

LANGUAGE AND ONLINE IDENTITIES

Forensic linguistics is at the cutting edge of the undercover policing of child sexual abuse on the open Internet and Dark Web, and language and identity is a fundamental part of this. The authors have drawn on their extensive experience in training undercover officers to develop innovative methods in identifying the creation and performance of online personas, crucial in detecting identity disguise online. This groundbreaking book demonstrates these methods through case studies, while also exploring the link between language and identity. By bringing together previously opposed positions in forensic authorship analysis, the book develops a novel theory of linguistic identity, which will resonate not just in forensic authorship research but in sociolinguistics more widely. This unique forensic linguistic project has real-life impact in assisting the police in their investigation of online abusers, and has impact for students and researchers of linguistics, through its contribution to the research of linguistic identities.

TIM GRANT is the UK's only Professor of Forensic Linguistics, at Aston University, providing academic research and expert investigative assistance to UK and overseas police forces as well as providing evidence for both prosecution and defence and in civil cases. He is a former President of the International Association of Forensic Linguists.

NICCI MACLEOD is a Senior Lecturer in English Language and Linguistics at Northumbria University. She has published in a number of peer-reviewed journals and edited collections, and provided expert reports to several police forces as well as to solicitors and private clients in criminal and civil cases.

LANGUAGE AND ONLINE IDENTITIES

The Undercover Policing of Internet Sexual Crime

TIM GRANT

Aston University, Birmingham

NICCI MACLEOD

Northumbria University, Newcastle



**CAMBRIDGE
UNIVERSITY PRESS**

Cambridge University Press & Assessment
 978-1-108-72003-8 — Language and Online Identities
 Tim Grant, Nicci MacLeod
 Frontmatter
[More Information](#)



CAMBRIDGE
 UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom
 One Liberty Plaza, 20th Floor, New York, NY 10006, USA
 477 Williamstown Road, Port Melbourne, VIC 3207, Australia
 314-321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India
 103 Penang Road, #05-06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment,
 a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of
 education, learning and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781108720038

DOI: 10.1017/9781108766425

© Tim Grant and Nicci MacLeod 2020

This publication is in copyright. Subject to statutory exception and to the provisions
 of relevant collective licensing agreements, no reproduction of any part may take
 place without the written permission of Cambridge University Press & Assessment.

First published 2020

First paperback edition 2024

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging-in-Publication data

NAMES: Grant, Tim (Professor of forensic linguistics), author. | MacLeod, Nicci, 1981– author.

TITLE: Language and online identities : the undercover policing of internet sexual crime /
 Tim Grant, Nicci MacLeod.

DESCRIPTION: Cambridge ; New York, NY : Cambridge University Press, 2020. |

Includes bibliographical references and index.

IDENTIFIERS: LCCN 2019038259 (print) | LCCN 2019038260 (ebook) | ISBN 9781108487306
 (hardback) | ISBN 978110820038 (paperback) | ISBN 9781108487306 (epub)

SUBJECTS: LCSH: Online sexual predators—Language. | Language and the Internet. |
 Instant messaging. | Online identities. | Forensic linguistics. | Sex crimes—Prevention.

CLASSIFICATION: LCC HV6773.I5.O58 G74 2020 (print) | LCC HV6773.I5.O58 (ebook) |
 DDC 363.25/953—dc23

LC record available at <https://lcn.loc.gov/2019038259>

LC ebook record available at <https://lcn.loc.gov/2019038260>

ISBN 978-1-108-48730-6 Hardback

ISBN 978-1-108-72003-8 Paperback

Cambridge University Press & Assessment has no responsibility for the persistence
 or accuracy of URLs for external or third-party internet websites referred to in this
 publication and does not guarantee that any content on such websites is, or will
 remain, accurate or appropriate.

Cambridge University Press & Assessment
978-1-108-72003-8 — Language and Online Identities
Tim Grant , Nicci MacLeod
Frontmatter
[More Information](#)

*For Kerrie, Andrew,
Flo, Stephany, George.*

Contents

<i>List of Figures</i>	<i>page</i> x
<i>List of Tables</i>	xi
<i>Preface</i>	xiii
<i>Acknowledgements</i>	xvi
<i>Table of Legislation</i>	xvii
<i>Table of Cases</i>	xviii
1 Introduction	I
The Problem	1
Online Child Sexual Abuse	4
Communicating Online	6
Who Are You?	8
Can You Be Whomever You Want?	10
Understandings of Language and Identity	15
Forensic Linguistics and the Interest in Idiolect	15
Language as Identity Performance	16
Identity Performance Online	21
A Sketch of a Resource-Constraint Theory of Identity	22
2 Data and Methods	26
Data	27
Genuine IM Data	27
Genuine Dark Web Data	28
Operational Data	29
Training Data	31
Experimental Data	32
Summary of Data	35
Methodological Approaches	35
Structural Analysis	37
Meaning Analysis	40
Interactional Analysis	43
Social Behaviour Analysis	50

viii	<i>Contents</i>	
	Ethics	52
	Research Ethics	52
	Operational Ethics	54
3	Experimental Results	58
	Introduction	58
	Identity and Variation in IM	59
	Variation in Structural Features	60
	Variation in Pragmatic Features	62
	Variation in Interactional Features	63
	Variation in Social Features	65
	Variation and Identity in IM: Conclusions	66
	The Experiments	66
	Design	66
	'Naïve' Identity Disguise: Experimental Results	67
	Conclusions	88
4	Training Identity Assumption	90
	Introduction	90
	Data	93
	Framework for Analysis	94
	The Structural Level	96
	Lexical and Morphological Analysis of the Target Persona	96
	Structural Level Performance Pre- and Post-Training	99
	The Meaning Level	102
	Speech Act Analysis of the Target Persona	102
	Speech Act Performance Pre- and Post-Training	105
	The Interaction Level	107
	Topic Development and Control by the Target Persona	107
	Topic Analysis of the Pre-training Chats	109
	Topic Analysis of the Post-Training Chats	110
	What Arouses Suspicion?	110
	Evaluation	112
5	Resources and Constraints in Abuse Identity Performance	114
	Language and Identity Revisited	114
	Online Communities of Practice and the Linguistic Individual	116
	Exploring Identity Resources in IM and Dark Web Chat Rooms	118
	Aspects of Performed Identity	119
	Abuse Communities of Practice	130
	Conclusions	137
6	Contexts for Linguistic Investigative Advice	141
	Evidencing That an Anonymous Writer Is a Known Author	142
	Issues of Admissibility	142
	The Uses of Comparative Authorship Analysis in Online Crime	148

	<i>Contents</i>	ix
	The Aim and Purpose of Comparative Authorship Analysis	153
	Author Search and Sociolinguistic Profiling	154
	The Purpose of Sociolinguistic Profiling	154
	Assuming the Identity of an Online Writer	160
	The Purposes of Police Undercover Authorship Synthesis	162
	Conclusions	164
7	Implications and Future Directions	166
	Introduction	166
	Implications for Supporting Operations	167
	Implications for Forensic Authorship Analysis	172
	Implications for Theorising Identity	175
	Conclusions	178
	<i>References</i>	181
	<i>Index</i>	193

Figures

2.1	Avatar screenshot	<i>page</i> 29
3.1	Overall findings in detection of identity assumption	68
3.2	Criteria on which judges base correct judgements	71
3.3	Impersonators' notes by linguistic level	76
3.4	Most frequently coded structural features, Participant 1	79
3.5	Most frequently coded structural features, Participant 2	79
3.6	Successful impersonation	80
3.7	Turn-initial capitalisation (TIC), Participants 2 and 1	80
3.8	TIC, Participants 1 and 3	81
3.9	Written speech in Participants 19, 20 and 21	82
3.10	TIC and emoticons, Participants 7 and 8	83
3.11	Speech acts, Participants 1 and 3	84
7.1	Linguistic leakage	169

Tables

2.1	Overview of datasets	35
2.2	Four domains of Computer Mediated Discourse (CMD)	37
2.3	Top four levels of taxonomy of structural features	38
2.4	Revised taxonomy of structural features	39
2.5	IM interaction terminology	46
3.1	Mean confidence in accuracy of judged switches	70
3.2	Preparation notes submitted by group	75
3.3	Openings, Participants 7, 8, 9	85
3.4	Usual turn length, Participants 4, 5, 6	86
3.5	Impersonation turn length, Participants 4, 5, 6	87
4.1	Four domains of CMD	95
4.2	Examples of victim's vocabulary and spelling choices	97
4.3	Variants of use in victim's vocabulary	97
4.4	Examples of UCOs' structural features prior to linguistic training	100
4.5	Examples of UCOs' structural features after linguistic training	101
4.6	Topic control in the historical chat log	108
4.7	Topic control in the pre-training chats	109
4.8	Topic control in the post-training chats	110

Preface

This book provides an account of our theoretical approach to identity and its relationship with linguistic performance, particularly within the context of online criminal interactions and communities. We report on findings from a project investigating the assumption of identities online, drawing on instant messaging (IM) data from a variety of contexts, both ‘real life’ and experimental, in order to expound the implications of our approach for the training and professional practice of online undercover investigators.

Chapter 1 provides a general introduction to the book, outlining the need for police investigators to engage in online identity assumption and justifying a linguistic contribution that can be made in terms of the variety of ways in which linguists and forensic linguists have approached questions of language and identity. We set out the context of online sexual abuse, in particular the abuse of children, and describe the current impact forensic linguistics is having in the training of investigators of these crimes.

Chapter 2 outlines the various sources of our data, which include logs of IM, conversations collected from genuine resolved cases of child sexual grooming, from a range of Dark Web fora dedicated to the topic of child sexual abuse, from undercover operations targeting the producers and disseminators of indecent images and videos of children, from role playing exercises that take place within a course for online investigators, and from a series of experiments we designed in order to systematically investigate key questions around language and identity performance. The chapter continues by describing our methodological approach to these data, starting with the micro-analysis at the structural level of language, including of vocabulary and orthographic features as we first began setting out in MacLeod and Grant (2012). We then move through our modifications of Searle’s (1969) Speech Act Theory and Gumperz’s (1982) approach to topic management inasmuch as they relate to IM and the operational context, before finally setting out our theory about how these stratified

levels of language interface in the performance of identities. Finally the chapter concludes with a discussion of ethics, both the research ethics of engaging with this area of criminal activity and the operational ethics of assisting the police in these contexts.

Chapter 3 expands on the experimental phase of the project, providing an account of its design and a thorough discussion of the results and their implications. The chapter explores the level of accuracy with which participants in IM are able to detect the substitution of one interlocutor with another, and the levels of confidence with which such decisions are made. It also addresses the effects of impersonator preparation on these scores. Finally, the linguistic criteria that people report having relied upon in making these assessments are scrutinised, and we are thus able to formulate opinions about which features are the most salient for the construction of one's linguistic identity.

These findings feed in to the linguistic identity assumption training we currently provide to undercover officers, which is the subject of Chapter 4. Here we focus in particular on how our component of the undercover training has been influenced by our own theories of identity performance as supported by our analyses. We outline the input we provide to trainees at the levels of linguistic structure, meaning and interaction, and describe the pro forma we provide for the analysis of online linguistic personae. We also report here on the findings of a small-scale experiment comparing trainees' competence at linguistic identity assumption before our training versus afterwards.

This theoretical underpinning to the concept of identity is the central topic of Chapter 5, where we examine particular facets of identity such as the performance of age, relationships and communities of practice. Keeping in mind our view of identity as being continuously negotiated through discursive practices, we focus here particularly on how age is treated as a relevant identity category by participants, and how relationships between adult offenders and child victims are performed. We probe the question of how these performances are situated in the wider social context, and in relation to other, more recognisable types of relationship that might be relied upon as resources.

Drawing together all the previous chapters, Chapter 6 explicates the ways in which forensic linguistics can continue to provide support to the work of undercover online operatives. Breaking down traditional constraints that have seen the work of the forensic linguist in this area limited to authorship analysis tasks, we sustain the argument here that we can usefully contribute to the strategy of authorship synthesis. While there is

Preface

xv

enormous potential for linguistic contributions to these policing tasks of identity assumption and infiltration, there are, of course, limitations. Both are discussed in detail in this chapter.

Finally, Chapter 7 discusses the implications of our work for the operational task of identity assumption, and for the more theoretical concerns around the linguistic individual and language and identity more generally. We conclude the chapter with our thoughts for the directions similar research might take in the future.

Acknowledgements

The authors would like to thank the policing partners on the project for unprecedented access to trainees and to unlimited genuine and operational data. We look forward to continuing to work closely with them in the future. We are also indebted to the three reviewers who gave such useful feedback on our initial proposal. The research underpinning this work was supported by the Economic and Social Research Council, Grant Reference ES/L003279/1.

Table of Legislation

Criminal Justice Act 1988, s. 160
European Convention on Human Rights, Article 6
Human Rights Act 1998, s. 1
Police and Criminal Evidence Act 1984, Code C
Protection of Children Act 1978, s. 1
Regulation of Investigatory Powers Act 2000, s. 26–29
Sexual Offences Act 2003, s. 14

Table of Cases

UK Cases

Criminal Practice Directions Amendment No. 8 [2019] EWCA Crim 495
Hodgson, D. *v.* [2009] EWCA Crim (31 March 2009)
Lundy *v.* R [2013] UKPC 28
Metcalf, R. *v.* [2016] EWCA Crim 681
PS *v.* LP [2013] EWCOP 1106
R *v.* Clarke (RL) [1995] 2 Cr. App. R. 425
R *v.* Horseferry Road Magistrates' Court, Ex p Bennett [1994] 1 AC 42
R *v.* Loosely [2001] UKHL 53
R *v.* Sang [1980] AC 402

US Cases

Daubert *v.* Merrell Dow Pharmaceuticals, Inc. (509 U.S. 579 (1993))
Frye *v.* United States (293 Fed. 1013 (1923))