

Institution: Aston University		
Unit of Assessment: 26 Modern Languages and Linguistics		
Title of case study: Investigative forensic linguistics: Equipping law enforcement to fight online sexual crime.		
Period when the underpinning research was undertaken: 2010 to 2020		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	Role(s) (e.g. job title):	Period(s) employed by submitting HEI:
Prof Timothy Grant	Professor of Forensic Linguistics	Aug 2007 to present
Dr Krzysztof Kredens	Senior Lecture in Forensic Linguistics	Aug 2007 to present
Prof Jack Grieve	Professor of Linguistics (now at University of Birmingham)	Jan 2012 to June 2017
Period when the claimed impact occurred: Aug 2013 to Dec 2020		
Is this case study continued from a case study submitted in 2014? No		
1. Summary of the impact We have equipped UK and international law enforcement to fight online sexual crime using investigative forensic linguistic methods through: - Training UK and international online undercover police and dark web investigators in linguistic methods; - Provision of linguistic software to UK and international law enforcement, which is being used in account takeover and infiltration operations, and in operations searching for online anonymous offenders; - Embedding investigative linguistic methods in online investigations, through the 9-month secondment of a post-doctoral researcher to a police intelligence unit; - Contribution to the identification and arrest of high-profile Child Sexual Exploitation and Abuse (CSEA) offenders.		
2. Underpinning research To improve the investigation and arrest of serious offenders we have researched individual authorial style in the context of online sexual exploitation and abuse. Through these projects we: - developed new ways to describe an individual's linguistic persona from an analysis of a chatlog. This work, advanced through the ESRC project <i>Assuming Identities Online</i> (AIO) (R1), assists undercover online officers (UCOs) in their performance of that identity (R2:Ch3). The descriptive framework includes: <i>structural descriptions</i> – incorporating word selection, syntax and spelling; <i>semantic-pragmatic descriptions</i> – incorporating speech acts descriptions; <i>interactional descriptions</i> – incorporating descriptions of topic control; and a <i>description of conversational strategies</i>, based in rhetorical moves analysis (R3). - developed training for UCOs based on the description and subsequent 'performance' of a linguistic persona and demonstrated that using this approach can reduce UCO's chances of detection as an impersonator from 75% to 25% (R2:Ch4; see Fig. 1); - developed a theoretical basis for understanding the linguistic persona through a <i>resource and constraint</i> model of identity performance (R2:Ch5); - developed and evaluated new techniques for authorship analysis of short texts usable for forensic investigation. These techniques are based on idioms or strings of words – known as <i>n-grams</i> R4.		

The solid line in Figure 1 (from **R2**:p.68) demonstrates how the training reduced the detection of UCOs engaged in online identity assumption in a field simulation experiment. It shows how prior to training, detections were as high as 75%, but after training in the linguistically informed methods, these detections of UCOs identity assumption fell to 25%.

Our experimental evaluations (and subsequent training) have further improved UCOs'

professional practice by bringing detailed understanding into the types of linguistic slips that will most likely compromise UCOs in their online identity assumptions, and through providing methods for building convincing linguistic legends adapted to specific online forums (**R2**). We have shown how offenders develop and use multiple, deceptive online identities (**R3**) and demonstrated important linguistic and behavioural differences between real offender-child grooming interactions, and those conversations which occur between potential offenders and the adult decoys of self-described paedophile hunting groups (**R5**).

All of these findings focus on the idea of an individual's language style and the expression of individuality through language, and link to our leading research in authorship analysis. **R4** demonstrates the novel technique of N-gram tracing and applies it to the historic problem of whether Abraham Lincoln or his secretary John Hay, wrote the famous 'Bixby Letter' in the

American Civil War. Figure 2 (from **R4**) illustrates word and character n-grams drawn from the Bixby Letter. N-gram tracing was inspired by the qualitative approach to authorship analysis applied in forensic linguistic casework and builds upon recent quantitative research in stylometry and forensic linguistics.

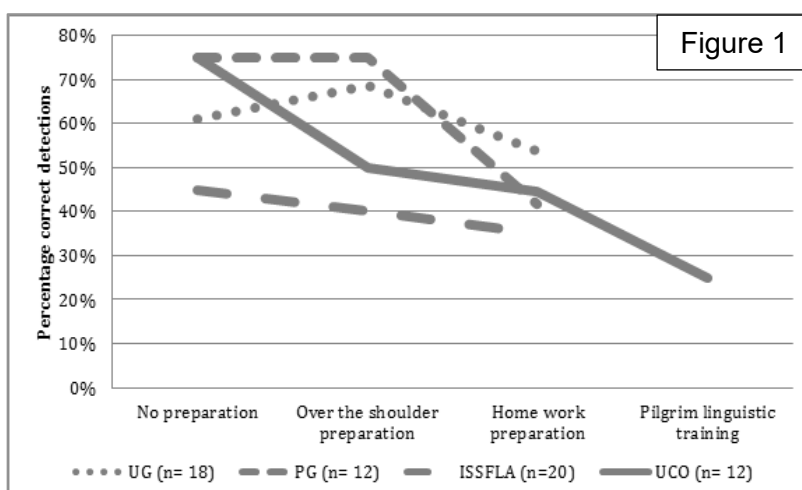


Figure 2		
Level	Length	Example
Word	1	i, have, been, shown, in, the, files, of, war, ..., field, battle
	2	I have, have been, been shown, shown in, ..., of battle
	3	I have been, have been shown, ..., field of battle
Character	1	i, _, h, a, v, e, b, n, s, o, w, t, ..., c, y
	2	i_, _h, ha, av, ve, e_, _b, be, ..., ba, tl
	3	i_h, _ha, hav, ave, _be, bee, ..., ttl, tle

The use of n-grams is not novel itself, but our approach in **R4** is distinctive in that it:

- is designed for application to short texts
- extracts from these short texts the comprehensive set of all word- and character- n-grams in a text; and
- only considers the presence and absence of these features, rather than their relative frequencies.

It is this technique that we have further developed and applied to the investigative problem of searching the open internet for the identifiable postings of offenders, based on the linguistic analysis of their anonymous dark web interactions.

3. References to the research

- R1. ESRC funded project Assuming Identities Online (ES/L003279/1) £418,888 August 2014 – July 2016. Project partners were the University of Warwick Security Ethics group and a software development SME, Lexagesys.

- R2. Grant, T.D. & MacLeod, N.M. (2020) *Language and Identity Online. The undercover policing of internet sexual crime*. CUP, Cambridge. ISBN-13 978-1108487306
- R3. Chiang, E. & Grant, T.D. (2019) Deceptive identity performance: Offender moves and multiple personas in online child abuse conversations *Applied Linguistics*. 40(4) 675-698. <https://doi.org/10.1093/applin/amy007>
- R4. Grieve, J., Clarke, I., Chiang, E., Gideon, H., Heini, A., Nini, A., & Waibel, E. (2018). Attributing the Bixby Letter using n-gram tracing. *Digital Scholarship in the Humanities*, 34(3), 493-512. <https://doi.org/10.1093/llc/fqy042>
- R5. Schneevogt, D., Chiang, E., Grant, T. (2018) Do Perverted Justice chat logs contain examples of Overt Persuasion and Sexual Extortion? A Research Note responding to Chiang and Grant (2017, 2018) *Language and Law = Linguagem e Direito* 5 (1), 97-102 <http://ojs.letras.up.pt/index.php/LLLD/article/view/4549>

The above are all equivalent or better than 2* publications – **R1** is a competitively awarded research council funded project, **R2** a significant monograph with a major publisher; **R3-5** are all journal articles placed in internationally recognised peer reviewed journals.

4. Details of the impact

With law enforcement partners we have:

- Trained online investigators in linguistic methods and have improved their investigation of internet sexual crime;
- Provided linguistic software to investigators, which has been used in online investigations and led to the arrest of serious offenders;
- Embedding investigative linguistic methods in online investigations, through the 9-month secondment of a post-doctoral researcher to a police intelligence unit;
- Contributed to high-priority investigations of online offenders including Matthew Falder, for which we were awarded Director's Commendations from the National Crime Agency (NCA).

Presentations delivered by Prof Grant to senior investigators of the NCA and to the Police National Undercover Working Group led to Prof Grant designing and providing bespoke linguistic training. This was delivered through Sorinteq, a specialist company, to groups of 15-20 UCOs three times a year over 6 years, totalling more than 250 UCOs. It was also attended by investigators from Germany and Sweden and delivered overseas, e.g. to 30 officers in the Royal Canadian Mounted Police (**S1:p1**), and from 2018 to the NCA, and UK Intelligence and Security Agencies (**S2**).

Funding (**R1**) enabled the full evaluation of the training and also the development of software to assist UCOs in identity assumption and legend building. The CEO of Sorinteq attests that, the training "...provided a new operational investigative tool that had previously been unrealized within [law] enforcement." and that feedback received from both trainer and trainee UCOs showed our training provided "...an excellent line of operational approach..." and the "integration of multi-disciplinary skills overlaid onto other techniques and capabilities." (**S1:p1**). He concludes, the "...software and training based in [Prof Grant's] research, has without doubt, improved the capability of policing during online undercover operations in the UK and abroad." (**S1:p2**)

West Midlands Police funded the secondment of an Aston Research Fellow to an intelligence unit two days a week for nine months (**S3**). She integrated forensic linguistic capabilities into the unit and progressed a project on risk assessments of anonymous online offenders. The head of this unit writes:

"This secondment has changed the way we examine language evidence in these cases and will have an on-going effect on our investigations."

"...we have been able to use [text removed for publication] skills to further a number of investigations. ...this has made a substantial contribution to these cases"

*"...at a national training event...I suggested investigative units in this domain should consider routinely employing or embedding academic forensic linguists within their units." (**S3**)*

Further endorsement of our work came from a previous Head of NCA Dark Web Intelligence, who comments (**S4**) specifically about **R2** and that this work shows:

"...forensic linguistic analysis is one of the behavioural approaches that can be exploited to enhance the prospects of attributing identity to offenders hiding behind on-line anonymity and to identify where an offender is exploiting multiple online personas. " (**S4**:para16).

Our work has included direct contributions to high-stakes investigations. Most significantly, Prof Grant and Prof Grieve contributed to NCA 'Operation CACAM', resulting in the arrest and prosecution of Matthew Falder. Falder operated anonymously online using dark web technologies and was described [in the media and NCA press releases as "one of the UK's most prolific paedophiles."](#) The Senior Investigating Officer (SIO) of this operation writes:

"[Text removed for publication]" (**S2**:p2)

He lists our two principal contributions:

- *"[Text removed for publication]"*
- *"[Text removed for publication]"* (**S2**:p2)

This contribution was acknowledged by the NCA Director of Threat Leadership through Commendations awarded to Prof Grant and Prof Grieve (**S5**).

Subsequent to the Falder investigation, and further to **R5**, we have developed n-gram tracing to create software to find non-anonymous online writings from anonymous online posts. We have used this capability in investigations against other high-risk offenders (**S2**) and it has led directly to their identification. The SIO of the Falder case notes that:

"[Text removed for publication]"

"[Text removed for publication]" (**S2**:p2)

Our training, software and linguistic expertise has broadly improved the capability of law enforcement in this domain nationally and internationally, including for hundreds of undercover officers. Our direct investigative work has helped identify and convict the most serious online sexual offenders. The ultimate significance of our impact is most profound for the thousands of victims who are now better protected from the harms online offenders inflict.

5. Sources to corroborate the impact

- S1. Letter from CEO of Sorinteq, a specialist training company about linguistic contribution to the Pilgrim training programme.
- S2. Letter from the Operations Manager of CEOP and SIO in Falder case.
- S3. Letter from the head of the West Midlands Police unit describing the benefits of secondment of Research Fellow into Police Unit
- S4. Media article reporting research with commentary from the then Head of Dark Web Intelligence at the NCA <https://jerseyeveningpost.com/news/2018/04/24/pioneering-research-could-help-catch-online-predators/>
- S5. Certificate of Commendation awarded by the Director of Threat Leadership of the National Crime Agency for investigative work carried out in the Falder case.