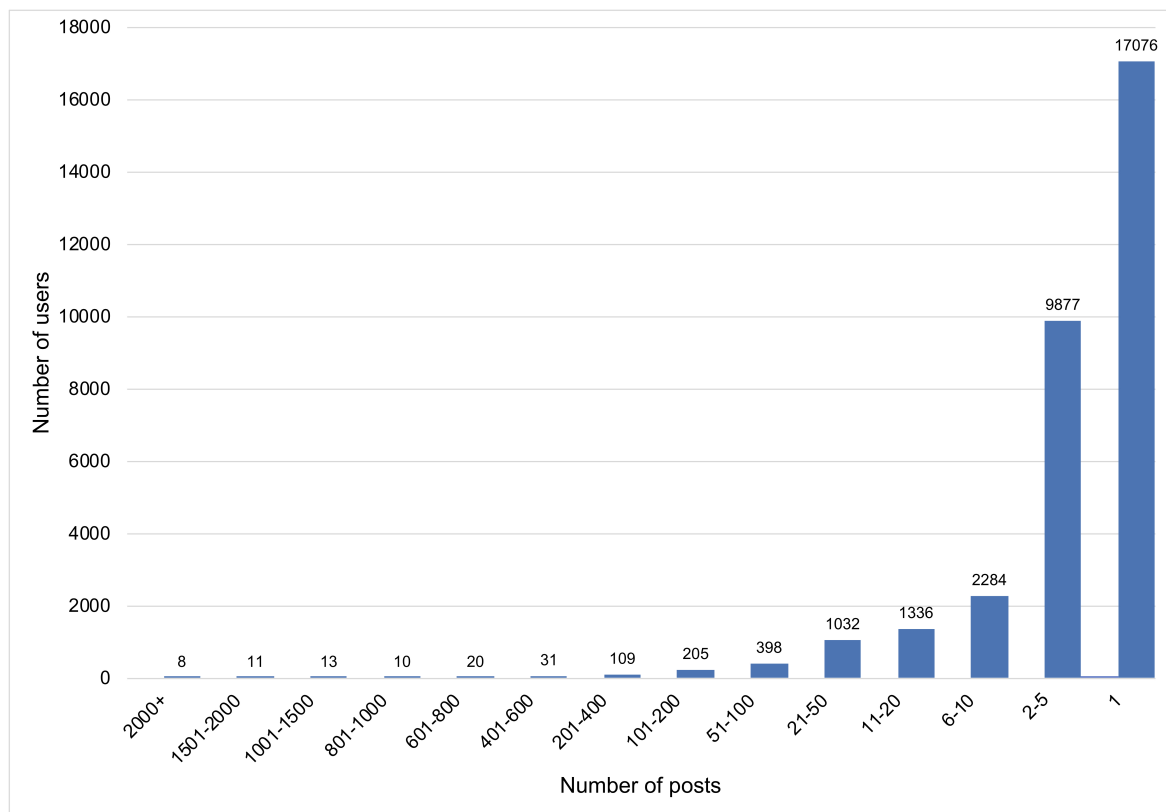


FIGURE 4.3: Distribution of number of posts by number of users in the full DarkWeb-CSA corpus.

4.2.4 Reliability and Generalisability

One problem with analysing usernames is potential ambiguity. Without being able to ask the individual who created the username about its exact meaning, there is always room for interpretation and ambiguity. To minimise the effect of those ambiguities on reliability and generalisability of the categorisation and related findings, several steps have been taken:

Where possible, simple Google searches were conducted to uncover potential multiple meanings of usernames. This was particularly important for terms and expressions unfamiliar to the researcher or when usernames included foreign language segments.

For terms and expressions with meanings likely related to criminal activities and paedophilia, the researcher consulted their supervisor and other researchers with expertise in the field of child sexual abuse. For specific acronyms used both in usernames and the content of forum posts which

the researcher was unfamiliar with, law enforcement officers with particular expertise in online child sexual abuse investigations have also been consulted to shed light on the meaning of such acronyms and abbreviations.

Further, literature from psychology and criminology in particular has been consulted to gain understanding of common terminology and slang used by paedophiles and child sex offenders in these contexts.

While this approach cannot fully counter the impact of ambiguous meanings when analysing the usernames in this work, it helps reduce its impact and aids in taking contextual information into account to ascertain the most likely meaning(s) for the usernames and username segments analysed.

4.3 Discussion of Findings

In the above sections, each username category has been described and discussed in detail and findings for each category have been introduced along a number of example segments and usernames taken from both the initial observation stage and the corpus-assisted analysis of the username segments of 10,222 usernames used by individuals communicating with one another in a dark web forum related to child sexual abuse.

Those findings have been linked to Van Leeuwen's (2013) work on social actor representation by identifying the sociosemantic categories for each identified username category. This link is illustrated throughout the chapter by examples of self-representation through examples of usernames and their segments in the corpus. Table 4.11 summarises those findings in one place and thereby provides an overview of the ways in which users in these CSA online communities choose to represent themselves through their chosen usernames.

The 20 identified username categories show that suspected CSA offenders operating on dark web fora draw on a number of different types of resources when crafting their usernames. These

Chapter 4. Study 1: Username Categories in Child Sexual Abuse Fora

Segment category	Subcategory	SA category	Examples (usernames or segments)
Age		Identification: Classification	<i>whiterabbit72, penny1989</i>
Animals		Categorisation: Identification Impersonalisation: Abstraction	<i>blackcat</i> <i>whitebear, SadPanda</i>
Brands		Exclusion: Backgrounding	<i>2CocaCola, heineken</i>
Childlove		Determination: Differentiation Appraisement: positive	<i>CLover, kidlove,</i> <i>childlover25</i>
Colours		Identification: Classification and Physical Identification	<i>white_bear, gingerbeard22</i> <i>blackninja</i>
Dark/Evil		Categorisation: negative Appraisement	<i>dark666666, evildevil1412, Antichrist</i>
Dark Web		Identification: Classification Categorisation: Functionalisation	<i>toruser123456, darkweb183</i>
Explicit Anonymity		Impersonalisation: Abstraction Personalisation: Categorisation	<i>AnonFrance, anon59,</i> <i>AnonymousFox</i>
Family Relations		Categorisation: Relational Identification	<i>Mom2Katelyn, pedodaddy, UncleGeorge</i>
Film, TV & Literature		Classification + Physical Identification Backgrounding	<i>1gandalf1, timmyturner, gru</i>
Food & Drinks		Backgrounding: Classification Nomination Backgrounding (adult) + Somatisation (child)	<i>Plum, Tomato, pizzaman</i> <i>Candy, Cookie</i> <i>cherrypoppy, Peaches Everywhere</i>
Gender		Identification: Classification	<i>anon-man, islandwoman, Pedogirl1992</i>
Names	Titles Users' own names Victim's names Fictional characters Celebrities	Nomination: Titulation Categorisation: Functionalisation Categorisation: Identification: Classification Nomination Classification + Backgrounding Classification + Backgrounding Identification: Classification Categorisation: positive Appraisement	<i>LadyLucy, DrJones</i> Occupation: Dr, Doc Gender: Mr, Miss, Lady <i>UncleGeorge</i> <i>VickyFan4Life, CuteLittleAnna</i> <i>RonWeasley, FreddyKrueger</i> <i>donaldtrump4561, stanlee</i> <i>nabokovisright or JoseffFritzl</i>
Nationality, ethnicity, location		Identification: Classification	<i>PolishGuy, englishdaddy</i>
Occupation		Categorisation: Functionalisation or Identification: Classification Differentiation	<i>PedLibrarian, Researcher</i> <i>totallynotFBI</i>
Paedophile		Identification: Classification Categorisation: Appraisement (+/-)	<i>newped1, Pedo4Life</i> <i>ProudPedoPapa</i>
Preferences	Self Victim	Identification: Classification Identification: Classification Physical Identification Objectivation: Somatisation	<i>Ygboylover</i> <i>luv10yogirls</i> <i>Blondlover</i> <i>cuntlover, ilovepussies</i>
Religion		Identification: Classification	<i>DarkMessiah, Matthew1914</i>
Sexual Orientation		Identification: Classification	<i>gayMatt, biped515</i>
Violence		Identification: Classification and Differentiation	<i>sadist1, ChildRapist</i>

TABLE 4.11: Functions of types of username segments by username category according to the inventory of social actor representation by Van Leeuwen (2013).

include demographic information such as age or gender, references to offline hobbies and interests, certain occupational backgrounds as well as other offline identities based on religion, ethnicity, nationality or sexual orientation. Users thereby also make use of a number of power resources as described by Newsome-Chandler and Grant (2023): through reference to occupations, kinship relations or technological expertise, these users demonstrate a certain expert knowledge that might grant them higher status, authority and power in these forum communities. The use of terms that directly refer to extreme violence also serve to demonstrate power, both against their victims but also

within the hierarchy present within the forum community. Use of topic-specific terminology or abbreviations/acronyms – such as *nepi* in *nepigrandad* or *loli* in *lolilove210* – also demonstrate subject expertise and can be regarded as a power resource in this context.

The by far most common form of SA representation within the username corpus is that of identification. Most commonly, users assign themselves to categories by means of classification, be that through age, gender, occupation, nationality or their sexual preferences. Other common forms are self-representation through physical identification, for example through the use of colours or reference to animals to describe their physical appearance. Members of this CSA related community further also choose to represent victims within their usernames by means of physical identification, such as through their physical appearance like hair colour. In addition to the above forms of identification, many users represent themselves through their usernames by means of relational identification, that is, through description of their kinship relation with the child victim or other abusers. As usernames, or names more generally, are one of the main ways in which individuals can be identified and distinguished from one another, it is not surprising that usernames using some sort of identification to represent the individual behind the name are so common. In a high-risk context like the one studied here, it is, however, surprising that many users opt for identification by using a proper name. This does, however, show how important conventional rules of naming still are in the online context, even in such highly private communities who seek connection while staying anonymous.

Community members also commonly categorise themselves through positive or negative appraisal: they choose usernames which directly and explicitly evaluate their social practices through the lexical choices made. These are users who self-identify, for example, through their 'love' relationship with a child (+) or by including adjectives such as evil or abusive (-) in their usernames to describe themselves. Users hereby relate their own actions to the wider societal context in which they are considered deviant or criminal, and these actions become parts of their own identity, important enough to be used to refer to themselves through username choice.

Others decide to self-represent by means of functionalisation and refer to themselves through descriptions of their occupation or other social practices they engage in, such as being experienced and proficient users of dark web and encryption technologies. This also includes members who use titles in their names to assign themselves to a social category. Others who use titles or proper names within their usernames might do so by means of nomination, to refer to themselves as a particular individual of a particular name.

Users also commonly self-identify by means of differentiation. Instead of assigning oneself to a particular group, these users represent themselves as belonging to the opposing group or category. This is particularly common for those who identify as either 'childlover' or 'violent' offender, which naturally form opposites and that distinction is of high importance for many members of this online community. Differentiation is also used by forum member to clearly distinguish themselves from law enforcement occupations, primarily through the use of negation. users in all these cases clearly identify not only through what they are, but more importantly through what they are not.

A number of forum members also make use of impersonalisation to represent certain actors. They use abstraction to refer to themselves in certain cases, such as when they are expressing explicit anonymity without any reference to an individual. More importantly, a strategy used commonly by forum members in this context is the representation of the child victim by means of somatisation, a form of objectivation where SAs are referred to through the use of body parts. In the current context, this almost always has a sexual component and is used to depersonalise and objectify the child.

Finally, users also make use of backgrounding techniques in some cases which happens when the author of a username decides to completely remove themselves as an actor within the username and only refer to other social actors or even objects. This is the case for some names in for example, the categories Brands, Victims' Names or Fictional Character.

It should be noted, that some of the SA categories assigned were not clear cases. For example,

titulation and (in)formalisation in the dark web CSA communications context can be seen as a form of categorisation, not nomination. Rather than referring to a specific individual, the user hereby positions themselves as, for example, being of a certain gender. Furthermore, self-identifying through someone's occupation in social actor terms is seen as functionalisation, a form of categorisation. However, users deciding to make their job or occupation the main part of their identity presented through the choice of username clearly strongly identify with this occupation which raises the question if in this context, identification through occupation does perhaps constitute exactly that in SA terms: identification. Van Leeuwen (2013) does himself make this note in his article and describes social categories as "historically and culturally variable" (p.54). The distinction is thereby highly dependant on context. In the case of the CSA related data in this work, this means that not only the social but also the community context has to be considered when assigning SA categories. The researcher needs to be aware of the wider context and be knowledgeable about community norms and values.

The original SA framework includes overlapping categories which are not always clearly distinguishable and individual social actor terms within usernames may belong to categories which appear to be opposites in the original systems network. The majority of usernames in the examined data set consists of several user segments belonging to different username categories. These segments sometimes belong to the same type of SA category, while others consist of segments referring to different SA categories. This common use of several ways to refer to oneself within one username can be understood as overdetermination in the SA sense: members of the CSA dark web community represent themselves as participating in more than just one social practice at the same time. One such group of users is that of members who, at the same time, identify as someone participating in the social practice of paedophilia or child sexual abuse and that of parenting, as seen in the examples of the users *pedodaddy* and *AbusiveDad*. Although members of the community will likely not understand this categorisation the same way as the rest of society, this form of overdetermination

could very well be described as a form of inversion (deviation) in SA terms: inversion is when an individual is described as simultaneously participating in two social practices which form opposites, and being a parent or parenting in society is widely understood as being the opposite of the practices child sex abusers participate in. Furthermore, overdetermination in form of connotation is present in a number of usernames of different categories. While users might describe themselves through naming choices that include functionalisations, classifications or physical identification – such as by identifying as librarian, Polish, male, or by reference to an animal – all those choices have connotations which invoke certain associations within the reader, which may be intended or unintended by the author. It can be assumed that members of these dark web communities related to child sexual abuse do intend those connotations to be shared with others of the in-group and as they form a community of practice, it is likely that members of the group share certain beliefs which would make them interpret those choices in the same or at least a similar way.

The results from this first study give a first impression of users' naming practices within online communities of suspected child sex abusers and how their linguistic choices might be linked to the context of child sexual abuse and dark web communications. The findings further suggest the importance and prevalence of some categories over others. Usernames including references to paedophilia, abusers' victim preferences or potentially identifying features such as proper names all appear to be popular choices among the community and therefore of high importance to the goals of both individuals and the community as a whole. To investigate these naming choices further with regards to their connection to users' identity performance beyond the usernames, the next chapter will focus on analysing usernames and related forum posts for only the frequent posters in the data set and for only the most common and relevant categories.

Chapter 5

Study 2: Usernames and Identity

Performance in Child Sexual Abuse Fora

The analysis of top segments and segment pairs in the DarkWebCSA-top corpus revealed a number of significant themes prevalent among those users. This suggests that including those types of segments in their usernames makes these users especially successful in this community and is relevant for gaining acceptance amongst their peers. This chapter takes the analysis a step further and is interested in if and how those users perform their self-assigned identities within their forum posts. To this end, forum posts for all 1837 users have been extracted into a corpus for further analysis. After annotating all username segments for these 1837 users, the usernames are grouped by category and then the forum post content is analysed by category group following a corpus-assisted approach. The focus hereby lies on the analysis of top keywords per subcorpus, thematic analysis of those keywords and social actor representations in the text data.

5.1 The Data Set: Username Groups Among Frequent Posters

The data set for this chapter is comprised of five subcorpora extracted from the DarkWebCSA-top corpus. These five subcorpora are based on significant username categories (see full description of selection process in section 3.1.1.3) and include the following subcorpora: *Childlove*, *Family*,

Names, Paedophile and *Violence*. Non-English forum contributions have been removed for the post content analysis leading to the subcorpora described in table 5.1 below.

Corpus	Language	Files	Tokens	Types	Lemmas
DarkWebCSA-en-violence	English	19	40689	5703	5108
DarkWebCSA-en-family	English	68	381097	19006	17304
DarkWebCSA-en-names	English	216	1712462	58228	53195
DarkWebCSA-en-childlove	English	64	1005121	35595	32824
DarkWebCSA-en-paedophile	English	45	337348	18057	16403

TABLE 5.1: Description of subcorpora analysed in study 2.

For each of the five subcorpora, one file in the subcorpus corresponds to one username with one line per forum post. This means that the *DarkWebCSA-en-paedophile*, for example, is comprised of forum posts by 45 users. The reference corpora against which these five subcorpora are analysed in this chapter are described in table 5.2 below.

Corpus	Language	Files	Tokens	Types	Lemmas
AmE06	English (Am.)	500	1001024	55577	48689
BE06	English (Br.)	500	996913	58627	52030
BNC2014-baby	English (Br.)	13	5044745	128067	152574
BNC2014-baby-eLanguage	English (Br.)	6	1002428	52027	49352
DarkWebCSA-top	Various	1830	20818195	493856	496871

TABLE 5.2: Reference corpora for study 2.

Details regarding the corpus tool employed for the analysis in this thesis as well as the choice of reference corpora are discussed in section 5.2.1.

5.2 Analysis and Findings

5.2.1 Corpus tool and corpus analysis steps

The software used for the analysis in the present study is LancsBox, a corpus tool developed by linguistics researchers at Lancaster University, UK (Brezina and McEnery, 2020). It is a corpus

toolbox that can be used by researchers to analyse their own and existing corpora. LancsBox has a number of build-in tools to analyse corpora and also provides a number of corpora included in the tool which can be analysed or, as done in this work, used as reference corpora when analysing other corpora.

LancsBox visualises and automatically tags corpus data in any language and offers options for analysing keywords and n-grams, as well as collocations and concordances. Further, many options for more elaborate searches, like regex searches, are available and researchers can use the tool to identify frequencies for keywords, bigrams or word classes and to examine more complex linguistic structures.

The main tools used for analysis in this project are the keyword and the KWIC tool. The keyword tool, called Words in LancsBox, allows for an in-depth analysis of frequencies of types, lemmas and part-of-speech (POS) categories. It further offers the option to compare corpora against each other to establish keyness of certain types or lemmas, as has been done in this work when comparing the subcorpora in question to a number of reference corpora.

The KWIC tool complements the functions of the keyword tool and generates a list of all instances of a search term in a corpus in the form of concordance lines so that keywords identified with the keyword tool can be further analysed and examined in context with the KWIC tool.

For the keyword analyses of the subcorpora in this study, the subcorpora have been compared against a number of reference corpora to show how unique the identified keywords are against other large English language corpora. As Gillings, Mautner, and Baker (2023) highlight, "[f]or the discourse analyst, reference corpora provide an important benchmark against which they can interpret the evidence gleaned from their specialised purpose-built corpora" (p.9). Thus, comparing the language used by suspected CSA offenders against one or more reference corpora can aid in identifying linguistic features which are typical for the deviant communities in question and can, in the case of the current study, help identify topics and terminology common for the studied subgroups

of users. The corpora used in this chapter are AmE06, BE06, BNC2014-baby, and BNC2014-baby-eLanguage. The first 3 constitute large corpora of British and American English text data. As users in the DarkWebCSA-top corpus may be from any language background but communicating predominantly in English, many varieties of English could be present in the data and therefore large English language corpora for both British and American English have been selected to cover a wider range of common English language use. The BNC2014-baby corpus also includes an eLanguage part which has been selected as a separate, additional subcorpus to be used as reference corpus as the DarkWebCSA-top corpus also consists of online language data. As there was no previously collected reference corpus available for the specific discourse type studied in this work, a combination of general English corpora – including digital language use – have been selected to best identify commonly used language in the studied dark web CSA communities. Possible limitations associated with the selected reference corpora have already been discussed in section 3.4.

In addition to the reference corpora available in LancsBox, all subcorpora have also been evaluated against the full DarkWebCSA-top corpus to examine which keywords stand out against the other users. This way, the author was able to analyse not just how users with a particular type of username compare to the general language use of English speakers, but also to compare what sets them apart from the other frequent posters in the data set. As there is to date no existing reference corpus available that covers a collection of textual data for the discourse type investigated in this thesis, the decision has been made to treat the DarkWebCSA-top as a specialised reference corpus against which to compare the subgroups of users in question. An in-depth analysis and discussion comparing the subcorpora against the DarkWebCSA-top corpus following a detailed concordance analysis is presented in Chapter 6. The full details on the reference corpora used can be found in table 5.3 below. Files containing corpus results from this study drawn on in the following section can all be found in Appendix I. The full subcorpora are also available for inspection in Appendix J.

Corpus	Language	Files	Tokens	Types	Lemmas
AmE06	English (Am.)	500	1001024	55577	48689
BE06	English (Br.)	500	996913	58627	52030
BNC2014-baby	English (Br.)	13	5044745	128067	152574
BNC2014-baby-eLanguage	English (Br.)	6	1002428	52027	49352
DarkWebCSA-top	Various	1830	20818195	493856	496871

TABLE 5.3: Reference corpora used for keyword analysis with LancsBox.

In the following sections, the five subcorpora are examined further. The focus lies thereby on the analysis of top keywords within the subcorpora, measured against the 5 reference corpora, and examination of concordance lines for keywords matching observed themes. For ease of reading, all concordance tables are provided in Appendix B at the end of the thesis.

5.2.2 Names

To this date, there have been no cultures or languages identified that don't use names. Plenty of linguistic and onomastic research over the years has shown that also in online spaces people tend to stick to the offline world convention of using some sort of proper name when self-naming. Kersten and Lotze (2020) analysed 500 British usernames and showed that 42.6% chose usernames including their real name (first, given or both); pseudonyms or nicknames which individuals also use offline are also common. Schlobinski and Siever's (2018) international project on usernames showed a similar trend across a total of 14 languages.

Interestingly, users in the present study also commonly chose to stick to this convention when constructing their usernames for use in dark web fora related to child sexual abuse.

The quantitative findings show that among the most frequent posters, 310 chose to include some sort of proper name as part of their username. When looking at these usernames more closely, some important distinctions need to be made within this subgroup.

While many users chose to include a first name as part of their usernames, this does not always link to the user's offline name or identity itself. For example, usernames such as *Kaitfan* and *Vicky-Fan4Life* show that the use of a first name as part of the username in these cases refers to a victim

of child abuse, either only known to the user or, more likely, a victim known to many other members of the community.

The amount of times that very common first names, especially male first names, are used as part of usernames in this community also shows an interesting effect: users pick a name that is most likely not their real name but a common enough first name to not be identifying but it is still picked to conform to the convention of using a proper name as (part of) an online name. Examples for this are the users *John*, *Oscar* and *Andy17*.

Other usernames which include people's given and family names are those where the user chose the name of a public figure as their username. This includes names of actors, politicians, writers or other celebrities, fictional characters or – particular to the context of this forum – the use of names of well-known paedophiles or child abusers. Examples here are *ThomasMann*, *VivianDarkbloom* and *nabokovisright*.

The question that arises from the above is whether users who choose to include a proper name in their username also express themselves differently or discuss different topics than those users who do not include such segments as part of their names. To gain a deeper insight into the linguistic behaviour and identity performance of users in this group, the following corpus results have been consulted. Table 5.4 shows the keywords identified in the *Names* subcorpus against the 4 reference corpora and the corpus for the top 1837 users.

Analysing these tables for common key themes led to five themes being discussed in the following sections as key for the usernames of forum members self-identifying by use of proper name.

5.2.2.1 Age (of consent)

One of the top identified themes in the post data for the *Names* subgroup is age. Especially against the full corpus of top users, numbers from 14-18 are key among the top keywords. Other keywords grouped in with this are *years*, *yr*, *age*, and *consent*.

Keywords against CSA-top	Freq.	Keywords against AmE06	Freq.	Keywords against BE06	Freq.	Keywords against BNC-2014-baby	Freq.	Keywords against BNC-2014-baby-eL	Freq.
j	663	url	4418	go	2084	url	4418	url	4418
16	934	download	1197	preteen	131	sexual	2399	sexual	2399
18	929	am	4011	situation	516	sex	2151	sex	2151
REDACTED	329	thanks	1479	anyways	134	child	2836	child	2836
rachael	270	girl	2560	can't	830	cp	870	girls	2260
REDACTED	259	sexual	2399	horny	130	download	1197	cp	870
consent	871	thread	990	oh	562	tor	809	age	2253
yr	273	girls	2260	hug	145	pedo	797	consent	871
REDACTED	253	cp	870	answer	474	girl	2560	girl	2560
12345	211	sex	2151	wouldn't	546	consent	871	tor	809
age	2253	child	2836	archive	198	thread	990	pedo	797
14	507	tor	809	sites	270	pedophile	661	her	11695
tess	183	pedo	797	ever	1184	adult	987	pedophile	661
j's	150	pedophile	661	seems	759	age	2253	she	8967
15	464	age	2253	browser	134	girls	2260	porn	579
2011	179	stuff	953	being	3028	porn	579	adult	987
acts	272	j	663	seeing	367	kids	1559	18	929
years	1919	16	934	das	127	password	555	16	934
s~	137	18	929	select	189	am	4011	password	555
intercourse	201	consent	871	updated	150	boys	1089	relationship	1048
imprisonment	137	kids	1559	everyone	657	mom	460	REDACTED	418
under	993	cute	611	anal	135	relationship	1048	attracted	499
vids\dsc	126	i	51426	vids\dsc	126	REDACTED	418	sexuality	445
dont	482	love	2840	mean	868	attracted	499	boys	1089
naughty	229	posts	529	couch	161	cute	611	sexually	416
married	287	think	4415	mentioned	277	posts	529	pedophilia	404
count	258	porn	579	etc	146	pedophilia	404	mom	460
sexual	2399	really	2889	imprisonment	137	little	4479	j	663
17	236	password	555	youtube	134	sexuality	445	download	1197
unsd	114	little	4479	und	131	18	929	children	2102
person	1403	dont	482	nn	125	pedophiles	391	pedophiles	391
adp	136	nice	930	much	2654	j	663	file	513
REDACTED	109	boys	1089	lisa	157	sexually	416	abuse	482
christy	107	post	1046	not	11705	16	934	boy	1132
act	524	thank	810	sad	230	preview	416	young	1392
o~	110	my	15567	still	2316	children	2102	adults	550
dude	175	me	10700	hell	288	welcome	647	REDACTED	329
criminal	208	REDACTED	418	niece	137	adults	550	feelings	489
ive	149	enjoy	635	REDACTED	122	boy	1132	older	563
url],[,],[,],[,],[,],password	98	link	643	kissing	145	feelings	489	pedos	309
stuff	953	bit	1167	pleasure	280	REDACTED	329	kids	1559
article	322	fuck	565	happens	256	person	1403	little	4479
totally	376	attracted	499	started	734	topic	458	thread	990
real	1180	welcome	647	though	1420	pedos	309	attraction	347
yeah	571	adult	987	gotten	135	attraction	347	community	579
welcome	647	ok	468	fix	167	videos	422	topic	458
natural	325	preview	416	cool	313	dont	482	preview	416
damn	306	lol	393	enjoying	199	agree	676	cock	309
whole	652	pedophilia	404	knowing	289	kid	529	younger	445
am	4011	videos	422	15	464	ass	336	naked	328

TABLE 5.4: Top 50 keywords for the Names subcorpus against all 5 reference corpora. (Cells stating REDACTED included keywords revealing specific usernames or forum names and have therefore been redacted from the public version of this thesis but can be found in the examiner appendices on the encrypted USB drive.)

A closer look at the concordance lines (with LancsBox's KWIC tool) for these terms reveals that users in this group are highly concerned with the (discussion of) age of consent. Most prevalent in the data is the discussion of what constitutes 'age of consent' in different parts of the world, including detailed information sharing of ages of consent across different countries and jurisdictions. This

also includes discussion of changes in legislation regarding age of consent and possible prison sentences related to sexual offences against children. Examples include the following three sentences from the randomly selected concordance lines:

- (1) The age of consent in Laos is 15, regardless of gender or sexual orientation.
- (2) In The Bahamas, the age of consent for opposite-sex activity is 16 [...]
- (3) [...] shall be punished by imprisonment (of five years to ten years).

Other mentions of age through use of numbers, often combined with the keyword *yr* or *years*, include discussions of specific preferences or experiences by individuals with children of that age, or their own sexual experiences at that same age. Sentences 4-6 exemplify this. So clearly, there is a theme here showing that users in this group heavily focus on discussing and question the concept of 'age of consent'.

- (4) I was probably actuality more interested in her 6 yr old sister in her tight yoga pants.
- (5) [...] mostly like girls from infants to around 6 years old at the max.
- (6) When I started trying to date at 14, I was always intimidated by girls my own age [...]

Social actors represented through those terms are almost always the child victims by reference to their age (and gender) through classification as primary form of identification, further highlighting the importance of age to this user group.

Occurrences of the key term *consent* specifically further show that users in this group are concerned with the question if children can consent to certain (sexual) activities and at what age.

5.2.2.2 Criminality and imprisonment

Related to the above is the theme of criminality and imprisonment. A look at concordance lines for *criminal* shows that these occurrences are typically co-occurring with the collocates *legislation*,

code, law, justice, act, liability and *complaint* (as seen in examples 7-9 below) – all collocates to do with criminality and legal terms, and all discussed directly in connection with sexual acts with and between children. This is further evidence that users who choose to include a proper name in their username are those members highly concerned with legal questions and issues regarding child sexual abuse.

(7) [...] the older partner would be exempt of **criminal** liability if three conditions are met: [...]

(8) The age of consent in the Netherlands is 16, as specified by the Dutch **Criminal** Code at Article [...]

(9) The government as forced to propose the **Criminal** Law Amendment Act 1885, which raised the age of consent from thirteen to sixteen [...]

A look at the concordance lines for the keyword *imprisonment* further shows that certain members of this subgroup also engage in detailed discussion of possible prison sentences attached to these crimes, but the data shows that this is only a handful of users in the corpus and therefore not representative of the whole group.

5.2.2.3 Names and community

Another key topic which stands out for this subcorpus is that a number of specific usernames and the name of a particular forum are among the top keywords. In addition, the keyword *community* itself features as a top keyword against the BNC-2014-baby-eLanguage corpus.

Looking at the concordance lines for these keywords, it appears that users who choose to self-identify through nomination, also frequently make use of proper names within their interaction. Usernames are used as proper names when interacting with one another, such as in this example by user *gayMatt* who uses the proper name from his username to sign off his messages in the forum:

(10) Welcome to [*community-name*]- glad you are here!! Best, Matt

5.2.3 Family Relations

This subgroup is comprised of users who self-identify through their familial relations, especially in relation to the child victim. This includes users who self-identify as a parent, grandparent, uncle or sibling of the child – with examples such as *Uncle Bob* and *pedodaddy* – as well users who indicate they are interested in sexual relations with children as a couple or family, such as in the username *PedoCouple*. A full list of users within this subgroup can be found in Appendix H.

Table 5.5 shows the top 50 keywords identified in the *Family Relations* corpus against the 4 English language reference corpora and against the DarkWebCSA-top corpus. Some clear themes within the keyword results have been highlighted by the author through a simple colour scheme and are discussed in the following sections.

5.2.3.1 Family and incest

Over all reference corpora, the theme that stands out from the data the clearest, perhaps not surprisingly, is that of family and incest. Keywords belonging to this group include: *daughter(s)*, *wife*, *niece*, *mom*, *kids*, *dad*, *incest*, *couple*, *family/families*, *uncle*, *husband*, *child* and *parents*, and tables with random concordance lines are available in Appendix B for all.

The relational terms *daughter(s)*, *wife* and *niece* are commonly used together with the possessive pronoun *my* or *our*, making clear the direct familial connection between users and other social actors. When referring to *mom*, *dad* or *uncle*, however, this is most commonly a reference to the parent or relative of a child that is spoken about, referencing a relation between two other social actors outside the self. This does show that not only is it important for users in this subgroup to discuss their own familial relationships with their partners and children, but connection to (potential) victims' families are also of high importance to them for relationship building with the child. Other mentions of *dad* refer to either the user's parent when they share narratives from the past, mostly when they were still children themselves, or as self-reference directly mirroring their username choice and

Keywords against CSA-top	Freq.	Keywords against AmE06	Freq.	Keywords against BE06	Freq.	Keywords against BNC-2014-baby	Freq.	Keywords against BNC-2014-baby-eL	Freq.
daughter	673	url	505	url	505	url	505	url	505
wife	443	pedo	355	pedo	355	pedo	355	pedo	355
chloe	121	daughter	673	daughter	673	daughter	673	sexual	456
REDACTED	98	cp	298	kids	695	sex	644	sex	644
REDACTED	97	girls	665	cp	298	sexual	456	daughter	673
niece	131	sex	644	sexual	456	cp	298	cp	298
gender	137	kids	695	mom	230	kids	695	her	4142
filename	107	am	936	am	936	mom	230	girls	665
liz	81	thanks	317	sex	644	wife	443	she	3562
claire	83	girl	530	girls	665	girls	665	wife	443
kylie	81	sexual	456	love	787	relationship	377	mom	230
ashley	101	i'm	826	i	14040	porn	170	child	528
ash	70	child	528	girl	530	child	528	relationship	377
REDACTED	66	i	14040	thanks	317	girl	530	porn	170
count	118	ok	173	wife	443	cock	163	kids	695
dee	69	love	787	cute	168	pussy	139	girl	530
mom	230	think	1233	my	4631	adult	207	cock	163
REDACTED	69	porn	170	cock	163	niece	131	older	230
kids	695	cute	168	porn	170	am	936	age	382
REDACTED	59	wife	443	dont	144	cute	168	pussy	139
kelly	66	dont	144	think	1233	mb	119	niece	131
dad	184	cock	163	guess	201	pgp	110	adult	207
incest	101	my	4631	pussy	139	attracted	131	naked	136
ly	49	really	763	download	144	filename	107	mb	119
emma	61	nice	266	adult	207	naked	136	young	411
couple	214	pussy	139	post	281	panties	105	pgp	110
family	358	download	144	myself	374	tor	105	attracted	131
meg	46	niece	131	relationship	377	adults	154	filename	107
families	85	enjoy	184	me	2814	older	230	younger	162
daughters	84	relationship	377	niece	131	incest	101	gender	137
bed	214	chloe	121	i'm	826	age	382	panties	105
pedo	355	post	281	lot	424	cum	103	sexually	107
lifestyle	52	thread	134	mb	119	REDACTED	98	tor	105
dont	144	mb	119	agree	188	sexually	107	incest	101
active	138	me	2814	thread	134	REDACTED	97	adults	154
started	349	myself	374	nice	266	her	4142	REDACTED	98
pgp	110	guess	201	pics	111	myself	374	cum	103
she	3562	pics	111	lol	110	feelings	137	REDACTED	97
uncle	74	older	230	pgp	110	my	4631	community	180
donna	47	lol	110	filename	107	dont	144	feelings	137
fantasies	84	pgp	110	maybe	342	download	144	it.	90
experiences	134	attracted	131	really	763	ass	103	active	138
she's	195	filename	107	tor	105	it.	90	ashley	101
husband	61	i've	325	enjoy	184	pics	111	hurt	117
thats	73	tor	105	don't	974	agree	188	parents	257
my	4631	panties	105	panties	105	thread	134	pedophile	83
older	230	cum	103	ashley	101	ashley	101	pedophilia	82
totally	118	bit	279	things	607	topic	121	pedos	80
relationship	377	incest	101	incest	101	younger	162	fantasies	84
times	309	agree	188	REDACTED	98	love	787	kiss	98

TABLE 5.5: Top 50 keywords for the Family Relations subcorpus against all 5 reference corpora. (Cells stating REDACTED included keywords revealing specific usernames or forum names and have therefore been redacted from the public version of this thesis but can be found in the examiner appendices on the encrypted USB drive.)

thereby clearly identifying with this part of their identity closely enough to also perform this role within the post text. Occurrences of both the keyword *parents* and *family* are used to either refer to the user's own family or the family of the victim, matching the findings for the other keywords above. In

all these cases, the user represents those actors by means of relational identification. The familial relation is thereby always foregrounded.

When talking about *kids*, on the other hand, this representation does not always occur as such. With representations such as *young kids*, *little kids* or *all kids*, these social actors are referred to in terms of aggregation, their belonging to a group but referred as a statistic rather than a group of known individuals. In a similar way, occurrences of the word *child* in this subcorpus refer most commonly to *a child*, a social actor represented by means of classification, as being a member of a certain social group, in this case children.

Finally, directly fitting with the theme of the usernames in this corpus, the keyword *incest* is openly and directly discussed by users of this group. These discussions involve the legality of incest, users' approval or disapproval of the practice and the question of what constitutes incest in the first place. By identifying as a parent or other relative of a child through their username and engaging in sexual abuse of those children, users within this group clearly do engage in the practice and this identity position is reaffirmed by their choice of topic as well.

5.2.3.2 Pronouns

Another clear trend in the keyword results for this subcorpus is the significant use of pronouns, especially in contrast to the English language reference corpora. The pronouns identified among the top 50 keywords against the 5 reference corpora include: *I*, *my*, *me*, *myself*, *she* and *her*. This not only indicated that users in this group spend a lot of time talking about themselves and their experiences, it further underlines their need for identification through familial relation. The pronouns *her* and *my* often refer to the child victim, making clear the strong identification the users feel through their relational identification. Other references with those pronouns relate to body parts, both their own and the child's, and are typically used when describing the abuse of the child.

5.2.3.3 Proper names

When comparing the subcorpus of users identifying through family relations to the corpus of top users, another clear trend becomes visible: users in this subcorpus frequently talk about specific individuals by use of proper names. The proper names identified among the top 50 keywords against the DarkWebCSA-top corpus are: *chloe*, *liz*, *claire*, *kylie*, *ashley*, *ash*, *dee*, *kelly*, *emma*, *meg* and *donna*. The names *ashley* and *chloe* also appear among the top 50 keywords against at least one of the four English language reference corpora.

From exploring the concordance lines of those proper names it becomes clear that almost all of these occurrences are references to a child victim (most likely not through their real name to retain anonymity and privacy). Many of these mentions are part of long narratives either within one post or across a number of posts. While they are constructed to read like personal experiences, it is likely that many of those narratives, or parts of them, are of fictional nature.

5.2.3.4 Self-identification as paedophile

Another keyword that pops up among the top 50 keywords for each of the 5 reference corpora is *pedo*. In reference to the eLanguage corpus, the keywords *pedophile*, *pedophilia* and *pedos* also appear among the top 50 keywords. Most of those occurrences constitute users' self-identification as paedophiles or the discussion of what it means and if it's the right label for them. Given that many of the users in this subgroup also include a form of self-identification as paedophile in their usernames, it is no surprise that these users would also refer to themselves in that same way in the forum discussions. The common use of *a pedo(phile)* indicates that this is a common representation by means of classification – just in the same way one would refer to someone as a member of another class dictated by society, such a gender, age or sexual orientation. This suggests that users in this group see themselves as part of a wider group in society, one that reaches far beyond the boundaries of this particular online community.

5.2.4 Childlove

The term childlove is used by paedophiles and child sex offenders who believe themselves to be in pseudo-romantic relationships with their victims and who want to distance themselves from those offenders they believe to be the more violent and bad type of paedophile. Many 'childlovers' believe that intimate romantic and sexual relations with children and teenagers are non-harmful or even beneficial for the victims. Others simply use the term to describe their preference for children and teenagers over adult sexual relationships. All have in common, though, that they clearly distinguish between those who love children and those who harm and hurt children. This group thereby forms the counterpart to the group of users in the *Violence* corpus which is discussed in section 5.2.5 below. Example usernames from the *Childlove* subcorpus include *childlover250* and *LovingDad*.

Table 5.6 shows the top 50 keywords identified in the *Childlove* corpus against the 5 reference corpora. Significant topics have been highlighted in different colours by the author and are further described and discussed in the section below.

5.2.4.1 Exchange of CSAM and security measures

Two themes that became immediately apparent from studying the top keyword table for this user group were the exchange of IIOC through sharing of links and files, and the discussion of security measures to avoid detection by law enforcement.

Keywords such as *tor*, *data*, *encryption* and *security* all occur in the context of advice regarding security measures. This is an interesting finding, as this theme does not directly emerge or link to the users' self-identification as 'childlover' in their naming choice.

5.2.4.2 The child-adult 'relationship'

The second theme emerging from the data is that of the child-adult 'relationship'. The keywords *relationship* and *relation* occur 664 and 194 times in this subcorpus, respectively. While mentions of

Keywords against CSA-top	Freq.	Keywords against AmE06	Freq.	Keywords against BE06	Freq.	Keywords against BNC-2014-baby	Freq.	Keywords against BNC-2014-baby-eL	Freq.
j	673	url	975	url	975	url	975	url	975
cindy	549	tor	719	tor	719	cock	761	tor	719
tender	470	cp	670	file	876	tor	719	cock	761
john	445	cock	761	cock	761	cp	670	file	876
suzy	309	j	673	cp	670	sex	1199	cp	670
belly	402	girl	1590	cindy	549	sexual	857	sexual	857
0	266	file	876	girl	1590	cindy	549	sex	1199
lover	568	cindy	549	j	673	child	1573	child	1573
ginny	254	am	2036	pedo	465	girl	1590	cindy	549
sperm	319	sex	1199	lover	568	lover	568	girl	1590
marylin	198	pedo	465	am	2036	pedo	465	her	7814
data	536	child	1573	sexual	857	tender	470	lover	568
arlene	189	lover	568	dont	435	file	876	pedo	465
father's	216	dont	435	pedophile	404	mom	436	j	673
girl's	337	tender	470	mom	436	j	673	tender	470
network	357	boy	1173	tender	470	boy	1173	boy	1173
george	251	pedophile	404	adult	598	pedophile	404	pedophile	404
daddy	404	boys	898	love	1716	boys	898	mom	436
jeremy	171	girls	1006	kids	816	adult	598	girls	1006
billy	191	little	3175	attraction	392	attraction	392	windows	409
machine	288	cum	364	daddy	404	cum	364	belly	402
encryption	333	attraction	392	sex	1199	belly	402	boys	898
pregnant	231	feel	1457	cum	364	daddy	404	she	5416
childporn	155	encryption	333	belly	402	encryption	333	cum	364
relation	194	ok	378	i	29691	sperm	319	adult	598
j's	148	thanks	541	encryption	333	pedophilia	309	attraction	392
dont	435	daddy	404	little	3175	suzy	309	father	546
father	546	belly	402	feel	1457	little	3175	encryption	333
mike	171	download	350	boy	1173	feelings	416	girl's	337
his	4468	your	5009	download	350	dont	435	suzy	309
tess	154	love	1716	&	318	REDACTED	289	pedophilia	309
command	172	suzy	309	pedophilia	309	girl's	337	sperm	319
womb	137	pedophilia	309	suzy	309	password	312	files	352
mary	162	i	29691	sperm	319	porn	298	relationship	664
install	203	sexual	857	password	312	attracted	330	computer	482
computer	482	REDACTED	289	REDACTED	289	encrypted	281	porn	298
babies	219	girl's	337	encrypted	281	relationship	664	REDACTED	289
ip	280	ip	280	myself	843	kids	816	security	432
software	278	password	312	cute	308	tails	282	feelings	416
encrypted	281	encrypted	281	feelings	416	windows	409	encrypted	281
momma	115	enjoy	415	girl's	337	pedophiles	265	daddy	404
f	133	kids	816	your	5009	ip	280	tails	282
##	111	attracted	330	00	266	files	352	attracted	330
key	359	fuck	366	pedophiles	265	penis	268	ip	280
mom	436	cute	308	files	352	computer	482	pedophiles	265
june	155	sperm	319	thanks	541	00	266	data	536
while	1726	tails	282	child	1573	ginny	254	password	312
torsafe	112	porn	298	penis	268	lea	255	00	266
security	432	penis	268	girls	1006	girls	1006	lea	255
kimmy	113	00	266	ginny	254	preview	266	penis	268

TABLE 5.6: Top 50 keywords for the Childlove subcorpus against all 5 reference corpora. (Cells stating REDACTED included keywords revealing specific usernames or forum names and have therefore been redacted from the public version of this thesis but can be found in the examiner appendices on the encrypted USB drive.)

relation almost exclusively refer to the sexual abuse of a child, *relationship* does in many cases refer to a romantic or sexual relationship with an adult partner and links to discussions about their partner's knowledge or approval of their own paedophilia. But references to sexually abusive relationships with

children still fall within this category as well. What is interesting about this choice of word is that it assumes that children are equal and consenting participants in these abuse interactions, something that clearly matches the users' name choices in which they identify as 'childlovers' who believe in precisely this ideology.

5.2.4.3 Proper names

Much like in the Names subcorpus, users in this subgroup refer to children being abused by their first name. Unlike in the other corpus, the top names in this corpus all belong to fictional characters in stories shared by only a small number of users. While these narrative form an interesting finding in itself and it would be worth studying such fictional narratives in the OCSA context further, these forum posts are not representative of the subcorpus. The same applied to the keywords in the corpus referring to a parent; they belong to the same narratives as the proper names.

5.2.5 Violence

This subcorpus is comparatively small but has been included in the analysis because of its importance for policing purposes. In some cases, in undercover operations against child sexual abuse offenders the most dangerous or violent offenders have to be identified. It is therefore of interest to analyse the language of those users who choose to indicate in their usernames their preference for violent acts and excessive harm to their victims. Among the 1837 most frequent posters, 22 included explicit references to violence in their usernames. Example names from this group include *ChildRapist* and *hurtcore_lover*.

Table 5.7 shows the keywords identified in the *Violence* corpus against the 5 reference corpora. When looking at the keywords, a few themes become immediately apparent: the exchange of CSAM, body parts, and violence itself. Those themes have been highlighted in colour in the table and are discussed in detail in the following sections.

Keywords against CSA-top	Freq.	Keywords against AmE06	Freq.	Keywords against BE06	Freq.	Keywords against BNC-2014-baby	Freq.	Keywords against BNC-2014-baby-eL	Freq.
jed	82	url	281	url	281	url	281	url	281
REDACTED	58	preview	104	preview	104	preview	104	file	105
font	56	password	95	password	95	password	95	password	95
mavis	39	jed	82	file	105	jed	82	jed	82
lena	31	file	105	jed	82	REDACTED	58	preview	104
hurtcore	44	thanks	96	REDACTED	58	cock	60	her	905
del	38	REDACTED	58	font	56	file	105	REDACTED	58
color="#3300ff	27	girl	151	cock	60	font	56	cock	60
groovestrap	22	cock	60	thanks	96	girl	151	font	56
knickers	27	font	56	girl	151	hurtcore	44	girl	151
color="#ff0099	21	videos	53	videos	53	videos	53	mouth	62
flesh	26	video	72	video	72	mavis	39	hurtcore	44
tara	24	hurtcore	44	hurtcore	44	kid	55	mavis	39
woman	63	mavis	39	kid	55	del	38	she	471
reup	27	del	38	mavis	39	ass	37	child	85
tiny	37	download	37	posted	43	pedo	33	pedo	33
preview	104	kid	55	del	38	posted	43	del	38
9-12f	19	posted	43	ass	37	boy	84	boy	84
password	95	pedo	33	download	37	mouth	62	kid	55
wire	19	ass	37	pedo	33	lena	31	lena	31
deleted	28	boy	84	post	65	video	72	videos	53
mouth	62	lena	31	mouth	62	vids	29	legs	41
piss	22	post	65	vids	29	cunt	30	vids	29
kid	55	little	236	deleted	28	cum	28	woman	63
pale	16	cunt	30	lena	31	color="#3300ff	27	ass	37
hot	55	vids	29	cunt	30	reup	27	knickers	27
thin	19	mouth	62	little	236	child	85	color="#3300ff	27
blood	21	please	56	pics	27	deleted	28	reup	27
section	35	cum	28	color="#3300ff	27	knickers	27	cum	28
preteen	21	hot	55	reup	27	download	37	porn	26
sadistic	15	pics	27	cum	28	little	236	deleted	28
anus	20	knickers	27	please	56	her	905	cunt	30
round	21	color="#3300ff	27	boy	84	porn	26	posted	43
thick	20	reup	27	hot	55	thanks	96	tara	24
child's	25	deleted	28	knickers	27	tara	24	tiny	37
suzi	13	child	85	rape	27	rape	27	flesh	26
barbed	13	fucking	27	tara	24	hot	55	rape	27
cunt	30	tara	24	porn	26	legs	41	groovestrap	22
broom	13	porn	26	groovestrap	22	pics	27	hurt	29
clitoris	15	rape	27	pussy	22	child's	25	girls	56
vids	29	groovestrap	22	color="#ff0099	21	groovestrap	22	child's	25
posted	43	legs	41	preteen	21	pussy	22	hot	55
banned	20	girls	56	shit	26	color="#ff0099	21	color="#ff0099	21
snuff	13	child's	25	piss	22	preteen	21	preteen	21
skin	22	fuck	28	anus	20	post	65	pussy	22
9yo	13	pussy	22	fucking	27	tiny	37	little	236
file	105	color="#ff0099	21	hurt	29	anus	20	pulled	29
tight	26	preteen	21	9-12f	19	flesh	26	lips	22
legs	41	piss	22	fuck	28	piss	22	anus	20

TABLE 5.7: Top 50 keywords for the Violence subcorpus against all 5 reference corpora. (Cells stating REDACTED included keywords revealing specific usernames or forum names and have therefore been redacted from the public version of this thesis but can be found in the examiner appendices on the encrypted USB drive.)

5.2.5.1 Violent acts and references to violence and body parts

Directly linking to the choice of username in this subcorpus, a number of top keywords for this group relate to violent acts or refer to violence and body parts. Keywords identified here are *hurtcore*, *rape*,

hurt and *blood*, as well as *cock*, *mouth* and *ass* as just some examples of references to body parts.

Examining concordance lines for *hurtcore*, it becomes clear that this term is used across this subgroup to describe a certain kind of CSAM. It is either used as a label for particular postings of links and files leading to CSAM, or it is used by users to question the categorisation of other's CSAM as *hurtcore*. Users in this subgroup make a clear distinction between this type of CSAM and others, and they make sure that others follow the rules of the forum and only share the right content within certain sections of the online forum.

The keywords *hurt* and *rape* are used to describe violent acts that are portrayed in CSAM shared among the users, or to discuss fantasies or experiences with children offline. *Blood* and references of body parts were all part of narratives shared among the users, or found in descriptions and discussions of videos shared within the community.

All those keywords directly link to violence and extreme cruelty against children and thereby match the users' name choices. Users who opt for usernames which indicate violence do indeed engage in or enjoy such acts.

5.2.5.2 Exchange of CSAM

The most prevalent type of keyword present in this set is perhaps that relating to the sharing of CSAM. Keywords such as *url*, *preview*, *password*, *file*, *porn*, *video(s)*, *posted*, *deleted* or *reup* are all keywords used by members sharing or requesting CSAM in this forum community. Clearly, users with a preference for violent sexual acts committed against children do frequent these forum communities to exchange CSAM with other like-minded individuals.

5.2.5.3 Politeness features

One final set of features stood out among the keywords identified for the *Violence* subcorpus and was present among the top 50 keywords against 3 of the 5 reference corpora: the politeness features *thanks* and *please* were present in the subcorpus 96 and 56 times, respectively. (Against the

DarkWebCSA-top corpus, *thanks* was keyword 51.) Those politeness features are used by members engaging in the exchange of CSAM. Among the random sample of concordance lines for the keyword *thanks*, 21 occurrences were users directly thanking another user for posting or sharing CSAM in the forum. Examples 11-13 are just some of the concordance lines found with that function. The sentence in example 13 occurred 10 times in this exact form across the random sample.

(11) **Thanks**. Nice series.

(12) Great fakes. **Thanks** to posters.

(13) **Thanks** for sharing.

This shows that it is customary among those users to thank each other for new contributions. Users use politeness features when requesting or receiving new materials, but also for purposes of face-work and a sense of community where contributions are acknowledged and advice is being shared. Examples 14 and 15 from the random concordance sample illustrate this.

(14) Also, **thanks** for the tip.

(15) **Thanks** for the warning.

For the keyword *please*, 8 occurrences in the random sample include apologies for low levels of English proficiency self-assigned by the users, which can be seen as a form of face-work where the user apologises for something that might be commented on later or might complicate understanding. Example 16 demonstrates this. Other uses of *please* are related to requests for CSAM, such as in example 17, or users requesting others to post their content in the appropriate sections of the forum, highlighting that there are certain rules in the community that users need to abide by or they will be called out and redirected by others. Examples 18 and 19 from the random sample are indicative of this.

(16) please apologize my bad english

(17) Yes, PLEASE post the version with sound if you have it.

(18) Please use the regular porn section for this

(19) Thanks for your post [username], but please stop posting non -hurtcore material in this section which is very clearly labelled for Hurt Core.

5.2.6 Paedophile

A large number of users in the present data set chose to self-identify as a paedophile (or hebephile, nepiophile) directly through their choice of usernames, such as the users *Pedo4Life* and *Hebephile66*. This in itself already shows that this part of the users' identity is important for them to highlight, even within a context where users are assumed to belong to this category through context as they have all chosen to enter this community and engage in discussions around child sexual abuse and share their experiences with paedophilia and CSA.

Table 5.8 shows the keywords identified in the *Paedophile* subcorpus against all reference corpora. The table shows the top 50 keywords each and colours have been introduced to highlight themes standing out from the data. Those themes are further discussed in the sections below

5.2.6.1 Exchange of CSAM and security measures

Among the top 50 keywords for this subgroup, a number of keywords were identified that belong to the discussion of CSAM exchange and online security in this online context. Users engage in discussions involving topics and concepts such as *ontopic*, *tor*, *os*, *hidden*, *ip*, *clearnet* or *anonymous*, all dedicated to discussions and advice sharing on the topic of online security with regards to the exchange of IIOC. Users share their experiences with certain software setups and help others in engaging in safe practices regarding their consumption and sharing of illegal materials.

Keywords against CSA-top	Freq.	Keywords against AmE06	Freq.	Keywords against BE06	Freq.	Keywords against BNC-2014-baby	Freq.	Keywords against BNC-2014-baby-eL	Freq.
republican	146	url	542	url	542	url	542	url	542
ontopic	140	cp	533	cp	533	cp	533	cp	533
os	186	tor	426	tor	426	tor	426	tor	426
cops	122	pedo	358	pedo	358	sex	701	sexual	491
kylie	80	am	1153	am	1153	pedo	358	sex	701
quote	200	sex	701	sexual	491	sexual	491	pedo	358
hidden	204	girls	606	sex	701	lea	218	child	650
fox	70	lea	218	kids	461	attracted	246	girls	606
jumps	70	attracted	246	lea	218	child	650	lea	218
cp	533	child	650	pedophile	189	pedophile	189	attracted	246
attracted	246	sexual	491	quote	200	porn	190	pedophile	189
lea	218	pedophile	189	os	186	os	186	porn	190
REDACTED	59	os	186	girls	606	boys	379	os	186
lazy	75	boys	379	attracted	246	kids	461	young	543
vid	128	girl	474	porn	190	girls	606	boys	379
brown	80	kids	461	cock	169	quote	200	cock	169
wife	187	porn	190	i	12462	cock	169	sexually	153
info	108	thanks	265	girl	474	am	1153	girl	474
virtual	82	quote	200	ontopic	140	sexually	153	republican	146
young	543	i'm	755	pedophiles	134	ontopic	140	ontopic	140
ip	114	cock	169	thanks	265	girl	474	quote	200
families	87	ontopic	140	hidden	204	hidden	204	pedophiles	134
pedo	358	i	12462	vid	128	pedophiles	134	hidden	204
prison	126	sexually	153	sexually	153	republican	146	sites	167
fake	89	vid	128	lol	126	vid	128	pedos	123
2013	78	pedophiles	134	republican	146	pedos	123	children	591
babies	85	thread	142	thread	142	daughter	256	clearnet	120
anonymous	88	lol	126	pedos	123	clearnet	120	pedophilia	117
daughter	256	pedos	123	love	565	young	543	cops	122
yf	57	think	1036	clearnet	120	adults	174	her	1956
sites	167	clearnet	120	pedophilia	117	pedophilia	117	adults	174
tor	426	hidden	204	think	1036	children	591	vid	128
jail	91	daughter	256	daughter	256	cops	122	windows	126
sexually	153	pedophilia	117	child	650	sites	167	older	186
REDACTED	41	ip	114	i'm	755	mom	114	daughter	256
am	1153	adults	174	post	243	ip	114	ip	114
tbb	50	info	108	don't	984	adult	159	prison	126
hdd	48	post	243	safe	197	thread	142	kids	461
decoy	42	cops	122	info	108	safe	197	files	120
guilty	80	love	565	mom	114	files	120	mom	114
jane	43	really	586	sites	167	boy	266	she	1585
offtopic	37	don't	984	boys	379	prison	126	age	278
clearnet	120	enjoy	151	ip	114	windows	126	wife	187
pedophile	189	children	591	agree	160	older	186	adult	159
boys	379	hi	112	my	3282	attraction	90	boy	266
busted	61	feel	445	cops	122	info	108	computer	147
quick	86	safe	197	feel	445	kylie	80	jail	91
lolita	64	download	105	adult	159	password	84	younger	123
lifestyle	42	young	543	download	105	feelings	111	attraction	90
dog	70	definitely	134	adults	174	computer	147	kylie	80

TABLE 5.8: Top 50 keywords for the Paedophile subcorpus against all 5 reference corpora. (Cells stating REDACTED included keywords revealing specific usernames or forum names and have therefore been redacted from the public version of this thesis but can be found in the examiner appendices on the encrypted USB drive.)

5.2.6.2 Paedophilia and Paedophiles

Perhaps not surprisingly, other keywords for this subgroup are those referring to paedophiles and paedophilia more broadly. Users thereby either refer to themselves directly by identification through

classification in social actor terms, or are representing themselves and others collectively as a social group. This directly matches their self-identification as paedophile through their chosen username and thereby makes this the key part of their online identity in this forum context.

Their paedophilia is highlighted further through the frequent use of the keyword *attracted*, which is used by members of this subgroup to discuss their sexual attraction to children, including the history of that attraction and whether or not they are also attracted to adults.

5.2.6.3 Law Enforcement and the Law

Finally, users identifying as paedophiles through their username choice are also highly concerned with discussions of topics surrounding law enforcement, prosecution and imprisonment. The keywords *lea*, *cops*, *prison* and *jail* have 218, 122, 126, 91 occurrences in the subcorpus, respectively. Posts including the keywords *lea* and *cops* are discussions of how to stay protected against detection by those actors as well as possible tactics used by law enforcement to track users, such as highlighted in examples 20-22.

(20) [...] being extra cautious WITHOUT becoming paranoid is the best way to keep LEA away.

(21) Can cops upload cp?

(22) The crazy thing is that police and LEA can trace those app

Examining the random concordance lines for *prison* and *jail*, it becomes clear that users in this subgroup are concerned with how to stay out of prison but also discuss their past experiences with imprisonment to share their experiences and advice with others, such as in examples 23-25 below.

(23) If you are meeting [...] with someone underage, you will probably be heading to jail.

(24) I have wasted a long time in prison for my sex crimes against children

(25) I went to **prison** for my sexual exploits with young boys

Discussions of how CSA offenders are or might be treated, are also of high concern to users in this group. Sentences 26-28 exemplify this. Other posts discuss prison sentences of out-group individuals, including famous cases of politicians or celebrities. Sentence 29 serves as an example for this.

(26) [...] I was scared for my life because I had heard how Pedophiles are treated in **prison** [...]

(27) Even if I do end up in **jail**, I am not afraid.

(28) I was raped and beaten and everything else you can possibly imagine being done to a child sex offender in **prison**.

(29) Republican Party leader [Name] [Location] pleaded guilty to six counts of raping his daughters and served 14 years in federal **prison**.

All this shows that users self-identifying as paedophiles are aware of the risks they are taking in engaging in child sexual abuse and are highly concerned with staying anonymous and undetected to avoid prosecution.

5.3 Discussion of Findings

In this chapter, 5 subgroups of usernames within the CSEA context have been analysed for key themes and topics within their forum contributions to relate those findings to their constructed identities through username choice. The analysis revealed that users who self-identify through proper names also uses these names within their forum posts to sign off messages and refer to each other by these names. They are highly concerned with community membership and reference their community frequently as a group they belong to. In addition, these users engage heavily in discussions regarding age of consent and the legality/criminality of sexual acts involving children.

Users who choose to self-identify through family relations within their usernames also heavily focus their forum conversations around family relations. Not only do they discuss their own relationship to their children but they also focus on other family relationships, namely such between themselves and the victim's family. These users identify through their their direct kinship relation with the child they are abusing or wishing to abuse. Furthermore, they engage in explicit discussion of incest, a practice they engage in and which is reaffirmed by this choice of discussion topic.

Those users who self-identify as childlovers focus their conversations on the perceived child-adult relationship in which they see the child as an equal and consenting party to their criminal acts, matching their username choice. They also engage heavily in discussions of CSAM exchange and online security.

Forum members who indicate their preference for violence through their username also perform this identity through their forum contributions. Part of their discussions further focuses on the distinction between 'hurtcore' and regular IIOC, and throughout the discussion and evaluation of CSAM is in focus. In relation to the strong focus on exchanging CSAM, users in this subgroup exhibit a frequent use of politeness measures which connect directly to requests and receipts of shared materials and indicate that users follow community customs and show appreciation for others' contributions.

Finally, those users explicitly self-identifying as paedophiles through their username choice match this identity construction by focusing their discussion on paedophiles and paedophilia, which includes discussions of history and definitions. Furthermore, these users commonly engage in discussions of security, avoiding detection as well as law enforcement tactics. They engage in sharing prison experiences and are highly concerned with security through their own awareness of the criminal behaviours they engage in.

From the start, it became apparent that there is some overlap between the groups due to users either choosing to pair username segments which belong to different categories – such as *pedo* and *daddy* in the username *pedodaddy* – as well as certain username segments being ambiguous

or being assigned to more than one category based on the more complex meaning behind the segment. This overlap can also be seen to some extent when analysing the top keywords and the forum text further.

While there are many differences between the subgroups, some share a number of features. Users in both the *Childlove* and the *Violence* corpus are engaging heavily in the exchange and discussion of CSAM, and users in both the *Paedophile* corpus and the *Family Relations* one do engage in discussions of paedophilia explicitly and refer to themselves as paedophiles in the data. Those overlaps, however, only occur for a few features or topics and can be explained by the partialness of identities expressed in those usernames.

Overall, the analysis showed that users do indeed perform their self-determined identity which is introduced through their choice of username within the community when interacting with other members. This does show that it is possible to infer certain characteristics from a user's naming choice.

Chapter 6

Study 3: Differences and Similarities between Types of CSA Fora Users

6.1 The Data Set: Selected Username Groups

This chapter is concerned with the analysis of five specific user groups based on their chosen type of username. Based on the annotation and analysis of the full data set in studies 1 and 2, for this third study the same five subcorpora that were the subject of the analysis in study 2 are analysed. These five subcorpora – *Childlove*, *Family*, *Names*, *Paedophile* and *Violence* – are then compared to the DarkWebCSA-top corpus as a reference corpus to identify patterns in the data and show how these five subgroups compare to the general user base of frequent posters in the studied fora. Following the comparative keyword analysis, the top 50 keywords per subcorpus build the basis of the analysis discussed in the following sections. The main characteristics for all five corpora are detailed in table 6.1 below.

Corpus	Language	Files	Tokens	Types	Lemmas
DarkWebCSA-top	Various	1830	20818195	493856	496871
DarkWebCSA-en-violence	English	19	40689	5703	5108
DarkWebCSA-en-family	English	68	381097	19006	17304
DarkWebCSA-en-names	English	216	1712462	58228	53195
DarkWebCSA-en-childlove	English	64	1005121	35595	32824
DarkWebCSA-en-paedophile	English	45	337348	18057	16403

TABLE 6.1: Description of subcorpora and DarkWebCSA-top corpus as reference corpus used in study 3.

For each subcorpus, one file corresponds to one username with one line per forum post. This means that the *DarkWebCSA-en-childlove*, for instance, is comprised of forum posts by 64 users. As can be seen in table 6.1, the subcorpora analysed are of different sizes based on the number of users in the group. While the corpora *Childlove*, *Family* and *Paedophile* are comparable in size with 64, 68 and 45 users, respectively, the other two subcorpora are significantly larger or smaller – the *Names* subcorpus includes 216 users, while the *Violence* subcorpus is comprised of data for only 19 users. These differences were taken into account when interpreting the results in this chapter (which are largely based on absolute frequencies within the full subcorpora) and this will be highlighted at relevant points in the analysis section.

6.2 Analysis and Findings

This section presents the findings derived from the detailed thematic analysis of top keywords against the full *DarkWebCSA-top* corpus for all five subcorpora introduced in the previous section as well as a detailed concordance analysis of keywords belonging to key themes that occur in several of the subcorpora's top keyword lists.

First, the thematic annotation of the top keyword lists is presented and thematic differences and commonalities between the subcorpora are briefly discussed. For this thematic analysis, the top 50 keywords for all 5 subcorpora against the *DarkWebCSA-top* corpus were annotated and analysed. The detailed thematic annotations with comments by the author can be found in Appendix C.

Then, the following sections discuss in detail the themes observed in the data and examine the corresponding keywords through a detailed concordance analysis with the help of the corpus analysis tools offered by *LancsBox*. For the concordance analysis of the relevant keywords, the analysis was conducted as follows: first, for each keyword discussed, the author chose 25 random concordance lines (or all concordance lines for the keyword if less than 25 occurrences) which have been extracted with the *KWIC* tool in *LancsBox*; then, patterns observed based on these

random examples are presented and instances from example concordance lines are picked from these random samples and discussed to highlight findings and exemplify details.

A table of the above mentioned 25 random concordance lines – the highlighted keyword together with a context of 10 on either side of the keyword – for each of the discussed keywords is provided in Appendix C for better readability of the main text. Examples selected from these tables are reproduced and discussed within the main text.

6.2.1 Top themes and keywords shared across subcorpora

As could already be seen in the thematic analysis in the previous chapter, there are a number of common themes across the five analysed subcorpora when comparing them to a number of reference corpora. Other themes seem only to occur for one or a few of the subcorpora, even when compared with a number of reference corpora, such as the Violence topic, which suggests that users' self-chosen names are directly linked to the content they share but also shows that there is an overlap between certain user groups.

In this chapter, the focus lies on comparing the top 50 keywords for each of the subcorpora when compared with the full DarkWebCSA-top corpus. Table 6.2 shows the full lists of the top 50 keywords for each of the subcorpora. Keywords highlighted in red occur in three of the keyword lists, yellow highlighted ones occur in two. Keywords shown as *REDACTED* are keywords which comprise forum names or usernames which have been redacted for anonymisation purposes but will still be discussed as part of Sections 6.2.2 and 6.2.4. (Detailed tables with top 50 keywords, absolute frequencies and annotated themes can be found in Appendix C.) Individual highlights from those full tables will be discussed in the remaining sections of this chapter and reproduced in part where discussed.

At a glance, it can be seen that the keyword *dont* is the only one present in more than two of the keyword lists, namely for the subcorpora *Family*, *Childlove* and *Names*. It is also one of the few

Family	Childlove	Paedophile	Names	Violence
active	0	2013	14	9-12f
ash	##	am	15	9yo
ashley	arlene	anonymous	16	anus
REDACTED	babies	attracted	17	banned
bed	belly	babies	18	barbed
chloe	billy	boys	2011	blood
claire	childporn	brown	12345	broom
count	cindy	busted	REDACTED	child's
couple	command	clearnet	act	clitoris
dad	computer	cops	acts	color="#3300ff
daughter	daddy	cp	adp	color="#ff0099
daughters	data	daughter	age	cunt
dee	dont	decoy	am	del
donna	encrypted	dog	article	deleted
dont	encryption	fake	christy	file
emma	f	families	consent	flesh
experiences	father	fox	count	font
families	father's	guilty	criminal	groovestrap
family	george	hdd	damn	hot
fantasies	ginny	hidden	dont	hurtcore
filename	girl's	info	dude	jed
gender	his	ip	REDACTED	kid
husband	install	jail	imprisonment	knickers
incest	ip	jane	intercourse	legs
kelly	j	jumps	ive	lena
kids	j's	kylie	j	mavis
kylie	jeremy	lazy	j's	mouth
lifestyle	john	lea	married	pale
liz	june	lifestyle	REDACTED	password
ly	key	lolita	natural	piss
meg	kimmy	offtopic	naughty	posted
mom	lover	ontopic	o~	preteen
my	machine	os	person	preview
REDACTED	mary	pedo	rachael	reup
niece	marylin	REDACTED	real	round
REDACTED	mike	REDACTED	s~	sadistic
older	mom	pedophile	sexual	section
pedo	momma	prison	silver	skin
REDACTED	network	quick	stuff	snuff
pgp	pregnant	quote	tess	suzi
relationship	relation	republican	totally	tara
she	security	sexually	under	thanks
she's	software	sites	unsd	thick
started	sperm	tbb	url],[,],[,],[,]password	thin
thats	suzy	tor	vids\dsc	tight
times	tender	vid	welcome	tiny
totally	tess	virtual	whole	vids
uncle	torsafe	wife	yeah	REDACTED
url],[,],[,],[,]password	while	yf	years	wire
wife	womb	young	yr	woman

TABLE 6.2: Alphabetised lists of top 50 keywords for all five subcorpora with keywords highlighted that occur in two or more subcorpora among the top 50 list.

keywords in this data set which does not neatly fit into one of the main themes identified. An initial concordance review for *dont* in all three subcorpora where it occurs could not identify a clear pattern for the use of this keyword and it was therefore not included in the further discussion of findings in this chapter.

Other keywords that occur in more than one keyword list are those belonging to the *Family* theme – *daughter, families, mom, wife* –, those referring to the discussion of security measures or CSAM, such as *ip*, and references to the paedophile identity or paedophilia, such as *pedo* or *lifestyle*. Other common ones are those keywords which constitute proper names, such as *kylie* or *j*, as well as references to victims' age group – *babies* – and the keywords *am, totally* and *count*.

All of the above will be discussed further in the following sections but the list already gives a first impression of which topics are representative of some of these subcorpora and how content across the corpora might overlap and relate, therefore showing a link between some types of user-names, such as *Family* and *Paedophile* which share seven top keywords when compared against the DarkWebCSA-top corpus.

Table 6.2 also shows another important detail which is also highly relevant in the policing context: the *Violence* subcorpus shares **none** of its top keywords with any of the other subcorpora, indicating that this corpus – and the user group whose content makes up the corpus – is distinctly different from the other analysed sub groups. This is also evident when looking at Table 6.3 which presents the comparison of identified themes within the top keyword lists for all 5 subcorpora: the *Violence* corpus includes 11 top keywords which are comprised of victim descriptions, a theme not found in any of the other lists, and 17 top keywords referring to body parts or explicit violence, something only seen in a small number in one other corpus (*Childlove*), where it refers to body parts only.

This shows that this subgroup of users is unique in that it explicitly discusses acts of violence and detailed victim descriptions, something which other groups do not do to the same extent. Both those themes are discussed in detail in Sections 6.2.6 and 6.2.10, respectively.

	Family	Childlove	Paedophile	Names	Violence
Age	1	0	1	10	2
Body	0	4	0	0	8
Community/Forum/Support	1	0	3	6	14
CSAM	4	0	4	6	12
Family & incest	13	5	3	1	0
LEA/crime	0	0	6	0	0
Names	15	19	5	10	7
Ontopic (paedophilia)	0	0	2	0	0
Paedophile	6	0	6	0	0
Preference	1	2	4	9	0
Pronouns	3	1	1	3	0
Relation(ship)	1	1	0	0	0
Security	1	15	19	0	0
Sex	0	0	1	5	0
Victim description	0	0	0	0	11
Violence & Torture	0	0	0	0	9
Other	5	3	0	6	1
Total*	51	50	55	56	64

TABLE 6.3: Comparison of themes derived from the top 50 keywords across all five subcorpora showing commonalities and differences between the corpora. The darker the yellow, the more subcorpora shared a theme. (*totals include overlapping thematic categories)

The following sections report the findings for common topics within the data drawn from the above tables. When discussing similarities between subcorpora, the focus lies on those topics and keywords which have clearly been identified in several of the subcorpora. Other keywords or themes indicate differences between the corpora which are relevant to answering the research questions posed in Section 1.2. When fitting into patterns already identified, less frequent features, themes or keywords might be used to supplement findings to get a broader picture of the topic discussed in relation to the relevant user group.

6.2.2 Use of proper names across user groups

Names are the one constant across all the studied subcorpora. No matter the topical focus of the user group, all make significant use of some forms of proper names in their discourse. With 19 and 15 top keywords constituting names in the *Childlove* and *Family* corpus, respectively, those two subcorpora show a particular focus on using proper names within their forum contributions.

	j	673	Child victim		chloe	121	Child victim
	cindy	549	Fictional character (child)		REDACTED	97	User
	REDACTED	470	User		liz	81	Child victim's parent
	john	445	Child victim; Fictional character (adult); Others		claire	83	Child victim
	suzy	309	Fictional character (child)		kylie	81	Child victim
	REDACTED	568	User		ashley	101	Child victim
	ginny	254	Fictional character (child)	Family	ash	70	Child victim
	marylin	198	Fictional character (child)		REDACTED	66	User
	arlene	189	Fictional character (adult)		dee	69	User's partner
Childlove	george	251	Fictional character (child)		REDACTED	69	Forum
	jeremy	171	Fictional character (child)		REDACTED	59	User
	billy	191	Fictional character (child)		kelly	66	Child victim
	j's	148	Child victim		emma	61	Child victim
	mike	171	Fictional character (child)		meg	46	User's partner
	tess	154	Child victim		donna	47	Child victim
	mary	162	Fictional character (adult)		jed	82	Fictional character (adult)
	f	133	User's partner		REDACTED	58	User
	june	155	Fictional character (child)	Violence	mavis	39	Fictional character (adult)
	kimmy	113	Fictional character (child)		lena	31	Fictional character (adult)
	j	663	Child victim		tara	24	Child victim
	REDACTED	329	User		woman	63	Fictional character (adult)
	rachael	270	Child victim		suzi	13	Fictional character (adult)
	REDACTED	259	User		kylie	80	Child victim
	tess	183	Child victim		REDACTED	59	User
Names	j's	150	Child victim		REDACTED	41	User
	s	137	Child victim		jane	43	User's partner
	REDACTED	109	User	Paedophile	REDACTED	64	Forum
	christy	107	Fictional character (child)				
	o	110	Child victim				

TABLE 6.4: Keywords per subcorpus identified as proper names together with their absolute frequency and type of proper name. (Cells stating REDACTED included keywords revealing specific usernames or forum names and have therefore been redacted from the public version of this thesis but can be found in the examiner appendices on the encrypted USB drive.)

But also the top keyword lists for the *Names*, *Violence* and *Paedophile* corpora contain 10, 7 and 5 names, respectively. This constitutes 1/5 of the top 50 keywords for the *Names* corpus and almost 2/5 for the *Childlove* corpus.

While the use of proper names is high in all subgroups, the kinds of names used and the social actors they refer to differ. Table 6.4 gives a complete overview of the proper names present in each subcorpus' top keyword list together with their absolute frequencies in the subcorpus and the type of proper name. Lower frequencies for top keywords in the smaller size subcorpora are as expected. Table 6.5 summarises those findings and shows the number of top keywords per type of proper name for each of the subcorpora.

	Childlove	Family	Names	Paedophile	Violence
Child victim	4	8	6	1	1
Child victim's parent	0	1	0	0	0
Fictional character (child)	10	0	1	0	0
Fictional character (adult)	3	0	0	0	5
User	2	3	3	2	1
User's partner	1	2	0	1	0
Forum	0	1	0	1	0
Others	1	0	0	0	0

TABLE 6.5: Number of top keywords per type of proper names by subcorpus

It can be seen that names referring to children – both real and fictional – make up the largest part of proper names used in the subcorpora. Names referring to child victims are 20 in total, while 11 names refer to fictional child characters introduced by members of the studied CSA fora. Those fictional child characters are, for example, a key feature of the contributions in the *Childlove* subcorpus, while real child victims are named in either of the groups, even if less frequently in some. The third type of proper names which receives a lot of attention by forum members are the users' own usernames. Other types of proper names are less common. The following sections discuss those different types of proper names in more detail.

6.2.2.1 Child Victims and Their Families

By far the largest amount of proper names present in the top keyword lists for the 5 subcorpora are names of child victims. A total of 20 names in the top keyword lists refer to a child which is or has been abused by a forum member (self-reported, not verifiable). While only one proper name in the top keyword lists for the *Paedophile* and the *Violence* subcorpus refers to a child victim, in the other 3 subcorpora 6-8 proper names in each of them refer to a child victim. A concordance analysis for the occurrences of child victims' names in the subcorpora leads to the following insights.

The top keyword referring to a child victim in the *Violence* subcorpus is *tara*. Concordances from the random sample show that in this case the child referred to is a child known from CSAM which is well-known among the CSA community, especially those with a preference for violence. Occurrences are either discussions or descriptions of materials that include the child, as can be seen from examples 30-32 below.

(30) Yes, a compilation of two **Tara** videos, thanks [username].

(31) 7.(Pthc) **Tara** 9yr- Masturbates Vibrator And Double Fingers- August 22nd, 2007 (5:49)

(32) There are several **Tara** videos where she orgasms: [...]

The child victim which is part of the top keyword list in the *Paedophile* subcorpus is *kylie*. Examples 33-35 from the corpus show that in this case the child in question is not only known to the user, but is actually their own daughter. Examining these examples further, it became apparent that this user is one of the users with an overlap in username categories and they also identify as a parent in their username, which explains this occurrence of a child victim's name as a top keyword in this corpus.

(33) Not to mention that it's been so valuable for **Kylie** to be able to go to her mom for advice

(34) And we consider our relationship with **Kylie** to be of that type, very much so.

(35) [...] I'd love to maintain a sexual relationship with **Kylie** into adulthood.

In the *Childlove* subcorpus, the following top keywords refer to child victims: *j/j's*, *tess* and *john*. As can be seen from the examples below, for all three children the users retelling these stories do not just include details regarding their sexual relationship with these children but also focus heavily on emotions and their relationship outside the sexual, which fits their self-assigned identity of 'childlover'.

(36) I'm always amazed how **J** can get me hard so soon after an orgasm.

(37) For the next forty minutes or so, **J** and I worked on his math skills.

(38) [...] where J's mother was in the kitchen just down the hall from **J's** room and he's as horny as any 12 year old boy can be.

(39) The question had brought back the shy 13 year old **Tess** from when I first met her.

(40) **Tess** was visibly more confident about the whole thing this time, [...]

(41) But what did happen during this period is that both **Tess** and me moved on from being in love with each other

(42) I just played there thinking of **John**, how we kissed and how he made me feel [...]

(43) Gently rubbing **john** enjoying feeling closer to him than ever before.

(44) I felt protective of **John** and it somehow made me feel good.

Six top keywords referring to 5 child victims are present in the *Names* subcorpus. Three of these top keywords – namely, *j*, *j's* and *tess* – are the same child victims as in the *Childlove* subcorpus and closer examination confirmed that this is due to a category overlap and the users sharing stories

about these 2 child victims are users who within their username self-identify as both childlover and through a name.

The other three top keywords for the *Names* subcorpus are *rachael*, *s* and *o*. Unlike the top keywords relating to child victims in the *Childlove* corpus, here the users almost exclusively talk about their emotional relationship with the children instead of sexual relationships. The two excerpts below exemplify this:

(45) Long distances had been separating me from **Rachael** before but never had it felt like losing her, [...]

(46) For me, I have fallen quite in love with both **S** (6ish) and **O** (11ish).

The *Family* subcorpus is the one with the most top keywords relating to child victims. Eight top keywords for this subcorpus refer to these children directly. Looking at the concordance lines for some of those instances indicate that these child victims discussed here are all female and they are either the user's daughter or niece. Contributions about these child victims include both sexual and non-sexual details of their relationships. Samples 47-49 are examples from one of the keywords to illustrate this. Further, the experiences shared about the users' relationships with these children are extremely detailed accounts.

(47) **Kelly** and I have been active for 6 years now.

(48) It never fails to make me and **Kelly** cry when we watch it.

(49) I never would have found the love **Kelly** and I share if I hadn't discovered this about myself.

In addition to the references to child victims, in the *Family* subcorpus one of the top keywords refers to a child victim's parent: *liz*. The concordance analysis for this keyword shows that the individual sharing those stories about this mother and her children uses this relationship with the

mother to get closer to the children – a form of grooming. Examples 50 and 51 are indicative of that behaviour.

(50) Now In the days leading up to Claire's birthday Liz helps me find the perfect gift for my LGF
[...]

(51) Liz's husband was scheduled to work that night and Liz wasn't comfortable bringing the kids
trick or treating by herself

Interestingly, while the person posting about *Liz* and her children does identify as 'uncle' in their username, it appears that this individual does not have an actual familial relationship with her and her children but is just a close friend of hers.

6.2.2.2 Stories of Abuse: Fictional Characters

Another large group of proper names identified in the subcorpora are those referring to fictional characters. For the analysis in this section, a distinction is made between fictional child characters and fictional adult characters.

In the *Childlove* subcorpus, 13 top keywords are names referring to fictional characters. 10 of these refer to children, while the other 3 refer to adult characters. Based on the names chosen, out of the 10 child characters, 6 of them are female (*cindy, suzy, ginny, marylin, june, kimmy*) and the other 4 are male (*george, jeremy, billy, mike*). For the adult characters there appear to be one male (*john*) and two female characters (*arlene, mary*). All these keywords are part of elaborate fictional stories regarding CSA and incest including large amounts of explicit, sexual content. A closer look at the corpus results shows that these stories have all been posted by the same individual and can therefore not be seen as representative of this subcorpus as a whole.

The same holds true for the *Violence* subcorpus: all 5 top keywords referring to fictional characters are from stories written by the same user, which is not representative of all users in this corpus.

Similarly, the one fictional character among the top keywords of the *Names* subcorpus also stems from posts by one single user only and should therefore also not be seen as representative of the subcorpus.

While these findings are not suitable to make claims about the type of user a subcorpus is comprised of, they do show that when users post such stories, these stories are long, elaborate and explicit and focus heavily on the characters which is how they end up among the top keywords in the first place.

6.2.2.3 References to Community, Users and User's Families

Proper names are also used by members to refer to themselves, other users, their own partners offline and the forum communities they belong to. Overall, 11 of the top keywords refer to a user, 3 names to a user's partner and 2 to a specific forum.

Examining the random concordance line samples for all the top keywords in this category, it becomes apparent that each keyword is used by only one user in the subcorpus and as such these are not representative of the group. However, overall, the findings from the concordance analysis show that there are some significant ways in which usernames and forum names are used **across** these highly relevant subgroups.

Usernames across subgroups are used as signatures for posts and especially posts including files or links to CSAM, or as (part of) passwords to download/open CSAM.

Users' partners get referred to by first name and posts including these social actors as keywords revolve around discussions of adult relationships and honesty/trust about paedophilia with the partner.

6.2.3 Exchange of CSAM and Security

A number of themes identified in the top keywords lists for the five studied subcorpora relate to the exchange of child sexual abuse materials (CSAM). This includes the *CSAM* theme which includes

the direct discussion of CSAM, sharing links to CSAM or uploading CSAM directly to a forum, as well as the themes *Security* – which is concerned with the discussion of security measures taken by forum members to avoid detection, such as the use of the TOR browser, or advice on not sharing identifying information – and *LEA/crime*, which includes examples such as direct discussions of law enforcement strategies, criminal charges for engaging in CSA or experiences with prison time.

CSAM				
Childlove	Family	Names	Paedophile	Violence
	filename	vids\dsc	cp	del
	count	url],[,],[,],[,],password	vid	reup
	ly	stuff	fake	preview
	url],[,],[,],[,],password	count	sites	password
		silver		deleted
		12345		hot
				vids
				posted
				banned
				file
				hurtcore
				section

TABLE 6.6: Top keywords annotated as CSAM theme by subcorpus

As can be seen in table 6.6, all but one subgroup do include top keywords regarding the distribution of CSAM. While in the *Family*, *Names* and *Paedophile* subcorpora 4-6 top keywords are annotated as such, the *Violence* subcorpus showed 12 top keywords belonging to this theme, equalling 24% of the top keywords for this subcorpus. This alone shows how prominent the exchange of CSAM is in the latter group of users. Interestingly though, that same group does not show any significant engagement with discussions of security measures, as none of the top keywords for this group belong to this theme (see table 6.7). Users identifying as childlovers by username appear to follow the opposite structure: while no CSAM-related top keywords appear for this group, 15 top keywords (30%) are part of discussions on security measures, showing that this group is highly concerned with staying anonymous and undetected by law enforcement.

Security				
Childlove	Family	Names	Paedophile	Violence
0	pgp		os	
data			hidden	
network			fox	
machine			jumps	
encryption			lazy	
command			brown	
install			info	
computer			virtual	
ip			ip	
software			fake	
encrypted			anonymous	
##			sites	
key			tor	
torsafe			tbb	
security			hdd	
			decoy	
			quick	
			dog	
			clearnet	

TABLE 6.7: Top keywords annotated as Security theme by subcorpus

The only two subgroups significantly engaging in discussions of guilt, prison time, criminal codes and law enforcement (according to the top keyword lists), however, are the *Names* and the *Paedophile* subgroups (see table 6.8). Self-identified paedophiles as the only subgroup shows high numbers of top keywords for all three discussed themes in this section and therefore appears to be highly concerned with security and avoiding detection, especially with regards to CSAM and their dark web activities.

LEA/crime				
Childlove	Family	Names	Paedophile	Violence
		imprisonment	cops	
		criminal	lea	
		article	prison	
			jail	
			guilty	
			busted	

TABLE 6.8: Top keywords annotated as LEA/crime theme by subcorpus

6.2.4 Forum Rules and Community Conventions

As can be seen in Table 6.3, for 4 out of the 5 subgroups keywords relating to the theme of *Community* are present among the top 50 keywords. While for the subcorpora *Family*, *Paedophile* and *Names* only a small number of keywords fall under this category – 1, 3 and 6 keywords, respectively – for the *Violence* subcorpus keywords relating to this theme account for more than a quarter of the top 50 keywords for this subcorpus. To gain better understanding of how the corpora differ for this theme, some highlights are discussed following a detailed concordance analysis for a number of identified keywords.

Community				
Childlove	Family	Names	Paedophile	Violence
	totally	dude totally yeah welcome damn article	tor quote 2013	font hurtcore color="#3300ff color="#ff0099 reup preview password deleted section vids posted banned file thanks

TABLE 6.9: Top keywords annotated as Community theme by subcorpus.

The table 6.9 above shows that the *Violence* subcorpus is distinct in that it features a number of top keywords which have both been annotated as *CSAM* and *Community*. This is because a number of top keywords for this corpus have a dual function: they are key features of sharing IIOC and links to CSAM while at the same time representing forum conventions which are needed for the proper functioning of the online group. This includes topics such as sharing, deleting or reposting content, including discussions of whether materials have been uploaded to the correct forum

sections. Examples 52-54 below exemplify this.

(52) If anybody managed to get those before the thread was **deleted** can you post them here?

(53) As [username] has deleted his files, can anyone please **reup** Part 3 of this?

(54) Please use the regular porn **section** for this normal CP.

Other keywords falling under this theme are politeness features, such as *welcome* and *thanks*. These are used to greet new members when joining the community and to thank other forum members for uploading and sharing CSAM, as well as for sharing advice and support:

(55) **Welcome!** Looking forward to hearing about your experiences and ideas.

(56) **Welcome** to [forum]! Thanks for contributing.

(57) Anyways, much **thanks** for taking the time to write all this out!

(58) Nice series. **Thanks** for sharing.

Primarily found in the *Names* subcorpus, keywords such as the 3 below add to the informal and collegial nature of the community and are used to keep that casual informality with one another and to show support and understanding for others.

(59) **Yeah**, these groups exist....

(60) **Damn**, I am so envious of you right now!

(61) **Totally** understand my friend. (*Family*)

(62) I **totally** agree with that and think all of us who [...] (*Names*)

In the *Names* corpus, *dude* is used primarily to refer to other forum members with whom the person is interacting:

(63) I mean, come on, **dude**.

(64) Good luck, **dude**.

In addition to the keywords labelled as *Community*, another set of keywords is of relevance when discussing community building, forum conventions and offenders offering support to one another: the direct use of proper names – in this case usernames – when interacting with other users in these fora. Table 6.10 below shows the number of top keywords per subcorpus which constitute a user’s username.

	Childlove	Family	Names	Paedophile	Violence
Username as top keyword	2	3	3	2	1

TABLE 6.10: Number of top keywords consisting of members’ usernames per sub-corpus

Compared to the other top keywords relevant for community building present in the *Violence* corpus, this is an interesting finding. Only one of the top keywords in this subcorpus constitutes a username, indicating that in comparison to other topics, the use of usernames – as, for example, term of address – appears to be less frequent and thereby not as relevant for the community of users in this subgroup.

6.2.5 Family Dynamics

Another important theme identified among the top 50 keywords for 4 out the 5 subcorpora is that of *Family & Incest*. While only accounting for 5, 3 and 1 of the top keywords in the corpora *Childlove*, *Paedophile* and *Names*, these types of keywords make up more than 25% of the top keywords in the *Family* corpus (see table 6.11 below). This shows users who include a family relation or kinship term in their usernames – often to refer to themselves – also, perhaps unsurprisingly, often talk about family relations, family members and incest when contributing posts to forum discussions. How this looks in detail and how it relates to the family terms used by users from the other 3 subcorpora is

discussed alongside a detailed concordance analysis below.

Family & Incest				
Childlove	Family	Names	Paedophile	Violence
father's	daughter	married	wife	
daddy	wife		families	
father	niece		daughter	
momma	mom			
mom	kids			
	dad			
	incest			
	couple			
	family			
	families			
	daughters			
	uncle			
	husband			

TABLE 6.11: Top keywords annotated as Family & Incest theme by subcorpus.

The top keywords relating to family and incest in the *Family* subcorpus can be grouped into 4 further categories. The first category are keywords directly referring to children, such as *daughter*, *niece* and *kids*. The concordance analysis shows that the context in which children are referred to is often directly related to paedophilia and CSA, but there are exceptions. The examples 65-70 below illustrate this. The examples who use the keyword *kids* to refer to children are either more general, as in example 65, or specific to the user's own children, as in example 66. The top keyword *daughters* is not always the plural, but often a misspelled *daughter's*, so referring to just one child (see examples 68 and 69). Keywords referring to children are frequently accompanied by the pronouns *my* or *our*, highlighting that users self-identifying as parents also frequently talk about their own children.

(65) I never had an attraction to **kids** at all until I became attracted to my daughter [...]

(66) [...], clean up the kitchen area afterwards while the **kids** play in the living room and then start getting ready

(67) My **daughter** and I have been lovers for 6 years now.

(68) What would you teach you sons, your **daughters** and at what age.

(69) My **daughters** first lingerie-ish piece of clothing was XXS, [...]

(70) I would have never acted on my desires if my **niece** had not made the first move.

Other keywords are references referring to the familial relationship to the child such as *mom*, *dad* or *uncle*. Examples for this category are given as examples 71-74 below. The concordance analysis reveals that most references to *mom* and *dad* are not references to the self but to other parents; only few instances relate to the users own identity as a parent. For the keyword *uncle*, 14 occurrences in the random sample of 25 concordances are *my uncle*, with many discussing the user's own relationship with their uncle and how they've been victims of CSA when they were young. Some discuss this as positive rather than negative experiences, though, helping them normalise and justify the abuse of children they commit.

(71) [...] and i want to invite her and her **mom** to watch the movies, i just dont know how to

(72) [...] I didn't know if her mom and **dad** were ok with that.

(73) The other **dad** and I definitely enjoyed the eye candy [...]

(74) [...] I was sexually involved with my **uncle** from age 5-16 and it was an awesome experience.

Interestingly, while most references to moms and dads in this subcorpus relate to other parents, they are still top keywords and show that talking about the relationship between children and their parents, even when it is not themselves, is highly important and representative for this group of users.

The next category are keywords directly referring to families and incest.

For the keyword *family*, while some are about the user's family when they were growing up or about other families, most occurrences are about the user's own family in which incest occurs and parents abuse their children sexually, as seen in examples 75 and 76. Others discuss their plans for such a family and their desire to start a family for this purpose, such as in example 77 below.

(75) We have a loving relationship as a **family** and I plan on keeping it that way [...]

(76) Anything you can do to "normalize" your **family** fun is going in the right direction.

(77) If we end up having a **family** dynamic that's not compatible with incest, [...]]

The keyword *families* refers to other families, either of like-minded individuals, as in example 78, or descriptions of observing and getting close to new families with the goal to interact and possibly abuse their children in the future (see examples 79 and 80).

(78) Plus once we started meeting with other **families** and sharing there were "no holds barred" so to speak.

(79) I have always loved going to the mall, watching **families** following the extra lovely girls around [...]

(80) There is a great chance that a lot of them will have **families**. Don't just target adults with kids.

Studying the random concordance lines for the top keyword *incest*, it becomes apparent that most of these cases are users discussing incest as something immoral or taboo, definitions of (types of) incest, and talking about what should or should not be allowed and why:

(81) [...] but they are also subjected to the perceived taboo of **incest**, which can make it worse for them.

(82) I think parent-child incest should be made legal for educational purposes.

Finally, a number of keywords refer to the relationship between parents or to parents/couples as co-offenders. Examples are presented as 83-86 below. The keyword *wife* mostly refers to *my wife*, discussing involvement of wives in the grooming or abuse of (own) children as well as difficulties with adult relationships or being open with their partners about their paedophilia. This is similar for the keyword *husband* in this subcorpus: discussions of whether the partner knows about their interests are common, but there are also mentions of a child victim's parent's husband who appears to be a risk to the user who has a close relationship with the mother and, as a common strategy by CSA offenders, uses his relationship with the mother to groom and get close to the child/children.

(83) My wife and I are planning to take the next step with [child's name]

(84) [...] so no real worry about getting caught. My wife actually thinks its cute when I wrestle around and tickle [...]

(85) Was your husband aware of any of this before he passed? (Very sorry

(86) The impending divorce appears to be on hold as her husband has finally started listening and talking through their issues [...]

The keyword *couple* is mostly used in *couple (of) times/years/weeks/etc*, so not in the sense anticipated here, but those occurrences referring to a couple as being two people in a relationship, this generally refers to either the user and their partner as co-offenders, or to the child-adult relationship as perceived by the suspected abuser:

(87) We are married pedo couple, with a little daughter.

(88) [...] closer than ever and really almost live like a married couple when my wife isn't home.

Family and incest in the top keyword lists for the *Childlove*, *Names* and *Paedophile* subcorpora are rarer but still make up a significant amount overall.

The analysis of the *Childlove* subcorpus identifies a focus on parents, but interestingly there is no overlap with the *Family* subcorpus regarding top keywords of this kind. Occurrences of the keyword *father* captured in the random concordance sample show that while some cases are discussions of father-child relationships in the real world, including sexual relationships, most cases are sections from fictional stories that involve sex scenes between children and their fathers (see example 90).

(89) [...] the relationship I had with my stepmom and my **father** and the relationship my sister had with them [...]

(90) Suzy loved it. She loved her **father**, and couldn't think of a better way to show it [...]

The keyword *father's* in this subcorpus refers to a father's ejaculate or body parts, most often the penis, all from fictional stories shared in this subcorpus. An example is shown as sentence 91 below.

(91) The little girl reached down, and grabbed her **father's** growing penis, [...]

The random concordance lines for *daddy* and *momma* in the subcorpus show that these are almost exclusively references to fictional characters in a user's story. The keyword *mom*, while having some real-world references, also belong to these fictional stories. The three examples below stem from these stories shared by a user of the *Childlove* subgroup.

(92) I wanted a baby that bad, maybe I should have **Daddy** make one in me.

(93) It was quite a surprise, when **Momma** led me into Her and Daddy's bedroom, [...]

(94) "**Mom?**" he inquired, "Could you get those condoms now, before I [...]"

The *Names* subcorpus only includes one top keyword for this theme: *married*. The concordance analysis shows that for users in this group, family is important, and they engage in discussions of adult-adult relationships but also child marriage/age of consent:

(95) She had never masturbated before we **married** and as far as I know she still has n't.

(96) Mary Hathaway (Virginia, 1689) was only nine when she was **married** to William Williams.

(97) [...] consent laws do not apply if the parties are legally **married** to each other.

For the *Paedophile* subcorpus, the random concordance sample for *wife* shows that this, in most cases, refers to the user's wife and many of these interactions center around the wife knowing about her partner's paedophilia, or CSA experiences in their own youth, such as in example 98. Others make clear that the wife is involved in the abuse as well. Sentence 99 is just one such example.

(98) My **wife** knows that I was sexual with my uncle [...]

(99) And my **wife** says he also has become quite a good lover.

Many occurrences of *families* in the *Paedophile* subcorpus appear to be references to families (fictional or not) who also share an interest in incest and CSA, such as the following two examples:

(100) Of the pedo **families** we knew and/or partied with, it seemed [...]

(101) [...] be lovely to live in a small community with several **families** who had this same approach [...]

While some occurrences refer to other people's daughters, a lot of mentions of *daughter* in this subcorpus are users sharing experiences about their own daughters and the abuse they have subjected them to, as can be seen in examples 102 and 103 below.

(102) the swinging and being part of the kink/BDSM community, our **daughter** grew up in an environment heavy with sexual activity.

(103) Since my **daughter** was the first child I had sex with, [...]

It appears a significant finding to mention here that the top keywords relating to family and incest in the *Paedophile* subcorpus overlap with those in the *Family* subcorpus, with the overlap in usernames, e.g. usernames such as *pedodaddy* belonging to both corpora, being responsible for this.

6.2.6 Suspected Abusers' Preferences and Victim Descriptions

Individuals interacting on dark web fora related to child sexual abuse also commonly talk about their sexual preferences. This includes details about their preferred child victims, such as age or gender, which was already shown to be an important part of users' identities when analysing what segments individuals' usernames consist of (see 4.2.2). When looking at the themes identified in the five subcorpora here, it can be seen that among the top keywords for 4 of the subcorpora keywords regarding preferences can be found. While the top keyword lists for the corpora *Family*, *Childlove* and *Paedophile* include 1, 2, and 4 such top keywords, respectively, the *Names* subcorpus has a total of 9 such words among their top 50 keywords, which makes up almost 20% of top keywords for this user group.

Users in the *Childlove*, *Family* and *Paedophile* subcorpora appear to be describing their preferred victims by means of rather broad categories such as *babies*, *older*, *young* or *boys*, often combining age and gender preference, such as in the two examples below:

(104) I love **babies** and toddlers. Boys and girls. (*Childlove*)

(105) I just cannot stop expressing my feelings for **young** boys. (*Paedophile*)

Users in the *Names* subcorpus commonly use more specific descriptions, especially regarding age. This is also related to this user group's strong focus on discussion of 'age of consent', as is discussed in detail in section 6.2.7 below.

Preference				Victim Descriptions
Childlove	Family	Names	Paedophile	Violence
girl's babies	older	16 18 yr age 15 years 17 14 under	young babies yf boys	tiny kid pale hot thin preteen round thick child's 9yo tight

TABLE 6.12: Top keywords annotated as *Preference* theme by subcorpus.

For the *Violence* subcorpus, however, no keywords from the top keywords list have been coded as belonging to the *Preference* theme. Instead, in this corpus we find another type of keyword which makes up more than 20% of the top keywords identified for this subcorpus: *Victim descriptions*. Unlike the other subcorpora, where no such keywords are found in the top 50, users belonging to the *Violence* subgroup often include detailed descriptions of their victims – including details regarding the child's skin (colour), size and weight – and do so in different contexts, predominantly the following two.

The first context concerns the exchange or distribution of CSAM. Here users include short descriptions about the child's appearance, age or gender, among other details. Example sentence 106 below illustrates this.

(106) [child's name] 2008 9Yo Last Cumshot In Mouth From Daddy (2:07)

However, most descriptions stem from users' fantasies, shared in the form of elaborate stories. The two examples below are taken from such stories and show how top keywords in this theme are used there.

(107) This child was pale all over, from the northern mountains, [...]

(108) And yet- my tiny, delicate 7– year old slave would now have to suffer [...]

6.2.7 Age

Mentions and discussions of age appear within the subcorpora in various forms. While for the *Family*, *Paedophile* and *Violence* corpora there are only 1, 1 and 2 references to age among the top 50 keywords against the full DarkWebCSA-top corpus, the subcorpus *Names*, which consists of users including a proper name in their username, shows 20% of their top 50 keywords belonging to the *Age* category.

Age				
Childlove	Family	Names	Paedophile	Violence
	older	16 18 consent yr age 14 15 years 17 2011	young	preteen 9yo

TABLE 6.13: Top keywords annotated as *Age* theme by subcorpus.

In the *Family* subcorpus, the keyword *older* occurs 230 times. Analysing the 25 random concordance lines for this keyword, it becomes apparent that the term is used in a few different ways. First, the term is used to describe a user's own age, such as in the following examples:

(109) [...] were into **older** guys like me [...]

(110) [...] and as I got **older** my AoA stayed the same.

In other cases, the term is used to describe another person's age in relation to the user's age, often in reference to a time where the user was a child or young person themselves, such as in:

(111) [...] when I was around 8. He was several years **older** than me.

These examples stem from posts where users discuss their own experiences with sex or sexual abuse when they were underage themselves. These experiences seem to be important in the context of incest, as users tend to use those experiences to legitimise and normalise the abuse of their own children or other children, grandchildren, nieces or nephews.

Other occurrences of the term *older* in the *Family* subcorpus are used to describe the age of the child being abused by the forum member or the age of children the user has a sexual interest in. Examples include the following:

(112) [...] there are times when I need something in-between the two. Older than Crystal but younger than Kelly. Right around the 6-11 age range [...]

Top keywords belonging to the *Age* theme in the *Names* corpus are plenty and even if not immediately obvious, they all belong to the same type of discussion theme among the forum members: age of consent. Those keywords that are numbers refer to ages of consent in different jurisdictions and for different types of sexual acts, *age* and *consent* directly stem from the trigram *age of consent* in the data, and *yr/years* are keywords directly following a number referring to age in most cases. Even the keyword *2011* belongs to this topic: it refers to the year of the United Nations Statistics Division (UNSD) data on the topic that users refer to in the discussion. Sentences 113-119 below taken from the random concordance samples for the above keywords exemplify these cases.

(113) Botswana The age of consent is 16 for females and 14 for males, but male homosexuality is punishable by 7 years' [...]

(114) The reason the age of consent was lowered from 17 to 16 in 2008 with the Order was to bring

(115) Age of consent was equalised in 1998.

(116) Namibia The age of consent is 16 for girls.

(117) [...] there was a cute little 6 yr old with her mommy.

(118) Furthermore, there is a close-in-age exemption of 3 years age difference for indecent acts between persons younger than 15.

(119) Aruba 18 18 15 18 2011 UNSD Australia 18 18 16 16 2011 UNSD 2 Austria [...]

The concordance analysis for the age-related top keywords in the *Names* subcorpus shows that users in this group are particularly interested in the concept of 'age of consent', how it differs in different regions of the world and how it has changed and continues to change overtime. It seems a logical topic to discuss for individuals interested in sexual acts with children. It is, however, interesting that the group of users who choose to include a proper name in their usernames are showing such a strong interest in the topic.

In the *Paedophile* subcorpus only one top keyword belongs to the theme *Age: young*. The concordance analysis shows that *young* refers to children in all cases, further specifying an already gendered actor in most cases.

(120) I just cannot stop expressing my feelings for young boys.

(121) [...] ideas if you want to indulge in some lust toward young girls and teenagers. Go to a competitive dance competition.

(122) I am strickly a nepi...nothing is hotter than young sexy babies and newborns

Top keywords belonging to the *Age* theme in the *Violence* subcorpus are limited: only two such keywords occur in the list – *preteen* and *9yo*. Both these keywords are used by suspected offenders in this group to describe child victims as can be seen from the random concordance line examples 123-126 below. Both these top keywords have also been annotated as *Victim Descriptions* and are discussed in the previous section 6.2.6.

(123) That way we could get our hairless preteen pussy legally.

(124) [...] it will enhance either the user's pleasure, or the **preteen** girl's punishment).

(125) I do not think she was **9yo**, previously her age was given as 4yo.

(126) I still say the pictures here are of the **9yo** brother and [child's name].

6.2.8 Pronouns

For 4 of the 5 subcorpora, keywords belonging to the *Pronouns* theme are found among the top 50 keywords against the DarkWebCSA-top corpus. Some of these are pronouns, others are keywords directly following a pronoun – like *am* – hinting at the common use of the pronoun *I* in the corpus. The full list of keywords coded for this theme can be found in Table 6.14 below. The only user group which does not have pronouns as part of their top keyword list is the *Violence* corpus, further showing how it is different from the other 4 subcorpora.

Pronouns				
Childlove	Family	Names	Paedophile	Violence
his	she she's my	dont ive am	am	

TABLE 6.14: Top keywords annotated as *Pronouns* theme by subcorpus.

Looking into the sample concordance lines for *his* in the *Childlove* corpus, it can be seen that this pronoun is often used to refer to a male child, a fictional character, in most cases. The sentence below is just one example:

(127) [...] as he collapsed on top of his mother; between **his** two horny sisters who greeted him with sloppy kisses [...]

In the *Family* subcorpus, the pronouns *she* and *my* are prominent. Uses of the pronoun *she* refer to the users' wives, girlfriends or daughters, in most cases. These highlight important people

in the users' lives, people also important to their paedophilic fantasies or lifestyle. The pronoun *my* is used together with a number of different nouns but the increased use of this pronoun compared to the DarkWebCSA-top corpus shows that users in this subgroup put a lot of emphasis on themselves and talk about themselves, their lives and their feelings more than users in other subgroups.

6.2.9 The Human Body

When sharing experiences, stories or fantasies with each other, it is common for users in these CSA-related fora to include detailed descriptions of the human body and to use explicit mentions of body parts. Among the top keyword lists of the examined subgroups in this chapter, two lists include keywords belonging to this theme. For the *Childlove* corpus, 4 top keywords against the full DarkWebCSA-top corpus fall into this category, and 8 top keywords for the *Violence* corpus.

Body				
Childlove	Family	Names	Paedophile	Violence
belly				flesh
pregnant				mouth
womb				anus
sperm				cunt
				clitoris
				skin
				legs
				tight

TABLE 6.15: Top keywords annotated as *Body* theme by subcorpus.

Looking at the random concordances lines, such as the example sentences 128-131, from the *Childlove* subcorpus, it is evident that all those keywords relate to incest and/or pregnancy in some way. For the keywords *belly*, *womb* and *sperm*, almost all occurrences stem from fictional stories about incest and child abuse. For the keyword *pregnant*, however, there are a number of examples that are not part of a fictional story and instead discuss pregnancy in children and how society's reactions to it have changed over time. See sentence 132 below for an example.

(128) "Thanks Mom," said the little girl, as she got up; enjoying the warm feeling of carrying her own father's sperm inside her belly.

(129) Daddy would deposit his sperm in Diane's womb, every night.

(130) [...] and he was about to ejaculate his vigorous baby-making sperm directly into the child's immature young womb [...]

(131) Even imagining getting his little girl pregnant was too much.

(132) When that girl got pregnant at 12 it was not news, a few years back a similar case was national news.

Examples from the top keywords for this theme occurring in the *Violence* subcorpus tell a different story: all top keywords annotated for this theme in this user group are descriptions of body parts and childrens' bodies and what is being done to them. With 16% of top keywords belonging to this theme, this clearly shows that such detailed descriptions of abuse and torture are typical for this type of user. Examples 133-135 below give an insight into how these keywords are used in the discourse. The reader may consult the concordance tables in the appendix if they wish to see further examples.

(133) Insert your cock in their mouth and piss in it.

(134) [...] I had all ready been punishing her vulva and anus with a combination of spanking with hand and ruler, [...]

(135) [...], she left the broom handle sticking out of her cunt and took the child's face [...]

6.2.10 Explicit Mentions of Extreme Violence and Torture

Closely related to the above theme for some of the studied forum users is the use of explicit mentions of violence and torture. The only subcorpus investigated here which includes top keywords belonging to this theme is, unsurprisingly, the *Violence* subcorpus. 11 top keywords (listed in table 6.16 below), more than 20% of top keywords for this subcorpus, belong to this category.

Violence	
Keyword	Frequency
groovestrap	22
flesh	26
9-12f	19
wire	19
piss	22
blood	21
sadistic	15
anus	20
barbed	13
broom	13
snuff	13
skin	22

TABLE 6.16: Top keywords from the *Violence* subcorpus annotated as *Violence* theme.

Some of these top keywords have also been annotated as belonging to the previous category, *Body*, as they are body parts discussed in the context of extreme violent acts. Examples 136 and 137 from the data are illustrating this type of gruesome content without having to reproduce concordances for all such top keywords. The random concordance tables for all the keywords in this list can be found in Appendix C if the reader wishes to consult more examples.

(136) as the wire landed on her chest and the barbed **wire** dug into her back.

(137) [...] I brought the **broom** handle down on the child half a dozen more times

6.2.11 The Paedophile Identity

Users in various of the subcorpora are extremely concerned with their identity as a paedophile. This shows in a number of keywords and keyword themes found in the top keywords lists but manifests differently in the different subcorpora. While the *Paedophile* corpus is the only one with top keywords belonging to the *Ontopic* category, both the *Paedophile* and the *Family* user groups include specific mentions of paedophiles among their top keywords. Given that this is in comparison with the reference corpus of all top users among the full data set, this shows that within these two user groups specifically, the identity as a paedophile has special weight.

Paedophile				
Childlove	Family	Names	Paedophile	Violence
	pedo		attracted	
	lifestyle		pedo	
	active		am	
	started		pedophile	
	fantasies		lifestyle	
	experiences		sexually	

TABLE 6.17: Top keywords annotated as *Paedophile* theme by subcorpus.

As can be seen in table 6.17, while there is some overlap between the groups regarding their top keywords for this theme (*pedo*, *lifestyle*), there are some notable differences. Judging from the keywords alone, users in the *Family* subcorpus seem to be engaged in sharing experiences with one another and discussions of fantasies, when their interest in children has started and whether they are *active* paedophiles, meaning currently engaging in the sexual abuse of children. Examples 138-142 below from the random concordance line sample illustrate this.

(138) There are no good and bad **fantasies**. There are only fantasies.

(139) With one notable exception, I have no consequences from my **experiences**. No criminal consequences, that is to say.

(140) She was 10 when I started having a sexual attraction to her, 11 before anything ever happened.

(141) I am a 47 yo formally active dad with my then 10-15 year old daughter.

(142) I am not active outside of viewing porn, [...]

Users in the *Paedophile* subcorpus appear to focus more on their sexual attraction to children and discuss this explicitly (see examples 143 and 144). Further, with the top keywords *pedo*, *pedophile* and *am* they put exceptional weight on identifying as paedophiles, which is directly in line with their identities self-assigned through their usernames. Examples can be found as 145-147 below.

(143) The fact of the matter is that if you're sexually attracted to little girls, [...]

(144) I am only attracted to little, prepubescent girls.

(145) [...] community really helped me come to grips with being a pedo.

(146) [...] fully explore who I am, and being a pedophile is part of who I am, [...]

(147) I'm very glad to hear your story, I myself am a conservative hebophile.

Both types of users have in common that they describe their paedophilia as a *lifestyle*, showing how much it impacts their everyday life and how much they identify with the label 'paedophile':

(148) it is so much easier to be part of this lifestyle if your partner is okay with it. (*Family*)

(149) Because your daughter is likely to adapt to the lifestyle better if she can share her feelings with her mother. (*Paedophile*)

Ontopic				
Childlove	Family	Names	Paedophile	Violence
			ontopic	
			offtopic	

TABLE 6.18: Top keywords annotated as *Ontopic* theme by subcorpus.

The frequent use of the terms *ontopic* and *offtopic* is also common in the *Paedophile* subgroup. Anything to do with paedophilia and sexual content including children is considered 'ontopic' and users engage in discussions of where to access materials, which includes mentions of security:

(150) I primarily use Tor for **ontopic**. But I also use Tor for searching controversial offtopic stuff.

(151) Never reveal any info about your **offtopic** identity. Maybe I should start using Tor for some other [...]

Sex				
Childlove	Family	Names	Paedophile	Violence
		acts	sexually	
		intercourse		
		sexual		
		act		
		stuff		

TABLE 6.19: Top keywords annotated as *Sex* theme by subcorpus.

Further, both the *Paedophile* and the *Names* corpora show explicit mentions to sex within their top keywords. All such keywords are listed in table 6.19. A concordance analysis for those keywords shows the direct link to the discussion of paedophilia and the paedophile identity as well as discussions around the legality of paedophilic relations.

(152) [...] awesome that the little ones that you are active with **sexually** know what a pedophile is and the definition and all

(153) She initiated our first little innocent **sexual** encounters after a long time of thinking and planning.

(154) sets a general age of consent of 15. However, sexual **acts** that do not include penetration may be performed from age [...]

(155) I have tried **intercourse** a few times but he is still quite young [...]

(156) The sex **stuff** is secondary. I refuse to objectify anyone sexually [...]

6.2.12 The Child-Adult Relationship

Some child sexual abuse offenders believe themselves to be in an equal relationship with their child victims. Coding the top keywords for all 5 subcorpora against the DarkWebCSA-top corpus showed that users in the *Childlove* and the *Family* corpora explicitly mention relationships frequently enough for the terms to show up in the top 50 keywords. The keywords *relation* and *relationship* appear in those two corpora 194 and 377 times, respectively.

Relationship				
Childlove	Family	Names	Paedophile	Violence
relation	relationship			

TABLE 6.20: Top keywords annotated as *Relationship* theme by subcorpus.

A concordance analysis of those occurrences shows: The key word *relation* in the *Childlove* corpus does indeed refer to such adult-child relationships as seen in the examples 157 and 158 below, however, these occurrences have almost exclusively been made by one user which alone does not allow for generalisability of this finding to 'childlovers' as a whole, even though it appears logical.

(157) [...] the hopeless romantic in me liked to believe that the **relation** between me and the boy will last forever [...]

(158) Although i now consider the sexual nature of our **relation** inappropriate, at that time it was a caring relationship [...]

For the *Family* subcorpus, however, the keyword has been used by plenty of different users with many appearing to fit the definition of the 'pseudo-romantic' relationship. The three examples below from the random concordance sample are by three different users in this group and illustrate how these relationships are described by users of this subgroup.

(159) Even though our brief relationship was pleasurable and consensual [...]

(160) As a dad that's in a long term relationship with my daughter, I can say [...]

(161) Obviously,he didn't know that my dad is having a relationship with me.

This shows that users identifying by means of their familial relationship with the child victim do often subscribe to this notion of the child-adult relationship which is a useful insight for undercover officers wanting to join such a group.

6.3 Discussion of Findings

The analysis of the 5 subcorpora in this chapter shows clearly that while similarities between user groups exists, each group is unique in some ways and those differences were highlighted through both a thematic analysis of top keywords and a subsequent concordance analysis for these keywords.

In the *Childlove* subcorpus, proper names are of particular importance. Two-fifth of the top 50 keywords are proper names. Fictional characters are prominent in posts by this type of user and contributions in form of fictional storytelling are frequent. At the same time, this was the only subcorpus not displaying any CSAM-related top keywords. These two findings together suggest that users who identify as 'childlovers' use those fictional stories for inspiration and sexual gratification where other types of users would consume CSAM. Interestingly though, this group of users is highly concerned with security measures and they are discussed extensively throughout the corpus, even though their behaviours within the online forum – sharing fictional stories rather than exchanging or

consuming CSAM – in themselves are not criminal. Further, forum members belonging to this subgroup focus their contributions on children, rather than adults, which matches their self-description through their usernames. This is evident from both the *Proper Names* and the *Family* theme. Users in this group also frequently mention a child's womb or belly, and talk about pregnancy in children, both in discussions of real life and in their fictional story contributions, often in the context of incest.

In the ***Family*** subcorpus, the focus for use of proper names again is on children but users also mention other adults involved. This is typically the user's partner, often as co-offender. Given that users in this subgroup identify via their familial relation to the child they are abusing, it is not surprising that some of their partners would be co-offending or at least know about the abuse happening. The high frequency of keywords such as *family* and *incest* in this subcorpus further highlights this. It further showed that users of this subgroup view their relationships with (their) children as consensual relationships – a detail useful for law enforcement agents when infiltrating such groups. 25% of the top keywords for the *Family* subcorpus were annotated as belonging to the *Family & Incest* theme, showing just how important it is for those individuals to discuss families, their own and others, with each other. A closer look at those keywords showed that those keywords referring directly to children are frequently accompanied by the pronouns *my* or *our*. This indicates users who identify as parents through their username choice also frequently talk about their own children when contributing to forum discussions. Looking at pronoun use by this group further, it was also shown that users in this subcorpus put a lot of emphasis on themselves, and discussions about themselves and their own experiences are much more frequent than in the DarkWebCSA-top corpus overall. Users in this subgroup also frequently refer to other parents, such as parents of their victims, but do not refer to themselves as parents as frequently. They do, however, use direct references to their own families frequently. The analysis of the keyword *uncle* revealed that users most often refer to their own uncles and many of those references are made in the context of describing their own sexual abuse by a family member when they were still children themselves. This also became evident when

analysing concordance lines for age-related keywords. Some individuals in this group use their own experiences to normalise and justify their own offences against children. When discussing their paedophilia, members of this subgroup focus on discussing fantasies and on sharing experiences, with the latter focusing on if and since when users have become 'active' paedophiles. Further, users in this subcorpus commonly see their sexual relationships with children as consensual rather than abusive.

In the **Names** subcorpus, proper names are also of particular importance. One-fifth of the top 50 keywords are proper names. The types of proper names used, however, are plenty – victims, users, fictional characters – and it appears users in this subcorpus don't follow a clear pattern like users in the other subcorpora. It is possible that this group does not represent a distinct user type because of the strong category overlap which means that within this corpus users with usernames annotated as *Family*, *Childlove* and *Paedophile* come together. There is, however, one way in which this group appears to be different from other user groups: age of consent is a key topic in this user group.

In the **Paedophile** subcorpus, proper names found among the top keywords focus on community, with 4 out of 5 top keywords referring to individual users or the forum community as a whole. This suggests community is more sought after in this group than any other investigated subgroup. Further, users in this subgroup are highly concerned with topics such as security, law enforcement, (avoiding) detection, prison time and CSAM, all topics where users are displaying a number of power resources as described by Newsome-Chandler and Grant (2023). This raises the question whether usernames which include segments that show self-identification as paedophile might be considered as claims to power themselves, whether by using such a name users show the community that they know how to be a paedophile without being detected. References to family, including wives, are frequent and are often discussed in the context of co-offending (with their partners) or meeting other like-minded individuals. And there is, perhaps unsurprisingly, a strong focus on being and identifying as paedophiles.

The **Violence** corpus is especially distinct from the other 4 subcorpora in several ways. It is the smallest of the subcorpora, highlighting that users with this type of usernames are generally much less frequent in the data set. This is further supported by the fact that this subcorpus shares none of its top keywords with any of the other 4 subcorpora. Moreover, the *Violence* subcorpus exhibits large numbers of top keywords for themes that are either non-existent (*victim descriptions*) or only occur in small numbers in the other subcorpora (*Body, Violence, Torture*). Explicit discussions of violence are common in this group of users and details regarding the child victims and the acts committed against them are frequently used in CSAM descriptions as well as fictional stories shared with other group members. In the *Violence* subcorpus, proper names among the top keywords frequently refer to adult, fictional characters – as opposed to (fictional) child victims. This shows that users who identify through their preference for violence focus strongly on themselves, the abusers, and their relationships. Children are purely objectified. 24% of the top keywords in the *Violence* subcorpus refer to the exchange of CSAM; keywords regarding security measures, however, are not part of the top keyword list of this user group. Another detail distinct to this subgroup is the frequent use of *thanks*. This is a politeness feature, in this context clearly used mainly out of convention but also appreciation of one another and building a sense of community. It is most commonly used in phrases such as *Thanks for sharing*.

As seen above, 4 out of 5 subcorpora studied in this work have been shown to have a number of distinct characteristics with regards to their users' forum contributions. It is important, however, to also look at the similarities between all or some of the groups to better understand how users' self-assigned identities through chosen usernames relate to their discussions within the groups. A number of such similarities are highlighted and discussed below.

While there are different foci of discussions in the various user groups, all talk about other social actors by use of proper names. This shows that, no matter how different the focus of discussion in

any of the given groups might be or however strongly they might be concerned with staying anonymous and avoiding detection, they all abide by certain interactional conventions when it comes to exchanges in online discussion fora. Across all subgroups, usernames are used as signatures for forum contributions, especially when these contributions include CSAM. At the same time, security measures, CSAM and paedophilia are also key themes in more than one subcorpus, which is perhaps predictable given that these are dark web conversations about paedophilia and CSA.

Further, users in more than one of the subcorpora show a significant use of relational terms and referring to families, such as *daughter*, *mom* and *wife*. Victims' age groups - such as *babies* – are also found as top keywords for more than one of the subcorpora. Users in the *Childlove*, *Family* and *Paedophile* corpora tend to express their preferences regarding their victims as broad age or gender categories. And individuals in both the *Family* and *Paedophile* subgroup focus a lot of their discussions on paedophilia, sexual attraction to children and their paedophile 'lifestyle', placing a particularly strong weight on their identity as a paedophile.

The similarities between several subcorpora can partially be explained through the overlap between user groups. Users with usernames such as *pedodaddy* are part of both the *Family* and the *Paedophile* subcorpus, for example, and could explain why both the themes *paedophile* and *family & incest* would be frequent topics of discussion among both subcorpora.

The analysis presented in this chapter clearly shows that there are distinct characteristics to at least 4 out of the 5 studied username groups. Themes identified in the top keywords match previous findings from studies 1 and 2, and have confirmed some initial findings indicating distinctive characteristics per group but also overlaps between the user groups. The findings further show an interaction between some common themes, such as *CSAM* and *Community*, or *Victim description*, *Violence* and *CSAM*. These findings confirm the usefulness of studying suspected CSA abusers' username choices for policing purposes: selecting the right username for the right context can be crucial in successfully portraying an authentic in-group identity.

Further links between the 3 studies and reflections on the findings, with regards to the literature and considering the aims of this work, will be discussed in final chapter 7.

Chapter 7

Discussion

The three previous chapters have explored a large dark web corpus of child sexual abuse interactions with forum contributions by 32,410 users. Subsets of this large corpus have undergone detailed thematic and concordance analyses. The work presented in those three chapters has aimed to answer the following research questions:

1. What types of usernames are present in the corpus and what kinds of resources do child sex abusers draw on when crafting these usernames?
2. How do users represent themselves in terms of social actor representation through their choice of username?
3. How are usernames linked to users' posted content and their identity performance in the context of child abuse interactions on the dark web?

The following section in this discussion chapter will summarise and discuss the findings from all 3 studies, and highlight how the different analyses carried out in this work have addressed and answered the research questions (RQ1-3). Connections between usernames and users' identity performance are explored and reflections on various aspects of the previously described findings are given in relation to the wider academic and societal context.

Finally, contributions to various areas of research as well as to the role of linguistic research in policing online spaces are discussed. Then, the application of findings and directions for future research are briefly explored, before finishing with a conclusion and some final remarks.

7.1 Summary of Findings

The first study has identified a number of common username categories within the corpus and has shown the usefulness of username segmentation as well as the application of social actor analysis to online self-naming practices. The analysis shows that users draw on a number of resources to create their username and specific considerations regarding anonymity in this criminal context impact those choices. Interestingly, both from a policing and an onomastic perspective, users in this high-risk/low-trust environment still opt for proper names as parts of their usernames in large numbers.

Through the manual segmentation process applied to the usernames and the subsequent annotation and thematic analysis of these segments, this first study has been able to show the types of usernames present in dark web fora related to child sexual abuse (RQ1). The detailed description of the 20 identified categories of usernames further shows the kinds of resources members of these online fora draw on when crafting their usernames (RQ1).

Applying Van Leeuwen's (2013) social actor framework to the study of usernames, this work could show that while self-representation through nomination and categorisation, especially functionalisation, classification and relational identification, occurs most frequently in the studied dark web fora, other forms of self-identification through backgrounding of the user and instead highlighting other social actors, often child victims by means of classification, nomination or somitisation, are also found (RQ2).

The second study further examined how users' naming choices relate to their forum contributions and interactions. Investigating 5 prominent username categories – *Childlove*, *Family*, *Names*,

Paedophile and *Violence* –, the analysis showed that, overall, users who identify in a certain way will also show linguistic behaviours related to that self-assigned identity and engage in discussions of topics which can be directly inferred from the username choice. The analysis also identified a number of additional commonly discussed topics for each of the subcorpora. For example, users who include proper names for self-representation also tend to engage in discussions of age of consent, legality/criminality of sexual acts involving children, and are highly concerned with community membership; users referring to themselves as 'childlovers' focus a lot of their forum contributions on the exchange of CSAM and discussions of security measures; users indicating their preference for violence in their username show frequent use of politeness features in their forum interactions. These findings show that while some topics can be inferred from a suspected child sex abuser's username choice directly, other behaviours are only understood by studying the user groups in more detail (RQ3).

The third study further investigated the 5 subcorpora by examining the top 50 keywords for each of the subcorpora against the DarkWebCSA-top corpus. A thematic categorisation for those top keywords and a detailed concordance analysis for those themes confirmed findings from the previous studies and gave further details as to what types of content is posted and discussed by users of each subgroup. While there are some similarities between the groups – such as the common use of proper names to refer to themselves and other social actors –, a number of key differences exist which show that users' username choices, especially for those with a username falling into the categories *Childlove*, *Family*, *Paedophile* and *Violence*, are indicative of some of the content discussed in the forum posts (RQ3). Users in the *Childlove* corpus put children at the centre of their discussion, use proper names more commonly than other groups, and frequently engage in fictional storytelling; users in the *Family* corpus discuss co-offending with partners, view their sexual relationships with children as consensual, commonly discuss their own and other families, and share personal experiences of CSA with one another; users referring to themselves as paedophiles frequently refer to

other users or the forum community directly, are highly concerned with security, commonly discuss law enforcement and share experiences of prison time; users in the *Violence* corpus explicitly describe and discuss sexual violence against children, fictional stories of abuse often center the adult, they engage heavily in the exchange of CSAM and further exhibit a high use of politeness features compared to other frequent contributors in the DarkWebCSA-top corpus (RQ3).

Through a corpus-assisted approach, this work was able to examine a large amount of username data and thereby be one of the first studies investigating the language in offender-to-offender dark web communications in the OCSA context. Findings from this research can be used to advance understanding of offender communities, child sexual abuse interactions as well as online naming, and can thereby add to existing linguistic and onomastic research. Further, the findings are hoped to also provide applicable knowledge to the policing of online CSEA.

7.2 Discussion: Usernames and Identity Performance in CSA-related Dark Web Communities

The work presented in this thesis has aimed to increase the understanding of dark web child sexual abuse interactions and to gain insights into suspected child sex abusers' naming practices and identity performance in this context. The three studies conducted build on one another and with each analysis step further details about users' self-representation and their emphasis on specific topics within their forum contributions have become apparent. The key finding throughout all three studies is that there is indeed a link between individuals' username choices and their performed identities in the forum contributions, which opens up avenues for further investigation as well as practical applications of these findings, for example in the policing context. Other findings from the three studies show key characteristics for usernames in this criminal context and indicate that users in these fora display different linguistic behaviours based on their self-assigned identities via usernames.

Username and forum contributions investigated in this work show that while there are different foci of discussion for the different subcorpora, they all exhibit a frequent use of proper names. This suggests that even in such a low-trust/high-risk environment (Grant and MacLeod, 2020), users tend to follow interactional conventions common to the language used in online discussion fora more generally, and also their usernames often follow certain conventions and patterns found in general interest discussion fora on the clear web (Kersten and Lotze, 2020; Aleksiejuk, 2017). This indicates that users of CSA-related dark web fora have a similar need for recognition and support as other users in online support groups regarding different topics.

Research on harmful, deviant or criminal online communities is still relatively new and research into CSA offender communities specifically is still rare, especially from linguistics. Literature on related online groups, however, has provided some insights which are comparable to the community studied in this work. Individuals belonging to the incel community also spend an extensive amount of time online interacting with like-minded individuals. Like individuals with a sexual interest in children, these individuals hold certain ideological views which are not compatible with mainstream, societal norms and values. Online communities of extremists also compare to the user group studied in this thesis, in that they get together online to share ideas and experiences and seek like-minded individuals to further normalise their views. Studies investigating such communities through linguistic approaches, also often resorting to corpus-based setups to enable the analysis of large language corpora, have shown that users in these communities often exhibit linguistic strategies used to sexually objectify (incels) and dehumanise (incels, extremists) others (Krendel, 2020; Pražmo, 2020; Rothenberger, Müller, and Elmezeny, 2018). This objectivation is used to justify acts committed against their victims.

In the dark web fora studied in this work this pattern also appears: users choose usernames that dehumanise and objectify children, such as through means of somatisation, for example, where the child is referred to through the mention of body parts. This shows that, while the targeted individuals

and underlying ideologies might be different, all these deviant online communities have in common that they make use of objectifying and dehumanising language to normalise and justify the horrific acts they commit against them. Such objectivation of children has also been identified by Woodhams et al. (2021) who studied convicted CSA offenders, further highlighting this strategy as something typical for these types of individuals.

Further links between the current work and previous literature confirm this common use of objectivation in online offender communities: in their work investigating dark web communications between suspected CSA offenders, Chiang et al. (2021) identify a number of characteristic moves for these types of individuals. One such move they describe is 'Image Discussion' where CSAM which was shared by forum members gets discussed by other members, often by reducing the children discussed to sexual objects. This relates to findings from this work which suggest that certain user types, such as those in the *Violence* subcorpus, frequently engage in discussions of CSAM, especially by describing the child victims in detail, often through somatisation.

Other moves identified by Chiang et al. (2021) also map onto findings from this work. 'Greetings' in the form of welcomes and through the use of direct references to other users is observed, politeness features linking to 'Image Appreciation' are commonly displayed by users in the *Violence* subcorpus, and top keywords identified as *community* theme link neatly to Chiang et al.'s findings on 'General Rapport'. Their identified move 'Offering Indecent Images' also can be found in this work through forum contributions annotated as *CSAM*, referring to contributions concerned with the exchange of CSAM. This shows that when combining findings about these types of offender communities from various linguistic studies, research can provide valuable, conclusive insights into how these communities function to aid law enforcement in the policing of such spaces.

Law enforcement agencies (LEAs) are often concerned with finding the most dangerous or most powerful users in criminal online communities to be able to prioritise those as targets in their policing work. Newsome-Chandler and Grant (2023) have recently uncovered a number of power resources

that individuals draw on when engaging in online forum discussions. Their work thereby specifically included a CSA dark web forum. Among those resources available to users to assert power in the forum interactions, several fit neatly onto the three themes discussed in this section. The *CSAM* theme which is all about users sharing those materials links to Newsome-Chandler and Grant's (2023) observation that 'CSAM possession' can be seen as an extralinguistic marker of power. Following the findings in this study, it can be seen that some linguistic markers involving top keywords discussed in this section could be linked to this power resource. The theme *Security* discussed in this section shows that some types of users are highly concerned with and engage in detailed discussions around security and avoiding detection, including various types of encryption and masking one's online footprint. These detailed discussions sharing advice and information about these topics show strong 'technical expertise' by the users, another power resource found by Newsome-Chandler and Grant (2023). Finally, both conversations about security measures and experiences with law enforcement, convictions and prison time can be interpreted as 'topic expertise through personal experience', where users claim personal experience of a given situation, such as a conviction for a sex crime against children, or as making a claim to being a 'subject of law enforcement/investigations', where a user might claim to have been arrested or convicted before for such a crime – two further power resources identified by Newsome-Chandler and Grant (2023). The only subgroup who exhibits all of these claims to power is the *Paedophile* subgroup, indicating that those users who self-identify as paedophiles through their usernames might be (seen as) more powerful in the forum interactions in this CoP. This raises the question if labelling oneself as a paedophile already in the chosen username might in itself be an expression of dominance and expertise, and thereby a claim to power.

Users exhibiting self-representation through classification such as age, gender or nationality, functionalisation such as occupation, relational identification such as mother or father, and nomination, including titulation, all have in common that their usernames include mentions or hints of their

offline identities. This is insofar surprising that users in these dark web discussion fora are typically highly concerned with anonymity and avoiding detection, and would refrain from sharing any identifying information. It appears, however, that suspected child sex abusers interacting on the dark web have a need for displaying some such markers of reality. A reason for this might be what is explained by Kersten and Lotze (2020) as the problem of "privacy vs. authenticity" (p.125): while users have a desire for privacy and anonymity, they simultaneously seek true connection with other like-minded individuals in an online community. This creates a dilemma for individuals in such spaces, as they are attempting to achieve these two opposing goals. In the current work, users seem to often solve this dilemma creatively through their choice of username, where they provide (often vague) hints towards their offline identities – just enough information to be accepted into the community, but not enough information to risk detection by law enforcement.

Even though only a rather small number of individuals in the DarkWebCSA-top corpus chose a username with direct reference to violence, the choice has been made to analyse this subgroup as one of the subcorpora in studies 2 and 3. Law enforcement agencies are always trying to find new ways to isolate the most dangerous, violent or powerful offenders, and users directly referring to extreme violence and torture in their usernames could potentially fall under this category. Hence, this group was deemed relevant for the context of policing and included in the analysis. The findings from study 2 and 3 show that while there are frequent mentions of violent acts in the subcorpus, these generally refer to fictional scenarios or as parts of descriptions or titles of shared CSAM. Based on the analysis of top keywords for this type of user, their focus lies on sharing fantasies, fictional stories, but also CSAM. For the policing context, this is interesting insofar that the creation of fictional stories or discussion of fantasies regarding violence against children in itself does not make this one of the most violent or dangerous groups of individuals. It is, however, unknown to the author whether any of the shared fictional stories are based on real events. Further, while it appears that users primarily share CSAM materials which have been created by non-members of

these fora, it is still possible that users in this group have created their own CSAM. However, the findings suggest that creating materials and committing such acts offline are not key to this user group. This suggests that users indicating their preference for violence through their usernames do not necessarily engage in these practices. But the fact that this group does engage in the exchange of the most horrendous CSAM out there, as evident from the concordance analysis, this is still a high priority group for policing. The findings from this work, which include showing that users of the *Violence* group frequently engage in exchange of CSAM paired with politeness features, may help officers policing and infiltrating such spaces to not only pick appropriate usernames to be accepted in the group but to also authentically portray an in-group member by making linguistic choices in their forum contributions which match the themes identified in this work. For example, users use politeness features when requesting or receiving new CSAM, but might also use these for purposes of face-work and a sense of community. Were an officer to enter such a forum and not engage in these practices, they might not be accepted fully and trust would not be able to be established with this community.

All three studies in this work could clearly show that there is a significant overlap between two of the prominent user groups: *Family* and *Paedophile*. Usernames such as *pedodaddy* or *ProudPedoPapa* are examples for this overlap in identities. In social actor terms, users with this sort of overlapping or combined self-assigned identity refer to themselves by both relational identification (*Papa*, *daddy*) but also through classification as paedophile (or subtypes). This type of inversion, where individuals describe themselves as simultaneously participating in two social practices, highlights a number of things. First, by stating that an individual engages in both parenting and paedophilia simultaneously, users might want to emphasise that, to them, these two seemingly conflicting social practices are compatible and do not actually oppose each others. By adding positive appraisal to the mix, such as *proud* in *ProudPedoPapa*, users further highlight that engaging in both these practices is something positive, that makes them proud and importantly something they would like

to talk about. Also, by combining these two partial identities within their one username, users are able to convey this information directly and may be able to quickly find other like-minded individuals. Further, users may include such reference to inversion to strengthen their reputation and assert dominance: claiming to be proficient as both a parent and a paedophile asserts dominance and power and could increase the user's rank within the hierarchy of the forum community. Other users who have similar interests may approach them for advice based on those details. Finally, users who may at least partially subscribe to society's view and evaluation of paedophilia as something negative, harmful or criminal may use this inversion tactic to show that in addition to this negatively judged part of their identity, they also engage in a social practice – parenting – which is regarded as positive.

The above links neatly to another phenomenon observed in the *Paedophile* subcorpus: while for some users who self-identify as paedophiles this might be purely descriptive, others add evaluations (appraisements in SA terms) to this identity. Once again, *ProudPedoPapa* is one such example. One reason for doing so might be that users are acknowledging the safe space they are in and can openly admit to the identity without fears of being judged for it. A dark web community as the one studied in this work where individuals with a sexual interest in children come together, follows different norms than the outside world and users are able to lean into that through their choice in username, where they can accurately portray their identity and gain support and confirmation. For some it might be the only, and first, opportunity to openly assert this identity to others.

Findings presented in this work clearly show that users differ greatly in the ways in which they refer to themselves by username. There are a number of distinct categories of usernames. Unlike findings by Woodhams et al. (2021), where the majority of suspected CSA abusers did not use usernames that were directly related to CSA or paedophilia, prominent groups of users studied in the current work clearly indicate a direct reference to paedophilia, childlove, child abuse and other forms of violence and abuse against children directly in their usernames. In their paper, Woodhams et al.

(2021) suggest that their findings might be due to their sample consisting of dark web users who have a heightened awareness for security and anonymity. These findings could not be confirmed in the sample from the eight fora studied here. Connecting the use of CSA- or paedophilia-related usernames does also not in itself risk detection or weaken security for the users, as was suggested by Woodhams et al. (2021). Users being members of a dark web forum discussing CSA and sharing CSAM already make themselves known as individuals with a sexual interest in children or an interest in CSA, and a username highlighting this does just confirm this but not add this information. Contrary to what Woodhams et al. (2021) suggest, users examined in the current work even use self-representation through such explicit references to exhibit their expertise, their experiences and their views regarding such practices; to them, including specific mentions of paedophilia or CSA in their username is not something to be avoided but used as a linguistic resource for their identity work and performance in the studied community of practice.

7.3 Contributions

7.3.1 Contributions to Linguistic Research into CSA

This research has shown the usefulness of applying a corpus-assisted discourse analysis approach to a large corpus of usernames and forum posts taken from online fora discussing child sexual abuse. While unusual for the study of individual names without direct co-text, this study shows that applying key elements of Van Leeuwen's Social Actor network to the study of usernames and identity performance in offender-to-offender CSA communications can further the understanding of linguistic behaviours and language use in these contexts.

Access to genuine data is rare and so is research into those communities. Few linguistic studies examining the language of offender-to-offender communications exist, and linguistic studies into online criminal networks are scarce. This work offers one of the few studies investigating the language used by suspected child sex abusers on the dark web and presents the first study of usernames in

this context. It thereby provides vital, initial findings to this area of linguistic literature. Through the manual segmentation process, which virtually anonymises the data, this work is also able to provide other researchers, especially in computational linguistics and natural language processing, with a data set for further analysis. The author hopes the data can be made available through Aston's Forensic Linguistic Databank (FoLD) in due course .

7.3.2 Contributions to Onomastics Research on Usernames

As the first study of usernames in the context of dark web communications between suspected child sexual abusers, this work directly contributes to onomastics literature in a new context. Adding to the work of Hämäläinen (2015; 2020) which shed light on users' naming practices in the context of online drug markets, this work provides the first look into a second criminal context and shows how understanding online usernames as a distinct naming category can aid in understanding not just those particular communities but also online naming as identity performance more generally. This work further shows how a linguistic tool like the social actor framework can be usefully applied to the study of usernames.

7.3.3 Contributions to the Role of Linguistics in Policing Online Crime

In the beginning of this thesis the benefits of employing linguistic analysis to aid the policing of online criminal behaviours has been discussed. Previous studies have successfully shown that such linguistic analysis has a positive impact on law enforcement practices in the CSA context, especially when it comes to undercover operations, and linguistic analysis can play an important role in furthering understanding of offender communications in online spaces related to child sexual abuse (see Grant and Macleod, 2016; MacLeod and Grant, 2017; Chiang et al., 2021).

The work presented in this thesis has aimed to build on this work by analysing another key aspect of suspected offenders' online identity performance: usernames. While this work presents only a limited and first analysis of username data in this context, it is a promising starting point in

better understanding child sex abusers' online identity work in those particular online communities by shining a light on the linguistic choices and resources available to members when creating their chosen identifiers and what those choices can tell us about their linguistic behaviours in the forum discussions.

The author believes that the findings presented in this thesis could have various practical implications for Law Enforcement agents policing online spaces related to CSA or undercover operatives planning to infiltrate such online CoPs. Study 2 and 3 show that users' naming choices are closely linked to the type of content they contribute in the discussion fora. When it comes to infiltration tasks, it might be beneficial for undercover officers to join as a new forum member under a name that particularly fits the exact policing context. If the task is to infiltrate a group of offenders who abuse their own children and share advice on this topic, an officer would do best by creating a username that also reflects this kind of membership by including a reference to their own familial relationship to a child as part of their username. If infiltrating online groups of offenders who primarily trade CSAM, especially that which is of the most violent type, often referred to as hurtcore, undercover officers have a higher chance of acceptance into the community if their usernames indicate their willingness to inflict significant harm on a child. It is the author's understanding that so far, undercover agents in the UK context are usually prevented from using such explicit names but this research and any future research on this topic could give LEAs the necessary evidence to support such naming choices in order to have officers perform more genuine offender identities in these online spaces.

Study 2 also provides us with information regarding the usernames in terms of social actor representation. The usernames studied here could be understood as either of the following three categories: 1) backgrounding or classification, 2) nomination, or 3) backgrounding (of the adult) and objectivation (in form of somatisation) of the child. This gives law enforcement officers who want to infiltrate a dark web CSA forum the possibility to pick a username according to their policing needs. If the forum to be infiltrated is one that focuses on violent content, perhaps a username that uses

somatisation would be a good choice to be more easily accepted within the group. For other types of groups, such as those where users self-identify through classification as 'childlover' or 'paedophile', officers could pick usernames that make use of the same form of classification.

Study 3 showed that users engage in discussion of different topics based on their type of username. This means organisations and institutions wanting to infiltrate such groups could consult such research to better understand not only what those common topics are (study 2) but also how exactly these topics are discussed (study 3). For example, users of different subgroups engage in sharing fictional stories, but knowing that 'childlovers' focus heavily on fictional child characters, while others focus more on the adult characters in the story, could be crucial in performing a genuine in-group identity.

Another interesting question that arises from this research is whether the findings could be used for predictions which might then help law enforcement with the policing of such online spaces. It would be interesting to see whether, for example, usernames that include a segment referring to violence are also more likely to discuss violence or be a member of a violence focused forum. Currently the number annotated in this work is still a small sample but with a much larger annotated data set, a logistic regression could be performed to explore predictions.

7.4 Application of Findings

Findings from this research that have practical applications include all those aspects applicable to law enforcement contexts. Not only can research like this into offenders' naming choices aid undercover officers in performing better, authentic identities during infiltration tasks, the findings also offer insights which suggest that with additional research in this area building on the insights shared in this thesis, law enforcement agencies might benefit from this type of linguistic and onomastic research for purposes such as identifying offender types by username which would allow them

to specifically target individuals that are deemed most dangerous or influential/powerful in these communities.

Another application lies in prevention work. Initiatives like the Project Dunkelfeld in Germany, where paedophiles are given a safe space to contact psychologist and psychiatrists to talk about their sexual interest in children with the goal of preventing offending, rely on individuals reaching out themselves (Beier et al., 2015). To reach paedophiles who are at risk of offending and to thereby prevent offending, usernames can be useful. Users who struggle with their identity as a paedophile and make that known through their choice of username may be possible targets for prevention approaches that start with the potential offender. Organisations, possibly with the help of law enforcement agencies, could target such individuals by directly contacting them through these fora and make them aware about initiatives like the Project Dunkelfeld, so individuals are targeted for prevention rather than approached only after they have committed an offence.

7.5 Future Research

The findings presented in this thesis also lead to new ideas and opportunities for further research, both in linguistics and onomastics but also in related disciplines which investigate data in the OCSA context.

One clear and important finding from the current research is that while there are different foci of discussions in the various user groups, they all follow the convention of using proper names as part of their usernames to some extent. The 5 types of users studied in studies 2 and 3 show a high use of proper names in the discourse. This partially are references to fictional characters but mainly involve discussing child victims. That is unlike the average user among the most frequent users and raises the question: what do other types of users focus on and what are those types? Attempting to answer this question through further annotation and analysis of the full data set forms a natural next step for further investigation.

Given the amount of foreign language data in the full corpus, especially for some of the large languages represented like German, Russian, Spanish and Portuguese, further analysis of those users and their naming choices would be a great addition to the existing research and literature on OCSA and username research more generally. Further, findings from this corpus-assisted discourse study could help to ultimately build a machine learning system for automatic social actor analysis, user name generation and role identification.

7.6 Conclusion

This work has aimed to increase the understanding of dark web child sexual abuse interactions and suspected offenders' identity performance in these spaces and to gain insights into suspected child sex abusers' naming practices to increase understanding and to aid law enforcement in online policing of such communities. It is the impression of the author that both those goals have been achieved.

The first study focusing on segmentation and categorisation of usernames in the online CSA context has shown that suspected child sex abusers make use of a number of resources – reaching from proper names, over self-identification as paedophile, childlover or parent, to expressions of victim preference or explicit violence against children. Users thereby often combine segments representing different identity features into one username to construct their unique identity which they want to share with the community.

The second study dives deeper into members' identity performances by investigating key terms within users' forum interactions based on identity categories established in the first study. With a link to Van Leeuwen's social actor theory, underlying intentions and beliefs based on group identity have been uncovered and provided valuable insights into the kinds of user groups prevalent in these online communities with a focus on child sexual abuse interactions and exchange of IIOC.

Study 3 takes a closer look at the 5 subcorpora and compares them against the DarkWebCSA-top corpus. Findings suggest that while the subgroups are similar in some ways, and some subgroups partially overlap in both their usernames and types of forum content, 4 out of the 5 studied subcorpora show there are distinct characteristics to these user types regarding their forum contributions.

The presented work relates to earlier works on identity expression and performance, and the usefulness of username segmentation to study different groups of users by identity category refers back to assumptions of partialness of self-expression, a key belief of much of the earlier referenced identity literature.

Linking back to literature from onomastics and more specifically username research, this work also underlines the importance of proper names for identity purposes – even in such low-trust and high-risk contexts as the one in the present study.

Though there is no way of knowing causality from this type of analysis, the findings from the three studies presented in this work do show that there is indeed a direct link between a user's choice of username and the content of their forum posts in the context of child sexual abuse communications.

7.7 Final remarks

Effectively combating child sexual abuse, including the distribution of child sexual abuse material (CSAM), requires a collective effort. National law enforcement, international agencies and non-profits working towards this goal have much to gain from working together with tech companies and CSA researchers from different academic disciplines. This research shows that academic research in linguistics enables much needed insight into users' naming practices and performed identities in such dark web spaces where a lot of the CSAM is shared.

Adding linguistic approaches and tools to other types of analysis provided by psychology, criminology and computer science can aid agencies in better understanding these communities to tackle

the issues arising from criminals operating in these high-risk and low-trust environments.

In addition to the value this research adds to the policing of CSEA online, this work offers insights into identity performance online and socio-onomastic issues related to usernames as a distinct form of name. Furthermore, it presents a new application for an established linguistic framework, social actor analysis, and examines the extent to which it can be usefully applied to forum interactions and investigations of online naming.

Through all the above, it is hoped that this work adds to the understanding of online offender-to-offender communications and provides evidence for the value of linguistic inquiry into this important context of online child sexual abuse and exploitation.

Appendix A

Ethics Approval Aston University Ethics Committee

16 June 2020

Professor Tim Grant
Student: Daniela Schneevogt
School of Languages and Social Sciences

Dear Tim,

Study title:	Analysing Group Identity and Social Actors in Criminal Online Communities with Linguistic and Computational Methods
REC REF:	#1660

Confirmation of Favourable Ethical Opinion

On behalf of the Committee, I am pleased to confirm a favourable opinion for the above research on the basis of the application described in the application form.

After starting your research please notify the University Research Ethics Committee of any of the following:

- Amendments. Any amendment should be sent as a Word document, with the amendment highlighted or showing tracked changes. The amendment request must be accompanied by a covering letter along with all amended documents, e.g. protocols, participant information sheets, consent forms etc. Please include a version number and amended date to the file name of any amended documentation (e.g. "Ethics Application #100 Protocol v2 amended 17/02/19.doc").

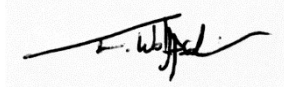
Amendment requests should be outlined in a "Notice of Amendment Form" available by emailing research_governance@aston.ac.uk.

- Unforeseen or adverse events e.g. disclosure of personal data, harm to participants.
- New Investigators
- End of the study

Please email all notifications or queries to research_governance@aston.ac.uk and quote your UREC reference number with all correspondence.

Wishing you every success with your research.

Yours sincerely

A handwritten signature in black ink, appearing to read 'J. Wolffsohn', is centered below the text 'Yours sincerely'.

Professor James Wolffsohn
Acting Chair, University Research Ethics Committee

Appendix B

Study 2: Concordance Tables

Appendix B. Study 2: Concordance Tables

Index	Left context	Keyword	Right context
285	manager of 46 and a boy of	14,	who had encounters over a longer time.[81]
147	q... q..... 2011 UNPD Panama 18 18	14	16 2011 UNSD Papua New Guinea 16
170	2008, raising the age of consent from	14	to 16.[11] There exist two close-in-age exemptions,
14	nor ever dream about existing in a	14	year old girl. And the door that
185	link] Coahuila 12 16 386 394 Colima	14	18 209 211 Distrito Federal 12 18
474	possible to marry the girl at age	14,	though I can't give you any examples
359	money pretty far." "I was 13 or	14	when I discovered my stepdad's porn stash.
260	whereas the age for heterosexual intercourse was	14.	The age of consent was equalised in
122	in all cases. Paraguay 14 16 Peru	14	14 Philippines 18 18 Poland 15 15
35	accurate for security reasons...but yes younger than	14	lol)"" Im like ""whoa! really? then she
28	you, you know Will be the first	14	th o'February since my early twenties that
68	The age of consent in Serbia is	14,	regardless of sexual orientation or gender. This
167	Article 143. ""Sexual intercourse with person between	14	and 16 years of age"".[3] Aruba The
341	cuddly Diana [,8. [,9. [,10. [,11. [,12.,13.	[],14.	Dreamy Delightfull Dolly Diana DianaH vids 1-7[],Sweet
7	said, I was once play wrestling a	14	year old female relative. It was *amazing*
353	""Jelly"". It was the nickname of the	14	yr old prostitute who was the main
369	probably not of legal age by now,	14	or 15 is my guess and from
24	I am only showing one preview not	14	of them. Preview=[URL],... ss.mp4.jpg Archive= RileyFamilyPicnic.7z-
67	age of consent in San Marino is	14,	per Article 173.[114] It is also illegal,
96	14 21 Chile 14 18 China 14	14	Hong Kong 16 16 Macau 17 N/A
348	of those old mags. When I was	14,	I found in my stepfather's stash a
379	bring their own kids along. Children age	14	and under attend free of charge when
172	raising of the age of consent to	14,	the laws against ""seducing"" minor girls were
356	with my little girlfriends Tess and Amber,	14	and 16. They are delightful, happy, beautiful
66	to 20 if the victim is under	14).	Further reading Criminal Code of the Russian
287	11 Ceuta (Spain) 12 Chad 13 Comoros	14	Côte d'Ivoire 15 Canary Islands (Spain) 16
372	counts as weird just as much as	14	inches does. Horse dick or Dachshund dick-
88	Herzegovina 16 16 Botswana 16 Illegal Brazil	14	14 Illegal for military gay men &
336	and 10. girls 6 and 12.. stepson	14.	I am very attracted to all.. yes..
375	step-father. Very interesting that when she was	14	she initiated sex with him only one
354	""When I started trying to date at	14,	I was always intimidated by girls my
361	She has an older sister, too. About	14,	just out of my AOP, but she
357	Trump as 13 and his daughter was	14	when he started hinting to his attraction

TABLE B.1: Random concordance lines for keyword 14 in Names subcorpus.

Index	Left context	Keyword	Right context
62	kind of sexual activity with minors under	15	(or minors under 18 who lack the
223	responsibility as well as sexual restriction from	15	to 14 years was proposed by the
227	the age of consent was equalised at	15	for all acts[citation needed]. But sex with
10	a misfire. Had to order several Korean	15	round mags. Though 13 round mags for
58	The age of consent in Sweden is	15,	as specified by the Swedish Penal Code,
149	16..... 2011 UNSD Slovenia 18 18 15	15	2011 UNSD Solomon Islands 18 18 15
316	ladies in attendance, ranging from 5 to	15	years old. Three pairs of sisters. (Two
24	and standard email. Quote from: Sierra on	15	December, 2013, 08:58:59 Lest there be any
146	14 15 2007 CEDAW (37th session) Niger	15	p 18 p..... 2007 CEDAW (38th session)
208	Europe 15 Switzerland Europe 16 Syria Asia	15	Taiwan Asia 16 Tajikistan Asia 16 Tanzania
265	The age of consent in Cambodia is	15	regardless of gender or sexual orientation.[8] The
244	the case of offenses with children aged	15	and 16, the relevant statute, Sexual act
22	nice. She massaged my neck for about	15	minutes and then asked me to come
250	12 Chad 13 Comoros 14 Côte d'Ivoire	15	Canary Islands (Spain) 16 Democratic Republic of
195	Haiti North America 18 Honduras North America	15	Hungary Europe 14 Iceland Europe 15 India
36	sexual interest in mid-to-late adolescents, generally ages	15	to 19Gerontophilia is a sexual preference for
27	apart. I think it lasted for almost	15	minutes before it just became awkward, or
30	(vid) [URL], HermogenesCanales Silver Angels Valensiya Stockings	15	[URL], HermogenesCanales MarvelCharm Ariana & Karina Unbreakable
279	The age of consent in Laos is	15	regardless of gender or sexual orientation.[55] Article
267	law prohibits ""debauchery"" with a child under	15,	and this term (the original Cambodian word
284	with a girl under the age of	15	shall be ""punished gravely.""[66] Oman Sexual activity
256	Africa is 16, as specified by sections	15	and 16 of the Criminal Law (Sexual
180	185 Jalisco 12 18 176 142-I México	15	18 273 271 Michoacán 12 16 240
361	of this. Oddly, it misses the first	15	minutes, but adds a boring extra 5
8	years in the military). When she was	15	and I was 19, we were forced
415	helps me a lot. No question." "Around	15.	Only a few pictures. And only those
296	one folder. Awesome Lesya issue!! Enjoy! Issue	15	of 16 [URL], [URL], I think I
234	relations with minors over age 15 (aged	15,	16 or 17) ""1 * where they are
292	works fine, just tested the download Ls-Land	15-	Little Duchess Issue 15 of 32 All
183	age of consent in Sint Maarten is	15,	as specified by the Criminal Code of
202	of the Congo Africa 18 Romania Europe	15	Russia Europe 16 Rwanda Africa 18 Sahrawi
185	married<13 years old 14 years old	15	years old 16 years old 17 years
38	pleaing to this at all 2: the	15	years- 60+ years are the min and

TABLE B.2: Random concordance lines for keyword 15 in Names subcorpus.

Index	Left context	Keyword	Right context
56	in 2003 it was returned again to	16	years[109] (which is the current age of
911	born in it's normal since she was	16	when she had me and my dad
459	2006 CEDAW (35th session) Samoa 19 21	16	18 2005 CEDAW (32nd session) San Marino
422	2001 CEDAW (24th session) Montenegro 18 18	16	16 2011 UNSD 15 Montserrat 18 18
449	16 18 2011 UNSD Qatar 16 18	<16	<18 2011 UNPD Republic of Korea 20
448	21 21 16 18 2011 UNSD Qatar	16	18 <16 <18 2011 UNPD Republic of
434	15 18 2011 UNPD Norway 18 18	16	16 2011 UNSD 1 Oman 18 18
230	18 18 Poland 15 15 Portugal 16	16	Puerto Rico 14 14 Qatar Must be
447	16 2011 UNSD Puerto Rico 21 21	16	18 2011 UNSD Qatar 16 18 <16
127	& Norfolk Is. 16 16 Northern Territory	16	16 Queensland 16 16 Sodomy (or in
436	1 Oman 18 18 2011 UNSD Pakistan	16	18..... 2013 CEDAW (54th session) Palau... q...
213	Morocco 15 Illegal Mozambique 16 N/A Namibia	16	16 Nepal 16 N/A Netherlands 16 16
423	CEDAW (24th session) Montenegro 18 18 16	16	2011 UNSD 15 Montserrat 18 18 16
152	Burma 14 Legal Burundi 18 Illegal Cambodia	16	N/A Cambodia 16 16 Cameroon 21 Illegal
440	2010 CEDAW (46th session) Paraguay 20 20	16	16 2005 CEDAW (32nd session) Peru 18
151	married Bosnia and Herzegovina 16 16 Botswana	16	Illegal Brazil 14 14 Illegal for military
188	16 16 Guinea 15 Illegal Guyana 16	16	Haiti 18 18 Honduras 15 15 law
437	q..... 2011 UNPD Panama 18 18 14	16	2011 UNSD Papua New Guinea 16 r
427	18 2011 UNSD Mozambique 18 18 16	16	2011 UNPD Myanmar 20 20 14... 2011
430	18 18 2011 UNSD Netherlands 18 18	16	16 2011 UNSD Netherlands Antilles 18 18.....
415	15 18..... 2006 CEDAW (34th session) Malta	16	16..... 2011 UNSD Marshall Islands 18 18.....
123	16 Aruba 16 16 Australian Capital Territory	16	16 New South Wales & Norfolk Is.
214	15 Illegal Mozambique 16 N/A Namibia 16	16	Nepal 16 N/A Netherlands 16 16 Netherlands
444	Philippines 21 21 18 18 2011 UNSD	16	Poland 18 18..... 2011 UNSD Portugal 18
247	16 16 Syria 15 Illegal Taiwan 16	16	Tajikistan 17 17 Tanzania 18 Illegal Thailand
13	the Workhouse) shows it starting when he's	16	and she's 12. The author's comment: "their
365	18 18 16 16 2011 UNSD Cuba	16	18 14 16 2011 UNSD Cyprus 18
216	Namibia 16 16 Nepal 16 N/A Netherlands	16	16 Netherlands Antilles 16 16 New Caledonia
331	18 Washington 18 18 West Virginia 16	16	Wisconsin 18 18 Wyoming 18 18 No
548	the perpetrator's spouse who is at least	16	years but less than 18 years of
336	(25th session) Angola 18 18 15 a	16	a 2013 CEDAW (54th session) Anguilla 18
196	the rest of the country. Jersey 16	16	Jordan 16 16 Kazakhstan 18 N/A Kenya
351	16 2011 UNSD Cape Verde 18 18	16	g 16 g 2006 CEDAW (36th session)

TABLE B.3: Random concordance lines for keyword 16 in Names subcorpus.

Index	Left context	Keyword	Right context
175	who would be my future wife a	17	year old beautiful Latin girl. She was
30	months-to-5 years prison. Greenland 15 15 Grenada	17	17 Guam 16 16 Guadeloupe 15 15
6	good Most of the play came later	[], 17.	When you begin enjoying enemas? How and
69	1 Sierra Leone 18 18 2011 UNSD	17	Singapore 21 21 <21 <21 2011 UNSD
54	New Mexico 17 17 New York 17	17	North Carolina 16 16 North Dakota 18
120	marriageable age is 15 for girls and	17	for boys.[53] Kyrgyzstan The age of consent
115	country- puberty- 12- 13- 14- 15- 16-	17-	18- 19- 20- must be married- no
60	2011 UNSD Democratic People's Republic of Korea	17	18..... 2005 CEDAW (33rd session) Democratic Republic
33	married Illegal Iraq 18 N/A Ireland 17	17	Isle of Man 16 16 Israel 14-16
119	Cambodia Cambodia 15 China China 14 Cyprus	17	East Timor East Timor 14 Egypt 18
152	to be alone with anyone younger than	17.	Three orders of protection were issued at
95	is allowed for teenagers aged 16 and	17	to engage in sexual intercourse as long
111	current ones.[44][45] The exemption for girls under	17	was recommended by the LRC and the
145	you ever masturbated a child to orgasm?	17-	Have you ever ""played doctor"" with a
204	Enrage) (17) Cosma Shiva Hagen Laden, Der	(17)	Cristi Harris Night Of The Demons 2
56	16 16 Tennessee 18 18 Texas 17	17	Utah 18 18 Vermont 16 16 Virginia
142	and married an 18 yr old at	17,	Had a couple kids by him, but
29	15 "Seduction" of a male younger than	17	could be punished by 3 months-to-5 years
15	dating a 23 year old and she's	17.	She's such a goody goody though, she'd
14	are what I like. So yeahhhh." "Jessica,	17,	USA: I've never met another girl that
170	again. So, for me somewhere between 14-	17.	The real appeal is having all my
46	16 16 Kentucky 16 16 Louisiana 17	17	Maine 16 16 Maryland 16 16 Massachusetts
61	2011 UNPD Dominican Republic 18 18 16	17	2011 UNSD Ecuador 18 18 <18 <18
136	boys between the ages of 14 and	17.	It is a homosexual phenomenon as in
143	my garage is three blocks away.' Question	17	You find out you have a coworker
91	16 Mauritius Africa 14 Mexico North America	17	Moldova Europe 16 Monaco Europe 15 Mongolia
12	tells me I've spent a little over	17	days here while logged in, geezus. And
177	or a 11 yo M and a	17	yo F or something really crazy like
45	Kansas 16 16 Kentucky 16 16 Louisiana	17	17 Maine 16 16 Maryland 16 16
19	the age of consent was lowered from	17	to 16 in 2008 with the Order
121	[URL], Updated download link PW: Johnny Ls-Land	17-	Forbidden Fruit Issue 17 of 32 All
149	are the first three. It is a	17	year old boy and his 13 year
141	character. I read it when I was	17.	I would love to find that story

TABLE B.4: Random concordance lines for keyword 17 in Names subcorpus.

Index	Left context	Keyword	Right context
44	of sexually explicit materials is regulated under	18	U.S.C. 2257, requiring ""original"" producers to retain
162	Vermont 16 16 Virginia 18 18 Washington	18	18 West Virginia 16 16 Wisconsin 18
3	not see the issue with folks under	18	posting here. Pretty sure that it isnt
16	been in 'the system' until I was	18.	In my case I've always believed that
166	Virginia 16 16 Wisconsin 18 18 Wyoming	18	18 No Results Filter: Loading... References Embed
68	for both heterosexual and homosexual sex at	18,	with some differences, such as the act
17	am aware she is about to turn	18	and I am intrigued to hear how
127	Romania 15 15 Russia 16 16 Rwanda	18	18 San Marino 14 14 Saudi Arabia
26	No LEA is out posing as an	18	y/o to get you into jail, they
132	Tajikistan 17 17 Tanzania 18 Illegal Thailand	18	18 Togo 14 Illegal Tonga 16 Illegal
14	My little girl is all grown up.	18	years old now and moved across the
69	activity with a child, ""(1)A person aged	18	or over (A) commits an offence if—
467	p..... 2006 CEDAW (34th session) Timor-Leste 15	18.....	2011 UNPD Togo 17 20..... 2012 CEDAW
128	15 15 Russia 16 16 Rwanda 18	18	San Marino 14 14 Saudi Arabia Must
175	a 2013 CEDAW (54th session) Anguilla 18	18.....	2011 UNSD Antigua and Barbuda 18 18
301	17 17..... 2009 CEDAW (44th session) Guyana	18	18 16 16 2012 CEDAW (52th session)
199	Belarus 18 18..... 2011 UNSD Belgium 18	18	<18 <18 2011 UNSD Belize 18 18
275	2004 CEDAW (31st session) Eritrea 18 j	18.....	2006 CEDAW (34th session) Estonia 18 18
278	18 18 15 15 2011 UNSD Ethiopia	18	18..... 2011 CEDAW (49th session) Fiji 21
296	Greece 18 18..... 2011 UNSD 9 Greenland	18	18 15 16 2011 UNSD Grenada 21
290	18 18 16 16 2011 UNSD Ghana	18	18..... 2011 UNSD Gibraltar 18 18 16
293	Ghana 18 18..... 2011 UNSD Gibraltar 18	18	16 16 2011 UNSD 8 Greece 18
211	Bermuda 18 18 2011 UNSD Bhutan 18	18.....	2009 CEDAW (44th session) Bolivia 14 16
74	that lowered the age of consent to	18	was accepted by 427 votes to 162.
424	16 18..... 2011 UNSD 1 Romania 18	18	2011 UNSD Russian Federation 18 18 16
12	with sometime, even if they're all over	18	now. I spent quite a few nights
279	18 15 15 2011 UNSD Ethiopia 18	18.....	2011 CEDAW (49th session) Fiji 21 21
351	2011 UNSD 13 Malaysia 16 and 21	18	and 21 16 18 2011 UNSD Maldives
437	q..... 2011 UNPD Senegal 18 18 <18	<18	2011 UNSD Serbia 18 18 16 16
883	mopeds and agricultural vehicles I think), alcohol	18,	marriage 16 (with parental permission) 18 otherwise.
225	<21 e 2008 CEDAW (40th session) Cambodia	18	18..... 2011 UNSD Cameroon 15 f 18.....
416	Poland 18 18..... 2011 UNSD Portugal 18	18	16 16 2011 UNSD Puerto Rico 21
165	18 West Virginia 16 16 Wisconsin 18	18	Wyoming 18 18 No Results Filter: Loading...

TABLE B.5: Random concordance lines for keyword 18 in Names subcorpus.

Index	Left context	Keyword	Right context
472	Charges are relatively low (up to 4	years	of prison), regardless of gender and ""obscene
1,801	first dip into the darkweb a few	years	back, (H2tC, as it happens) found just
934	Hello all, I am struggling for many	years	with feelings. Always had interest in sweet
290	victory over his father for a few	years	yet to come but he can win
761	age of consent was set to 21	years	and in 1973 reduced to 18 years.
736	coitus with a person younger fifteen (15)	years	of age or who sexually abuses such
746	Inability of person of less than 10	years	to comprehend Within the meaning of the
317	such a rush. He is only 5	years	old but I have never cared or
1,529	only take those every 5 to 10	years	or so. CP has been my primary
1,845	rebellion, and certainly, not in my 'mature'	years.	Indeed, we carried on meeting even in
1,816	was heartbroken when he died a few	years	back, reasonably suddenly, a loss I feel
739	an intercourse with a person younger 18	years.	Between adult persons, payment receipt or provision
273	rapes in France jump 18% in five	years""	rapecrisis.org.uk/statistics.php ""Rape Crisis England & Wales headline
1,604	envy your position and view for four	years.	Did they sunbathe often during that period
275	servicesMaking Child PornographyTrafficking in persons under 18	years	To be honest it's hard to imagine
729	term of imprisonment of one to ten	years.	(2) Whoever commits lewd act with a
730	lewd act with a child under fifteen	years	of age or induces a child to
814	time under the age of sixteen (16)	years;	and c) that the girl or person
1,891	in living off the grid for many	years,	this way you have multiple perspectives, and
1,573	work (that know me for a few	years)	suspect I do not fall for woman
308	like girls from infants to around 6	years	old at the max. I am mainly
276	a bad way within just a few	years.	Back when I was 18 the only
722	shall be punished by imprisonment (of five	years	to ten years).[13][14] History The Belgian Criminal
818	in São Tomé and Príncipe is 14	years,	regardless of sexual orientation and/or gender.[46] Senegal
1,915	and haven't talked about it for many	years	now. I do notice when I talk
825	consent. The penalty of imprisonment exceeding 10	years	is the sentence for anyone who has
1,910	to underage girls. This was like 10	years	ago. It made me think of a
1,895	can be up to 1 or 2	years),	then closing them. What cops can NOT
265	all the items I gathered over the	years,	they're not many but still a bunch,
733	sexual intercourse"" aimed at persons under 18	years	with any promise or provide of payment,
805	age of consent in Kenya is 18	years,	as children are defined under 18 and
331	post this but over the last 5	years	or so I have been struggling with
1,802	photos of my kids over swo many	years,	many naturist photos both at the club

TABLE B.6: Random concordance lines for keyword *years* in Names subcorpus.

Index	Left context	Keyword	Right context
123	peaking up the skirt of a 4	yr	old Hispanic girl who'd dropped her panties
16	make this one more simple. The 8	yr	girl you consider to be a little
37	an 18 yr old, but plays 13	yr	old Jasira. She is a sexually awakening
13	in a circle jerk around the 3.14	yr	old girl in the middle offering her
15	to her three brothers. The bird-watching 11	yr	old girl, Robin, he takes the bus
6	at the number of 10 and 11	yr	old girls taking nude selfies or playing
199	it. Dorothy explains that the little 3	yr	old has stopped growing and has been
47	looking good than with a certain seven	yr	old suddenly looking very good. There are
19	probably actuality more interested in her 6	yr	old sister in her tight pink yoga
27	as well. One pretty, skinny little 8	yr	old brunette in a dark blue one-piece
271	or have gotten pregnant by a 12	yr	old boy. If still possible- abortion.If not-
8	trying not to mentally undress a 7	yr	old with her mum watching is a
174	when I was engaged to a 15	yr	old at age 18, incest was in
204	wished that I could marry that 10	yr	old, actually shorter and lighter than her
75	worker for a family with an 11	yr	old autistic girl named Glory played by
2	to slide every inch right up it"	"yr	so right. for sure her shitter needs
147	gives her book review, followed by 7	yr	old boy getting a bit of sex
83	me. My heart was with the 15	yr	old at home at the time. But
32	I wish I could marry a seven	yr	old girl." "Must be a little easier
4	It would feel fucking great to sink	yr	hard dick into her. Ive got a
190	foxed me out at 27, the 4	yr	old girl who was such tease to
80	a marriage between a 16 and 19	yr	old was still doable, even it raised
158	bad. I wish I had a 6	yr	girl on my lap right now. Cuddle,
102	scenario- my new celly is an 18	yr	old boy and he has learned that
1	those apart and ram it in" "Yes	yr	right, allowing you to slide every inch
97	those CP pics of that oiled 5	yr	old with the caption ""For Pollyfan"". Not
216	Jesus." "If I had married the 10	yr	old girl suitor-ess when she proposed to
214	get it for her, like a 5	yr	old too short to reach the kitchen
167	seen them. How much does a 3	yr	old remember when they are 4? They
250	toward me a 10 and a 12	yr	old girl, dressed very seductively in short
104	yr old girl's butt and a 18	yr	old boy 's- I mean 8 yr
9	out that i can legally fuck 16	yr	olds, but if i was to film
230	and suddenly she reverts to a 12	yr	old getting her cherry popped on top!

TABLE B.7: Random concordance lines for keyword yr in Names subcorpus.

Index	Left context	Keyword	Right context
1,590	jazz. One set had the girls from	age	8 to 15 in white leotards with
375	of Russia was adopted in 1996, the	age	of consent was proclaimed to be the
698	a few states) and 15, while the	age	at which there are no restrictions for
684	over fifteen and under eighteen years of	age,	relying on the superiority arising from any
547	Data not available 1 The minimum legal	age	at which marriage can take place is
572	Coahuila, Hidalgo, and Zacatecas. The Minimum legal	age	for marriage remains the same respectively for
557	years whereas in Nunavut, the minimum legal	age	is 18 years. 6 Persons less than
958	as long as a person over the	age	of 21 does not exploit a 14-
1,002	the age of 18, provided that the	age	gap between them is less than 3
54	and that this is normal, regardless of	age.	I wish this myth would disappear because
650	12: in 1890, the Parliament raised the	age	of consent to 14.[17] The punishment for
559	pregnancy) the marriage can take place without	age	restrictions. 9 To marry at age younger
877	permanently or as a temporary, practical expedient."	"Age	of Consent Laws in Europe: [URL], Map
355	subjects a minor under 15 years of	age	to sexual intercourse or makes him/her submit
954	at 13 in 1863. In 1942, the	age	of consent for homosexual acts was set
304	some political parties no matter what the	age	of consent is set at it will
160	abuse to have sex before a certain	age	and well, as you said, teenagers being
429	temporary visit. This is regardless of the	age	of consent in the country where the
865	all sex outside of marriage irrespective of	age,	as in the case of Yemen.[54][55] Prostitution
942	of 14 for sexual intercourse. Finland The	age	of consent in Finland is 16, as
1,129	consent is 16. A girl under the	age	of fourteen (14) cannot consent to sexual
177	in mind that she is at an	age	when she barely has understood that she
617	her consent and who has attained the	age	of fourteen years but has not yet
289	on the premise that children under the	age	of sixteen do not have the mental
1,144	also 48 References Consent by country Country	Age	of consent Afghanistan Afghanistan Must be married
326	LEA: ""Drax"". viewtopic.php?f=30&t=4734#p33400 [URL],... its-guilt" "Moldova The	age	of consent in Moldova is 16, per
1,087	due to the Libyan Civil War. Madagascar	Age	of consent is 14 years, according to
1,081	female homosexuality may be illegal.[28] Lesotho The	age	of consent is 14 years, regardless of
651	""seducing"" minor girls who were over the	age	of consent. In 1886, a law was
263	and she is too (even at her	age)	she tells me she wont ever find
242	might just enjoy babies for a certain	age	window but I was curious what is
217	hard to imagine someone of such young	age	to be neckdeep into the darkest side
630	and that other person has attained the	age	of 14 but has not yet attained

TABLE B.8: Random concordance lines for keyword *age* in Names subcorpus.

Index	Left context	Keyword	Right context
101	independence was introduced, making the age of	consent	of 15 for all sexual acts, regardless
124	were removed. In 1995, the age of	consent	was set at 14 for all acts.
30	of little to no value for the	consent	topic. I read through the rest of
675	under fourteen years. Nepal The age of	consent	in Nepal is 18 regardless of gender
165	Section 80, which specified an age of	consent	of 21. The Criminal Law (Consolidation) (Scotland)
1	of life for them. There is no	consent.	There is picking and choosing and betting
123	15 for all. Slovenia The age of	consent	in Slovenia is 15, as specified by
122	thus effectively making an equal age of	consent	for any sexual practice regardless of sex
125	was amended to raise the age of	consent	to 15 years and added the condition
128	maturity to that minor. The age of	consent	rises to 18 if there is deception
16	someone between 12 and the age of	consent	(statutory rape). The former can yield sentences
50	for is 12yo for the age of	consent	Just one little question that I have
27	offender laws, youth rights, and age of	consent	reform. Also I think grass roots activism
121	conduct. Since 2006, an equal age of	consent	came into force, regardless of sexual orientation
440	states, or other jurisdictions, the age of	consent	may be lower than the age at
294	in 1969 with the same age of	consent	as vaginal sex. Also introduced in 1969
43	by definition. Before someone is able to	consent	to sexual interaction with another person they
48	has to find themselves a definition for	consent	and ensure that it is applicable to
258	9.2 History of the Canadian age of	consent	10 Cayman Islands (United Kingdom) 11 Caribbean
220	to the Islamic law, marriage requires parental	consent.	Consent of the bride herself, as well
413	and development: for instance, the age of	consent	in Finland is 16, but the law
183	orientation.[153] Isle of Man The age of	consent	in the Isle of Man, a Crown
191	males).[159] In 1995, the sodomy age of	consent	became 18 (under the Sexual Offences (Jersey)
164	Since 1 December 2010, the age of	consent	in Scotland is 16, regardless of sexual
404	calculated.[28] Defenses and exceptions The age of	consent	is a legal barrier to the minor's
257	are subject to interpretation. The ages of	consent	in the countries of Central America range
249	Browse Online data Browse Contact" "Age of	Consent	Laws in North America: [URL],... th Age
428	certain age limit (generally 18) in which	consent	of the younger person was obtained through
230	bride and groom with or without parental	consent	in Baja California, Baja California Sur and
273	Bahamas In The Bahamas, the age of	consent	for opposite-sex activity is 16 and the
416	and Suriname have a higher age of	consent	for same-sex sexual activity. In both the
261	must be of legal age to give	consent,	although exceptions to the age of consent
7	to time and I would never, EVER	consent	to a DNA swab or even a

TABLE B.9: Random concordance lines for keyword *consent* in Names subcorpus.

Index	Left context	Keyword	Right context
89	law. Article 175 of the Federal District	Criminal	Code refers to the previous Article 174,
4	the fact that they consider me a	criminal	until proven otherwise seems pretty fucked up
31	is subject to a multifaceted regulatory framework.	Criminal	legislation is in force at the Commonwealth,
136	of 18 years are considered minors, the	Criminal	Code does not criminalize sex with people
166	drama-thriller explores the moral ambiguities of a	criminal	act that occurred 15 years ago between
26	posted to the account.” According to the	criminal	complaint, investigators believe Harding uploaded some of
121	is 16, as specified by the Lithuanian	Criminal	Code §151.1,[66] introduced by the bill of
184	but the fact that I am a	criminal	just waiting to be caught. On the-
27	his police car. Ray said in the	criminal	complaint that location data from Harding’s police
47	Code of the Russian Federation[110] (in Russian)	Criminal	Code of the Russian Federation[111]— Chapter 18
33	par. (1) shall not be subject to	criminal	liability if he/she is similar to the
62	an age of consent of 21. The	Criminal	Law (Consolidation) (Scotland) Act 1995 lowered the
111	(Sexual Offences) Act 2017, which amended the	Criminal	Law (Sexual Offences) Act 2006.[47] Sentences are
46	the victim is under 14). Further reading	Criminal	Code of the Russian Federation[110] (in Russian)
157	The world wants to make any non-conformity	criminal	not just eccentric, to make us think
55	16, as specified by the Swiss Federal	Criminal	Code, Article 187(1). However, there exists a
1	area is dangerous because cartels and potentially	criminal	illegals use that area to cross over
135	the older partner would be exempt of	criminal	liability if three conditions are met: the
6	take care of you sooner than any	criminal.	Again, Im not looking to argue the
67	acts was lowered to 18 under the	Criminal	Justice Act 2001.[155] In 2006, under the
83	Curaçao is 15, as specified by the	Criminal	Code of the Netherlands Antilles (which Curaçao
35	is 16, as specified by the Dutch	Criminal	Code at Article 245,[84] and Article 247
32	rather than open and free public debate.	Criminal	legislation is complemented by a further tier
30	age when children are punishable for their	criminal	actions because they are deemed to have
88	Legislative Assembly has powers to legislate in	criminal	law. Article 175 of the Federal District
48	between men of any age was a	criminal	offence in Serbia, although the law has
169	disruption the police harassment presents to their	criminal	business, meet and vow to use all
66	decision and by repealing and updating 140-year-old	criminal.	laws of Gibraltar.[151] Guernsey (including Alderney, Herm
29	between right and wrong in all matters	criminal.	Take the United Kingdom for example where
94	(in 2015—from 13 to 16). The International	Criminal	Court Statute does not provide a specific
113	age of consent to 18 years.[49] The	Criminal	Law Amendment Act 1935 raised the age
172	never hurt them, that makes them a	criminal	first and a sick guy second. It’s
93	The government was forced to propose the	Criminal	Law Amendment Act 1885, which raised the

TABLE B.10: Random concordance lines for keyword *criminal* in Names subcorpus.

Index	Left context	Keyword	Right context
2	case. I mean ""the minimum[...] is life	imprisonment""	is quite stupid. What's the maximum then?
87	shall be liable to a term of	imprisonment	of one to ten years. (2) Whoever
70	even with consent, shall be punished with	imprisonment	from two to four years This does
1	facing prosecution and up to 2 years	imprisonment,	under the Regulation of Investigatory Powers Act,
69	or without consent, shall be sentenced to	imprisonment	for a period of 12 to 15
7	a minor, shall be punished by an	imprisonment	sentence of three months to three years.
25	a term up to three years or	imprisonment	for the same term." 2. The same
23	authority or influence, shall be punished with	imprisonment	of one to three years. History Same-sex
26	other grave consequences,— shall be punishable by	imprisonment	for a term of three to five
37	body is liable to a term of	imprisonment	of not more than eight years or
52	of sixteen (16) years(...)- is liable to	imprisonment	for seven years.[25] Dominican Republic The age
75	or (b) on conviction on indictment, to	imprisonment	for twenty years. (3) Where a person
90	age of 15, shall be punished by	imprisonment	for up to 8 years, unless the
36	intercourse, and is liable on conviction to	imprisonment	for seven years. A female adult is
22	ways, the offender shall be punished with	imprisonment	from eight to twelve years, in the
73	is liable on conviction on indictment, to	imprisonment	for life. 19. (1) Subject to section
12	under 14 years, to 3 year of	imprisonment,	and up to 21 years. If the
104	upon a child shall be punished with	imprisonment	of six months to five years. However,
55	or mentally handicapped, shall be punished with	imprisonment	from two to four years. Articles 163
83	by arrest of about six months or	imprisonment	of one year to three years.""[12] Belgium
76	is liable on conviction on indictment to	imprisonment	for life. Three close-in-age exemptions exist, as
137	10 years for distribution. Authorizes fines and/or	imprisonment	for up to 30 years for U.S.
126	liable to a term of nine years'	imprisonment	at hard labour. 2. The said term
27	a term up to five years, or	imprisonment	for a term up to three years.[132][133]
8	body is liable to a term of	imprisonment	of not more than eight years or
34	will have a capital punishment or life	imprisonment.	Bangladesh N/A Illegal Barbados 16 Illegal Belarus
111	Decree n. 7, 1985] ,The penalty of	imprisonment	for a term not exceeding twenty years
116	is punishable with up to 5 years'	imprisonment.	Hong Kong In Hong Kong the age
106	used by the offender.[71] The punishment is	imprisonment	from three to nine years, with or
109	an offence and liable on conviction to	imprisonment	for a period not exceeding six years
49	body is liable to a term of	imprisonment	of not more than sixteen years"". Before
6	age of 16 shall be punished by	imprisonment	for up to 5 years. (2) The
11	same penal code, reads: ""A penalty of	imprisonment	for a term not exceeding 10 years

TABLE B.11: Random concordance lines for keyword *criminal* in Names subcorpus.

Index	Left context	Keyword	Right context
79	I want to find a like minded	community	to which I can honestly and respectfully
85	Best, Matt Great intro- welcome to the	community!	wiggler for penis Indeed! When I spend
205	dick its a phenomenon in the pedo	community	HALUCINATION they are so wanting to see
436	some friends and become part of a	community.	There is no magic formula that will
106	from now. I mean, we as a	community	can either explain the world to you
300	want to take the heat from the	community	or fellow employees. Just about all employers
262	and never has been a media sharing	community"".	Or think we really don't mean it.
96	to help me, I really appreciate the	community..."	"What sleeping stuff? I tried ambien this
103	cornation of an amazing mainstay of this	community.	Returning to the castle after a long
114	one of those guys who keep this	community	together! Take care;" "Ayyy, thanks guys Usually
109	my life that I share with this	community).	There are things like Rachael's age that
210	binds us together into a Family /	Community!	Thank you Maestro, I've just been admiring
288	staff is managing the spirit of the	community.	It's their job. You said you agreed
86	the introduction! Great introduction! Welcome to the	community!!	Thank you for introducing yourself! Jealous! I
97	into this in public doesn't help our	community.	My point exactly. Don't take it so
101	everything that was going on in the	community.	One couldn't deny the pride in her
93	have said? Thanks again for this great	community."	"Update: So thinking where to start.... I
267	one way to do that. Family friendly	community	events is another. If people ask why
110	some of the amazing people of our	community	who took their time and listened to
229	say that an individual family within the	community	couldn't socialize nude with them (and possibly
111	to be active. This is a support	community	and not some sort of pedo facefook,
95	have been looking for such a great	community	for a very long time. I think
285	sort of an extended introduction to the	community,	which deserves a bit of a spotlight,
261	absolute need for the families in our	community	at least to be educated about this
112	our very own real world and this	community	appart, after all this is the land
122	of it. We are like a closed	community	since the public body of thought is
274	to inflame people needlessly. We want a	Community	of Support. So conforming means setting ourselves
193	this area, documents of core beliefs and	community	values? Of course I would assume all
208	this Community safe. A strong and safe;	Community	and System-Service is our combined-responsibility. I wish
268	take me some time, but if the	community	would like or if staff would think
84	Hutcherson in that one. Welcome to the	community!	I enjoy the horror genre as well.
280	has adopted as his extended family /	community	than he does about getting his rocks
138	because we can't come together as a	community	and its not just my opinion its

TABLE B.12: Random concordance lines for keyword *criminal* in Names subcorpus.

Index	Left context	Keyword	Right context
8	new and you clearly havent ""got"" what	<i>community-name</i>	is about. Back to the topic: I
220	I think, in the spirit of the	<i>community-name</i>	Make a Wish Foundation, every wish must
148	of what I have contributed here in	<i>community-name,</i>	mine also!) I can identify with getting
82	boards that are out there (including the	<i>community-name).</i>	Still, I'm frightened to be honest about
65	on-topic or 'normal'. Take care;" "D ear	<i>community-name,</i>	I'd like to thank all of you
62	any genre) and love photographing nature. Until	<i>community-name,</i>	I never had any interactions with female
54	welcome to <i>community-name</i> . Best, Matt Welcome to	<i>community-name-</i>	so glad you are here! Best, Matt
55	are here! Best, Matt Welcome to the	<i>community-name</i>	family! Well to <i>community-name</i> ! I look forward
3	many levels. "To me ""pedo"" here at	<i>community-name</i>	is just a word to incorporate the
36	sessionid in the URL, or a cookie.	<i>community-name</i>	clearly does not use the former method,
20	of a above average caliber here at	<i>community-name</i>	that this wouldnt be an issue. Lo
80	nickname they use to log into the	<i>community-name.</i>	So I decided to open a thread
63	think about. Thanks for sharing! Welcome to	<i>community-name!!</i>	I have never tried the Sims games-
12	I got lucky last year and found	<i>community-name.</i>	This place has kept my head even
40	lot of the threads and responses on	<i>community-name</i>	and have learned a lot about how
51	about it. I want to join the	<i>community-name</i>	Community as I haven't had the opportunity
66	at you in big chunks! Take care;"	<i>"community-name</i>	is listed on the harry71 onion spider
17	Im definitely apathetic in my agnosticism. OG	<i>community-name</i>	fo lyfe mah nukka. Even if I
52	biggest contribution I can make to the	<i>community-name</i>	Community is to offer support, kindness, and
26	those who are interested. Most members of	<i>community-name</i>	who watch CP will watch this sort
81	completely refrain from it. Since I discovered	<i>community-name</i>	the frequency of jokes when I'm asked
58	A welcome addition indeed! Matt Welcome to	<i>community-name-</i>	glad you found the community! Matt Welcome
71	highly respected member who's been around since	<i>community-name</i>	1.0, it could be me or even
4	TF more than almost anyone here at	<i>community-name</i>	for a time but he has proven
41	sitting in front of your computer on	<i>community-name,</i>	holding up your phone to text someone
411	other guys, and lasses, occasionally. I found	<i>community-name</i>	originally as a non-member, feeling it wasn't
290	myself stuck with virtual friends here at	<i>community-name,</i>	for which I am very much grateful."
11	pretty tight knit group. Was just perusing	<i>community-name</i>	with the website in the background when
39	here and love how well put together	<i>community-name</i>	is. Thank you Staff Team for all
34	develop in a healthy direction- sites like	<i>community-name</i>	are desperately needed. I don't have much
263	to construct a demographic of the typical	<i>community-name</i>	member. I am guessing that exclusive pedos
78	we're lucky and there's someone around the	<i>community-name</i>	who knows a little more about what
311	a part of a great place like	<i>community-name.</i>	Thanks mystique! I'll do my best to

TABLE B.13: Random concordance lines for keyword referring to specific community name in Names subcorpus. The community name has been replaced with a placeholder here for security reasons.

Index	Left context	Keyword	Right context
35	and known something was going on. My	daughter	is also attracted to young children. I
453	daughter. Donna (thats what I call my	daughter	here) did not seem very surprised so
85	breast, body and pussy pics to my	daughter	and I never told a soul. It
309	of it than me at least, my	daughter	still makes me hard almost every time
67	further once and threw out that our	daughter	should be naked too and we should
62	tits. Love it! I don't have a	daughter	yet, but I've wondered if when I
279	coast is clear so to speak my	daughter	loves jumping in bed with me. I
45	nuclear family of Mom, Dad, Son, and	Daughter	who have a healthy sexual family relationship
68	aroused so why would be blame our	daughter	for being aroused too and acting on
63	would we use books to teach our	daughter	sexual anatomy when I have a penis
132	you have some guy describing fucking his	daughter,	or molesting his niece, making child porn
210	would love seeing someone playing with my	daughter,	even sexually. This is one of my
170	couldn't live without, being dancing and my	daughter	the other two. I love all kind
195	would handle myself if we got a	daughter.	And of course we did. The first
206	cant cross. I'm still closer to my	daughter	then many might think is ""normal"", but
175	was." "My wife was playing with our	daughter	long before me but nothing really heavy
280	experience. I was at dinner with my	daughter	and we were sitting outdoors at a
5	now. Granted, she was older than your	daughter	is, and under different circumstances, but it
269	phase when I would watch my sister's	daughter.	It just seemed to easy to take
149	be sure if they are admiring my	daughter	or if its just a typical viewing
281	going on her mind. "Well having a	daughter	roughly that age tells me there is
184	on TV have sex in them. My	daughter	is a big fan of sitcoms like
185	"Nowadays my biggest turn-on is that my	daughter	initiates our ""sessions"" as many times as
92	I was involved with my best friends	daughter	and went a bit further than you
225	some of their panties. She had a	daughter	that was a preteen and found panties
46	same as what pastpedodad did here. His	daughter	grew up in a home where sexual
189	urges until I became aroused by my	daughter	and involved in parent-child incest. But now
197	me and the way to protect my	daughter	is for me dying. Thats sad, but
159	like that, so she assumed that our	daughter	was masturbating in some way but hadn't
246	wouldn't want to be tempted with a	daughter	if you have one but you know
212	daughter. Or maybe a friend for my	daughter	where we could all enjoy each other.
430	was not against doing something with our	daughter,	and besides she had her own pedo
419	tell you one of the ways my	daughter	has changed. Yesterday we had our first

TABLE B.14: Random concordance lines for keyword *daughter* in Family Relations subcorpus.

Index	Left context	Keyword	Right context
16	with a child. My relationship with my	wife	is very egalitarian emotionally and we do
280	to be true, even though my own	wife	was part of a self contained community
441	close, intimate bonding (making LOVE to my	wife	or daughter) or recreational ""Hell, it feels
291	bad to even look at? We both,	wife	and I, have guilt at times. I
226	a little guilty about my feelings, my	wife	was like, ""Well, let's see if she
246	the single most important aspect of my	wife	and I deciding to seek out other
270	your willingness to stand by them. My	wife	grew up in a commune type environment
255	I bet she feels the way my	wife	and daughter do, which is that being
232	time to time. The main thing my	wife	and I agreed on after that was
281	the farm I used to take my	wife	and daughter to. The owners would never
275	the opportunity. I don't regret it, my	wife	doesn't regret it and most of all
273	to where she was constantly asking my	wife	about it. It was my wife who
406	person who guides her through confusion. My	wife	and I are both very sex positive
279	there were times either me or my	wife	saw something which might have been a
160	deal with first. I can't imagine the	wife	is going to have a great time
301	life I've had one secret that my	wife	doesn't know. Actually no one in the
124	what I'm planning on doing to my	wife	than I do about anything I ever
39	taking care of my nephew. My brother's	wife	was a lazy piece of shit. I
151	all over me just before bed. My	wife	put our son down while I took
282	between pedos and beastie folks. Both my	wife	and daughter have had k9 experiences and
245	child. Counting my own family, since my	wife	grew up in that situation. I think
225	having sex with kids." "Very true. My	wife	isn't into scat (altho she has tried
238	we met at the nude beach. My	wife	occasionally talks to a few other people
398	There was nothing impulsive about it. My	wife	and I had countless conversations, trying to
144	And though I'm quite certain that my	wife	is strongly on the anti-side she has
290	don't mean talking them into it. My	wife	can really get in to BDSM at
23	bikini. I gestured my thanks to my	wife	who gestured back that it was cool
88	me." "This is what happened to my	wife,	although she would not agree that she
33	there's no way I can get my	wife	to agree to having an incestuous relationship
119	to never give up. I met my	wife	through random chance, not going out looking
394	them don't know much about it. My	wife	says when she was Kylie's age very
309	that she would normally ask me. My	wife	has always been very good with the
251	dad starts fucking his daughter because his	wife	won't give him enough sex, is a

TABLE B.15: Random concordance lines for keyword *wife* in Family Relations sub-corpus.

Index	Left context	Keyword	Right context
56	(not fully). Unfortunately, three years ago my	niece	decided to kill our friendship and asked
103	of thing I did with my young	niece	and most was fun play and some
26	in the woods near the house. My	niece	gets scared so I tell her she
35	in my child love support post my	niece	isn't really my niece. To understand how
36	support post my niece isn't really my	niece.	To understand how our relationship started its
42	uncle". The kind of uncle that my	niece	and nephews could look up to. The
47	the watch for anything that causes my	niece	harm. It took her over a year
92	the time. And to know that my	niece	seemed to be enjoying it. Kind of
71	I wanted to meet up with my	niece	and have an honest conversation about what
74	niece and her young daughter. My actual	niece	had been my lover for 3 years
80	I was once very obsessed with my	niece.	It got to the point that at
48	Furthermore she knows that I love my	niece	more than anything and will literally do
51	for advice concerning my relationship with my	niece.	I gave you negative karma because your
76	niece and her daughter. I thought my	niece	was going to rip me a new
93	have had the opportunity to babysit my	niece	a few times. We would get into
44	course I love my job and my	niece	and nephews so yeah id lose all
37	vague. Now before I entert meet my	niece	or her mom I used to hang
10	Before mom and stepdad left the room	niece	was seated leaning forward but as soon
38	a beautiful little girl. This is my	niece	Claire the girl who stole my heart.
123	place. She told me that Chloe, my	niece	had showed her that it felt good.
117	uncomfortable feeling now that I've seen my	niece	naked. Her budding breasts are so inviting,
39	ask to say good bye to my	niece	and nephews. Liz rolls her eyes and
45	lol blessing because when I'm with my	niece	she knows what her kisses do to
82	first experience I remember was with a	niece.	It started when she was about 4
57	in the past. I think if my	niece	and her family would not have cut
111	my visit was my 3 year old	niece.	Man I love that kid and she
2	It's in the Lit section, titled My	Niece.	I'm the friggen pope compared to you.
67	how I had been intimate with my	niece	and with another girl several years ago.
60	I was actually very sexual with my	niece	from when she was about 1 1/2
55	a lot of guilt and pain. My	niece	told her mom that I had molested
65	then the one I received from my	niece	a few years ago. She was about
96	girl. OMFG. PapaBear" "I have seen my	niece	get very aroused during horseplay with me
116	one else, I had never seen my	niece	completely naked since she was like 4

TABLE B.16: Random concordance lines for keyword *niece* in Family Relations sub-corpus.

Index	Left context	Keyword	Right context
69	her child be naked like this. Her	<i>mom</i>	had actually been ""molested"" (her word) by
1	to fuck them." Love to married a	<i>mom</i>	like this "Sorry if posted, I didnt
16	the corner of the couch (so thoughtful	<i>mom</i>	makes up for stepdad). For the most
29	punishment coming my way. Littlepolly I know	<i>mom</i>	has had ""the talk"" with her before
6	husband's notion of you being a great	<i>mom</i>	who would never, ever hurt her child
30	fuck it up (pardon my French). Now	<i>mom</i>	has already flung my handling of my
18	and that's when things got interesting. Before	<i>mom</i>	and stepdad left the room niece was
206	instigated by her desires, not ours. Her	<i>mom</i>	and I are well aware how things
146	a strain on their relationship. The way	<i>mom</i>	explained it was they hadn't gotten to
57	if the mom suspects me. Does their	<i>mom</i>	radar go off around pedo 's? I
45	it's another thing she likes when her	<i>mom</i>	tells her to get her pumpkin bucket.
136	for a long time before that that	<i>mom</i>	and dad had liked to have ""adult
151	a shitty deal for everyone. PapaBear" "Helicopter	<i>Mom</i>	will try to kill me, I'm fairly
117	You need to be careful of the	<i>mom</i>	and dad though. Make sure you think
149	so that is my shield. PapaBear" "Her	<i>mom</i>	has mentioned in passing that she gets
14	he'd be as accepting as mom). Now	<i>mom</i>	helped me pick out the perfect bday
36	she wouldn't catch. In all this time	<i>mom</i>	has never brought any of it up
34	Because of this I'm way closer to	<i>mom</i>	than anyone else in that family (not
24	niece myself but so far it seems	<i>mom</i>	has touched on all the important parts
153	turned on I was to see her	<i>mom</i>	with other men and she liked being
118	a park or something like that. With	<i>mom</i>	of course but that is okay. Also,
135	went to Dee and asked if she	(<i>mom</i>)	thought dad would be okay with it.
38	room for her so I agreed with	<i>mom</i>	and dad's permission of course and the
167	for a pizza with Amanda and my	<i>mom</i>	so they can chat, I take a
161	dad was her biological father, because her	<i>mom</i>	was having sex with so many men
148	After a particular incident at school my	<i>mom</i>	started asking me where I got my
119	she be jealous of me and her	<i>mom</i>	together? It was that last question popping
44	I entert meet my niece or her	<i>mom</i>	I used to hang out with my
42	so with help from crutches and get	<i>mom</i>	she was able to rest outside. I
15	so I sit next to her and	<i>mom</i>	has her sit in my lap so
172	years of difference, so I asked amanda	<i>mom</i>	when did she and ricky met, and
164	tale ending, having her reunited with her	<i>mom</i>	to supposedly begin a new life and
131	curious and asking questions (mostly to her	<i>mom</i>)	last fall when she started fifth grade.

TABLE B.17: Random concordance lines for keyword *mom* in Family Relations sub-corpus.

Index	Left context	Keyword	Right context
240	go into the music world and what	kids	see in music videos. I have to
182	you were as just a couple of	kids	trying to figure it out. Combine that
67	fun so I invited her and her	kids	to join my brother his new girlfriend
378	3 or 4 days. Nice that the	kids	can't read yet, I can browse text-only
638	time after that. Now that I have	kids	I know that yea I am real
531	about all that happened other than the	kids	involved getting suspended for a short time.
83	it's time for me to leave. The	kids	come flying out to say goodbye and
558	it could n't. Playing with adults and	kids	when I was young helped shaped who
236	get to join. He said if the	kids	are old enough they will like the
35	happen naturally. It's been my experience that	kids	are naturally curious about pretty much everything.
357	say nice things about her to the	kids	either, but I do believe in the
239	it are really smarmy. But it makes	kids	want to know more about sex because
256	that see that little bit extra in	kids	believe it is somewhat natural. But there
367	and why they work, and seeing the	kids	accomplish things on their own for the
5	the older generation were sexually active as	kids	through playing "doctor" and other such games
73	get enough candy to make sure the	kids	got enough. Once we get back to
546	he was talking about sex with little	kids.	The same left that had no issue
616	was in that movie about the young	kids	making home movies. I wonder if I
543	However, that doesnt make we dont worry.	kids	even if they try and bad at
257	the best approach. I would rather my	kids	ask me about anything than someone at
373	later. It wasn't much longer before the	kids	did wake up and come to our
532	favor adults having sex with really young	kids,	but I believe those of us who
135	knit community of pedo's who share their	kids.	I know there are communities like this,
371	both out in no time. With the	kids	sound asleep I managed to convince Meg
372	tired and concerned that one of the	kids	might wake up and walk in on
545	times and theres no getting off. All	kids	are different, so i cant say how
66	Naturally I couldn't let Liz keep her	kids	from the Halloween fun so I invited
45	together was for a family barbecue. The	kids	had their swimsuits for the pool and
221	where to start! Cumming in a little	kids	wet diaper is the best if you
664	bookstore. shared it with the entire neighborhood	(kids)	Wow! 1977- bought my third copy of
360	I'm also a parent, and making my	kids	do things they don't want to do
289	of that desire. I just love beautiful	kids	and I think that is a common
319	guy that always seems to be around	kids.	I am not sure of any individual's

TABLE B.18: Random concordance lines for keyword *kids* in Family Relations sub-corpus.

Index	Left context	Keyword	Right context
165	can discuss our play whenever. Girl talk.	Dad	doesn't relate the same way. Membership approved."
118	the young lover grows up and busts	dad	or uncle or grandpa or trusted friend?
123	our part of the country when the	dad	was transferred by his employer." "For what
129	all someone went to work with Tara's	dad,	someone knew vicky as the girl next
119	She seemed to enjoy activities with her	dad	and she was certainly old enough to
100	aunt didnt talk to her or her	dad	for the rest of the year, they
124	about having the same interests. The other	dad	and I definitely enjoyed the eye candy
6	her bring them home to share with	dad.)	A son that will want to take
101	the bottom line is, even though her	dad	was pretty careful most of the time
45	tv? i could be wrong again lol	dad	film little naked daughter video i swear
40	react to that. It's true that my	dad	and I have been sexual with his
110	had a long sexual relationship with her	dad	that ended badly. But not because she
35	jumped and fell out of the bed.Luckily,my	dad	caught me on time before I hit
4	me. She actually suggested it and said	"dad	all the cool guys do it." I
113	with another families and swapping kids. The	dad	says it is a huge turn on
19	we've chosen the path of just friends	(dad	has a gun and has made it
3	was "You two be careful. If your	dad	sees you he's gonna kill you both."
104	But after Dee had her abortion her	dad	started seeing another woman who he married
32	and ran into the kitchen where my	dad	was cooking lunch. "Mavie?What's wrong?Why are you
2	her they'd have sex with their own	dad	if they weren't so scared of trying
10	up knowing I no longer had my	dad	really was the day the final remains
7	what happen to this boy? maybe his	dad	was caught or he killed the boy
144	she was born I was the proudest	dad.	She was so cute in my eyes.
27	to me because I am a ""chill""	dad.	She has no idea how attracted I
122	this thread, since becoming an active pedo	dad	I have been bugged by the assumption
146	me. She has to talk to her	dad	about it and make sure it's ok
9	chuck it in her general direction? [My	dad	seriously did that when I was 19
38	but me that have something for my	dad.	Since you guys are just outsider I
15	Granted dad is my niece's second step	dad	it doesn't make it any easier. How
5	up and have some fun. As a	dad	that's in a long term relationship with
127	never been 100 percent sure that her	dad	was her biological father, because her mom
133	that like her... sometimes i think her	dad	look at her funny also i find
149	Chloe let out that she teased her	dad	when she was about Ashley's age. She

TABLE B.19: Random concordance lines for keyword *kids* in Family Relations sub-corpus.

Index	Left context	Keyword	Right context
38	nudity. At least I am not. My	daughters	have seen me many times nude. My
1	for me Friday evening. I have two	daughters	10 and 13, my youngest was playing
76	I would love to be my future	daughters	boyfriend, I would never deprive her of
51	Queen. As a single dad raising 2	daughters	I find it rather challenging, and do
46	on is to keep you and your	daughters	from getting hurt." "I don't understand your
54	you like to raise your sons and	daughters	if you just started a family. Forget
19	blond hair and are really cute. My	daughters	have baby sat both of their kids.
53	he wants to do more with my	daughters."	"In an ideal world (in your world
21	through something like this. I have two	daughters	and almost without exception, all their girl
11	On the "other man" front... Both my	daughters	started being sexually active when they were
8	lives with his wife Lucy and their	daughters	Sierra (12 when I met her) and
15	had some connection to me (nieces, friends	daughters,).etc. Two of the girls told their
59	just like homosexuality, or autosexuality. LovingDad" "My	daughters	first lingerie-ish piece of clothing was XXS,
55	What would you teach you sons, your	daughters	and at what age. What do you
10	it, I could see all of my	daughters	messages. Her slut GF was sending her
47	she cried. I do not need my	daughters	consent to make decisions about what is
6	going to notice us chatting about my	daughters.	We didnt do anything together generally because
58	of mine (we met occasionally and ""shared	daughters""")	had the incredible bad luck of the
57	kept it secret from my wife (and	daughters	mum) for years, until, a week after
7	turn to producing instead by offering my	daughters	a job first!! My youngest has already
52	should not be separated. I love my	daughters	and sex should be natural. I don't
22	more than once to sneak into my	daughters	room during a sleepover, but did not
72	and things went well. The regular daddy	daughters	day occurred, with one minor exception. Ashley
49	walking with her 14 to 16 yo	daughters	(tourist) ""hey! Can I barrow your daughter
84	sports. Things other dads do with their	daughters	that is completely non-sexual. Again, with sex
50	the Paper backs were of daddy's and	daughters."	Did PSC exist on clear net circa
64	whether dads would want to share their	daughters	with other pedos. Writing that made me
45	the one taking the lead, not your	daughters.	Their actions may appear as though they
28	I look at the girls in my	daughters	grade 7 class compared to the girls
2	ok with an older man meeting my	daughters	to see if they would ever form
60	the decisions. LovingDad" "I only share my	daughters	with my wife and each other.[],I'm just
67	hard was seeing one of the host's	daughters	on her back in bed (actually it
70	it clear that they are both my	daughters	and that I love them the same.

TABLE B.20: Random concordance lines for keyword *daughters* in Family Relations subcorpus.

Index	Left context	Keyword	Right context
5	her or any of my nephews. My	uncle	title comes from the fact that I'm
61	and turned red but Chloe said ""hey	uncle	Jim, wanna join us, it's fun"". I
39	I was 10, I stayed with my	uncle	after school twice a week. My Uncle
42	Always been a bi pedo as my	uncle	got me started with all this, but
30	have become the mirror image of my	uncle	in a different age. To this day
7	summer break. She's helping Phil and her	uncle	plant a backyard garden. I heard over
8	garden. I heard over to meet the	uncle	and catch up with my neighbors in
6	with these kids parents. I became her	uncle	last Halloween (she looked so amazing in
15	porn in the 7 age range. My	uncle	was happy to supply it. I'm not
10	"Growing up I had an abusive alcoholic	uncle.	What I learned from him is exactly
11	be. As soon as I became an	uncle	I voted to be the ""cool uncle"".
2	got here. I was molested by my	uncle	from a very young age, probably from
17	I had a wonderful time with my	uncle	for several years as well as some
14	to fail in that goal.Â I'm an	uncle	to 4 wonderful kids (3 boys and
48	if it was my dad not my	uncle...	But anyway no judgement...on others as I
54	to seeing your posts ND4BH {NorthernDad}for{[BabyHeart]} "Hi	Uncle,	nice to be around people with the
26	where to put my little cock... My	uncle	simply watched the whole thing and stroked
29	it, but now I boast about my	uncle	and my early sexual exploits in online
9	first comment about me being a cool	uncle	to her kids. We hang out for
22	knows that I was sexual with my	uncle	and about my involvement with my niece
16	16 we had a visit from my	uncle	and his daughter who was probably around
18	rooted in my own involvement with my	uncle	and incest with my siblings growing up.
1	because they feel as if admitting an	uncle	or parent loving them in this unique
32	16 years of having sex with my	uncle,	who was also a voyeur and loved
4	seedy side that used to excite my	uncle	most which was mens toilets and forests
3	as soon as he had gone my	uncle	told me to ""suck his dick before
28	exhibitionism. He was involved with a different	uncle	for many years. What I have never
12	uncle I voted to be the ""cool	uncle"".	The kind of uncle that my niece
13	be the ""cool uncle"". The kind of	uncle	that my niece and nephews could look
19	of my long term relationship with my	uncle	and view him as a monster. I
33	I had a sexual relationship with my	uncle	from toddler to teen and actually loved
27	to several cousins in the family. My	uncle	just happened to be the most creative.
25	a bathrobe. We went inside and my	uncle	pointed at her and said, ""Do you

TABLE B.21: Random concordance lines for keyword *uncle* in Family Relations sub-corpus.

Index	Left context	Keyword	Right context
48	maybe I can better relate to my	husband.	I am not antagonistic about it like
16	gets down. I then apologize to Liz's	husband	for how she was wrapped around me
17	get my swimsuit on. Liz and her	husband	opted not to bring their swimsuits so
13	to stop fighting. She explains that her	husband	is special. He's female to male trans
8	I'd ever be interested in having a	husband	rather than a wife short of having
19	Behind me is Lucas followed by Liz's	husband	then Michael and Liz herself. This is
61	is older she is certainly allowed boyfriend,	husband	and family with me no longer her
29	reaches the living room first because her	husband	stopped at the bathroom. She gets an
34	a happy father, and a hard working	husband.	But on the inside, im sad to
56	any other pedo men, but this pedo	husband	certainly doesn't take his pedo wife for
2	about. Period. Granted, the fact that your	husband	knows of your sexual orientation is an
58	my company. She will NEVER tell her	husband	of these secret feelings! Lol. I feel
23	up to the birthday Liz and her	husband	started fighting. Liz fell out of love
25	and I'm told that Liz and her	husband	are fighting in their room and Claire
32	appears to be on hold as her	husband	has finally started listening and talking through
12	hang out with Liz she and get	husband	had been fighting so we hang out
11	we split the ways between Liz her	husband	and myself). With the candy now deemed
51	place interests me. From the way my	husband	talks about his interest in kids, it's
15	like that to be greeted by Liz's	husband.	I tell Claire to go get ready
38	would not bring that up to my	husband	even though I am quite sure he
28	at this point that Liz and her	husband	finish in their room and start heading
55	stepdad from a young age. Unlike your	husband	though, I did not have an inclination
36	she would want to find a pedo	husband	at this age. I hope she understands
26	Once I've done that Liz and her	husband	come out of their room and head
20	going to fall. As Liz and her	husband	help me try and calm her I
22	lips before heading to her seat. Liz's	husband	is the first to notice the snail
14	on the kitchen and living room. Her	husband	gets the boys' room set up and
54	this lifestule with her but i allow	husband	to do it, so i cant talk.
6	orders of magnitude easier to sell your	husband	on the concept of gradually sexualizing your
31	then assures me that she's keeping her	husband	distracted from noticing what's going on. It
21	have to worry about Liz or her	husband	looking back and seeing plus due to
49	been if I didn't know how my	husband	is. But I'm still very unsure of
1	she makes some pedo a very happy	husband	if she likes guys at all. No

TABLE B.22: Random concordance lines for keyword *husband* in Family Relations subcorpus.

Index	Left context	Keyword	Right context
69	way to start a conversation with a	child.	It's important to keep it innocent if
273	of your capabilities of providing for a	child.	And why would your child ever know
382	bibical Alpha, the man. I believe the	child	is healthier with consistant parents or family.
482	I would fuck me. Thai—— "The Bangkhud	child	is exposed to sex at a very
307	I have gotten another phone call from	child	services since my last update confirming that
326	I love the genuine and sincerity a	child	brings in any relationships. They are beautiful
281	do things together. I would treat the	child	attraction as secondary and get yourself out
395	of growing up in a cult where	child	sex was a regular part of the
360	on adult, child on child, adult on	child."	Do you think it could be the
293	people that are unfortunately caught with a	child.	They never tell the back story of
380	the relationship is long over and that	child	becomes a religious zealot or reads a
388	parent had been played with as a	child.	Counting my own family, since my wife
389	typical." "Since my daughter was the first	child	I had sex with, I tend to
333	own oppinion from rape and loving a	child...	two very different things.... I believe that
280	admitted to participating in any type of	child	abuse, then what you told her may
291	Home schooling to me means that the	child	is restricted from experiencing both the bad
309	to deal with messy diapers. Assuming your	child	is laying on his/her back with you
341	your best friend."" He means if the	child	is enjoying sex and does not feel
403	and would never try to approach a	child	outside of a consenting situation in which
300	decisions but it is different in every	child	and every situation. In the scenario as
398	I think gets easier the older the	child	is. Even though my wife actually grew
306	advice here on how to force a	child	to do something she doesn't want to.)
354	(No Nick Attached)- 18:2252A(g): ENGAGING IN A	CHILD	EXPLOITATION ENTERPRISE 18:1591(a), (b)(1): SEX TRAFFICKING OF
278	think I would, if possible, buy a	child	sex doll. I have no idea if
409	Has anyone had a problem wit a	child	growing up and then hating you for
319	want to. I wouldn't say ""if the	child	is willing"", but ""if the child asks
316	have to do is give your own	child	the camera and tell them to go
285	it as being hypocritical to be a	child	lover and hate when you hear of
415	at it is that it should be	CHILD	centered and not just for the the
428	being pretty obvious about it." "Actually,because of	child	porn I found this site. I used
425	showed who i was, not in the	child	lover way, but in the im no
289	me is a poison and putting a	child	in the hands of these so called
377	the odds are on our side. Is	child	sexual abuse a ""grossly underreported crime"", according

TABLE B.23: Random concordance lines for keyword *child* in Family Relations sub-corpus.

Index	Left context	Keyword	Right context
211	consider the sex I had with my	parents,	and others, as abuse." hey there. I
77	a wonderful thing for us adults and	parents.	This is a great community but I
74	some magazines that she found in her	parents	closet. Mostly straight sex with some lesbian
83	The kids were just weird. And the	parents	were maybe no better. Almost a bit
36	last drop, from what must be, their	parents	or someone who loves them. The gentleness
56	work with love me, and even the	parents	love it when I am with their
25	obviously want nothing to do with their	parents	and get their self esteem via their
71	a bit, some in front of their	parents	too! I've never acted on it for
80	that to be very dangerous. Would the	parents	expect you to contact the kids during
70	in a good place, being that the	parents	are ok with everything. Just take it
104	A good kid should do what his/her	parents	tell them but at the same time
10	the conversation away from that because her	parents	are literally in the next room plus
134	first it was pretty shocking to see	parents	sharing their daughter with other swingers, which
43	to truly dictate what could work for	parents."	"I'm going to try to remain as
124	and their parents, or at least good	parents.	I almost always know what my son
27	whole time and it was terrific. Her	parents	were fine with me even though I
105	couple staying together for their kids- unhappy	parents	lead to unhappy children. But I am
72	Even if THEY approach me? Honestly, now-a-days	parents	are on such a look out for
28).etc. Two of the girls told their	parents-	one soon after our experiences and one
35	LEA. However, if both girls or their	parents	somehow connected now and shared stories, I
97	and how good that feels. But our	parents	didn't know any of this. At least
178	will follow her (and mostly with her	parents)	around the mall. We talk about when
185	a kid who was playing with my	parents.	When we would meet up with the
19	anyway so you tend to meet many	parents	that continue to come to you for
85	it would be important to tell your	parents	about your desires. And you mention that
150	to be alone and their girl friends	parents	also like the thought of me being
34	in this group who did tell her	parents	about it when she was 16 or
184	certain hours a day. What about the	parents	that work off hour days?" I've had
62	of me as toddler made by my	parents.	It's better than nothing. Thanks for your
181	would probably travel around the states teaching	parents	about their children. Become a child-whisperer! I
45	have been as if there was their	parents	there. If the conversation were to go
136	to have a loving relationship with her	parents	but obviously they are not as active
173	or 13 year old kids here. The	parents	went apeshit and the school made a

TABLE B.24: Random concordance lines for keyword *parents* in Family Relations subcorpus.

Index	Left context	Keyword	Right context
320	guest house and I would pay your	family	a fair wage. Brian's eyes lit up.
52	mentioned this in other posts. My brother's	family	knows that I am a pedo and
16	generally favor a backdoor approach where your	family	just ""ends up"" being incestuous, with no
33	I ignored everything I learned from my	family.	I'm not looking for a quick squeeze
4	change for me. I can't see my	family	accepting the relationship Kelly and I have.
63	known ""molesters"" to several cousins in the	family.	My uncle just happened to be the
57	I think if my niece and her	family	would not have cut me off, I
64	am considering cutting off relationships with my	family	so I can be isolated and pursue
164	future and this time she wants the	family	to go with her. I would rather
178	more too. In fact we often have	family	showers, it is rare that the kids
55	incest had been going on in my	family	for generations. However, it was all covered
137	in our bed and just be a	family.	Or he may do oral on me
169	to speak to most of her own	family	(parents, siblings, grandparents are all out- just
172	because she will no longer use my	family	for babysitting and she knows I have
93	sex part is a secret, our whole	family	is like a secret society. She is
3	up the father/daughter appearance in front of	family	and friends though. You say your daughter
115	not really active with anyone outside my	family	there is no suspicions. The times that
50	It would have been bad for my	family	if I had continued. Having said this,
59	This news has not made it my	family,	friends and work, but I think someone
183	Marques comes from a respected middle class	family	and his father, a Brazilian, is an
51	a pedo-mom and had a pedo friendly	family.	Also, if I had been a single
6	than the janitor that provides for his	family.	Success isn't what you have, it's how
22	to get to the destination for the	family	vacation? Let's pull over at this gas
17	simply think that incest is a beautiful	family	paradigm to raise children under and I
68	I kept getting different stares from extended	family	members. I had no idea who knew
7	trying to do the same to young	family	members or my mum's friends kids trying
67	except in the case of a willing	family	or single mother or father who wants
53	At some point, I think my own	family	and others will find out about my
167	tell her before that shutting out her	family	was the right thing to do, but
13	likely to succeed method of starting a	family	tradition of incest. For me, it's not
49	is if it's part of the overall	family	structure. The mom is into it, the
100	me see a deficit in the old	family	love and acceptance category. Interesting how that
225	of drifted away from our circle of	""family""	friends a few years ago although we

TABLE B.25: Random concordance lines for keyword *family* in Family Relations sub-corpus.

Index	Left context	Keyword	Right context
59	or so I see nothing wrong with	incest	at all. Now with me being close
67	lesbian porn more than straight porn. And	incest	can be exciting to see or read
76	incest. I know all the stuff about	incest	being in the Bible, blah blah blah.
54	the Taboo series. That was the first	incest	porn I found and watched them all.
44	mine was only undercover and not discoverable.	Incest	is any sexual behavior between relatives at
27	really was dumbfounded. With my history of	incest	and sex with adults as a child
46	I did. 8. Do you agree with	incest?	Generally yes 8.1. Would you have sex
57	many. Within the bible are stories of	incest,	slavery, murder and every other type of
31	Say for the sake of argument, that	incest	is legal for the purpose of ""education"".
41	point it's become something of a meme	(Incest	is wincest), people also openly joke about
64	It is a book of murder, rape,	incest	and some of the most sordid events.
75	has never told the stepmom about his	incest	relationship with Dee. There is a lot
74	waters where there were no laws about	incest,	pedophilia, etc. Families would go there every
56	I know there are others. As an	incest	lover, I think its great. If she
66	or wrong. So to me it isn't	incest	or taboo." Congrats Cacman. Good luck on
15	not to trigger the negative taboo of	incest	which is so deeply ingrained. So i
2	used for the sexual gratification of others.	Incest	puts a double whammy on things too.
23	up, there was a tremendous amount of	incest	going on in my family that included
65	never thought of it as taboo or	incest	or anything else. It was just fun
1	is uncalled for. @HS I'm sorry your	incest	experiences were unpleasant for you. I fully
4	in life, which includes continuing our relationship.	Incest	between a father and daughter can work.
78	happens is, the families engaging in pedophila	incest	have to stay quiet and remain underground
9	brother, daughter, son). There are people into	incest	who are against incest with underage participants.
73	some freenet sex site. It was called	""INCEST	ISLAND"" or something like that. It was
85	for non-interventional laws like the ones covering	incest	in New Jersey and another state I
28	considered though. What if instead of emphasizing	incest	between parents and children, it just became
22	this because I have a lot of	incest	in my past and I have had
82	touching a kid is 'wrong' and how	'incest'	is wrong, and all that. Has anyone
62	the families. I just feel that family	incest	is a cool thing and it would
77	the founder grafting his own views about	incest	and pedophilia onto the basic Christian type
14	uncomfortable. Finding a wife who's open to	incest	is 99.99th percentile. My wife is 99th
43	it as I can find. I think	incest	between consenting adults will eventually be decriminalized.
49	my daughter I joined a bunch of	incest	themed groups onclearnet, mostly as a lurker.

TABLE B.26: Random concordance lines for keyword *incest* in Family Relations sub-corpus.

Index	Left context	Keyword	Right context
261	after the next thing, calling him a	pedo.	As many times he said 15. Also
98	group of people that I was a	pedo	and attracted to children. It freed me
126	she was also at the very least	Pedo	tolerant. I am not active outside of
172	don't think someone needs to be exclusively	pedo.	I am a bisexual adult that enjoys
37	that there may be a softening toward	pedo	behavior and sexual contact with younger people.
163	to know if your mother is a	pedo.	"I am reasonably certain I do not
21	harming anyone what's wrong with being a	pedo?	I like young girls but have I
36	arose out of my suppression of my	pedo	desires)." "I recently performed an inventory of
183	I've been trying not to laugh at	pedo	humor in the media anymore, or at
28	my spirit animal). But my first true	pedo	crush would have to be my neighbor.
179	never be a famous advocate for the	pedo	lifestyle. With doctors like this one and
162	that you may be gay or a	pedo	or whatever. But that does n't mean
43	their daddy... I denied I was a	pedo	and that my past activities were just
93	not mention that I am a ""proud	pedo""	except to my brothers and sisters on
170	honest, I don't really think much about	pedo	stuff unless I get in a situation
280	it hit me, maybe he is a	pedo	too, and that is why he loves
32	grips and accept that I was a	pedo	much earlier in my life and then
18	pedosexuality. This naturally went against the anti	pedo	upbringing leading to many depressed nights. I
34	that I discovered that I was a	pedo.	I could not figure out why I
73	desire to make a connection with another	Pedo	for some time. By far it's better
151	little someone in my minds eye, my	pedo	penis becomes quite sensitive. I love toddlercon
192	think she still suspects that I'm a	pedo	too though she didn't bring that up.
195	friend who I suspect may be a	pedo	as well. Many years ago while helping
109	mid teens I realized I was a	pedo	and knew it was something to be
187	I did learn that someone was a	pedo	and that they were the kind of
52	acted as a major catalyst for my	pedo	desires. I now think about girls 24X7.
23	done that. Just because you are a	pedo	doesn't mean you are going to hurt
199	out many times if I am a	pedo,	and I have always denied it. At
203	boys too.. What do everyone think? welcome	pedo-	what a great name. I can imagine
114	kid. A knock at the door- another	pedo	had come over to see me and
100	feel when finally I can post on	Pedo	Literature .This is my first story,"" Loverboy", a
200	if she ever found out about my	pedo	side. We quickly moved in together, and
50	that I am not guilty of my	pedo	actions, but simply feel bad about getting

TABLE B.27: Random concordance lines for keyword *pedo* in Family Relations sub-corpus.

Index	Left context	Keyword	Right context
18	whether autopedophilia is normal behavior among other	pedos,	I've known quite a few in my
48	I'm not surprised there's some cross-over between	pedos	and beastie folks. Both my wife and
24	some degree at times. Who knows, maybe	pedos	release some kind of different pheromone? Dogs
43	a pedo, why are not millions more	pedos?"	This was propaganda to ensure military for
44	theirs. I see we or some, us	pedos,	are no better conversing with open minds
17	with an overly positive view of what	pedos	are really like as a result. Or
15	of childhood sexual experiences some politically active	pedos	have put forth proposals for making child-child
28	nothing to do with sexuality. I think	pedos	are in every walk of life. You
61	and halter tops. I can't imagine how	pedos	on nude beaches or in nudist/naturist camps
66	have serious psychiatric problems). Where I live,	pedos	are worse than garbage. There is not
10	time when I met him I thought	pedos	were rare and hard to find now
31	to do with an attraction to kids.	Pedos	exist in all walks of life so
42	Children are very controlling and jealous. Do	pedos	have this family? A family of Father
64	same time, the fact that we are	pedos	has morphed us to who we are
1	here: Just like normies the are good	pedos	and bad pedos. Good pedos seek out
22	people that I would swear they were	pedos.	So yes, I suspected many guys. None
68	not have sexual encounters with kids. All	pedos	are the same: Monsters who rape children.
14	happening. Given the favorable social circumstances many	pedos	here atleast the ones who're incest-inclined would
70	demanding, loud attitude like I read in	"pedos	in prison" section. It will ring. And
60	have NOW. "I don't object to other	pedos	being interested in my daughter. That doesn't
69	not sure why I tend to like	pedos.	My ultimate fantasy is to have a
53	them. I am looking for fellow female	pedos	to talk with and share advice. would
76	would be hard to safely meet other	pedos	in real life. I guess that's why
79	sick people out there. Obviously not all	pedos	are as loving as most of us
4	and would never harm a child. Bad	pedos	seek self gratification at the expense of
20	at was full of sex offenders and	pedos.	It was small, less than 500 inmates
77	into sex take on a group of	pedos.	Most of whom, if new like me,
65	worse than being a murderer. Some societies	pedos	are treated with a little bit of
8	and from that point on saw all	pedos	as that man. Rapists wanting to hurt
25	a lot of bad experiences with other	pedos	morals and how most see girls as
67	even the public knowledge that there are	pedos	who do not have sexual encounters with
33	I think that is the majority of	pedos.	I may be wrong but most just
29	be some government action going on with	pedos	online. I just think it is way

TABLE B.28: Random concordance lines for keyword *pedos* in Family Relations sub-corpus.

Index	Left context	Keyword	Right context
22	life now and I am a proud	pedophile)." "	"I started sexually servicing my uncle when
56	made it to middle age as a	pedophile	and never offended. I've had opportunities I
54	"I sometimes am confused about the label	pedophile.	When I hear that word I think
49	between 16 and 25 would be a	pedophile.	People thought a pedophile was a fat,
62	very confused. I am married to a	pedophile	man, but I'm not a pedophile myself.
51	can't qualify for disability being a diagnosed	pedophile,	sex-offender, bipolar, arthritic, man over 50. Sucks
59	would say careful in declaring you are	pedophile,	personally I really don't like the word
57	child ever know that you are a	pedophile	unless you are active in that?" I
41	have been pro-life like forever and a	pedophile.	It has always been an internal contradiction
60	non-pedophile myself who is married to a	pedophile.	At first, I was scared of the
9	think there's nothing wrong with being a	pedophile,	but we unfortunately get represented by the
53	and dislikes. I think that when a	pedophile	is caught in the act, maybe using
3	inside I may have always been a	pedophile,	I just didn't realize it. I say
61	non-pedophile. However, I was raised by a	pedophile	and I'm married to another. And while
19	acts as a teen and then adult	pedophile.	The teen memory was actually very fun.
28	than you had ever expected. Having this	pedophile,	handsome and ""normal"" looking by societal standards,
6	watch it. But I'm not exclusively a	pedophile	so some adult porn still turns me
38	Survey For PSC-ers Who Had Experiences With	Pedophile	As Kids And Not Yet Explored Kids""
58	she would accuse me of being a	pedophile-	I was really surprised when she then
39	them and create bad reputation for the	pedophile.	community where besides of them are considerate
23	the fact that I was probably a	pedophile.	I really had not thought about it
43	the same. It's been ages." What's a	pedophile	favorite candy? Lolipop There's one category which
52	you made a great post. Being a	pedophile	doesn't necessarily mean that you will have
26	Title: You've always wanted to marry a	PEDOPHILE	Because that's the vibe I've been getting
10	not know if Michael Jackson was a	pedophile,	but what if We are the world
46	makes you a good Father. As a	pedophile,	there will always be some things you
21	would not divulge that I am a	pedophile."	"It's not that unusual for kids to
44	sexual experiences or abuse to becoming a	pedophile."	"Thanks for the responses.(except the first
47	test and that I was not a	pedophile,	as you could score at under 8,
55	I am quite sure she isn't a	pedophile.	Yes we have rekindled some of those
48	arrested for simply saying they are a	pedophile?	thank you Sierra, was that for me?
33	again is hard to play ""burn the	pedophile""	when thinking about those things. The focus
18	took was one simple truth: being a	pedophile	doesn't make you a monster. It's a

TABLE B.29: Random concordance lines for keyword *pedophile* in Family Relations subcorpus.

Index	Left context	Keyword	Right context
64	I am a fairly new convert to	pedophilia	I can't really say I used to
39	god and the bible into discussion about	pedophilia.	There is nothing more perverse than the
4	don't touch children [URL], Treating Pedophilia [URL],...	pedophilia	[URL], [URL], Pedophilia: Interventions That Work [URL],...
7	about pedophilia: There is no cure [URL],...	pedophilia"	"I don't know wonderboy, maybe more a
20	lot of resistance to the idea of	pedophilia	being considered a valid sexual orientation, but
62	Maybe I have some insight coming to	pedophilia	late, strictly because of an opportunity instead
43	do with anything in life let alone	pedophilia.	Geez not trying to be critical here
54	the envelope"" makes it easier to consider	pedophilia	as another step in someone's sexual evolution
65	LGBTQ group who advocates being inclusive of	pedophilia	is doing themselves a great disservice. We
3	say they don't touch children [URL], Treating	Pedophilia	[URL],... pedophilia [URL], [URL], Pedophilia: Interventions That
58	is that she grew up believing that	pedophilia	was not as big a deal as
15	fought with my upbringing which told me	pedophilia	was wrong and all pedophiles are monsters
13	some friends with shared interests (outside of	pedophilia	unless you know for a fact they
61	to talk about everything that comes with	pedophilia.	[],I was sexually active as a kid
6	It's a Challenge [URL],... lenge.html Pessimism about	pedophilia:	There is no cure [URL],... pedophilia" "I
44	My suggestion is that you keep your	pedophilia	very private. I would love to talk
12	actually helped come to terms with his	pedophilia.	He swore he liked teens and had
14	make my history project the history of	pedophilia.	I used many sources to learn the
52	after all.[],A german professor found out that	pedophilia	apparently is just an ""alternative wiring"" in
57	grafting his own views about incest and	pedophilia	onto the basic Christian type beliefs. So
21	is "really important" to teach children about	pedophilia	and pederasty in the classroom because it
66	victimization will always be the downfall of	pedophilia.	Until we can change the parameters of
73	has also managed to record the online	pedophilia	community and concluded that each year the
67	sexuality that are, by today's definition, non-consensual	(pedophilia,	bestiality, etc), would erase all of the
9	subtly. Test the waters by talking about	pedophilia	with them and gauge their reactions through
30	subject. As far as public discussions on	pedophilia	go, like at work, ,etc. I tend
70	there are people who do not practice	pedophilia.	She calls them "law abiding pedophiles". According
49	into any subject regardless if it is	pedophilia	related or not. All religions are based
46	don't know how people feel about your	pedophilia.	You have gone a step further that
19	sounded too stupid. If you meant the	pedophilia	in general, they just said things like
25	I am going to therapy for the	pedophilia	thing in general and I hope to
60	checking those for updates. My views on	pedophilia:	It is okay to have feelings towards
77	struggling to come to terms with my	pedophilia.	I have been watching cp for years...

TABLE B.30: Random concordance lines for keyword *pedophile* in Family Relations subcorpus.

Index	Left context	Keyword	Right context
3,200	up her back and shoulders up to	her	neck laying on top of her and
624	that if she slips I can catch	her.	This put her ass level with my
164	sole purpose of gaining more access to	her	little sister I'd say it was wrong.
1,699	with the girl and any adults in	her	life. I don't know your personal situation
964	off his cock and naked body to	her.	(Maybe there was more to the story...)
652	lets out a long whispered ah as	her	legs start squeezing more. Can feel the
608	last second she jumps and I catch	her.	She wraps herself around me the way
584	kicking. I hug her tightly and tell	her	I have to go. Suddenly Claire leans
582	careful. She then has her finish wrapping	her	legs around me to save my balls
378	steer the conversation away from that because	her	parents are literally in the next room
680	first. Once I've done that Liz and	her	husband come out of their room and
592	possible that it'll be very easy on	her.	She giggles and tells me she knows
24	plans on going to college to get	her	teaching degree. She did go through a
117	bathed her and even took showers with	her,	I never thought of her in a
114	I can't say I even thought about	her	like that until her mother passed. All
72	in the forums (or in front of	her	for that matter). We don't condone force
579	where Claire's right foot is and tells	her	to watch out cause she could hurt
194	creaking and stopped dead in our tracks.	Her	friend walked right past my room and
646	gently with my right hand. She lifts	her	head up to look at me and
148	indiscretion when I turned things sexual with	her.	She simply said she loved me and
102	Tara also, yet I have cum to	her	stuff many times. I like vagina and
543	to reject Sierra citing that most girls	her	age develop crushes on adults and when
104	as my hand made it's way between	her	legs and I started rubbing her pussy,
576	stand up. Claire immediately goes to wrap	her	legs around me but with Lucas on
433	pretty easy to understand my attraction to	her.	"First off sorry if this subject has
587	bring me to the ground all on	her	own. So there I am sprawled on
546	let her push me out. I tell	her	that I love her as a sister
674	that I would do anything to protect	her	even risk my own life and safety
343	that I shouldn't have such as telling	her	about my underage neighbor and how I
50	didn't have to continually reinforce that in	her	like I've read that most people have
545	is doing and that I won't let	her	push me out. I tell her that
186	5 or 6 and I would bathe	her	or shower with her, she sometimes wished
565	back that I'm not afraid. I correct	her	by telling her that I'm brave and

TABLE B.31: Random concordance lines for keyword *her* in Family Relations subcorpus.

Index	Left context	Keyword	Right context
1,067	masturbate in front of them. Some of	my	most fulfilling erotic moments were having younger
2,786	did cause a stiring in me. Luckily	my	wife came and picked her up before
224	was the day the final remains of	my	childhood were over. And in some ways
17	don't think I said in any of	my	posts that he wasn't welcome here, just
335	look into getting something like this for	my	niece although I want to sit her
2,284	You're a lucky guy If either of	my	children ever let me know that they
1,407	anyone wants to talk about it more,	my	tc is in my profile. lol,, its
126	I have spoken to guys before about	my	fantasies about this very situation and that
34	Kelly, and as you can tell from	my	"Have I deprived her of anything" thread,
124	I have two daughters 10 and 13,	my	youngest was playing out with her friends
1,743	contribute with all that i can from	my	own situation and with some logical sense.
499	adult. She stands up to just about	my	belt. She's skinny but not too skinny.
1,223	matter.I'm worried that I will be bothering	my	dad if he knew this.I know that
1,713	girls? I always wanted to be with	my	mom, but never had the pleasure.... I
498	my niece Claire the girl who stole	my	heart. She is looking up at me
115	mind, I'd rather take the secret to	my	grave. I can understand that, but I
1,638	excitement and pleasures of sexual perversion via	my	dad's porn stash. By the age of
2,181	the better. I just finished wiping out	my	porn collection, all of it not just
48	are all the time though. I guess	my	main question is, how participative are they
649	I thought this would be fitting for	my	25th post in the community: âThere are
1,314	void. A longing." "As most people know	my	sister is adopted (from out of country).
2,238	swing her hands through the falling water.	My	cock was pressed nicely between her ass
753	being upset. I wrapped a towel around	my	waist, but was naked underneath. When I
2,167	future. If you've not read any of	my	other threads then for context I will
2,183	and plan to never meet again. In	my	current recently-divorced state of mind I am
632	make me feel good too so with	my	free hand I guide her hand to
90	to show her mine so I pulled	my	shorts and underwear down and she played
2,152	have a hard time resisting myself around	my	daughter a lot of the time and
1,717	relationship would be awesome! Ideas? I;ve reached	my	first milestone here at 27 posts... but
697	curse lol blessing because when I'm with	my	niece she knows what her kisses do
1,546	sexual kid. Actually with each one of	my	childhood friends I experimented. I only regret
58	the boy department. Care to send one	my	way? I would have returned it. I
1,741	No one to share this with and	my	fantasies are getting dangerous it the way

TABLE B.32: Random concordance lines for keyword *my* in Family Relations subcorpus.

Index	Left context	Keyword	Right context
27	talk sessions and mommy/daughter activities with both	Ashley	and Emma. Some of you right now
79	about lost it at this point. My	Ashley,	who had never done this before seemed
7	first time I see them both since	Ashley	told be about them. I'll let you
19	want to let you all know that	Ashley	and I are doing great and have
1	Sierra (12 when I met her) and	Ashley	(16-17 when I met her). When I
17	attractive in me. This is what made	Ashley	start thinking that it was ok to
23	I choose to tell you the things	Ashley	and I do but this is usually
28	a bad idea but I do trust	Ashley	100% in keeping this just between us.
14	all of this I'm still dumbfounded by	Ashley.	I never thought she'd be this advanced
2	athletic build. I didn't see much of	Ashley	as she was working at a local
50	for baths and bed came around and	Ashley	decided she wasn't ed to take her
18	had other ideas that she shared with	Ashley.	She told her that it was her
48	daughters day occurred, with one minor exception.	Ashley	seemed a bit competitive, she hasn't done
3	can both see the obvious conflict if	Ashley	decides she does want to spend her
47	she was sure I would say yes.	Ashley	did tell me that my brother has
22	the daddy daughter things together with both	Ashley	and Emma, I choose to tell you
15	or teen actors they see on tv.	Ashley	told me that she used to touch
4	if she walks in on you and	Ashley	one night? I'm also really getting curious
29	her when it occurred to me, ""shit,	Ashley	isn't in her room. She's in my
56	She's 12 and a real hottie. So	Ashley	had questions and wanted to know if
55	into bed and very quickly noticed that	Ashley	had her butt into my groin but
100	him I was still freaking out about	Ashley	and that I never thought I'd ever
36	making breakfast and went upstairs to wake	Ashley	up. When I walked into her room
30	said thanks but she always sleeps with	Ashley	when she's over. I told her that
85	them that it was my house and	Ashley	my daughter so I don't have to
45	did tell her how I dealt with	Ashley	and she thanked me for saying what
21	try to include activities other than with	Ashley	from time to time so you are
8	Now that my daughter (we'll call her	Ashley)	is showing signs of starting to mature,
67	and squeezing as she moans in pleasure.	Ashley	loves what I've been doing lately more,
96	went to bed after tucking Emma in	Ashley	was already in my bed. I did
5	your wife for your daughter? What if	Ashley	decides she never wants to be with
98	so caught up in my relationship with	Ashley	that even though I've thought a bit
101	situation. I guess I was focused on	Ashley	and I and how we'd progress. Now

TABLE B.33: Random concordance lines for keyword *Ashley* in Family Relations sub-corpus.

Index	Left context	Keyword	Right context
14	of having this type of relationship with	Chloe,	it's going great. Ash loves her uncle
76	She understood that. I think she and	Chloe	had their own fun under the sheets
63	sure that's a reason she continued. Then	Chloe	did something I not my brother expected.
121	old girl with the legend, ""Sweet Little	Chloe,	Your Ultimate Child-Lover"" and searches inside for
104	still going and also his relationship with	Chloe.	My jaw dropped again. I told him
1	like getting dressed that day. The lovely	Chloe	Moretz when she was in Kick Ass.
61	moment!" "Hey friends So interesting day yesterday.	Chloe	my niece and my brother Jim (we'll
90	was bluffing all this time. I said	""Chloe,	I love you more than you'll ever
28	fingers inside Chloe's pussy, fingering her, and	Chloe	had her finger on Ashley's clit. Both
6	Chloe. She told me that her and	Chloe	have been talking about ""sex and stuff""
84	very similar with me, I've looked at	Chloe	growing up and noticed how hot she's
32	be clear, I don't feel guilty about	Chloe,	her and I haven't done anything and
18	naked"", what do I do? I told	Chloe	that she could sleep in Emma's room
12	want to knw about my brother and	Chloe,	they are doing great! Their relationship has
77	sister in law come back home tonight.	Chloe	won't be spending the night and my
19	room where she slept till about 9.	Chloe	never even felt Ashley and I come
62	was normal for a few hours until	Chloe	started flirting with my brother and teasing
111	fingers around his hard cock, Charles helped	Chloe	slowly masturbate his large 8 inch cock,
116	child's body. Fucking a little girl like	Chloe	was always his dream, his ultimate lover,
26	her to her own bed and saw	Chloe.	Now either Chloe sleeps naked at home
99	too. Ash said she didn't know what	Chloe	meant so she asked for an explanation
46	theirs. Ash said no of course but	Chloe	told her she would try to see
110	arm around her little shoulders, Charles pulled	Chloe	to him, kissing her child's mouth, slipping
115	supple legs apart, amazed at how flexible	Chloe	was, Charles began to kiss and lick
79	law were here last night to pick	Chloe	up. My sister in law was so
113	his cock, almost made Charles explode. ""Oh	Chloe,	suck daddy's cock baby"" he moaned while
15	my brother called saying that he needed	Chloe	to spend the next few days with
27	own bed and saw Chloe. Now either	Chloe	sleeps naked at home or it's not
71	private place so I discussed everything that	Chloe	and I had talked about. I also
87	had my pleasant conversation with Ashley and	Chloe	about their games and Chloe's comment to
2	rid of Chloe's boobs in ""Hugo"", though.	Chloe	developed very nicely at a young age.
89	all of this and then I said	""Chloe,	I know you and your dad have
57	that they got into that position and	Chloe	put a flashlight between them so she

TABLE B.34: Random concordance lines for keyword *Chloe* in Family Relations sub-corpus.

Index	Left context	Keyword	Right context
1	gaming with her she introduced herself as	Liz	and she absolutely fell in love with
25	home I feel the need to message	Liz	explaining that I didn't think Claire did
34	good bye to my niece and nephews.	Liz	rolls her eyes and tells me I
54	the ground. Claire is still scared and	Liz	notices rather than scold or get upset
26	picking her up in the first place.	Liz	then points out that she's fine with
52	followed by Liz's husband then Michael and	Liz	herself. This is where the infamous snail
40	what do I do on my face.	Liz	just turns around and says think you
15	to look forward to hanging out with	Liz	and the kids again but sadly life
5	the mental scarring from past abuse in	Liz	She point blank refused my offer and
37	the marriage the kids as well as	Liz	and I are getting hungry so we
9	3 kids are gravitating towards me so	Liz	apologizes and comments that they usually don't
20	said. The next text I get from	Liz	is an apology and an invite to
12	apartment. Before we get there I have	Liz	stop at Walmart so I can get
2	back to the apartment complex to meet	Liz	and her husband in person. We would
3	after the kids went to bed. Eventually	Liz	starts showing me pictures of her kids
4	babysit if they were as good as	Liz	made them out to be. This was
11	ultimately ruined the night for the kids.	Liz	and I gather up Claire Lucas and
56	better and she shakes her head no.	Liz	then helps me shift my hand so
6	was scheduled to work that night and	Liz	wasn't comfortable bringing the kids trick or
46	it's time to set up the rooms.	Liz	focuses on the kitchen and living room.
57	way I don't have to worry about	Liz	or her husband looking back and seeing
44	get most of the boxes moved when	Liz	has to go pick the kids up
55	starts worrying she's going to fall. As	Liz	and her husband help me try and
45	with me to the new house and	Liz	agrees to letting them. The drive to
41	while Claire continues to cling to me.	Liz	notices where Claire's right foot is and
16	or hang out like I used to.	Liz	ended up falling into a spiral of
49	shouldn't be he doesn't care. Soon after	Liz	comes out with a bag containing the
62	the days leading up to Claire's birthday	Liz	helps me find the perfect gift for
63	everyone else is and I'm told that	Liz	and her husband are fighting in their
22	so we head back to the apartment.	Liz	puts Michael to bed. While she's doing
36	The next time I hang out with	Liz	she and get husband had been fighting
14	gummies which we split the ways between	Liz	her husband and myself). With the candy
18	a stop to it. When I noticed	Liz	was pushing everyone away I reached out

TABLE B.35: Random concordance lines for keyword *Liz* in Family Relations subcorpus.

Index	Left context	Keyword	Right context
71	can guess they want to but ultimately	Claire	wants me to herself this time. She
1	beautiful little girl. This is my niece	Claire	the girl who stole my heart. She
22	I will come back with a present.	Claire	agrees and Lucas asks if he can
64	much room as Liz Lucas Michael and	Claire	are all on the couch so I
2	to the car closely followed by Lucas.	Claire	opts to stay by my side and
28	and and runs off to play while	Claire	continues to cling to me. Liz notices
65	so she tells me to just lift	Claire	up take her spot and have her
69	for everyone however money is tight and	Claire	ordered the most expensive meal like me.
12	and underwear. I chuckle and comment about	Claire	being a little monkey which she giggles
66	so that's exactly what I do. As	Claire	is playing Liz and her husband head
42	and dried off. Once everyone finished eating	Claire	went back to the pool to play
82	I have made it very clear to	Claire	that those touches are a bonus to
68	we are doing from her husband. Eventually	Claire	notices something hard poking her ass. So
76	slowly slides in with my other finger.	Claire	is starting to get flushed in the
80	that I'll be able to be with	Claire	for a couple weeks at the very
83	any kind of damages he's done to	Claire	with his insecurities I've been very busy
13	to stop being clingy and get down.	Claire	then pouts and gets down and I
78	my pocket. When we were making out	Claire	must have stuffed her wet panties into
72	boys come out in their pajamas while	Claire	comes out in a big t shirt
49	carrying her she thanks me for helping	Claire	feel safe. As we hike through the
62	first or do I check in on	Claire.	Ultimately I chose to make my presence
79	was planning on spending more time with	Claire	soon however with a coworker suddenly quitting
53	she did with her kids let alone	Claire.	She explained about how she was raped
11	phone. As I get ready to leave	Claire	runs at me and gives me a
31	be a surprise. When they get in	Claire	tackles me to the ground as soon
3	to meet up with my brothers family	Claire	wants me to be her escort around
61	husband are fighting in their room and	Claire	is in her room sad. I weigh
26	for more food. Next thing I know	Claire	is hugging me from the right and
29	tell her I have to go. Suddenly	Claire	leans in and kisses me on the
24	drive through to get food. I notice	Claire	is extremely quiet and scared so once
35	gets the boys' room set up and	Claire	and I start setting up her room.
63	do it. Once the game is in	Claire	wants me to sit with her while
17	So I sit down with them and	Claire	sits in my right leg Lucas on

TABLE B.36: Random concordance lines for keyword *Claire* in Family Relations sub-corpus.

Index	Left context	Keyword	Right context
7	time last year, maybe a little earlier,	Kylie	started getting curious about sex and started
78	years older than they actually are. And	Kylie	tells me they definitely are more interested
58	realized it would be really useful for	Kylie	to have another woman be available to
52	only a year or two older than	Kylie	is now. And of course also from
62	if Tracey became a sexual partner for	Kylie,	that might make her less valuable as
57	girls since beginning to have sex with	Kylie.	The answer is yes, for sure. I
45	let things play out at the pace	Kylie	decided. Jane and I both knew the
5	years ago the wife and I decided	Kylie	was old enough to leave alone when
60	have already discussed the possibility of sharing	Kylie	with Tracey and Don. Frankly Jane and
50	we remain cautious and continue to let	Kylie	decide on what we do together. Even
10	Jane) and I discovered that the more	Kylie	saw and heard about sex, the more
46	to go and wanted to see how	Kylie	would like it. Anyway the reason I
51	we have been totally in sync about	Kylie	and I am 110 per cent aware
53	swinging and ""taboo"" sex. You ask how	Kylie	reacted to being introduced to sex by
33	she had no objection to us introducing	Kylie	to sex." "I don't know if this
54	of course there were stops and starts.	Kylie	would give us signals she was ready
8	but the more she explained, the more	Kylie	wanted to know. At a certain point
77	last summer when we started playing with	Kylie	but now I notice so much more.
17	to go beyond flirting if that's what	Kylie	wanted. Its still a hypothetical question though,
81	about the possibility. When we finally brought	Kylie	into the last discussion and they became
63	less valuable as someone outside the family	Kylie	could confide in. So we are being
22	my turn-on was basically just knowing that	Kylie	was curious about sex. The idea that
56	different way. We never forced or tricked	Kylie	into doing something she didn't want to
42	but now that we are involved with	Kylie	we have to be doubly careful." "I
23	wanting to pass that philosophy along to	Kylie,	played a big part in our decision."
72	the lines of, ""If we can bring	Kylie	along the way my dad did with
2	the background ProudPedoPapa. Hey maybe one day	Kylie	will be interested in playing with them
47	could see with my peripheral vision that	Kylie	was staring at his cock. I couldn't
69	have been thrilled to show them to	Kylie.	Maybe such clips are out there on
44	her point blank about that and when	Kylie	said yes is when I started to
68	maybe the day after) we waited till	Kylie	went to bed then told them the
59	and she's really the perfect person since	Kylie	has known her for her entire life.
13	more. Our activity is structured around when	Kylie	is in the mood. Jane and I

TABLE B.37: Random concordance lines for keyword *Kylie* in Family Relations sub-corpus.

Appendix B. Study 2: Concordance Tables

Index	Left context	Keyword	Right context
362	you are browsing the internet (non-onion websites).	Tor	hops.jpg Download and Start the Tor Browser
446	are several IRC servers that run over	Tor	(.onion addresses) that you can use. They
97	magnet link. I'd avoid mixing clearnet and	tor	activities as much as you can. I
116	extension that is widely in use among	tor	users. Being creative, an adversary could develop
108	people would download legal content without using	tor	thus revealing their real IP. That IP
179	probably lose 0.9% by requiring a secure	TOR	connection that they can't or don't know
521	Truecrypt volume. Run the following commands sudo	tor	-f /media/truecrypt1/torrc Next we have to make
385	to a third party location versus through	Tor.	Another example is if software or malware
252	played back via Flash, and can't protect	Tor	Browser users, a browser based on Firefox.
200	real. But when you do it tru	tor,	than it is funny and safe This
468	(for CentOS): Add The Tor Repository Add	Tor	to the list of repositories, a source
199	it's easier to find a girl on	tor.	Okay thank you for the offer. Okay
437	9.2 Configuring Web-Browsers and Applications to Use	Tor	9.2.1 Configuring Firefox to Use Tor 9.3
61	browser. (any link not opened in a	tor	browser can instantly reveal your IP and
478	reduce the RelayBandwidthRate and RelayBandwidthBurst to prevent	Tor	from using too much of your bandwidth.
487	file." "if you use a vpn or	tor	you hide Your ip but its okais
497	rooted all trafique in my house to	tor	network and encrypted this is verified by
479	using too much of your bandwidth. Restart	Tor	Restart Tor to make it see the
424	Tunnel JonDonym through Tor Tunnel Proxy through	Tor	Tunnel Retrosahre through Tor Tunnel SSH through
668	your using a tablet to get onto	tor,	what tablet d you have and what
185	hides my location? Before I knew about	Tor	and stuff I used to go on
233	and bust you again next month. Securing	Tor	Browser: The default configuration of Tor Browser
467	and configure Tor (for CentOS): Add The	Tor	Repository Add Tor to the list of
499	use to you. You may consider using	tor	everywhere and using fake information. Funding You
14	too many risks. While since having discovered	tor	and later the hidden parts, my methods
523	http proxy variable. It could also be	Tor	not working well so try updating a
6	had to abandon p2p, I came to	tor,	now we can go to frenet and
93	various creative ways to fuck yourself on	TOR.	These are short posts revealing how simple
208	Ab und zu gibt es Aktualisierungen fÄ¼r	TOR	und fÄ¼r Ubuntu. Vor einiger Zeit kam
428	UDP over Tor 1 Tunnel VPN through	Tor	Tor enforcement TorChat 1 Free Software, Libre
488	adresse mac spoof. You need a anonabox	(tor	routeur) for rooted all your trafique internet
439	all applications are preconfigured to run through	Tor.	Other uses includes encrypting your files, sending
330	your hidden service in /media/truecrypt1 3. Start	Tor	using the command sudo tor -f /media/truecrypt1/torrc

TABLE B.38: Random concordance lines for keyword *tor* in Childlove subcorpus.

Index	Left context	Keyword	Right context
399	Data Securely 4.3 File Slack 4.3.1 Hiding	data	in file slack space 4.4 Alternate Data
37	an unencrypted disk. Make sure all incriminating	data	is always encrypted. Using strong encryption you
82	for hide your ip and encrypt your	data..	viewtopic.php?f=204&t=5076= link for post for check if
4	and whether to pursue or pass the	data	along to other relevant agencies. The old
217	Doing this, the attacker can capture all	data,	inject data, or log information such as
220	systems. Generally it consists of a computer,	data,	or a network site that appears to
244	to make the access and organization of	data	more efficient. Deleted Data A common misconception
251	official page. Shadow data is the fringe	data	that remains on the physical track of
10	drive will assign different cells to store	data	to prevent wearing out any particular cells.
15	to convict you. 35 passes of random	data	sounds quite excessive to me. Quote from:
1	mention tracking vehicle movement. Since the acquired	data	is stored even if the vehicle is
7	for sure what is done with the	data,	and where it is stored. I know
50	is more than enough to prevent any	data	recovery, even without wiping. And a single
207	Follow TCP Stream Reassembling packets to view	data	such as images and getting detailed view
38	from a truecrypt/veracrypt encrypted drive renders the	data	unrecoverable for any practical purposes. No offense
46	services MAY be able to recover some	data	of a platter provided it's completely intact
212	or to try a device into sending	data	to it. Fragmentation— IP fragmentation is the
214	poisoning) is a computer hacking attack, whereby	data	is introduced into a Domain Name System
36	RF are available for buildings housing sensitive	data.	Mixers are a form of obfuscation, if
9	few minutes without power will render any	data	from ram unrecoverable at room temperature. One
61	The colder you make ram the longer	data	stays around in it. This technique can
206	cleartext, meaning you can read all the	data	that is contained within the packets. This
163	you should now delete the actual user	data	from the registry as well. You should
8	about thermite, seems to do miracles in	data	destruction Leave the ram powered down for
191	the Source IP address to send the	data	back to whomever sent it in the
51	is more than enough to destroy the	data	if you don't want to destroy the
14	proper tool ?[],And where do potentially compromising	data	eventually hang out that is out of
186	Your computer will first be requesting the	data	on port 53 (DNS) to request an
205	content on the webpage). If you submit	data	you will see a POST request meaning
218	the attacker can capture all data, inject	data,	or log information such as IP addresses
178	telecommunications network that allows computers to exchange	data.	There are two types of networks: a
80	the search, seizure, or interception of the	data	sought. In some cases, additional legal safeguards
151	resent so there is no loss of	data	during transmission. UDP on the other hand

TABLE B.39: Random concordance lines for keyword *data* in Childlove subcorpus.

Index	Left context	Keyword	Right context
100	overview of the types of tunnels and	encryption	protocols VPN's use. VPN Protocols Point-to-Point Protocol
216	Decrypting LG and other Android Full Disk	Encryption	(FDE) oclHashcat includes support for decrypting PBKDF2-HMAC-SHA1
267	an attacker with the main hard drive	encryption	key to be able to find out
4	as well. And while I did use	encryption,	I doubt that it would have stood
298	traffic of all other Tor users. Use	Encryption	& Bitcoins, Be Hard to Trace Encryption
256	in terms of the grand scheme of	encryption	options. It's also not deniable. Anybody looking
66	agencies like the NSA can probably break	encryption	but they won't for the reasons stated
273	core linux apps, the GPL license, etc.	Encryption	[URL], [URL], [URL],- A good IM client
28	without being able to break very secure	encryption...	the way it works is pretty simple..."
293	Hdd encrypted.. Luks encryption + 3 other	encryption	Is the minimal!" "Hello to all i
257	the United States if I remember correctly.	Encryption	isn't foolproof, it's a deterrent- something that
45	porn. [URL], Canada Laws? general right to	encryption	Although the Canadian Charter of Rights and
254	NOC, rebooting automatically dismounts encrypted partitions, so	encryption	will stop nosy sysadmins from poking around
77	its way out. TrueCrypt is an open-source	encryption	program. It works by creating 'volumes', which
299	Encryption & Bitcoins, Be Hard to Trace	Encryption	might be crackable using highly advanced computers
1	everything that had to do with tor	encryption	and most of all cp. I decided
65	to worry about. We are assuming that	encryption	will protect you. Again, intelligence agencies like
315	all os is now with a good	encryption	for all os if you like the
50	serve to limit certain methods of circumventing	encryption	which are clearly disproportionate or prejudicial. Evidence
2	Look into real anti virussoftware, look into	encryption,	look into how not to leave evidence,
17	resort to truecrypt, just the built in	encryption.	I am not trying to hide it
8	It leaks metadata and has an unproven	encryption	scheme relying on obscurity. Of course one
287	hidden service machine. If they capture your	encryption	passphrase, your data will be compromised. Recently
242	last year of the dangers of consumer	encryption	programs, and British Prime Minister David Cameron
71	cannot be forced to give up your	encryption	passphrase in a criminal proceeding (but that
237	not necessarily mean better encryption. Use GnuPG	encryption	in your routine e-mails. The more encrypted
333	shared folder between Host/ Guest OS') Clearnet:	Encryption:	[URL], For video editing: [URL], For re-encoding
279	keys must be tried to crack your	encryption.	Linux- Linux is an operating system. It's
173	of 12 tested keyboard did not use	encryption.	Article, torsafe JS off [URL], Torsafe, JS
10	suddenly only protected by the relative weak	encryption	of your wireless keyboard. That is if
82	worth discussing quantum computing here. All modern	encryption	systems rely on the fact that factoring
258	will make your adversary work harder. Every	encryption	scheme people have devised has eventually been
20	the decryption is a part of the	encryption	process it can not be circumvented. Congratulations

TABLE B.40: Random concordance lines for keyword *encryption* in Childlove subcorpus.

Index	Left context	Keyword	Right context
7	and not a shred of proof, perfect	security	and secrecy... As I understand it she
89	we consider open source necessary for good	security;	we have for decades. Public security is
113	findings at the Black Hat Europe 2016	security	conference in November and the 33rd Chaos
369	a risk, you need to implement physical	security	measures that deal with the possible threat.
34	is a US based company a simple	security	letter would make all their promises null
54	for belonging may overcome the physiological and	security	needs, depending on the strength of the
170	header before sending the data. Specifically, Encapsulating	Security	Payload (ESP) and Authentication Header (AH) are
11	my major objection against pedofunding. Disregarding inherent	security	concerns. Quick comment about the Captcha, it
365	3.2 Radio Leakage, TEMPEST, etc. 3.3 Physical	Security	3.4 Traffic Correlation 4 Installing Ubuntu 5
220	into logical groups, subnetting can improve network	security	and performance. For example, most home devices
264	is up to the manufacturer to provide	security	when sending jobs to a printer.. Read
58	to buy a separate computer for extra	security.	Not sure if that is good enough
223	emails, checking your bank account, application data	security,	or transmitting insecure data over a secure
406	our campaign page." "Tails Update Fixes Numerous	Security	Flaws Tails, the live operating system of
69	how much children care about their online	security.	Children these days live in a world
124	3.2 Radio Leakage, TEMPEST, etc. 3.3 Physical	Security	3.4 Traffic Correlation 4 Installing Ubuntu 5
354	security field is that 'obscurity is not	security'.	This basically means that you can't assume
277	Continuity 5.1 Security Concerns with Backups 5.2	Security	Concerns with Sleep and Hibernation 5.3 Ensuring
132	to lots of software but it's a	security	risk because you don't know who is
147	very in-depth guide about security for activists	Security	for Activists- [URL], Security at Resist.ca- [URL],
160	to work across various qubes without compromising	security.	The innovative Template system separates software installation
285	the Tor anonymity network, Debian GNU/Linux and	security	by isolation. DNS leaks are impossible, and
42	security sets a very dangerous atmosphere where	security	is not taken seriously. I am a
404	Back in 2015, a group of internet	security	specialists identified the risk of harmful software
125	Free Software 12.2 Encryption 12.3 Privacy/Anonymity 12.4	Security	12.5 Tech Collectives/Resources 12.6 Software For Your
273	a balance between these three aspects of	security	to ensure the proper use and control
187	largely because of Chrome's sandboxing and plug-in	security.	Comparatively, Internet Explorer implemented (lacking industry standard)
128	a risk, you need to implement physical	security	measures that deal with the possible threat.
389	But if you dont Use The good	Security	Protection.. you Are in jail Tomorrow.! Yess
325	data online and more.. viewtopic.php?f=204&t=4885&p=19542#p19542 For more	security	tails and one hdd encrypted new and
338	your disk appears to only be a	security	issue if someone gets to your physical
82	dirty ass (28 photos, 10 MB)" "Strong	Security	KeePass supports the Advanced Encryption Standard (AES,
280	confidential or secure. In essence, privacy is	security	and confidentiality means that third party individuals

TABLE B.41: Random concordance lines for keyword *security* in Childlove subcorpus.

Index	Left context	Keyword	Right context
116	boys. I don't know what type of	relation	you are looking for, but any type
68	the child will look back on a	relation	that makes me a opt for conservatism.
28	discussed before. I am in a committed	relation	with my boyfriend. We know each other
51	that relation by starting your own sexual	relation.	Now the only viable option is to
7	last, and i desired a more lasting	relation.	While i can't be certain what the
95	will be very difficult to make a	relation	work in the long run. Though reality
101	to clarify where you stand in this	relation.	It seems there you both consider worth
124	you is is ok to start a	relation	with a child. So expect any psychologist
151	current bf. It wasn't exactly a planned	relation,	but we are now together for over
153	not wish to enter into a sexual	relation	with a young girl. Now you are
166	relation, you brought Nathan in to your	relation.	Not only that, you have made it
23	#1 in the bedroom? Sharing in your	relation	with Nathan may be his way to
155	wish to make her complicit in this	relation.	You may find that she considers you
2	would not be time to build a	relation	in which I'd feel comfortable to have
159	in the past, I stopped after a	relation	ended very abrupt and outside mine or
87	my life I had a rather complicated	relation	with emotions. I never was good at
152	have in fact inserted itself into the	relation.	You stated she knows you used to
163	it isn't illegal. If you have a	relation	that is illegal it can end up
178	the possibility of an existing long term	relation.	Looking purely at the AoC i would
154	not only you enter into a sexual	relation	with a minor but wish to make
16	opinion that the appropriate conditions for such	relation	are not feasible.Â The end result is
191	intensity of my adult attraction has no	relation	with my attraction to little girls. I
98	how to combine my activities with my	relation	with F. I guess F and I
58	to hearing about your long and loving	relation	with him. Just wanted to wish you
76	The difference in the nature of the	relation	follows from (in my opinion) there should
123	that I have tried to recreate my	relation	with John. And yes he was the
176	to push those involved in such a	relation	away. It will accomplish nothing but to
19	weapons in communication. While an adult child	relation	does not have to be harmful, they
54	be a ticking time bomb in your	relation.	Have you considered the consequences to your
122	a chance to salvage some of the	relation,	a chance you will get to know
4	like that. I was worried about our	relation.	Looking back it was a very surreal
77	sex is based on the existing parent/child	relation.	And I wonder would you make the
25	This was by no means a loving	relation	between an adult pedophile and his young

TABLE B.42: Random concordance lines for keyword *relation* in Childlove subcorpus.

Index	Left context	Keyword	Right context
78	rates in the U.S. bear no fixed	relationship.	Between 1999 and 2009, for example, New
484	being in any sort of long term	relationship	with an adult girl or a woman.
211	who seek advice on the voluntary intimate	relationship	with a family member. If your goal
209	of hard knocks. If you develop a	relationship	with girls relative to your own age
203	I become conservative is with a hands-on	relationship,	which usually doesn't end well for anybody,
4	Would your ex also consider taking her	relationship	with your daughter further? "Sweetcheeks 01 Well,
193	whatever we can to be in this	relationship	together. Again thank you very much for
169	the other, but still I maintain good	relationship.	Those group dont know them eachother. I
89	into thinking that you have an equal	relationship.	It is very easy to be fooled
166	waiting his AOC to engage more close	relationship...	I understand this I dont want to
27	Have you considered the consequences to your	relationship	with your gf and daughters if discovered?
20	I wanted to die. I knew my	relationship	and my life was over. Except it
173	come afterward. Being gay, and having a	relationship	with a young men still has many
127	day i can open myself to a	relationship	with one, i don't know if some
176	that. QUOTE: Do you want a sexual	relationship	Well yes, I would like to sleep
14	those you will have to define sexual	relationship	as well. While fondling might be seen
284	I would give anything to consumate our	relationship,	she wants it so badly. Just the
192	someone who is in a committed legal	relationship	like I am and how they make
2	a child and then have a positive	relationship	many years down the road with that
274	ever happen. As for every genuine loving	relationship,	god knows how many children are being
190	a life-long friendship, instead of an actual	relationship	because of the fact that we only
199	it was their choice to have the	relationship	in the first place later on in
142	a 45 yr old man to start	relationship...	I mean that I'm avoiding getting to
187	be capable of being in a regular	relationship	that was legal considering my attraction. Well,
174	had provocative manner because of a complex	relationship	thought, he's still a young boy. Getting
147	doesnt mean he want to have a	relationship	with me, actualy being 16 means a
160	that I able to get a close	relationship	and I know how. My problem is
259	out things started changing rapidly. Our open	relationship	enabled her, of course, to see other
201	this conversation–power and control in a sexual	relationship	is becoming more and more paramount. I
99	can work. While i am in a	relationship,	he doesn't know about my preferences and
167	something, I don't know. Maybe a complete	relationship	with a 16 years old may fix
316	commit or marry an adult woman the	relationship	is not going to last long, except
359	to have with you is the father-daughter	relationship.	Don't screw up. And if you don't

TABLE B.43: Random concordance lines for keyword *relation* in Childlove subcorpus.

Appendix B. Study 2: Concordance Tables

Index	Left context	Keyword	Right context
38	Author: unknown REDACTED (Boy,	Hurtcore,	Insects, Pain) Lolly Torture Erogames For girllovers
28	wrong section, this section is clearly for	hurtcore,	and these are not hurtcore! Ronald Franco
27	am surprised to see Tara posted as	hurtcore,	I know that Bart was rough on
32	was not hurtcore, and this is the	hurtcore	section! This is sometimes called "The Club"
37	Sergdriver REDACTED "Short	Hurtcore	Stories (Russian) REDACTED
26	I suggest she should be in the	Hurtcore	section! Yes, a compilation of two Tara
44	in the wrong section. It is no	hurtcore	and should be posted in the regular
31	From comments above it certainly was not	hurtcore,	and this is the hurtcore section! This
29	clearly for hurtcore, and these are not	hurtcore!	Ronald Franco and Karen, and not hurtcore,
40	holes.[]],I've never yet hurt a baby, but	hurtcore	and crying babies gets me soo hard[],.
25	Hurtcore section with stuff that was not	hurtcore,	and so should have been here in
33		Hurtcore	: "" "" mix of hurtcore
24	time complaining about people filling up the	Hurtcore	section with stuff that was not hurtcore,
23	file: Tara, and in my opinion not	hurtcore,	so in the wrong section! The first
35	mix of hurtcore & gore: "(Girl, Bondage,	Hurtcore,	Spank, Whipping, Piss Torture) Author: Bozo Da
43	may be it should be posted in	hurtcore.	Preview: [URL], File: [URL], Password: !!WHIPMASTER!! This
19	by a Mod! It was in the	HurtCore	section when I wrote that! Use Topic
1	been into setting before now" Is this	Hurtcore?	"Sorry if this is not the right
6	like this." WRONG SECTION- THIS IS NOT	HURTCORE!	POST IN "REGULAR PORN"! HOW OFTEN DOES
7	post markuss, but please stop posting non	-hurtcore	material in this section which is very
4	the pictures do not appear to be	HurtCore,	so should not be in the HurtCore
21	belongs in Regular Porn, it is not	HurtCore.	Yes, the good old days before the
2	PASS luv2beat&hurt49859!" None of this looks like	HURTCORE	to me "Risky i guess but it
15	the Regular Porn section, it is not	HurtCore.	The girl is not hurt or unhappy,
13	slow anyway. Thanks Carwerdmas, nice but not	hurtcore.	It belongs in the Regular Porn section!
8	discussions of 18 (twinks) and younger of	Hurtcore	materials (videos & pictures) as well as
16	agony) before I personally rate it as	hurtcore.	I don't see Grettel indicating any of
9	This is Kait Sugar. THIS IS NOT	HURTCORE!	While you are to be thanked for
3	safe host, not minfil.org! This is the	HurtCore	section, and very clearly labelled as such.
12	preview is Demetria, and it is not	hurtcore.	I think the 3 pics are of
10	and stop posting stuff that is not	hurtcore	in the HurtCore section! This and others
11	stuff that is not hurtcore in the	HurtCore	section! This and others that you have
22	do a search, MK, remains down. Not	hurtcore,	so it should be in the Regular

TABLE B.44: Random concordance lines for keyword *hurtcore* in Violence subcorpus.

Index	Left context	Keyword	Right context
4	bag over his pretty little head while	rape	the fuck out of the pretty boytoy!
3	in child porn. Any time it's hardcore	rape	of a boy or girl with crying,
26	love and affection, punish her soundly, and	rape	her thoroughly. Per your suggestion, I had
22	excitement fired me, and I started to	rape	her thin, undeveloped body with all the
12	death but not all Lust killers love	rape	some of my friends love the process
10	Bound"" "Author: Dr. Fisch (Hurtcore,	Rape,	Toddler, Piss Drinking, Girl, Blood, Snuff)
15	strong and horny. It's no wonder we	rape	their little cunts and fucking hurt them.
11	kill the boy only without molesting or	rape	him in first so bad news for
2	really hot about beating them unless there's	rape."	Please tell me there's a video where
16	the letter A. Act friendly and then	rape	them. banned because he asked to be
14	fetish, yes, but also more." "Yes, baby	rape.	The baby videos on toybox for example."
9	this forum please respond. I feel like	rape,	abuse together wit mom and dad. I
24	at the time. Back then stopping to	rape	a young girl in her bed while
6	the hottest for me, would love to	rape	them hard man!!"
18	all your posts! "I believe he mean""	Rape	her"" LOL" Old but still good. Pass
13	be deleted" Grab a little fucker and	rape	it hard. banned because he asked to
1	pay good money to see a woman	rape	a kid like this. If I could
20	vids That looks amazing 2hours of baby	rape	and pure little ass Pass works perfect
19	one for torture." She is a good	rape	meat! "Yes right old videos and nothing
5	Non Consensual Gay Stories Was alot hot	rape	and snuff stories there. Thanks" Mmmm thanks
7	Orgasm! (5:56) 12.Tara8Yr2007-Buttfuck-ShortClip (00:33) 13.Tara 7yr- ass	rape	(07:10) 14.Tara 8yr- March 16th Masturbate and
17	Great Thread, did you heard about the	rape	videos fro sale in India? maybe someone
23	take it out on you. I'm gonna	rape	the shit out of you, and I'll
25	Password: REDACTED Story: ped (10) M/g bdsm	"rape"	Story: Directions for use of the 9-12f
8	the best give him sleeping pills and	rape	him in his sleep "If you are
27	I looked puzzled "you know, beat or	rape	them to death!" my heart leapt at
21	hard-working), pistoning buttocks. Slow this delightful anal	rape	may have been, but hard it certainly

TABLE B.45: Random concordance lines for keyword *rape* in Violence subcorpus.

Index	Left context	Keyword	Right context
25	so tenderly while she was intending to	hurt	her child so much "Just a bit
7	section which is very clearly labelled for	Hurt	Core. This belongs in "Regular Porn". Correct!
15	be in the Regular Porn section on	Hurt	REDACTED. There is a section for requests.
13	is dead, apparently permanently. It is not	Hurt	REDACTED's fault that there is currently nothing
20	I play with their holes.[]I've never yet	hurt	a baby, but hurtcore and crying babies
11	is not HurtCore. The girl is not	hurt	or unhappy, and the boy certainly is
18	jerked over a newborn. Fucking wanted to	hurt	the little slut, but mom would find
2	around to best age to start to	hurt	them" any idea why i cant get
14	Choice is very limited at the moment,	Hurt	REDACTED (not only HurtCore, has a Regular
8	tell us what GoodWin's Password was on	Hurt	REDACTED? And to be on the safe
5	problem" "I be love to torture and	hurt	especially little boys (a little bitch sometime
9	on the safe side, zash's Password on	Hurt	REDACTED too? Yes please, English subtitles always
12	completely out of order. It is not	Hurt	REDACTED's fault that REDACTED is dead,
6	for best orgasm ever..." "love real adult	hurt	and snuff especiaally by gore sites and
26	was happening, we were actually going to	hurt	and torture some little girls to the
23	rate. "I want you and I to	hurt	her now," the mother said. I wasn't
21	the rest I'm still fucking. Yeah, I've	hurt	a kid -once-, because I went overboard,
28	old ass, bleeding, I bet that really	hurt,"	she kept reaching out and revolving the
4	to read here about your fun to	hurt	your cute boy and his cocklet too
22	her gag. My cock swelled until it	hurt.	The next nine strokes were hard and
24	the girl, you could tell it really	hurt	her, but before she regained her breath,
10	are to be thanked for posting on	Hurt	REDACTED, please take note of the Board
16	's! I love it when boys are	hurt	severe great vid! I got so hot
19	we rape their little cunts and fucking	hurt	them. banned because he asked to be
17	a kid. Made some cash without getting	hurt	(much). Thanks for sharing." "Thanks for all
29	on the barbed wire board trying to	hurt	herself more as this seemed to be
3	a child and pedo sadists like to	hurt	children it make so much fun the
1	hard Preview [URL],... PTAiMB.jpg File [URL], Password	hurt"	"wow love it, around to best age
27	love nettles" She said almost breathlessly "They	hurt	like hell" She stroked the bunch over

TABLE B.46: Random concordance lines for keyword *hurt* in Violence subcorpus.

Index	Left context	Keyword	Right context
13	some knickers on to soak up the	blood,	she fetched some large style cotton knickers
17	she bit the child's lip viscosly drawing	blood,	she left the broom handle sticking out
3	and arteries for spurting so young hot	blood	for heavy erected cock, please apologize my
19	pushing her onto the barbs loving the	blood	that was smoothing between them. I noticed
4	(Hurtcore, Rape, Toddler, Piss Drinking, Girl,	Blood,	Snuff) Sergdriver
9	was distantly aware of her hot virgin	blood	dripping from my heavy hairy ball bag
20	the shock of the intrusion. I saw	blood	all around the pole, it had really
15	it, we're gonna make a bit of	blood	now sweetie" the woman handed me a
6	the withdrawn cock covered in their toddler	blood.	It's difficult to make out the pics
18	did the same and we soon had	blood	flowing from her face, her lips were
14	on, it fascinated me the way the	blood	soaked into them, she commented on hoe
10	lifted her panting body. Her rich virgin	blood	smears his long cock, balls and matted
5	and cutting off the skin. Love the	blood	and the screaming.[]But I really prefer an
16	girls little nipple and a spot of	blood	appeared the kid sobbed, the woman giggled.
21	face shattering most of the other phials,	blood	flesh and glass shreds came cascading from
8	from her cracked voice were incredible! My	blood	boiled in my veins, a new excitement
2	when needles are stuck into them, when	blood	is poring out, when a child's arm
1	when needles are stuck into them, when	blood	is pooring out, when a child's arm
11	tender young skin brushed with bright virgin	blood.	She was panting, sobbing each time he
12	and clitoris hood to have flooded with	blood	to become exceedingly sensitive. She could still
7	The already traumatised skin burst in places,	blood	spraying out. My tiny slut squealed and

TABLE B.47: Random concordance lines for keyword *blood* in Violence subcorpus.

Index	Left context	Keyword	Right context
44	and down the half of my steel-like	cock	shaft still protruding from her daughter's wildly
5	it over my body and round my	cock	and have the most amazing wank." "Baby
34	match my own deep moan as my	cock	slid into her tight anal sheath to
16	I see a nurse holding a baby	cock	and cutting off the skin. Love the
38	vaginal sleeve. It felt so good! My	cock	was now pushing deeper. The previous obstruction
50	on her face as he edged his	cock	back between her little legs and lodged
2	so young hot blood for heavy erected	cock,	please apologize my bad english" "so faantastic
46	feel the boy's little privates. His little	cock	had shrunken inwards and his little balls
49	as he rubbed the tip of his	cock	back and forth along her hairless little
41	at the thought of how sore my	cock	would be after, Look around. Suddenly, Suzi
52	thing down onto his cock, willing his	cock	to enter her tight little cunny while
19	start with the letter I. Insert your	cock	in their mouth and piss in it.
22	put a sexy little hairless pussy or	cock	with the women then I find them
30	around in the air, I felt my	cock	harden at the sound of her terrified
9	how's about: U: Unload on his baby	cock.	"We can always send private messages I
1	I dont need to just ass and	cock....	i need to see this bitching crying.
20	up. [],Love how they naturally suck on my	cock	and giggle when I play with their
45	calloused hands, and sawing his massively thick	cock	in and out of her horribly distended
4	on my face, the rest on my	cock.	Then smear it over my body and
51	pulled the young thing down onto his	cock,	willing his cock to enter her tight
47	opened wide too when he saw the	cock	Jed carried. Jed was rightfully proud of
31	keening coming from behind her gag. My	cock	swelled until it hurt. The next nine
59	image of spreading her open with his	cock,	of watching her face as he tore
35	and time again by my fat hard	cock,	her horribly stretched bowels cramping painfully up.
36	cunt up. Pulling my burning, shit covered	cock	out of her now wide open and
8	when I saw Baby Heart listed, my	cock	sprang to attention." This is an alphabetical
43	again. I slap her face harder. My	cock	swells again. Not a sound this time.
7	her in the ass, then wipe your	cock	on her teddy bear. 3. A man
53	scream and thrashed on his lap. His	cock	could feel the tightness of her around
28	to deep– fuck her throat, slamming my	cock	into her abused gullet as hard as
27	but she held still. I withdrew my	cock	to her lips, allowing her to breathe
6	and their little tongue licking against your	cock	head is awesome. This little one is
29	until only the head of my thick	cock	was in her mouth. I wanted her

TABLE B.48: Random concordance lines for keyword *cock* in Violence subcorpus.

Index	Left context	Keyword	Right context
61	suck the material of her jeans, my	mouth	was flooded with her urine, I swallowed,
32	the girl's sweet tasting fluids filled her	mouth	with delightful flavour. Her tongue moved up
7	the cutie in white with the open	mouth."	"Life for us pedos can be a
41	kid, specially onto her face, "Open your	mouth"	she barked, and the child opened it
5	8.(Pthc) Tara 2008 9Yo Last Cumshot In	Mouth	From Daddy (2:07) 9.2007 Tara 8yr ass
8	in. The feeling of that wet baby	mouth	is awesome and a few pumps and
48	fascinated as the childs piss flooded Lenas	mouth	and cascaded down her tits. Lena must
54	started to kiss me, real deep open	mouth	kisses, I didn't like the taste of
60	anus, there was one masturbating into her	mouth.	The kid was dripping in spunk, it
38	as a lovely sweet juice filled his	mouth.	He savoured her orgasm and let her
34	body. His heart was suddenly in his	mouth	as he stared at her slender young
4	clown (13:03) 3.Tara-(Pthc) 2007 8Yr- Ass To	Mouth	(12:21) 4.! New! (Pthc) 2006 Tara 7yo
57	and glass shreds came cascading from her	mouth.	The child passed out after screaming her
6	the last 3 too with so swollen	mouth	make me hard so much welcome for
9	pedos shooting their cum in a baby	mouth,	or in a torn baby hole. banned
52	enjoying moving the shit around the girls	mouth,	tasting and re-eating her crap, the child
16	squirt of thick white cum into her	mouth.	My cum ran out of her nose,
15	of my thick cock was in her	mouth.	I wanted her to taste it. I
31	rising to her knees to lower her	mouth	to the child's pink sex. Becky gasped
37	allow his mouth to kiss her. His	mouth	peeled her cheeks apart and lift fastened
43	on the little girl's face, tonguing her	mouth	and licking in her eyes and up
39	clothes. Her eyes were sullen as her	mouth	worked to bring spit onto her tongue.
42	her mother had finished pissing in her	mouth,	she said "Now for the last, but
24	came, but a sweet taste filled her	mouth	while her mother groaned and turned her
14	and my cock slid into her tight	mouth,	and continued to stretch its way down
33	feel it swell within a hot little	mouth.	Loved to have a timid little tongue
45	lust, she started spitting into the kids	mouth,	it was as if she just didn't
47	girl, she kissed her full on the	mouth	trying to open it and push her
22	milk flowed freely into her eager little	mouth.	Her brother's coming had stopped all that
28	little mouth. "Just take it into your	mouth	and suck gently upon it," Mavis breathed
56	she popped the rest into the child's	mouth,	telling her not to bite them as
35	the indent that was her pre-nubile pussy	mouth,	a little slot no larger than his
19	a thief when younger, living hand to	mouth	for years, only his wits and running

TABLE B.49: Random concordance lines for keyword *mouth* in Violence subcorpus.

Appendix B. Study 2: Concordance Tables

Index	Left context	Keyword	Right context
1	is crying I dont need to just	ass	and cock.... i need to see this
13	2006 Tara 7yo masturbates and dad fingers	ass	(4:30) 5.(Pthc Pedo) Extreme- Tara 8Yo Girl-
20	cry twice? A. Fuck her in the	ass,	then wipe your cock on her teddy
2	her forehead. That dog has eaten better	ass	then me "I use to have no
18	(5:15) 26.vid- Tara 8yr- vibrator and cums,	ass	fuck- July 14, 2007 (3:56) 27.Tara's Oral
6	wanted to see a baby boy get	ass	raped is an extreme turn on for
5	boy like he's a toy. Holding his	ass	up and spreading it for the man
21	deep. Wow! Make him lick your dirty	ass.	banned because he asked to be deleted
16	Real Orgasm! (5:56) 12.Tara8Yr2007-Buttfuck-ShortClip (00:33) 13.Tara 7yr-	ass	rape (07:10) 14.Tara 8yr- March 16th Masturbate
32	viewtopic.php?f=55&t=6716 Me i would say 1yo up	ass	and 4-5yo up pussy How about you??
8	does. So fucking frustrating. Needs less creepy	ass	whispering from the guy and more slamming
11	he starts to fuck her up the	ass	he turns around and switches off the
3	then the camera panning to a boy's	ass.	She spreads it open with her hand
28	me means putting my dick in her	ass,	pretty much until the hilt, and thrusting
22	beautiful cum running down the baby boy's	ass	crack, cocklet and balls.],Thanks for posting. banned
9	he is going to cum in her	ass,	and he does, and she then has
23	here Old Mami! "Great color in that	ass,	thanks" "Thanks Great Thread, did you heard
36	boy's bottom. He licked the cute little	ass	cheeks, then spat on the sore looking
10	He fucks her balls deep in the	ass,	and at one stage the girl says
17	9yo- she was such a good little	ass	slut (new unmasked) (7:41) 17.Tara 10yo 2009-
24	shoved his whole length up the kid's	ass.	You'll notice that the kid's ass closes
4	straight up fucking the boy in the	ass	while the kid screamed and cried. We
12	by a clown (13:03) 3.Tara-(Pthc) 2007 8Yr-	Ass	To Mouth (12:21) 4.! New! (Pthc) 2006
37	talking "Look at that five year old	ass,	bleeding, I bet that really hurt," she
25	kid's ass. You'll notice that the kid's	ass	closes shut the minute the guy pulls
7	trouble fitting all the way in her	ass,	which is the better hole." It's longer
19	of it but I mostly got my	ass	handed to me so that wasn't very
15	(2:07) 9.2007 Tara 8yr ass fuck and	ass	vibrator (3:23) 10.New Tara!!! (5:01) Spam Pedoland
29	accomplice... he could easily fuck the girl's	ass	at this stage. "If you can't film
31	2hours of baby rape and pure little	ass	Pass works perfect just don't have the
34	or torn. He wanted a little virgin	ass,	nice and pink and soft. The boy's
27	and the fakes. And then gets his	ass	handed to him when other vids show
26	guy pulls out. It's just how the	ass	works. If you want to know what

TABLE B.50: Random concordance lines for keyword *ass* in Violence subcorpus.

Index	Left context	Keyword	Right context
81	end, but I guess I'll see. Also,	thanks	for the tip!" I don't find myself
48	Agree with badunclejoie. Enjoy pics like this.	Thanks	for sharing. Always been a big fan.
61	a kid, but I love these vids.	Thanks	for sharing." Great fakes. Thanks to posters.
57	Is there place where you posted it?	Thanks.	See these previously. Good vids. Thanks. "Saw
70	post more. Thanks for sharing. OUCH! And	thanks	for sharing. "Damn, lots of vids. Nice
66	with security. Please re-up to another host.	Thanks.	Nice series. Thanks for sharing. I'm sure
72	quite good. But that was long ago	Thanks."	
12	have it, or if I want it!	Thanks	baobab55. If his dick really is as
52	say these are hot? Now I did!	Thanks	for sharing. "I'm holding my nuts right
59	don't know how old the boy is.	Thanks	for sharing." Saw this before. Great vid.
56	sharing." Great blowjob. Very enjoyable to watch.	Thanks	for sharing. Love your avatar. Think I
78	matters to be extremely interesting. Anyways, much	thanks	for taking the time to write all
83	towards babies. Long may this site run.	Thanks."	"A hot selection of pics, all new
84	banned because he asked to be deleted"	"thanks	HowardCL, i thought i was the only
27	girls were sniffing, but I doubt it.	Thanks	for this one rimbeau, new to me"
2	Some previews: [URL], [URL], [URL],... MJ [URL],	Thanks	Mio15! The boy is sorta lucky with
82	them on that other site too, but	thanks	for gathering them together as a set.
95	[URL], Password: !!WHIPMASTER!! Preview: [URL], Working Mirror:	Thanks	to Corporal. Oct 14 [URL],... 5fc1546778 Password:
15	are pictures, presumably intended as video previews.	Thanks,	but preview "File not found". Didn't take
28	the Gabby video most commonly posted, so	thanks	for the additional two. What on earth
13	general opinion is that she is Russian.	Thanks	Kingpin, now that I have seen the
50	He was used like a rented mule.	Thanks	for sharing. More good stuff. Thanks for
62	these vids. Thanks for sharing." Great fakes.	Thanks	to posters. Please post more. "Love BDSM,
93	but I mean more videos or pictures.	Thanks"	
88		"Thanks	to the OP and thanks to Loki
24	not think that applies to this one!	Thanks	for the warning. I scanned both the
25	I hope that is correct! Great post,	thanks	EggplantsGhost, this is the first I have
63	Made some cash without getting hurt (much).	Thanks	for sharing." "Thanks for all those ""bad""
86	Old Mami! "Great color in that ass,	thanks"	"Thanks Great Thread, did you heard about
36	stuff...hot stuff! Thanks for sharing. Good stuff.	Thanks	for sharing. "I thank you for this.
85	Grand mamas! Its definitely an old mummy!	Thanks	a lot very great hope you have
11		Thanks	for the post padoga, the greatest number
71	of vids. Nice to hear the story.	Thanks	for sharing." "Girl A, though she looks

TABLE B.51: Random concordance lines for keyword *thanks* in Violence subcorpus.

Index	Left context	Keyword	Right context
11	"Thanks dude, i must be blind, lol" ""	Please	RE-UP this... . . . Great way to do lil
8	hot about beating them unless there's rape."	Please	tell me there's a video where the
50	911896.jpg [URL]=[URL], [URL File: [URL], Password: !!WHIPMASTER!!	Please	use the regular porn section for this
7	it's impossible to sync them up.=("Yes,	PLEASE	post the version with sound if you
28	for sharing." Great fakes. Thanks to posters.	Please	post more. "Love BDSM, fake or real.
10	someone have the full version of this	please?	[URL],... erture.jpg" Great job man "I have
32	worry about your new passions or desires	(please	apologize my so bad english) sure take
34	and the so sweet red juice too...	please	apologize my bad english" "like to be
17	in my belly" 00:01:25,001-->00:01:28,000	"Please,	not too hard daddy" 00:01:56,001-->00:01:58,000
39	free user, re- upload to another host	please." ????????	Nice but to old the girlies! I
24	zash's Password on REDACTED too? Yes	please,	English subtitles always improve a video for
47	re-up: [URL], Password: !!WHIPMASTER!! Next time send	please	a PM for a re-up. Reup 03.July
49	jism. She gulped it down eager to	please	and thereby avoid further punishment. "Not bad,
6	so much thanks for your postings here,	please	apologize my bad english..." "YESSS love to
31	much wounds and useless struggle and squirm,	please	apologize my bad english" sure so awesome
22	gape in her asshole! More like this	please!	Thanks qvext for posting this one with
51	and moaned. Thank You, GrooveStrap! Mr.***** , *****@*****.com	Please	post the Topic Link. Preview: [URL], File:
14	the normal download sequence starts, in English!	Please	radekradek1 repost this to a different host,
41	about Female Genital Mutilation (pictures, videos, experiences,...)	Please	send me a message. Also in german."
5	best job in the world for ever	Please	apologize my bad english welcome" Thank You
30	volunteer! Thanks for sharing. I like it.	Please	post more. Thanks for sharing. OUCH! And
36	spasm brought us the best orgasm ever...	please	apologize my bad english... I be love
35	are play toys only is my devise	please	apologize my bad english, thxs." be love
46	[URL], Password: !!WHIPMASTER!! Reup: [URL], Password: !!WHIPMASTER!!	Please	sent me next time when you want
44	do it with your daughters on request."	Please	Reup anybody please more circumcision of girls
1	have anything of this, photo or video	please	post it here! You would make me
37	we want is to compromise our safety.	Please	will posters bear this in mind when
21	it. Thanks for your post markuss, but	please	stop posting non -hurtcore material in this
2	and torture both boy and girls, childrens,	please	apologize my bad english thxs..." so fantastic
52	through the knickers. Chapter 6 to follow	Please	make a preview Preview: [URL], File: [URL],
53	have upload these Video at Screambitch too.	Please	use this link: [URL], Password: REDACTED These
40	don't know thanks? "I can do this,	please	contact me" All what you want and
3	young hot blood for heavy erected cock,	please	apologize my bad english" "so faantastic sound

TABLE B.52: Random concordance lines for keyword *please* in Violence subcorpus.

Index	Left context	Keyword	Right context
33	files (contains all my cp, lolicon and	ontopic	Word documents) in a 15 gigabyte encrypted
64	Yeah this is an unhealthy obsession with	ontopic	safety indeed. I've been obsessed with every
74	of paranoia. Addiction is the emotional attachment.	Ontopic	is my captor. I am addicted (emotionally
75	captor. I am addicted (emotionally attached) to	ontopic	(my Captor) whenever I feel paranoid. This
102	am not going to look at any	ontopic	stuff while traveling aboard. But I don't
127	to have a completely separate computer for	ontopic	and offtopic activities which is probably even
78	exhilarating!" "I have a busy life, so	ontopic	is the only hobby I engage in.
111	still afraid there are traces of my	ontopic	activities on my unprotected HOS (as in
44	witness testimony of somebody claiming you have	ontopic	on your computer, a cop inspecting your
70	I might develop an unhealthy obsession towards?	Ontopic	hobbies have recently made me unhealthy obsessed
80	for anonymity reasons). During the past year,	ontopic	was the only hobby I engaged in.
15	reducing the chances of you accidentally uploading	ontopic	onto offtopic sites (like uploading CP onto
20	question. I've been using Tor exclusively for	ontopic	for such a long time that I
42	buying anti virus software (like everything for	ontopic	activities) with a prepaid credit card not
62	every day life. Except my OCD towards	ontopic	is a bit different from my previous
105	a high anonymity proxy for any clearnet	ontopic	related searches. FYI FBI is given tens
110	files, hiberation files, and etc of their	ontopic	activities on their OS and the only
97	if you are just running all your	ontopic	activities on directly on your Windows/Mac OS
12	As for meeting up with people on	ontopic	clearnet sites, many of those sites are
81	only hobby I engaged in. I think	ontopic	is so addicting not only because of
59	last thing I want is to have	ontopic	stuff in the folders were offtopic stuff
1	VM with TBB Start Tor When browsing	ontopic	sites, I stare at a cop and
17	COMPUTER GAMES which are much safer than	ontopic.	I spend so much time on the
60	the past year I've been obsessed about	ontopic.	I was obsessed with every aspect of
21	so I've been using Tor exclusively for	ontopic.	Tor is an excellent anonymity tool. It
46	(having a hidden os dedicated entirely to	ontopic	and having all traffic in the hidden
101	encryption. I am going to keep all	ontopic	stuff at home in an encrypted external
2	because we all want is to discuss	ontopic	stuff and obtain more fap material. If
89	copy and pasting your offtopic stuff onto	ontopic	sites or vice versa. I only started
95	damaged and becomes inaccessible. I have my	ontopic	files in an encrypted folder on my
10	on other hobbies and less time on	ontopic	hobbies. Which I am doing, and quite
30	windows 7 host os, but all my	ontopic	activites are conducted from a windows xp
8	kind of backup for all of your	ontopic	porn.[] ,You would basically just create an encrypted

TABLE B.53: Random concordance lines for keyword *ontopic* in Paedophile subcorpus.

Index	Left context	Keyword	Right context
82	weather it has the same issues as	TOR,	from bouncing traffic through node to node,
28	to tell if an app is using	tor.	DownThemAll (dTa) uses Tor because whenever I
1	see cp, and now I'm here on	tor."	"One thing I like about PSC is
59	all modern web browsers (Notwithstanding browsers like	Tor.)	With all other OOP / Scripting languages,
55	safe/why isn't it safe to use the	Tor	browser to be looking at Tor things
29	app is using tor. DownThemAll (dTa) uses	Tor	because whenever I use it, the url
195	can see everything I am doing on	Tor.	I asked various security experts and computer
361	my laptop and then I started using	Tor	for surfing Hidden Wiki and the deep
111	out with p2p. I went straight to	Tor	and onionland (i was already using Tor
162	I think the best way to promote	tor	is doing it anonymously over the clearnet.
72	my head at the moment." "Sandboxie: [URL],	(Tor	safe?). Does anyone know if it's any
298	on the Internet will become impossible. Since	Tor	is one of the few anonymous corners
138	who used Tor didn't get busted because	Tor	is damn near impossible to trace. So
142	and paste an onion site's link onto	Tor	browser, but instead I almost end up
58	and stop the shame. Both work with	tor.	Click the cases link in consenting juveniles.
148	I've been using Tor exclusively for ontopic.	Tor	is an excellent anonymity tool. It is
5	thing, maybe by like running everything through	tor	similar to tails set up. Oh and
320	a darknet image host?" "Noticed a few	Tor	sites require Cookies for log ins.[],Didn't log
155	automatically"" on Content <Options<Tools for	tor.	[URL], ""So when these old pictures show
160	offtopic tor users the less bad reputation	tor	gets, so the safer ontopic tor users
137	with their real ip. People who used	Tor	didn't get busted because Tor is damn
305	some time) How does it hurt the	TOR	network?" "So can anyone of you guys
299	would gladly support the government's censorship of	Tor	to fight CP. Too bad they get
304	04 April, 2013, 07:31:02 There are other	tor	sites like 4pedo or ff&ff which are
409	use Tails with noscript enabled or the	TOR	browser through the TOR browser bundle. Use
315	close my Virtual Machine and reopen the	Tor	Browser VM image while tor browser is
128	I'll stick exclusively to hidden services on	tor	browser. IIRC Vampiric from Forbidden Fruits always
364	their real ip. People who only use	tor	and other anonymising networks are over 9000
25	rar file) NOTE: External applications are NOT	Tor	safe by default and can unmask you!
74	it, and so thanks!" "I've read that	Tor	does not support UDP, or TCMP or
362	can find me since I only use	Tor	for CP. I never browse for cp
313	me at an unnecessary risk while using	Tor.	I don't want to spend \$100 plus
83	operates in a very similar way to	TOR,	and all traffic is encrypted, so its

TABLE B.54: Random concordance lines for keyword *tor* in Paedophile subcorpus.

Index	Left context	Keyword	Right context
89	all since I can't create my Decoy	OS	yet. I've tried using the Hidden OS
63	only reason I mistakenly created a hidden	OS	was because I thought I had to
58	do everything ontopic exclusively from your Hidden	OS!	Remember if you think any of us
5	with a VM and a hidden cp	os?	also, if you run your first (home)
53	proof set up possible (having a hidden	os	dedicated entirely to ontopic and having all
1	try tails and to clean up the	OS	and HD I used to run tor
54	and having all traffic in the hidden	OS	run inside a Tails virtual machine)etc.
73	to transfer offtopic stuff from my hidden	OS	to one of my friend's offtopic usb
12	of it? (Or is Tails a hidden	OS?)	Then having TrueCrypt on my drives. Then
23	Virtual Box for hosting the XP guest	os.	I have another windows XP virtual machine
87	on Windows as well as any other	OS.	[],Except windows is virus prone,[],VM solves that.
66	gb. If I didn't use the hidden/decoy	OS	set up and just stuck with full-disk
83	ontopic activities on directly on your Windows/Mac	OS	(which is definitely not recommended). I don't
128	for the Hidden OS and the original	OS	is deleted. I would suggest creating a
157	because I've viewed CP on my host	OS	without a virtual machine before, so god
11	good and secure setup? Having a hidden	OS	within a VM on Win7? Tails on
119	is why I must have the Decoy	OS	successfully installed. I hope you guys had
2	it possible to make tails a ""permanent	os""	for your hidden partitian and if so
111	Hidden OS uses for creating the Decoy	OS.	I need to urgently accomplish this because
99	have regained access to my original Hidden	OS	and then used my back up copy
84	this cannot be done on the hidden	OS	(to my knowledge). Anyone know how to
50	on the erased partition (where my hidden	OS	was cloned from) and then encrypt that
139	password to my one and only encrypted	OS.	An anonymous individual instructed that I should
88	am wrong." "I currently have my Hidden	OS	properly installed. The only problem is it
121	but by running everything inside a Hidden	OS,	then you don't have to worry about
127	drive on your HDD for the Hidden	OS	and the original OS is deleted. I
19	Virtual Box for hosting the XP guest	os.	I have another windows XP virtual machine
75	copy and paste files from my hidden	OS	partition to their usb flash drive (which
62	To be honest I think a hidden	os	is kinda pointless for me even though
104	(HOS) is possible since the original Hidden	OS	has already been deleted. So in order
112	etc of their ontopic activities on their	OS	and the only way I think I
138	order to get access to the Hidden	OS,	but now I am stuck with only
21	22:53:12 I have a windows 7 host	os,	but all my ontopic activites are conducted

TABLE B.55: Random concordance lines for keyword *tor* in Paedophile subcorpus.

Index	Left context	Keyword	Right context
67	this case, an OS), is a private,	hidden,	encrypted container that the LEA cannot prove
108	is easy, but i cannot remove the	hidden	OS's bootloader. I might be going to
32	where these settings are located?" "Quote from:	Hidden	Soul on 24 February, 2013, 15:38:19 It
1	kid at the kids house, especially with	hidden	nanny cams and all that. Hate to
14	already was. If you can see some	hidden	logic in her acts that I just
8	I prepared myself for her in a	hidden	spot where I knew she would see
183	properly (i.e. you must understand how a	hidden	volume works before using one). You also
27	safe, wouldn't it?" "Thanks for your reply,	Hidden	Soul. Quote from: Hidden Soul on 22
2	is the best operating system for the	hidden	partition on an encrypted drive? You know
7	do to legitimize their love? They stayed	hidden	for a while and then started doing
129	suggest encrypting your entire HDD using the	Hidden	OS method. Using Tails to access ontopic
41	dwelling /b/rothers (Anon) who introduced me to	Hidden	Wiki." Everyone who thinks kids are not
6	hidden os with a VM and a	hidden	cp os? also, if you run your
110	knowledge). Anyone know how to remove the	hidden	OS's bootloader so that I can bring
124	deniability. I don't think decrypting my current	Hidden	OS and then trying to regain access
80	the read and write operations on the	hidden	OS are actually done on the C
28	for your reply, Hidden Soul. Quote from:	Hidden	Soul on 22 October, 2012, 01:52:49 I
139	safe I would still use create a	Hidden	OS. Read about it Hidden OS and
3	make tails a ""permanent os"" for your	hidden	partitian and if so when upgrades come
31	stuff when you set up a new	hidden	os on your new hard drive." "I
63	address known. But that would be the	hidden	service that is comprimised, not Tor itself.
34	reason. In fact that particular link for	Hidden	Wiki never worked for me. The OP's
71	passwords. Let alone the password to your	hidden	os. tl;dr Only reveal your password to
111	the encryption on this laptop (decoy and	hidden	OS) so that i can bring it
123	could have regained access to my original	Hidden	OS and then used my back up
136	is deleted. I would suggest creating a	Hidden	Operating system rather than just the standard
181	of fighting child ""abuse"" images." "Quote from:	Hidden	Soul on 14 February, 2013, 14:47:06 truecrypt
60	clearnet sites, or[],B. You connect to a	hidden	service who's entry node is also comprimised
172	operates on the clearnet and not a	""hidden""	darknet work like tor? Why is a
79	A quick question about hidden OS. My	Hidden	system partition is on he C drive
147	instructed that I should go through the	Hidden	OS procedures from scratch (which is very
174	find cp on non-tor sites." "Quote from:	Hidden	Soul on 21 March, 2013, 01:32:15 The
145	computer and now the Bootloader for the	Hidden	OS is gone. I think I shouldn't

TABLE B.56: Random concordance lines for keyword *hidden* in Paedophile subcorpus.

Index	Left context	Keyword	Right context
51	Like Tor's developers." "hahaha not afraid of	Anonymous	at all. I use to be a
7	crimes if you reveal personal info. Or	Anonymous	starts cyberbullying you to the point where
9	or ""My age is Anonymous"" Yeah, answer	""Anonymous""	as your answer to any question designed
6	completely anonymous strangers. We are all completely	anonymous	strangers. So you mustn't be friends with
48	Most n00bs think Tor magically makes them	anonymous.	WRONG. Tor only hides your ip address.
20	She also got my email (which was	anonymous).	I then got this horrible and threatening
10	as CP. I am being framed by	anonymous	hacker! I encrypted my computer since a
56	am so mad-skilled at computers and being	anonymous	on the internet!! Catch me if you
64	to anyone so that I am consistently	anonymous.	I think the paranoia aka irrational fear
22	a couple of other posts- Sexual Addicts	Anonymous	is a good place to get help.
47	sense rules to safety." Long-time member of	Anonymous.	Been using Tor since 2006 for various
66	that I am so use to being	anonymous	on the clearnet SHOULD HAVE made the
50	Nonetheless, good enough that there are likeminded	anonymous	individuals." "Downloads are too slow. It is
34	any personal info is dangerous. Posting as	anonymous	all the time is the safest. With
46	encrypted). The protocol is designed to ensure	anonymous	communication between the hidden server and the
59	be friends with you guys. I am	Anonymous	friends with you guys. Me and other
26	the day was going to Sexual Addicts	Anonymous	(SAA) meetings. It was really quite liberating
44	at me bro and OTHER members of	Anonymous	using Tor for CP. CP will forever
58	truecrypt and whatnot) and how to stay	anonymous	as possible. In fact, as the old
31	use FDE." probably getting ddos attacked by	anonymous	moralfags. or maybe you are using an
32	pedophilia thoughts on Tor, torchat and other	anonymous	services. Not clearnet unless it is routed
61	it doesn't tell users to use an	anonymous	or fake email address. I suspect some
45	at any given moment. Also due to	Anonymous	DDos'ing the site 24/7. The site is
52	I use to be a part of	""Anonymous"".	Technically I still I'm because I'm always
30	other explanation? Indeed. Good modus operandi. Stay	anonymous	at all times and never reveal information
60	info, you are more or less completely	anonymous	on teh internetz lol. Hope this helps.
17	when I attended SAA meetings (Sex Addicts	Anonymous).	My sexual issues were so much deeper
16	you know. I have told people at	anonymous	Sex Addicts meetings and they always accepted
54	which is fancy way of saying, ""Stay	anonymous	at ALL times and NEVER reveal information
25	CP. Instead, I am attending Sexual Addicts	Anonymous	meetings and trying to get control over
39	on facebook. Anyone who's trying to stay	anonymous	on the and yet has a facebook
33	[URL], she has already been identified by	Anonymous.	I wonder if the LEA will reward
49	about my obsession with cp, pedophilia, anonymity,	Anonymous	and other onionland related stuff. They think

TABLE B.57: Random concordance lines for keyword *anonymous* in Paedophile sub-corpus.

Index	Left context	Keyword	Right context
318	and provide the spirit we need as	pedo	in this world. Here we are all
153	its true. I don't know if the	pedo	side of me was subconsciously repressed or
52	that their may be a longer term	pedo	play opportunity. If I really got to
99	guess that means I might be a	pedo-	haha). I've really missed the community. I
162	noted the dangers of being with another	pedo	IRL, where you push each other into
100	share them with good pedo's and other	pedo	families. Pedo families could meet and have
90	world countries who we can introduce to	pedo	love." "I think there will always be
163	of them around. Luckily I'm a 'good	pedo'.	But I'll tell nobody." Society won't get
176	of this thread, since becoming an active	pedo	dad I have been bugged by the
75	marry a woman pedo and raise and	""pedo	family"". In this fantasy, I would turn
146	So not only am I a nasty	pedo,	I'm the worst kind of pedo. Or
102	have not been accused of being a	pedo	by a stranger, however, I certainly have
150	topic too. [URL]," "Hi Long term male	pedo	here, I love all ages and both
18	on the situation may also vilanize the	pedo	when someone finds out or they share
58	listing, do not reveal your interest in	pedo	until you feel it's very safe. First
119	writer could ever imagine. I believe in	pedo	radar as a possibility, and I believe
9	bed. Yeah a babies body covered in	pedo	cum is such a huge turn on
130	also for us, adults). If you're a	pedo	family, I'd say before about the age
118	consciousness that physics will one day discover.	Pedo	radar is definitely possible from a physics
113	impossible by definition to be a chaste	pedo,	since just by being a pedo, your
23	even imagine there are like 1,000 real	pedo	sites but what do I know, I
80	have eventually hit pay dirt with a	pedo	woman or community. Like many in this
248	I haven't talked about Tor or anything	pedo	related to him. I hope he forgets
268	February, 2013, 06:15:15 Thanks for this post,	Pedo.	Some people are not willing or able
83	their odds of connecting with a woman	pedo	if they lived in a clothing optional
74	back in time and marry a woman	pedo	and raise and ""pedo family"". In this
260	sure she didn't suspect I was a	pedo	lol) but now I seriously regret I
275	posting, browsing, and doing other pedo OL	pedo	activities that it is interfering with my
234	fewer, users of those ""evil"" CP sites."	"Pedo	hate mail is all over the clearnet.
79	find a woman who was also a	pedo	who loved and accepted me. I'd love
243	a rat's ass about this whole Western	pedo	hysteria. So they aren't affected by this
251	IP address[],[URL], I connect to these .onion	pedo	sites with at least one extra proxy
111	group of people that I was a	pedo	and attracted to children. It freed me

TABLE B.58: Random concordance lines for keyword *pedo* in Paedophile subcorpus.

Index	Left context	Keyword	Right context
56	not kissing ass either lol" "Meeting another	pedophile	interested in young boys like me would
175	I said that I was not a	pedophile,	just a sevineophile. Because of it was
180	sleep. The fact that I was a	pedophile	came to me at a young age.
186	should have mentioned that when I say	pedophile,	I am referring to the true definition
188	however, I was never approached by a	pedophile.	I had friends who, after the fact,
187	of everyone in the world is a	pedophile	(an article on wikipedia states that men
110	and accepting that I would be a	pedophile	for the rest of my life. It
54	and still living life as an exclusive	pedophile,	I really don't have that many friends
10	I realized that my life as a	pedophile	will never be the same because I
6	amongst our own kind. Others in the	pedophile	community think that we are below them
3	daycare center and the husband is a	pedophile	and I know because we have shared
184	act upon your urges to be a	pedophile,	if you are attracted children then by
12	and parents. I mean after all the	pedophile	is the ""monster"". Though this does remind
19	him and tell him I am a	Pedophile	as well and that if he has
126	into their faces, chests, or genitalia! Our	pedophile	guests come to our amusement parks not
67	day, I was also the kind of	pedophile	that might have tried to have fun
99	apologize. First, said the idea of a	pedophile	militia is ridiculous. To which slipperyboy responded
2	are active with sexually know what a	pedophile	is and the definition and all that.....pretty
66	a major role in me becoming a	pedophile.	Back in the day, I was also
79	seeing that your a member of a	Pedophile	support forum, but I feel it's important
176	a Washington DC escort service. Republican racist	pedophile	and United States Senator Strom Thurmond had
183	It is not illegal to be a	pedophile.	You could meet up with millions of
49	the entire world that I was a	Pedophile	that pretty much cemented the fact that
21	I work knows that I am a	Pedophile	and I have no friends at work
60	around telling all your friends you are	pedophile,	you look at child pornography, and you
189	You also ask me as a ""new""	pedophile	if I am attracted to other young
179	Fleischman of Brown County, now facing felony	pedophile	charges in Wisconsin: -[URL], The criminal complaint
125	as a teen pretty much solidified my	pedophile	identity, namely that of a non-exclusive pedo
116	a pedophile man, but I'm not a	pedophile	myself. He's tried explaining it to me.
102	shouting match so I'm locking it. "A	pedophile	ring could be anything an anti could
16	when I realized that I was a	Pedophile.	I think kids like to take pictures
57	use to hate that I was a	Pedophile	and I was like why the hell
101	the time he recognized me as a	pedophile	and didn't see that as something bad,

TABLE B.59: Random concordance lines for keyword *pedophile* in Paedophile sub-corpus.

Index	Left context	Keyword	Right context
3	that its defined by. I think all	pedophiles	are selfish that are active with kids.
67	and I am sure that there are	pedophiles	amongst even the law enforcement agencies that
19	and the other Tor developers are actually	pedophiles.	One of my pedo friends told me
16	lots of people do not understand what	pedophiles	go through or how we feel. When
8	if they really wanted to and since	Pedophiles	are the scum of the earth to
64	how the experts of Pedophilia are not	Pedophiles	themselves. I would think that we know
29	that it's wrong to label people as	pedophiles,	for reasons I stated above." "I was
89	at least attracted to adults unlike some	pedophiles.	I am only attracted to little, prepubescent
130	for sure. I wouldn't include children as	pedophiles,	maybe potential but probably just not include
91	OP's question would definitely be a NO.	Pedophiles,	those who are primarily erotically and emotionally
59	money is raised—is the Starlet Stage, where	pedophiles	can pay \$100 per sex session to
21	is a software designed by pedophiles for	pedophiles	to hide ip addresses from FBI agents.
96	was always interested in chatting with other	pedophiles,	but I hesitated to do so because
88	trolls or LEA. Trying to incite real	pedophiles	to snoop down to their depravity. Then
97	using that announcement to bait more unwary	pedophiles	onto Tor or giving us the false
132	say 1% of those worlds population of	pedophiles	is in the high class or 3,528,592.
86	up accidentally starting flamewars. Spiteful flames against	pedophiles	are a dime in a dozen on
46	an anti could think of as regarding	pedophiles.	The ring a pedophile gives to his
39	molesting children while virtue signalling ""not all	pedophiles""	help me or anyone else on this
93	would be funny if they were actually	pedophiles	in disguise lolololol. Pretending to be antis
62	that matter so how the hell are	pedophiles	ever going to be accepted into this
94	he asked this: """"Do you think what	pedophiles	do to kids is wrong? Do you
72	favorites." "I don't think they can detect	pedophiles	or actually know what a pedophile really
33	state. Is this a common theme among	pedophiles?	How do others come to terms with
4	we are judged enough for simply being	pedophiles	and if I came out as a
57	with it. I also don't feel like	pedophiles	NEED CP or direct contact to be
124	want to get off. Those are not	pedophiles.	Nonetheless, I think that pedophiles will always
70	stand out like a sore thumb) All	Pedophiles	are Child Molesters just like all Child
100	convenient scapegoat. Jews, homosexuals, gypsies, and now	pedophiles.	I don't see how America can even
78	think its hilarious when pedophiles bash other	pedophiles	because that just makes it obvious that
36	grows into adulthood. I believe that as	pedophiles	and hebophiles we have a responsibility to
35	same labeling and shaming by society that	pedophiles	recieve. Doing my research has helped me
61	and most likely more sane than non	pedophiles.	I am all for wanting to be

TABLE B.60: Random concordance lines for keyword *pedophiles* in Paedophile sub-corpus.

Index	Left context	Keyword	Right context
115	up. They are one of the reasons	pedos	have a bad reputation. I'll have to
39	families (and they don't have to be	pedos	themselves- just believe parent-child sex as good)
31	of pro-pedo girls, cam girls that accept	pedos,	or anything of that nature? I really,
46	tell ourselves we're pedo, we act like	pedos	so it's often difficult or impossible to
10	even that has some risk though and	pedos	I find in general are very tight
62	care about the well-being of my fellow	pedos,	and it pains me to see any
30	in a non sexual way... these ""non	pedos""	are attracted to the same girls I
15	the positive aspects of pedophilia, where real	pedos	can tell their stories, not child rapers,
23	according to certain medical definitions, 100% of	pedos	are, since it's part of the criteria
22	former FBI busted for CP. Proving many	pedos	become undercover cops. Besides looking at cp
14	posting on LC to get the other	pedos	on board. we will need STRICT moderation
59	teens (due to the overwhelming amount of	pedos)"	"Well firstly, chatroulette is basically out of
42	I think we are, 1st and foremost,	pedos.	And humans are, 1st and foremost, animals.
17	CP because of their ""good"" intentions. Catching	pedos.	Sometimes this leeway is given to civilians
101	there. I bet you most antis are	pedos	deep down inside. You ever heard of
6	as well as excoriate and ostracise abusive	pedos	and selfish grooming pedos." "I think I
58	trying them anymore. There's waaay too many	pedos	flooding them, and too few teens (due
98	Pretty funny site. Yep, claims that only	pedos	use truecrypt. lolwut? Would you be using
113	You could meet up with millions of	pedos	in times square with no repercussion's as
102	Self-loathing indeed. Most antis are probably self-loathing	pedos	I bet. I am just not sure
99	my pedophilia, but hanging out with other	pedos	over Tor changed that. Especially my fellow
91	shared itetc. I guess these SMART	pedos	had their drives fully encrypted, used Tor,
76	my previous post. The IWF treats all	pedos	as the ultimate evil. Their bitter stereotypes
97	start using Tor for CP. The more	pedos	move from clearnet p2p to Tor, the
74	cybercriminals. Blackhats are the most skilled. Hence	pedos	are caught more often and in greater
84	sense, but I think the majority of	pedos	get busted because of their lack of
111	sex with kids. What about us female	pedos?	We like to cum with children too.
108	LOVE to hear from some fellow female	pedos	here who may have had similar experiences.
83	a shit about cp, aren't paranoid of	pedos,	and etc. I am sorry for making
43	show; it might even be possible they're	pedos	themselves. If that's the case, they can't
7	and ostracise abusive pedos and selfish grooming	pedos."	"I think I was looking for cartoon
81	as evil. Thankfully truly disgusting cp and	pedos	are the minority. And we should unite
8	way but because a large number of	pedos	are treating children the way most adults,

TABLE B.61 : Random concordance lines for keyword *pedos* in Paedophile subcorpus.

Index	Left context	Keyword	Right context
222	are born with a predisposition towards being	attracted	to kids. Very much like how some
104	It's not necessarily his status that she's	attracted	to, but the attraction is the result
112	believe me the teenaged girls are immensely	attracted	to a powerful presence. People will still
63	I this way and why am I	attracted	to kids only. I realized I was
1	being the max age that I am	attracted	to. "I have just started to get
46	get me very excited" I am mostly	attracted	to boys and they never seem to
101	biologically programmed to seek protection and be	attracted	to men who can provide resources. This
11	for some odd reason and that being	attracted	to babies and or wanting to have
64	to kids only. I realized I was	attracted	to children when I was in High
97	non sexual way... these ""non pedos"" are	attracted	to the same girls I am, but
96	with other people. How many men are	attracted	to men with boobs and wearing skirts?
28	girl she will grow up and be	attracted	to other girls because a man had
18	aware of my Pedophilia I was mainly	attracted	to young boys because they were easier
26	some of the others am not exclusively	attracted	to children in fact I would say
2	cock get so hard if I wasn't	attracted	to them. The cock wants what the
16	my own children so I am more	attracted	to my friends kids and family members
24	older teens I'm queer (meaning I'm sexually	attracted	to what I'm sexually attracted to whether
120	to convince myself that I could be	attracted	to women, only to find myself infatuated
95	something with you my friend. They're all	attracted	to people they shouldn't be. They feel
241	cock?? I agree, not the least bit	attracted	to boys/men but definitely open to experimenting
245	me if I would have been sexually	attracted	to a son if we had one
233	to be a pedophile, if you are	attracted	children then by my definition you are
235	them subconsciously you know that you are	attracted	to them. I know the number of
89	the matter is that if you're sexually	attracted	to little girls, you are not like
84	always someone amazed that I was not	attracted	to my own kids. I am not
220	the history of it?" "I always was	attracted	to the mainly of adults ages with
116	something with you my friend. They're all	attracted	to people they shouldn't be. They feel
158	without feeling any guilt about being sexually	attracted	to them. Of course our whole lifestyle
127	your identity on this site. "Trump being	attracted	to 13 and 14 year olds does
94	remnant of older biological imperatives) and men	attracted	to girls who are clearly nowhere near
242	one label because I am still sexually	attracted	to women of all ages. One of
123	want to believe that I would be	attracted	to Elizabeth. I can imagine how this
229	Early Pubescent Children (Tanner Stages 2-3)Pedohebephilic Typeâ€”Sexually	Attracted	to Both Specify type: Sexually Attracted to

TABLE B.62: Random concordance lines for keyword *attracted* in Paedophile subcorpus.

Index	Left context	Keyword	Right context
1	The crazy thing is that police and	LEA	can trace those apps and can easily
127	anti virus software could have sent the	LEA	malware to their database. Don't want anti
29	publicly discuss these weaknesses. This is helping	LEA	us. Keep these highly sensitive discussions
13	not to say you should ever underestimate	LEA.	They have plenty of resources. You must
26	more questions out of you. Then the	LEA	can brainwash you into believing that they
52	my own ISP at home. Then if	LEA	got your PC, they would keep your
9	paranoid is the best way to keep	LEA	away. Paranoia is unhealthy." "Ironically most of
30	Don't want this info publicly available to	LEA.	It is helping LEA CENSOR our free
18	your thoughts? Correct. Also remember this. LEA,	LEA	everywhere!! Be as paranoid as you can
144	microsoft and any software developers who help	LEA	oppress and surveillance the Internet, FUCK YOU
149	or even your computer get searched by	LEA:	however keep using encryption just in case.
159	I am crossing my fingers that the	LEA	aren't going to raid me for that
85	the special tricks I use to avoid	LEA	in real life (to prevent LEA from
71	I would also add don't fall to	LEA	FUD. "I use to do that too.
204	are just unhealthy and paranoid thinking that	LEA	can trace them through Tor or some
125	googling without Tor protecting their IP. Therefore	LEA	already know they were searching for cp
31	publicly available to LEA. It is helping	LEA	CENSOR our free speech. My 2 cents.
61	of people here that could be from	LEA	but hide it well because they might
131	the key for your private container. The	LEA	cannot determine if you have a private,
68	the Internet. IRL I actually chat with	LEA	whenever possible. I gotta get more comfortable
130	same time. So, you can give the	LEA	the key that opens up your public
32	Don't want this info publicly available to	LEA.	It is helping LEA CENSOR our free
40	because I am not constantly worried about	LEA	showing up on my doorstep looking for
76	agree with the laws they enforce. But	LEA	hunt us. And the hunted always wanna
38	it is possible that the person is	LEA	or a vigilante. I'm not sure how
17	are your thoughts? Correct. Also remember this.	LEA,	LEA everywhere!! Be as paranoid as you
116	safe, is probably another feature used by	LEA	to expose your real IP, so I
191	10-15 minutes at a time may allow	LEA	to figure out my real ip. I've
179	Doing this because I'm afraid google and	lea	are monitoring ALL SEARCH results. 1984 Big
36	the point where it is impossible for	LEA	to find me. Also I must be
24	reasonable) ""i am 29 years old"". Then	LEA	asks ""what country to do you live
11	bully. Anon doxed some deep web pedos.	LEA	did n't. Anon hacked LEA and even
41	by months of hard work by dedicated	LEA	teams, and 2. Fucking idiots. I swear,

TABLE B.63: Random concordance lines for keyword *LEA* in Paedophile subcorpus.

Index	Left context	Keyword	Right context
85	posted as anonymous rather than pseudonyms. 4)	Cops	will 100% see public posts. When pm'd
64	address of the uploader or downloader. Therefore	cops	cannot trace the origins of the downloads
63	January, 2013, 22:25:41 Full disk encryption frustrates	cops.	Especially when you have a hidden OS
119	to get material. It's just funny that	cops,	antis and journalists seem better informed than
99	youtube video linked from onionland where the	cops	will capturing your real ip address when
72	not worth it. Why would you want	cops	to raid your house and accumulate thousands
91	it doesn't allow any CP. Otherwise, the	cops	could take it down easily. Maybe it
118	in the HCandy portal are infiltrated by	cops,	don't be naive boys." "How far do
120	seen as exploited kids, blah blah blah.	Cops	have their black lists, no need to
108	I think there is a ZERO chance	cops	can find me based on writing style.
115	This Lolita City Thread says otherwise, Can	cops	upload cp? [URL], [be careful. May contain
76	cops arrest him, he said to the	cops	""I looked at cp on the bus
78	in must give up passwords to the	cops	(yeah I live in North America). The
88	was mentioned on any onion site, the	cops	see that search being made on a
81	assume every ontopic site is run by	cops.	No exceptions. 4,, Ask for security advice
70	a game and therefore toying around with	cops	(giving out hints to make it more
67	fair in the love of war. The	cops	could be using that announcement to bait
83	guard. By the time you realize the	cops	where behind it the whole time, they
65	they would probably turn you into the	cops	or something like that." "Exactly. I wish
34	for me. Afraid somebody might call the	cops	on me. Luckily all this happened 2-3
1	those are MUCH more dangerous because the	cops	don't need digital forensics evidence of your
100	then have somebody turn you into the	cops.	With this set up, you can have
23	then having them tip us to the	cops.	If you use Tor, encrypt your entire
47	the site was already run by the	cops.	Or the owner, theMayor, is just batshit
55	definitely not turn you in to the	cops	if they find out that you were
21	also look through my home's windows for	cops.	I am not going to reveal any
40	if you were actually WANTED by the	cops,	there is no need for windows to
48	seamlessly with the fact that you mention	cops	can use traps where they post fake
59	had an article about a flash drive	cops	could use to access locked Windows machines.
61	privacy rights), just don't talk to the	cops	so you don't incriminate yourself. Thus you
50	in the act. As long as the	cops	don't have the intention of trading CP
8	their job to assume you are guilty.	Cops	work with a very simple algorithm either
19	the law yourself, and never talk to	cops!	Remember, cops are thugs. They belong to

TABLE B.64: Random concordance lines for keyword *cops* in Paedophile subcorpus.

Index	Left context	Keyword	Right context
65	am caught again I will never leave	prison	and I am okay with that. I
25	again and I do go back to	prison	for sex acts with young boys then
33	I have wasted a long time in	prison	for my sex crimes against children. I
60	and then I would go back to	prison	for buying such a thing that I
8	with young children even after being in	prison	for such actions. Like if I take
39	done to a child sex offender in	prison.	I was the scum of the earth
83	to read on the thread ""Pedos In	Prison""...	On this thread, I am reading through
66	got caught, but did not go to	prison,	that the outcome is pretty awful. (It
100	daughters and served 14 years in federal	prison.	Republican St. Louis Election Board official Kevin
67	pretty awful. (It goes without saying that	prison	is 1000 times worse). I also had
29	careful and want to stay out of	prison?	A. Not sure" "Everyone that I know,
77	prefers to pay for their time in	prison	than collect their taxes and have them
74	you know, you can get 10 in	prison	because you accidentally have a picture of
13	the time. "When I was first in	prison	I was scared for my life because
96	probably going to get butt raped in	prison.	What if I use Linux within a
23	out for that guy, he was in	prison	for molesting young boys and that is
72	for you... If I was facing the	prison	time that you are, I would be
32	"I have been in and out of	prison	ever since I was a very young
82	was given a choice of going to	prison	for not cooperating with police, or destroying
50	that I will be going back to	prison	and I know that this time I
110	to serve six to 18 months in	prison.*	Republican anti-abortion activist Howard Scott Heldreth is
14	had heard how Pedophiles are treated in	prison	and at first the name calling and
49	and friends. When I got out of	prison	all those friends that I had before
48	the show Dance Moms." I went to	prison	for my sexual exploits with young boys
64	me since I have been out of	prison	and to be honest I am okay
69	toy." "I googled ""how to survive in	prison	as a sex offender"" and it came
59	me and I ended up going to	prison	and then when that happen kids came
89	happens and I deserve to go to	prison	for it. I feel like it really
70	some of the posts. I'm reading the	prison	stories right now. This forum really is
1	pure heaven!" "I have been out of	prison	for awhile and being a lifetime child
103	He was sentenced to 12 years in	prison	for repeated sexual assault of a young
99	is serving a 37-year sentence in federal	prison	for sexually abusing 8- and 10-year old
41	all. I keep risking going back to	prison	but when I really don't see adults

TABLE B.65: Random concordance lines for keyword *prison* in Paedophile subcorpus.

Index	Left context	Keyword	Right context
47	bargain (you could still end up in	jail	for up to 10 years). My solution
49	for him, spam ""FBI virus!!"" ads threatening	jail	sentence unless \$ is paid already beaten
54	Even if I do end up in	jail,	I am not afraid. I am a
61	treatment. CP addicts end up rotting in	jail.	Hopefully my advice helps cp addicts. protip:
14	underage, you will probably be heading to	jail.	3) The odds of finding a pedo
9	over. Spend the next 20 years in	jail	for breaking some silly laws. CP laws
52	Deprived from kids forever. Getting raped in	jail.	The social stigma. Sex offender is worse
12	get. Ageplay will not land you in	jail.	Even expressing that you are attracted to
25	be nice, but wouldn't get you in	jail.	In life, we sometimes have to take
63	want to get caught and thrown in	jail	for the horrors they inflict upon innocent
30	my toys, but he never went to	jail	for it. Again, might be a subject
84	GOP closet pedophiles. He was sentenced to	jail	after soliciting a 14-year-old boy, offering \$20
56	""evil"" drug dealers and send them to	jail	for PROFIT. The war on CP and
55	even fabricate evidence to put you in	jail.	Recall a guy got arrested for uploading
23	I will most likely not go to	jail	and the experience will be more rewarding
71	They want to throw more people into	jail	to make more profit for the jail
4	I hope she gets out before the	jail	collapses around her from age." you are
46	be given a choice of going to	jail	for not cooperating with LEA, and hindering
68	sexual abuse feel, but putting people in	jail	for many years even for that kind
90	the charge be if not given out?	Jail	time for a few years in certian
24	their process towards putting you in a	jail	cell. How many innocent people have been
10	the join-the-FBI test, they'll throw you in	jail.	But there have been many former FBI
74	ridiculous ways to send more people into	jail.	They want to loot the convict's property
43	may or may not already be in	jail	(i.e. Vicky). But there is a tremendous
28	are things people ought to go to	jail	for, however I believe in law and
36	choice*, it's not some kind of a	jail	he's forced into. And if he wants
44	I have a criminal record (not serious,/no	jail	time, but still a record), which actually
88	away from me and put me in	jail	and I knew in my heart that
53	doubt I will ever end up in	jail.	Even if I do end up in
42	for having to continue sentence after one's	jail	release date in the forms, et al,
1	and all that. Hate to be in	jail	the next day after they watch the
6	consent and BAM you are thrown in	jail	for distributing CP. Such madness stems from
51	anonymity Rather an hero than rot in	jail.	Srsly. Wondering. Why are former cp convicts

TABLE B.66: Random concordance lines for keyword *jail* in Paedophile subcorpus.

Appendix C

Study 3: Thematic annotations and concordance samples

This appendix is provided on an encrypted, password-protected USB drive for the examiners only.

This appendix contains the detailed annotations of the top 50 keywords for each of the 5 subcorpora analysed in study 3. The provided files also contain the random concordance line samples for each of the top keywords in the respective subcorpus .

- Thematic annotations for *Childlove* subcorpus

File path: USB > Thesis Appendices > Corpus Results > Concordance Analysis - Childlove.xlsx

- Thematic annotations for *Family* subcorpus

File path: USB > Thesis Appendices > Corpus Results > Concordance Analysis - Family.xlsx

- Thematic annotations for *Names* subcorpus

File path: USB > Thesis Appendices > Corpus Results > Concordance Analysis - Names.xlsx

- Thematic annotations for *Paedophile* subcorpus

File path: USB > Thesis Appendices > Corpus Results > Concordance Analysis - Paedophile.xlsx

- Thematic annotations for *Violence* subcorpus

File path: USB > Thesis Appendices > Corpus Results > Concordance Analysis - Violence.xlsx

Appendix D

Full Data Set: DarkWebCSA

This appendix is provided on an encrypted, password-protected USB drive for the examiners only.

In this appendix, the full data set, both the raw file and the cleaned version, are provided for the examiners.

Raw data

File path: USB > Raw Data

Cleaned data

File path: USB > Thesis Appendices > Full Corpus

Appendix E

Full list of 32,410 usernames and respective post counts

This appendix is provided on an encrypted, password-protected USB drive for the examiners only.

This appendix contains a file with the full list of 32,410 usernames and their respective post counts.

File path: USB > Thesis Appendices > Username Lists > Full List and Counts

Appendix F

Python scripts for data extraction

This appendix is provided on an encrypted, password-protected USB drive for the examiners only.

This Appendix includes the following Python scripts for data extraction:

- Python script: Username extraction

File path: USB > Thesis Appendices > Scripts > username_extraction.py

- Python script: Post text extraction

File path: USB > Thesis Appendices > Scripts > post_text_extraction.py

Appendix G

Full list of 10,222 usernames in reduced set and respective username segments

This appendix is provided on an encrypted, password-protected USB drive for the examiners only.

This appendix includes a file containing the full list of 10,222 usernames in the reduced set and their respective username segments.

File path: USB > Thesis Appendices > Username Lists > Usernames and Segments 10222

Appendix H

Username Lists per Subcorpus

This appendix is provided on an encrypted, password-protected USB drive for the examiners only.

This appendix includes the list of usernames in the 1837 top users corpus as well as the following username lists by category extracted from the full top user corpus of 1837 usernames:

- Username list Family Relations

File path: USB > Thesis Appendices > Username Lists > Subcorpora > family.txt

- Username list Violence

File path: USB > Thesis Appendices > Username Lists > Subcorpora > violence.txt

- Username list Names

File path: USB > Thesis Appendices > Username Lists > Subcorpora > names.txt

- Username list Childlove

File path: USB > Thesis Appendices > Username Lists > Subcorpora > childlove.txt

- Username list Preference:Age

File path: USB > Thesis Appendices > Username Lists > Subcorpora > pref-age.txt

- Username list Preference:Gender

File path: USB > Thesis Appendices > Username Lists > Subcorpora > pref-gender.txt

- Username list Paedophile

File path: USB > Thesis Appendices > Username Lists > Subcorpora > paedophile.txt

Username list of top 1837 users:

File path: USB > Thesis Appendices > Username Lists > Subcorpora > topusers.txt

Appendix I

Top keywords and bigrams by subcorpus

This appendix is provided on an encrypted, password-protected USB drive for the examiners only.

This appendix contains the following files:

- Top bigrams per subcorpus

File path: USB > Thesis Appendices > Corpus Results > Bigrams

- Annotated keywords per subcorpus against reference (top 50)

File path: USB > Thesis Appendices > Corpus Results > Annotated keywords subcorpora

Appendix J

CSA Subcorpora

This appendix is provided on an encrypted, password-protected USB drive for the examiners only.

This appendix includes the following subcorpora:

- Top 1837 Users Corpus (DarkWebCSA-top)

File path: USB > Thesis Appendices > CSA Subcorpora > Full Top Users

- Childlove subcorpus

File path: USB > Thesis Appendices > CSA Subcorpora > Category > Childlove

- Childlove subcorpus, English posts only

File path: USB > Thesis Appendices > CSA Subcorpora > Category > English Only > Childlove-en

- Family Relations subcorpus

File path: USB > Thesis Appendices > CSA Subcorpora > Category > Family

- Family Relations subcorpus, English posts only

File path: USB > Thesis Appendices > CSA Subcorpora > Category > English-Only > Family-en

- Names subcorpus

File path: USB > Thesis Appendices > CSA Subcorpora > Category > Names

- Names subcorpus, English posts only

File path: USB > Thesis Appendices > CSA Subcorpora > Category > English Only > Names-en

- Paedophile subcorpus

File path: USB > Thesis Appendices > CSA Subcorpora > Category > Paedophile

- Paedophile subcorpus, English posts only

File path: USB > Thesis Appendices > CSA Subcorpora > Category > English Only, Paedophile-en

- Preference:age subcorpus

File path: USB > Thesis Appendices > CSA Subcorpora > Category > Preference-Age

- Preference:age subcorpus, English posts only

File path: USB > Thesis Appendices > CSA Subcorpora > Category > English Only > Preference-Age-en

- Preference:gender subcorpus

File path: USB > Thesis Appendices > CSA Subcorpora > Category > Preference-Gender

- Preference:gender subcorpus, English posts only

File path: USB > Thesis Appendices > CSA Subcorpora > Category > English Only > Preference-Gender-en

- Violence subcorpus

File path: USB > Thesis Appendices > CSA Subcorpora > Category > Violence

- Violence subcorpus, English posts only

File path: USB > Thesis Appendices > CSA Subcorpora > Category > English Only > Violence-en

Bibliography

- Aldridge, Judith and David Decary-Hetu (2015). "Sifting through the net: Monitoring of online offenders by researchers". In: *European Review of Organised Crime* 2, pp. 122–41.
- Aldrin, Emilia (2014). "Choosing a Name = Choosing Identity?: Towards a Theoretical Framework". In: *XXIV ICOS 2011, International Congress of Onomastic Sciences, Barcelona, Spain, September 5–9, 2011*. Generalitat de Catalunya, pp. 392–401.
- Aleksiejuk, Katarzyna (2016). "Internet Personal Naming Practices and Trends in Scholarly Approaches". In: *Names and Naming. Multilingual Matters*, pp. 3–17.
- Aleksiejuk, Katarzyna (2017). *Names on the internet: towards electronic socio-onom@stics*. University of Edinburgh.
- Alzahrani, Rajab Gamaan Muhsen (2013). *A Corpus-Based Critical Discourse Analysis of the Ideological Representations and Legitimation in the Salafi Discourse in Saudi Arabia (1980-2000)*. Lancaster University (United Kingdom).
- Angouri, Jo (July 2015). "Online communities and communities of practice". In: *The Routledge handbook of Language and Digital Communication*, pp. 323–328.
- Arguello, Jaime et al. (2006). "Talk to me: foundations for successful individual-group interactions in online communities". In: *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, pp. 959–968.
- Babaii, Esmat and Maliheh Rashed (2017). "Representation of social actors in Chinua Achebe's novel "Things Fall Apart" and its two Persian translations". In: *Translation & Interpreting* 9.2, pp. 151–165.

- Baker, Paul (2006). *Using corpora in discourse analysis*. A&C Black.
- Baker, Paul et al. (2008). "A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press". In: *Discourse & Society* 19.3, pp. 273–306.
- Barany, Amanda and Aroutis Foster (2024). "Context, community, and the individual: Modeling identity in a game affinity space". In: *Virtual Learning Environments*. Routledge, pp. 85–102.
- Barber, Kate (2021). "9. From the Outside Looking in: The Risks and Challenges of Analysing Extremist Discourses on Far-right and Manosphere Websites". In: *Vulnerabilities, Challenges and Risks in Applied Linguistics*. Multilingual Matters, 131–146.
- Barker, Meghanne and Michael Prentice (2017). *Intertextuality and interdiscursivity*. Oxford University Press.
- Baxter, Judith and Kieran Wallace (2009). "Outside in-group and out-group identities? Constructing male solidarity and female exclusion in UK builders' talk". In: *Discourse & Society* 20.4, pp. 411–429.
- Bechar-Israeli, Haya (1995). "From <Bonehead> to <cLoNehEAd>: Nicknames, play, and identity on Internet Relay Chat". In: *Journal of Computer-Mediated Communication* 1.2.
- Beier, Klaus M et al. (2015). "The German Dunkelfeld Project: A pilot study to prevent child sexual abuse and the use of child abusive images". In: *The journal of sexual medicine* 12.2, pp. 529–542.
- Benjamin, Victor and Hsinchun Chen (2012). "Securing cyberspace: Identifying key actors in hacker communities". In: *Intelligence and Security Informatics (ISI), 2012 IEEE International Conference on*. IEEE, pp. 24–29.
- Beuchot, Alberto and Mark Bullen (2005). "Interaction and interpersonality in online discussion forums". In: *Distance Education* 26.1, pp. 67–87.

- Booth, Amy (2023). "Fractured In-Group Identity (Re) Negotiation in an Online White Nationalist Forum". In: *Applied Corpus Linguistics* 3.3, 100062.
- Bowman-Grieve, Lorraine (2013). "A psychological perspective on virtual communities supporting terrorist & extremist ideologies as a tool for recruitment". In: *Security Informatics* 2.1, p. 9.
- Braun, Virginia and Victoria Clarke (2006). "Using thematic analysis in psychology". In: *Qualitative research in psychology* 3.2, pp. 77–101.
- Brezina Vaclav, Weill-Tessier Pierre and Anthony McEnery (2020). *#LancsBox v. 5.x. [software]*. Available at: <http://corpora.lancs.ac.uk/lancsbox>.
- Brindle, Andrew (2016). *The language of hate: A corpus linguistic analysis of white supremacist language*. Routledge.
- Brookes, Gavin and Tony McEnery (2020). "Corpus linguistics". In: *The Routledge handbook of English language and digital humanities*. Routledge, pp. 378–404.
- Bucholtz, Mary and Kira Hall (2005). "Identity and interaction: A sociocultural linguistic approach". In: *Discourse studies* 7.4-5, pp. 585–614.
- Burns, Carolyn M et al. (2008). "The emotional impact on and coping strategies employed by police teams investigating internet child exploitation". In: *Traumatology* 14.2, pp. 20–31.
- Carter, Pelham et al. (2021). "Comparing manual and computational approaches to theme identification in online forums: A case study of a sex work special interest community". In: *Methods in Psychology*.
- Chiang, Emily (2019). "Rhetorical moves and identity performance in online child sexual abuse interactions". PhD thesis. Aston University.
- Chiang, Emily (2024). "'I read the rules and know what is expected of me': The performance of competence and expertise in 'newbie' offenders' membership requests to dark web child abuse communities". In: *Discourse, Context & Media* 57, p. 100744.

- Chiang, Emily and Tim Grant (2017). "Online grooming: moves and strategies". In: *Language and Law / Linguagem e Direito*.
- Chiang, Emily and Tim Grant (2018). "Deceptive Identity Performance: Offender Moves and Multiple Identities in Online Child Abuse Conversations". In: *Applied Linguistics*.
- Chiang, Emily et al. (2021). "Linguistic analysis of suspected child sexual offenders' interactions in a dark web image exchange chatroom". In: *International Journal of Speech Language and the Law* 27.2, 129–161.
- Choo, Kim-Kwang Raymond and Russell G Smith (2008). "Criminal exploitation of online systems by organised crime groups". In: *Asian journal of criminology* 3.1, pp. 37–59.
- Chua, Cecil Eng Huang, Jonathan Wareham, and Daniel Robey (2007). "The role of online trading communities in managing internet auction fraud". In: *MIS Quarterly*, pp. 759–781.
- Cornetto, Karen M and Kristine L Nowak (2006). "Utilizing usernames for sex categorization in computer-mediated communication: Examining perceptions and accuracy". In: *CyberPsychology & Behavior* 9.4, pp. 377–387.
- Craven, Samantha, Sarah Brown, and Elizabeth Gilchrist (2007). "Current responses to sexual grooming: Implication for prevention". In: *The Howard Journal of Criminal Justice* 46.1, pp. 60–71.
- Danet, Brenda and Susan C Herring (2007). *The multilingual Internet: Language, culture, and communication online*. Oxford University Press.
- Darics, Erika and Veronika Koller (2019). "Social actors "to go": An analytical toolkit to explore agency in business discourse and communication". In: *Business and Professional Communication Quarterly* 82.2, pp. 214–238.
- Di Cristofaro, Matteo (2023). *Corpus approaches to language in social media*. Taylor & Francis.
- Donlan, Lisa (2020). "Constructing authorial pseudonyms and authorial identity in online fanfiction communities". In: *Internet Pragmatics* 3.1, pp. 95–115.

- Elstein, Solly and Krzysztof Kredens (2020). "Occupational stress in forensic linguistic practice." In: *Journal of Applied Linguistics & Professional Practice* 17.1.
- Engström, Robin and Carita Paradis (2015). "The in-group and out-groups of the British National Party and the UK Independence Party: A corpus-based discourse-historical analysis". In: *Journal of Language and Politics* 14.4, pp. 501–527.
- Esser, Arne C. (2021). "How does the language of corpora from radicalized communities discovered on Parler compare to online conversations on Twitter regarding the 2021 Capitol riots and election fraud?" PhD thesis.
- Fairclough, Norman (2013). *Critical discourse analysis: The critical study of language*. Routledge.
- Fairclough, Norman, Jane Mulderrig, and Ruth Wodak (2011). "Critical discourse analysis". In: *Discourse studies: A multidisciplinary introduction*, pp. 357–378.
- Flowerdew, John and John E Richardson (2017). *The Routledge Handbook of Critical Discourse Studies*. Taylor & Francis.
- Frank, Richard, Bryce Westlake, and Martin Bouchard (2010). "The structure and content of online child exploitation networks". In: *ACM SIGKDD Workshop on Intelligence and Security Informatics*. ACM, p. 3.
- Gatson, Sarah N. (2011). "Self-Naming Practices on the Internet: Identity, Authenticity, and Community". In: *Cultural Studies Critical Methodologies* 11.3, 224–235.
- Gee, James Paul (2005). "Semiotic social spaces and affinity spaces". In: *Beyond communities of practice language power and social context*. 214232.
- Gee, James Paul and Elisabeth Hayes (2012). "Nurturing affinity spaces and game-based learning". In: *Games, learning, and society: Learning and meaning in the digital age* 123, pp. 1–40.
- Gillings, Mathew, Gerlinde Mautner, and Paul Baker (2023). *Corpus-assisted discourse studies*. Cambridge University Press.

- Graham, Lindsay T and Samuel D Gosling (2012). "Impressions of World of Warcraft players' personalities based on their usernames: Interobserver consensus but no accuracy". In: *Journal of Research in Personality* 46.5, pp. 599–603.
- Grant, Tim (2012). "TXT 4N6: method, consistency, and distinctiveness in the analysis of SMS text messages". In: *JL & Pol'y* 21, p. 467.
- Grant, Tim (2022). *The idea of progress in forensic authorship analysis*. Cambridge University Press.
- Grant, Tim and Nicci Macleod (2016). "Assuming Identities Online: Experimental Linguistics Applied to the Policing of Online Paedophile Activity". In: *Applied Linguistics* 37.1, 50–70.
- Grant, Tim and Nicci MacLeod (2020). *Language and online identities: The undercover policing of internet sexual crime*. Cambridge University Press.
- Gross, Benjamin M and Elizabeth F Churchill (2007). "Addressing constraints: multiple usernames task spillage and notions of identity". In: *CHI'07 Extended Abstracts on Human Factors in Computing Systems*. ACM, pp. 2393–2398.
- Halliday, Michael AK (1985). *An introduction to functional grammar*. London: Edward Arnold.
- Halliday, Michael Alexander Kirkwood and Christian Matthiessen (2013). *Halliday's introduction to functional grammar*. Routledge.
- Hämäläinen, Lasse (2015). "User names of illegal drug vendors on a darknet cryptomarket". In: *Onoma* 50, 45–71.
- Hämäläinen, Lasse (2022). "From Bonehead to @realDonaldTrump: A Review of Studies on Online Usernames". In: *Names* 70.2, pp. 35–53.
- Hämäläinen, Lasse, Ari Haasio, and J. Tuomas Harviainen (2021). "Usernames on a Finnish online marketplace for illegal drugs". In: *Names* 69.3.
- Hassa, Samira (2012). "Projecting, exposing, revealing self in the digital world: Usernames as a social practice in a Moroccan chatroom". In: *names* 60.4, pp. 201–209.

- Heritage, Frazer and Veronika Koller (2020). "Incels, in-groups, and ideologies: The representation of gendered social actors in a sexuality-based online community". In: *Journal of Language and Sexuality*.
- Herring, Susan C (2007). "A faceted classification scheme for computer-mediated discourse". In: *Language@ internet* 4.1.
- Holt, Thomas J (2010). "Exploring strategies for qualitative criminological and criminal justice inquiry using on-line data". In: *Journal of Criminal Justice Education* 21.4, pp. 466–487.
- Holt, Thomas J, Kristie R Blevins, and Natasha Burkert (2010). "Considering the pedophile subculture online". In: *Sexual Abuse* 22.1, pp. 3–24.
- Holt, Thomas J, Adam M Bossler, and Kathryn C Seigfried-Spellar (2015). *Cybercrime and digital forensics: An introduction*. Routledge.
- Jaki, Sylvia et al. (Nov. 2019). "Online hatred of women in the Incels.me forum: Linguistic analysis and automatic detection". In: *Journal of Language Aggression and Conflict* 7.2, pp. 240–268.
- Janbek, Dana M and Paola Prado (2012). "Rethinking the role of virtual communities in terrorist websites". In: *CTX: Social Media in Jihad and Counterterrorism*. 2(4).
- Jaspal, Rusi (2020). "Content analysis, thematic analysis and discourse analysis". In: *Research methods in psychology* 1, pp. 285–312.
- Jones, Kyle M L et al. (Jan. 2016). "Community of practice or affinity space: A case study of a professional development MOOC". In: *Educ. Inf.* 32.1, pp. 101–119.
- Kersten, Saskia and Netaya Lotze (2020). "Creating a self-image: Face-work and identity construction online". In: *Journal for Media Linguistics*.
- Koller, Veronika (2012). "How to Analyse Collective Identity in Discourse-Textual and Contextual Parameters." In: *Critical Approaches to Discourse Analysis across Disciplines* 5.2.

- Kredens, Krzysztof, Kate Haworth, and Sarah Atkins (2022). "The ethics of forensic language data". In: *Fourth European Conference of the International Association of Forensic and Legal Linguistics*. Aston University.
- Krendel, Alexandra (2020). "The men and women, guys and girls of the 'manosphere': A corpus-assisted discourse approach". In: *Discourse & Society* 31.6, pp. 607–630.
- Krendel, Alexandra, Mark McGlashan, and Veronika Koller (2022). "The representation of gendered social actors across five manosphere communities on Reddit". In: *Corpora*.
- Krone, Tony (2005). *International police operations against online child pornography*. Australian Institute of Criminology.
- Lawson, Robert and Mark McGlashan (2017). "'You need to become prime dick': a corpus-based analysis of self-help discourses and gender constructions in online seduction communities". In: *116th annual American Anthropological Association*.
- Litvinova, Tatiana et al. (2018). "Using Corpus Linguistics Tools to Analyze a Russian-Language Islamic Extremist Forum". In: *Internet Science*. Springer International Publishing, 54–65.
- Lorenzo-Dus, Nuria and Matteo Di Cristofaro (2018). "'I know this whole market is based on the trust you put in me and I don't take that lightly': Trust, community and discourse in crypto-drug markets". In: *Discourse & Communication* 12.6, pp. 608–626.
- Lorenzo-Dus, Nuria, Cristina Izura, and Rocío Pérez-Tattam (2016). "Understanding grooming discourse in computer-mediated environments". In: *Discourse, Context & Media* 12, 40–50.
- Lorenzo-Dus, Nuria and Anina Kinzel (2019). "'So is your mom as cute as you?': Examining patterns of language use in online sexual grooming of children". In: *Journal of corpora and discourse* 2.1, pp. 1–30.
- Lorenzo-Dus, Nuria, Anina Kinzel, and Matteo Di Cristofaro (2020). "The communicative modus operandi of online child sexual groomers: Recurring patterns in their language use". In: *Journal of Pragmatics* 155, pp. 15–27.

- Lumsden, Karen (2019). “‘I want to kill you in front of your children’ is not a threat. It’s an expression of a desire’: Discourses of online abuse, trolling and violence on r/MensRights”. In: *Online Othering*. Cham: Springer International Publishing, 91–115.
- MacLeod, Nicci and Tim Grant (2017). “‘Go on Cam But Dnt Be Dirty’: Linguistic Levels of Identity Assumption in Undercover Online Operations Against Child Sex Abusers”. In: *Language and Law=Linguagem e Direito* 4.2, 157–175.
- MacLeod, Nicci and Tim Grant (2021). “Assuming Identities Online: How Linguistics Is Helping the Policing of Online Grooming and the Distribution of Abusive Images”. In: *Rethinking Cybercrime: Critical Debates*. Ed. by Tim Owen and Jessica Marshall. Cham: Springer International Publishing, 87–104.
- Malesky, L Alvin and Liam Ennis (2004). “Supportive distortions: An analysis of posts on a pedophile Internet message board”. In: *Journal of Addictions & Offender Counseling* 24.2, pp. 92–100.
- Maloney-Krichmar, Diane and Jennifer Preece (2002). “The meaning of an online health community in the lives of its members: Roles, relationships and group dynamics”. In: *IEEE 2002 International Symposium on Technology and Society (ISTAS’02). Social Implications of Information and Communication Technology. Proceedings (Cat. No. 02CH37293)*. IEEE, pp. 20–27.
- Manley, Eric, Heidi Levitt, and Chad Mosher (2007). “Understanding the bear movement in gay male culture: Redefining masculinity”. In: *Journal of Homosexuality* 53.4, pp. 89–112.
- Martellozzo, Elena (2015). “Policing online child sexual abuse-the British experience”. In: *European Journal of policing Studies* 3.1, pp. 32–52.
- Marwick, Alice E and Robyn Caplan (2018). “Drinking male tears: Language, the manosphere, and networked harassment”. In: *Feminist Media Studies* 18.4, pp. 543–559.
- Mihás, Paul (2023). “Qualitative research methods: approaches to qualitative data analysis”. In: *International encyclopedia of education*, 302–313.

- Mitchell, Kimberly J, Janis Wolak, and David Finkelhor (2005). "Police posing as juveniles online to catch sex offenders: Is it working?" In: *Sexual Abuse: A Journal of Research and Treatment* 17.3, pp. 241–267.
- Moskowitz, David A et al. (2013). "Physical, behavioral, and psychological traits of gay men identifying as bears". In: *Archives of Sexual Behavior* 42, pp. 775–784.
- Naruszewicz-Duchlińska, Alina (2003). "Pseudonimy internetowe (nicknames) jako forma autoreklamy". In: *Prace Jezykozawcze* 5, pp. 85–98.
- Newsome-Chandler, Helen and Tim Grant (2023). "Developing a resource model of power and authority in anonymous online criminal interactions". In: *Language and Law/Linguagem e Direito* 10.1.
- Nielsen, Marie H, Line Aaskov, and Jørgen E Larsen (2022). "When virtuous paedophiles meet online: A sociological study of a paedophile community". In: *Sexualities* 25.5-6, pp. 598–614.
- Nini, Andrea (2023). *A theory of linguistic individuality for authorship analysis*. Cambridge University Press.
- Paquette, Sarah and Franca Cortoni (2022). "Offense-supportive cognitions expressed by men who use Internet to sexually exploit children: A thematic analysis". In: *International journal of offender therapy and comparative criminology* 66.6-7, pp. 647–669.
- Pellicone, Anthony and June Ahn (2014). "Construction and community: Investigating interaction in a Minecraft affinity space". In: *Proceedings of the Tenth Conference for Games+ Learning+ Society*, pp. 187–193.
- Perito, Daniele et al. (2011). "How unique and traceable are usernames?" In: *International Symposium on Privacy Enhancing Technologies Symposium*. Springer, pp. 1–17.
- Prażmo, Ewelina (2020). "Foids are worse than animals. A cognitive linguistics analysis of dehumanizing metaphors in online discourse". In: *Topics in Linguistics* 21.2.

- Quayle, Ethel and Max Taylor (2011). "Social networking as a nexus for engagement and exploitation of young people". In: *Information security technical report* 16.2, pp. 44–50.
- Raátz, Judit (2011). "Nick as self-attributed name". In: *Nouvelle revue d'onomastique* 53.1, pp. 183–211.
- Rasti, Alireza and Rahman Sahragard (2012). "Actor analysis and action delegitimation of the participants involved in Iran's nuclear power contention: A case study of The Economist". In: *Discourse & Society* 23.6, pp. 729–748.
- Riesigl, Martin and Ruth Wodak (2001). *Discourse and Discrimination*.
- Rogan, Hanna (2006). "Jihadism Online - A study of how al-Qaida and radical Islamist groups use the Internet for terrorist purposes". In: *FFI/Report* 915, p. 2006.
- Rothenberger, Liane, Kathrin Müller, and Ahmed Elmezeny (2018). "The Discursive Construction of Terrorist Group Identity". In: *Terrorism and Political Violence* 30.3, 428–453.
- Saidi, Firas et al. (2017). "Approaches to analyze cyber terrorist communities: Survey and challenges". In: *Computers & Security* 66, pp. 66–80.
- Sánchez Moya, Alfonso (2018). "Exploring digital discourses on intimate partner violence: a socio-cognitive approach". PhD thesis. Universidad Complutense de Madrid.
- Scheidt, Lois Ann (2005). "The performativity of naming: Adolescent weblog names as metaphor". In: *Health of the Discipline, National Communication Association (NCA), Boston MA. URL (consulted March 2006): http://www.loisscheidt.com/working_papers_archive/The_Performativity_of_Naming.pdf*.
- Schlobinski, Peter and Torsten Siever (2018). *Nicknamen international: zur Namenwahl in sozialen Medien in 14 Sprachen*. Peter Lang AG.
- Schneevogt, Daniela, Emily Chiang, and Tim Grant (2018). "Do Perverted Justice chat logs contain examples of overt persuasion and sexual extortion? A research note responding to Chiang and Grant (2017, 2018)". In: *Language and Law=Linguagem e Direito* 5.1, pp. 97–102.

- Steel, Chad MS et al. (2023). "Self perceptions and cognitions of child sexual exploitation material offenders". In: *International journal of offender therapy and comparative criminology* 67.10-11, pp. 1017–1036.
- Stommel, Wyke (2007). "Mein Nick bin ich! Nicknames in a German Forum on Eating Disorders". In: *Journal of Computer-Mediated Communication* 13.1, pp. 141–162.
- Taylor, Gary W and Jane M Ussher (2001). "Making sense of S&M: A discourse analytic account". In: *Sexualities* 4.3, pp. 293–314.
- Torres-Soriano, Manuel Ricardo (2013). "The dynamics of the creation, evolution, and disappearance of terrorist internet forums". In: *International Journal of Conflict and Violence* 7.1, p. 165.
- Van Dijk, Teun A (2006). "Discourse, context and cognition". In: *Discourse studies* 8.1, pp. 159–177.
- Van Langendonck, Willy (2007). *Theory and typology of proper names*. Mouton de Gruyter.
- Van Leeuwen, Theo (2013). "The representation of social actors". In: *Texts and practices*. Routledge, pp. 41–79.
- Vendius, Trine Thygesen (2015). "Proactive Undercover Policing and Sexual Crimes against Children on the Internet". In: *European Review of Organised Crime* 2.2, pp. 6–24.
- Wang, Yubin et al. (2016). "Identifying users across different sites using usernames". In: *Procedia Computer Science* 80, pp. 376–385.
- Wenger, Etienne (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge university press.
- Westlake, Bryce G and Martin Bouchard (2016). "Liking and hyperlinking: Community detection in online child sexual exploitation networks". In: *Social science research* 59, pp. 23–36.
- Westlake, Bryce G, Martin Bouchard, and Richard Frank (2011). "Finding the key players in online child exploitation networks". In: *Policy & Internet* 3.2, pp. 1–32.
- Wodak, Ruth and Michael Meyer (2001). "What CDA is about—a summary of its history, important concepts and its developments". In: *Methods of Critical Discourse Analysis*.

- Wodak, Ruth and Michael Meyer (2009). "Critical discourse analysis: History, agenda, theory and methodology". In: *Methods of critical discourse analysis* 2.1, pp. 1–33.
- Wodak, Ruth and Michael Meyer (2015). *Methods of critical discourse studies*. Sage Publications.
- Wood-Doughty, Zach et al. (2018). "Predicting twitter user demographics from names alone". In: *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, pp. 105–111.
- Woodhams, Jessica et al. (2021). "Characteristics and behaviors of anonymous users of dark web platforms suspected of child sexual offenses". In: *Frontiers in Psychology* 12, 623668.
- Wright, David (2020). "Corpus approaches to forensic linguistics: Applying corpus data and techniques in forensic contexts". In: *The Routledge Handbook of Forensic Linguistics*.
- Young, Kimberly (2008). "Understanding sexually deviant online behavior from an addiction perspective". In: *International Journal of Cyber Criminology* 2.1, p. 298.
- Zhang, Wei and Stephanie Watts (2008). "Online communities as communities of practice: a case study". In: *Journal of Knowledge Management* 12.4, pp. 55–71.