Northumbria Research Link

Northumbria
University
NEWCASTLE

UniversityLibrary

5000

Assuming Identities Online: How linguistics is helping the policing of online grooming and the distribution of abusive images.

Dr Nicci MacLeod, Lecturer in English Language and Linguistics, Northumbria University nicci.macleod@northumbria.ac.uk

Professor Tim Grant, Professor of Forensic Linguistics, Aston University

Abstract

In this chapter we seek to elucidate the potential of linguistic analysis in the undercover pursuit of criminals online. We examine the relationship between language and online identity performance and address the question of what linguistic analysis is necessary and sufficient to describe an online linguistic persona to the extent it could be successfully assumed by another individual. We compare the performance of trainee undercovers officers (UCOs) at online impersonation of a specific individual before receiving any linguistic training to their performance after such input.

We also report on a series of experiments in which participants engaged with each other over Instant Messaging (IM), before an impersonation situation was engineered. Information was collected about what level of accuracy and confidence individuals can detect the substitution of one interlocutor with another, what linguistic criteria give rise to such suspicions, and how individuals prepare for impersonation tasks. We also draw on other data sets including from police training and operational settings where identity disguise is employed online in the context of investigations into child grooming and the distribution of abusive images.

Introduction

There is no doubt that child abuse is one area of criminal activity that has been made easier and less risky by technological advances. The sexual grooming of children, i.e. the preparation of children for sexual activity (Chiang and Grant 2017), is a widespread issue, and one that has escalated in line with the advancement of the World Wide Web. Increased access to large numbers of like-minded individuals *and* potential victims at the click of a button has led to figures suggesting that 60% of children in the UK have been sexually solicited online (Internet Watch Foundation 2013). Compounding these statistics, the anonymity afforded by the Internet means decreased levels of perceived risk involved in such activities. The rise of the Dark Web, a heavily encrypted and thus anonymous means of accessing online content, has rendered traditional methods of offender identification, such as tracking geolocation and IP addresses, wholly ineffectual. Thus, online policing of child abuse has been described as being in crisis, with undercover operatives in dire need of alternative methods for pursuing the identification and prosecution of offenders.

Our involvement with the UK-wide 'Pilgrim' training programme for undercover online officers (UCOs) arose from the 2007 policing operation against members of a worldwide nexus of child sex abusers. Following the arrest of the ringleader Timothy Cox, Operation Chandler involved the assumption of his identity by UCOs who then interacted with his network in order to identify offenders, secure arrests, and rescue victims.

As well as assuming a particular offender's identity, UCOs are often required to pose online as victims or potential victims in order to effect arrests of dangerous child abusers. Consider this scenario: a child has been identified by a parent or guardian as being at risk from an online would-be sex offender. The child and the perpetrator have been engaging in sexualised conversation over Instant Messaging (IM) for some time, and the child believes she is in a relationship with the adult man with whom she has been chatting. The parent alerts the police, and the child is removed to a place of safety. Clearly the police wish to identify the perpetrator and secure an arrest, but for obvious reasons the child cannot be placed back in the dangerous situation, even if she was willing to be. Instead, an officer must operate undercover online, posing as the child, *synthesising* her identity, to arrange a meeting where the police can close in and make their arrest.

The task is cognitively demanding and fraught with ethical challenges. It is time sensitive, in that any delay in appearing online may alert the offender to police involvement – so time for preparation is limited. Furthermore, UCOs may be required to take over each other's operations due to shift patterns, sickness or leave. Similarly, a UCO may be required to engage in multiple operations simultaneously. The wary perpetrator, already likely to be on high alert to the possibility of being apprehended, must not have their suspicions aroused that the person they are talking to is no longer the child that they have been grooming for sex for a number of months.

During the conversations UCOs must avoid operating as *agents provocateur*. That is to say, they must anticipate and deflect any subsequent accusations of having incited illegal behaviour from their targets during the course of their communications (see Martellozzo, 2012). The Police and Criminal Evidence Act 1984 (PACE) further constrains the UCOs, as officers must also avoid conducting an illegal interview, and The Regulation of Investigatory Powers Act (2000) (RIPA) s. 26-29 requires that all their undercover actions be legally authorised and warranted.   (+75)

Forensic Linguistics

Forensic linguistics refers to the contributions of linguistic analysis to the delivery of justice. There are a number of questions for scholars to address at this interface between language and the law, and the field is generally viewed as comprising three main streams of work. First, there is a concern with the peculiarities of the written language of the law and issues around the comprehensibility of these texts for lay readers. Texts of interest to forensic linguists working in this area have included police cautions (e.g. Rock, 2012), jury instructions (e.g. Dumas, 2012), contracts, statutes,

and wills (e.g. Tiersma, 2001). Such work has sometimes resulted in the redrafting of legal texts such that they can be more easily understood by those for whom justice relies upon them being able to understand them. The second body of work within forensic linguistics focusses on interactions in the legal system, such as investigative interviews (e.g. Heydon, 2005) and courtroom examination (e.g. Ehrlich, 2001). Scholars have taken interest in how a bilingual context affects the dynamics of these interactions (e.g. Berk-Seligson, 2002) and in the experiences of lay participants whose linguistic and/or cultural background differs to that of the legal system in which they find themselves (e.g. Eades, 2002). The role of the interpreter in legal interactions has proved a further focal point for research in this area (e.g. Hale, 2004), as has the linguistic experiences of particular categories of vulnerable witnesses within the judicial system, such as children or those with communicative disabilities (e.g. Aldridge, 2010), or victims of sexual crimes (e.g. MacLeod, 2016). The application of work in these areas has included the publication of handbooks for legal practitioners to assist in communicating with clients of particular backgrounds (e.g. Eades, 1992) and the delivery of research-based training packages to institutional agents such as police interviewers (MacLeod and Haworth, 2016), representing a direct and mutually beneficial engagement between professionals and researchers, and undeniable societal impact.

Lastly, and of most relevance for the current discussion, there is the provision of investigative assistance and/or evidence based on linguistic analysis. Linguists are consulted by both sides in civil and criminal matters on an increasing scale, and are often called upon to present their findings in court. The work undertaken by forensic linguists in this vein is diverse, ranging from commenting on linguistic aspects of trademarks (e.g. Butters, 2008) through to providing opinion on a suspect's linguistic competence and ability to participate in an interview (e.g. Pavlenko, 2008), to ascertaining meaning in forensic contexts (e.g. Grant, 2017). In the UK context the most common question linguists are called upon to answer is 'do these texts share an author?', or 'who is the most likely author of this disputed text from this candidate pool of suspects?' ('comparative authorship analysis' - see Grant, 2010). At the early stages of an investigation, police may wish to employ the services of a forensic linguist to comment on the social background of the author of an anonymous text, in order to narrow down the pool of suspects or provide intelligence for media appeals ('sociolinguistic profiling'). As the discipline has increased in visibility, more attention has been paid to refining the methods forensic linguists use to address their questions (e.g. Nini, 2015; Wright, 2013, 2017).

The point of departure for sociolinguistic profiling and comparative authorship analysis as described above is an understanding of the causes of consistency and variation in language production. This understanding also underpins the new forensic linguistic task of authorship *synthesis*, a task that UCOs must engage in in the types of operation described in the introduction to this chapter. The authors have become involved in assisting investigators with this kind of identity assumption (see MacLeod and Grant,

2017, Grant and MacLeod, forthcoming), providing a clear evidence base for covert policing methods. Part of the Pilgrim course involves a simulated operation, in which trainee UCOs prepare and perform the identity of a specific child victim from a resolved historic case, and engage in conversation via IM with the 'offender', played by one of their trainers in another part of the building. Trainees are instructed to arrange a meeting with their target, and to elicit particular details 'on record' such that the criteria of 'arranging or facilitating the commission of a child sex offence' as set out in the Sexual Offences Act 2003 s. 14 are satisfied. As well as attempting to produce convincing linguistic identities, the trainees must also collate and remember a clear picture of the child's life based on the content of the historic chat logs, such as hobbies, school and family life.

The authors' role in this exercise is to (a) provide a half day of linguistic training input prior to the simulation and (b) evaluate the trainee's performance from a linguistic perspective once the simulation is complete. We report later in the chapter on the effectiveness of our input for trainees' identity assumption.

Language and Identity Online

There are a number of linguistic components of an individual's identity that need to be emulated if a UCO is to successfully assume their identity online. Among the phenomena focussed on in the computer mediated discourse analysis (CMDA) literature are features at the structural level, such as spelling and punctuation rules, and an individual's use of abbreviations, initialisms, emoticons, and so forth. Individuals may also vary pragmatically, in that they differ in terms of their preferences for particular speech acts, levels of indirectness, and so forth (note that this is also subject to contextual variation). Individuals have also been shown to vary in terms of their interactional patterns, such as average turn length, propensity for breaking turns over more than one transmission, and introduction and response to particular topics. These considerations are best demonstrated in the following table, adapted from Herring (2004: 18).

Table 1: Domains of CMDA (adapted from Herring, 2004:18)

	Phenomena	**Issues**	**Methods**
Structure	typography, orthography, morphology, syntax, discourse schemata	Genre characteristics, orality, efficiency, expressivitiy, complexity	Structural/Descriptive Linguistics, Text Analysis
Meaning	Meaning of words, utterances (speech acts), macrosegments	What the speaker intends, what is accomplished through language	Semantics, Pragmatics

| Interaction | Turns, sequences, exchanges, threads | Interactivity, timing, coherence, interaction as co-constructed, topic development | Conversation Analysis, Ethnomethodology |

Training Identity Synthesis

The linguistic input in Pilgrim comprises a three to four-hour session covering aspects of vocabulary, orthography, pragmatics and topic development, and the relevance of these key concepts for the practicalities of adopting an alternative persona online. Trainees are then introduced to a pro-forma developed to assist them in analysing and describing a linguistic persona. They have two hours to prepare for the roleplaying activity described above. The historical chat log is some thirty pages long and comprises several chats between the victim and the offender.

As well as noting down the target persona's tendencies regarding spelling and vocabulary (based on MacLeod and Grant's (2012) taxonomy), Pilgrim participants are provided with space to record their observations on her pragmatic, discursive and interactional behaviour. Having been introduced to speech act theory (Austin, 1962; Searle, 1969) and the illocutionary nature of language, trainees are encouraged to record details of this aspect of the target user's language throughout the chat log. Gumperz's (1982) approach to tracing topic development within a conversation is also introduced, and trainees are asked to categorise turns as either introducing a topic, as maintaining a topic, or as rejecting a topic, based on an analysis of their cohesive ties with the preceding discourse.

In order to evaluate our input in this regard, we compared the performance of trainees before receiving any linguistics training with their performance afterwards. In the 'Before' set there are five chats with a single trainee author and one with two trainee authors, with the point at which one takes over from the other clearly marked in the text. In the 'After' set there are six conversations, all of which are produced by two trainee authors with the point at which one takes over from the other clearly marked in the text (the instruction to swap over is explicitly provided by the instructor around half way through the conversation). The length of the conversations ranges from three to nine sides of A4, spanning around an hour and a half for each conversation.

For reasons of space we shall limit our discussion here to the trainees' performance in relation to two structural features and their overall performance in relation to topic control.
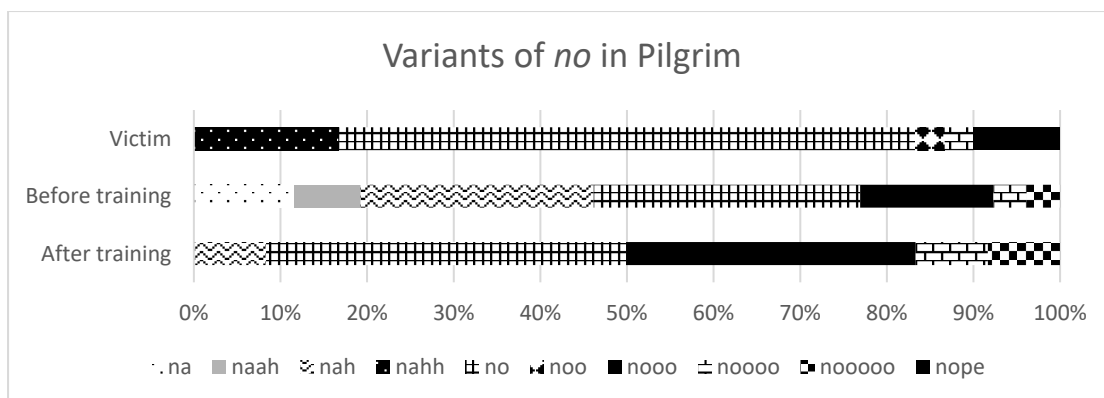
Figure 1: Variants of 'no' in Pilgrim

As Figure 1 shows, the victim's preference is for the standard *no*, although she phonetically stylises the item in around a third of instances, her favoured stylisation being *nahh*. Prior to the authors' input, the group of trainees makes use of no fewer than seven variant spellings. The most frequent variant matches that of the victim (*no*), but this does not account for as large a proportion of occurrences as it does in the victim's historical data. Furthermore, the second most frequent variants in the 'before' set – *nooo* and *nah* – do not appear in the victim's chat at all. Lastly, there are no occurrences of *nahh* or *nope* in the 'before' set, which together account for over a quarter of the victim's occurrences of the variable. The trainees' proportional use of the standard *no* increases after training, bringing the number closer to that evident in the victim's chat. However, the second most frequent variant here is *nooo*, which, as mentioned earlier, does not occur at all in the victim's historical transcript. The broad effect of training on these variable spellings, then, appears to be that while trainees' awareness of the use of standard forms becomes raised, they nevertheless continue to struggle with choosing appropriate rarer forms.

The same is not true for another variable spelling that appears numerous times across the data, that of 'what'. Of sixteen occurrences of the item in the victim's chat history, she invariably spells it *wht*.
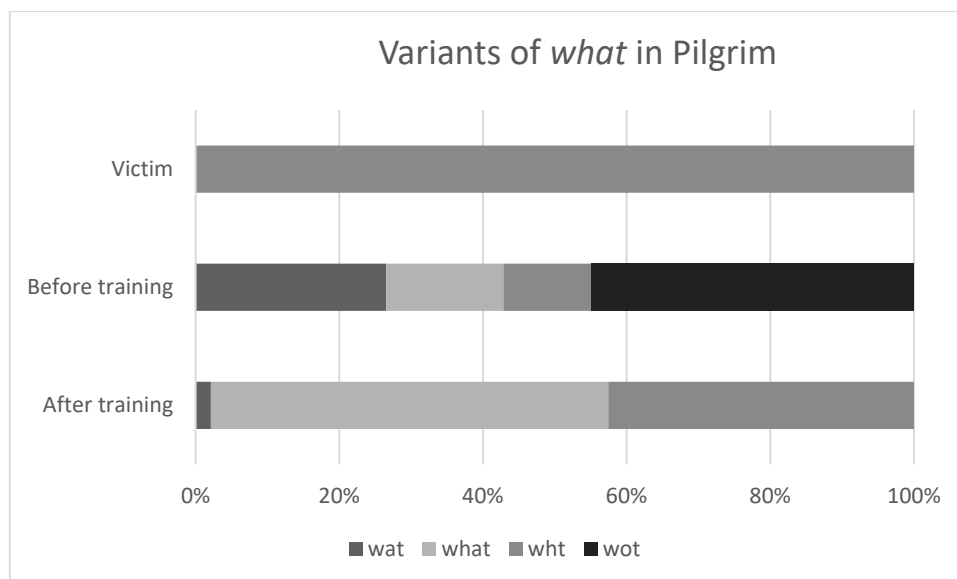
Figure 2: Variants of 'what' in Pilgrim

Evidently the training input has made improvements to trainees' performance when it comes to this variable feature – as well as losing completely the variant *wot* – previously the most frequent variant and yet nowhere to be found in the victim's chat, they increase their use of the variant *wht* – most frequent in the victim's chat – from 12% to 43%. It is nevertheless disappointing to note that the standard variant *what*, though never present in the victim's chat, accounts for the majority of occurrences in the trainees' chats post-training.

The post-training set of interactions showed a reduction in structural level identity performance errors, and none were commented on by the instructors in their evaluations in the post-training condition. There was a notable reduction in linguistically focussed challenges in the post-training condition, and remaining challenges focused either on factual inconsistencies or behavioural differences observed by the instructors.

The table below shows the topic management patterns across all the trainees both before and after training.

Table 2: Topic management in Pilgrim

Interlocutor	Topics initiated	Topics developed	Topics declined
Victim	Sexual contact with offender Own sexual arousal and use of objects Own appearance on cam Offender's motive for blackmail	Video conversation – clothing and sexually explicit content Offender's 'hacking' Sister Offender's instructions for	Teasing (sister in room) Sexual activity involving sister

	Offender's location, health, appearance.	sexual behaviour on camera	
Before training	Sexual activity (one occurrence) Meeting including location, time and travel arrangements. Offender's phone number and location Offender's experience of 14-year-olds Offender's clothing, description, name. Offender's arousal Being nervous	Cameras Victim's own clothing Some sexual talk	Sexual activity Victim's arousal Victim's masturbation, oral sex, anal sex Putting webcam on.
After training	Sexual activity (much more apparent across all trainee UCOs) Offender's arousal Present from offender Victim's mother as reason for not answering phone Travel arrangements Offender's clothing School uniform Victim's virginity School in morning Being nervous Offender's phone number	Travel and plans for the evening Proposed sexual activity Location of meeting Victim's location.	Sharing webcam Possible phonecall

A willingness to engage in conversation about sexual topics may seem counter-intuitive and dangerous to an investigator impersonating a person under the age of consent, but can be a fundamental component of impersonation. Examination of patterns of topic management in the trainee chat logs prior to linguistic input demonstrate significant differences as compared to the victim. In the historic chat log the child introduces sexual topics and sexual activity on several occasions. In all but one case the trainee UCO failed to do this. For some trainees, not only did they not

initiate sexualised conversation, but they declined it when it was instigated by the 'offender'. This natural reluctance to engage in online sexual activity whilst performing as a 14-year-old girl needs overcoming in these tasks, yet some trainees find this difficult to achieve. An explicit learning objective of the simulation exercise is to facilitate officers doing this more easily whilst staying within their authorisation, and avoiding going further than activity and discussions evident in the historic chat.

A further feature of the pre-training chat is the nature and quantity of initiated topics which are of operational interest to the UCOs. These types of operation are extremely demanding, and UCOs need to focus on a number of tasks simultaneously. They must try to obtain information about the offender which might identify them; they must be explicit about the child's age; and they must establish intention to engage in sexual activity in order to ensure that the act falls within the terms of the Sexual Offences Act 2003 s. 14. Furthermore, they are instructed to try and arrange a location to meet away from other children who might be endangered by a sexual predator, and attempt to get a description of the offender so that they can be easily recognised at the meeting place. All these tasks may create points of inconsistency with the child's previous conversations, and it is part of the skills that the officers develop to work these new topics naturally into the conversations while simultaneously maintaining a linguistic identity consistent with the child's.

After training the officers showed consistent improvement in this area and particularly appreciated how the language analysis can protect against accusations of acting as an agent provocateur. As can be seen in Table 2 the points of inconsistency typically involved introduction of the operational issues as discussed above. In both pre- and post-training chats an additional operational issue was that the trainees did not respond to attempts by the 'offender' to start a webcam conversation, activity in which for obvious reasons they could not engage during a genuine operation.

The aims of the trial reported on here were to evaluate the usefulness of the linguistic model for adopting a persona and to determine when and on what basis instructors, playing the role of the offender, would detect or become suspicious of identity assumption by the trainee operative. By taking a principled linguistic approach to authorship synthesis training we can hope to ensure that different aspects of identity performance can be analysed and then emulated, thus complementing the skills of UCOs to enable them to provide more convincing and less detectable identity assumption in their investigation of a broad range of crimes, including the online sexual abuse, exploitation, and grooming of children.

Experiments

The data described in the previous section are ideally suited to testing the effectiveness of our linguistic training for the task of identity assumption. However, we had further questions about the relationship between language and identity that these data could not answer for a number of reasons. The texts themselves are relatively short, and furthermore, while we were able to carry out a straightforward comparison of the before and after sets, the data did not allow for a systematic examination of the

effects of preparation on identity synthesis. Nor did they allow for a deeper engagement with the question of what arouses suspicion among naïve interlocutors that their conversational partner has been substituted. Lastly, despite their 'inauthentic' status, the content of the chats was nevertheless highly sensitive, and consistent with practitioners endeavouring to create as 'true to life' a scenario as possible. While the often graphic details of sexual abuse may be familiar and comfortable territory for UCOs, this is not necessarily true for the conference attendees and journal readers to whom we wished to ultimately disseminate our research findings, and to whom we believe ourselves to owe a duty of care. It is for all these reasons that a series of experiments was designed to address the question of what linguistic analysis is necessary and sufficient to describe an online linguistic persona to the extent it could be successfully assumed by another individual.

Participants took part in three 15-minute online conversations using the now defunct chat client Yahoo! Messenger. In the online conversations, participants acted in one of three roles, before switching roles and repeating the process two more times per session. The three roles were as follows:

'Judge': Judges were seated alone at a PC and asked to engage in online chat with their allocated interlocutor over IM. They were provided with a list of topics in the event that conversation ran dry, but were told not to feel limited by this. Judges were told in advance that at some point during the chat their interlocutor was going to be replaced by an impersonator. Judges were then asked to complete a form indicating where they believed the switch occurred, how confident they were in this, and the criteria on which they based their decision.

'Interlocutor': Interlocutors were seated with their allocated impersonator, out of sight of their judge, and asked to engage with their judge over IM. They were not asked to play any role but simply to engage in conversation. They were instructed that at some point during the chat they should hand over control of the keyboard to the impersonator with whom they had been paired. At this point, they were asked to complete a form with information about when the switch took place, to enable us to ascertain judges' accuracy.

Impersonator: The impersonator's job was to take over from their allocated interlocutor at some point during the online chat. Depending on the condition, impersonators had either (i) no knowledge of the interlocutor's online style ('No Preparation' condition); (ii) some knowledge based on watching them type ('Over the Shoulder' condition); or (iii) extensive knowledge based on taking away and scrutinising their historical chats ('Homework' condition). Impersonators were instructed to try to provide as convincing an impression of their interlocutor as they could in the circumstances, and were told that those who succeeded in fooling their judges would be entered into a prize draw.

The experiments generated a number of data sets, including the chat logs themselves which could be analysed for effective impersonation, and 'judge's comments about what had aroused their suspicion that a switch had taken place. One phenomenon to emerge from analysis of the former was what we have termed linguistic 'leakage'. This is demonstrated in Figure 3 below.
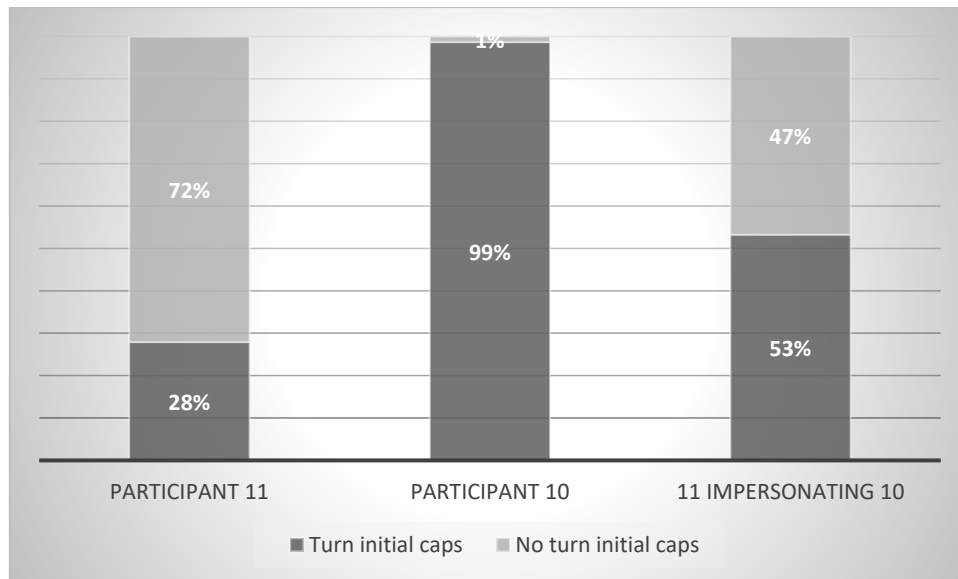
Figure 3: Turn initial capitalisation, participants 10 and 11

Clearly Participant 10 is, on the whole, a turn initial capitaliser, while Participant 11 tends *not* to capitalise at the start of turns. When 11 attempts their impersonation of 10, it does appear that they have picked up on this difference, and have attempted to introduce more capitalisation into their style – moving from just 28% up to 53%. However, they have failed to completely suppress their habit of non-capitalisation – a feature which is incredibly rare in 10's style, accounting for only 1% of turns overall. We thus see a 'hybrid' identity emerge, which has incorporated some features of the target identity but retains some residue of the 'home' identity. One practical outcome of these observations has been a recommendation to trainees to study their own linguistic style in preparation for identity assumption tasks, so that they are cognisant of what they must repress, as well as what they must emulate.

Moving on to the comments provided by 'judge's about the features that alerted them to potential impersonation, it is interesting to note that comments at the structural level – i.e. those relating to spelling, punctuation, vocabulary and so on – account for the vast majority of these observations. We see comments such as 'there was a decrease in capitalisation' and 'a doubling of letters for stylisation' and so on. This provides strong justification for the training model we currently offer, where the structural elements of language are placed front and centre. However, it is important to note that a good deal of comments relate to the interactional level – judges note changes such as 'the sentences were longer' and make comments about the types of topic that their interlocutors talked about. Thus, effective impersonation extends beyond an accurate performance of low-level features. This provides a rationale for ensuring trainees analyse every level of the linguistic persona before attempting to assume it online.

Conclusion

With the continuing advance of technology and increased opportunities for anonymous communication online, it is clear to see that the future will present an expanded requirement for forensic linguists to assist with online policing tasks – not just identity assumption but also, for example, the infiltration of online criminal communities. By ensuring that such assistance is underpinned by empirical testing and robust theory, we can safeguard the overriding rationale of forensic linguistic work – the use of linguistic analysis to assist in the delivery of justice.

References

Aldridge, M. 2010. Vulnerable witnesses in the criminal justice system. In M. Coulthard and A. Johnson (eds) *The Routledge Handbook of Forensic Linguistics.* Abingdon: Routledge, 296-314.

Austin, J.L. 1962. *How to do Things with Words.* Oxford: Clarendon.

Berk-Seligson, S. 2002. *The Bilingual Courtroom: Court Interpreters in the Judicial Process.* Chicago: University of Chicago Press.

Butters, R., 2008. A linguistic look at trademark dilution. Santa Clara High Technology Journal 24:3.

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

Dumas, B. 2012. Language of jury instructions. In C. Chappelle (ed.) *The Encyclopedia of Applied Linguistics.* Chichester: Wiley.

Eades, D. 1992. *Aboriginal English and the Law: Communicating with Aboriginal English Speaking Clients: A Handbook for Legal Practitioners.* Brisbane: Queensland Law Society.

Eades, D. 2002. Evidence given in unequivocal terms: Gaining consent of Aboriginal young people in court. In J. Cotterill (ed.), *Language in the Legal Process.* London: Palgrave, 161-196.

Ehrlich, S. 2001. *Representing Rape: Language and Sexual Consent.* London: Routledge.

Grant, T. 2010. Text messaging forensics: Txt 4n6: idiolect free authorship analysis? In M. Coulthard & A. Johnson (eds) *The Routledge Handbook of Forensic Linguistics.* Abingdon: Routledge, 508-522

Grant, T. 2017. 'Duppying yoots in a dog eat dog world, kmt': determining the senses of slang terms for the Courts' *Semiotica* 2017: 216, 479-495.

Grant. T. and MacLeod, N (forthcoming) *Language and Online Identities: The Undercover Policing of Internet Sexual Crime.* Cambridge: CUP

Gumperz, J., 1982. *Discourse Strategies*, Cambridge: Cambridge University Press.

Hale, S. 2004. *The Discourse of Court Interpreting.* Amsterdam: John Benjamins.

Herring, S. 2004. Computer-mediated discourse analysis: An approach to researching online behavior. In S. A. Barab, R. Kling, & J. H. Gray (Eds.), *Designing for Virtual Communities in the Service of Learning*, New York: Cambridge University Press, pp. 338-376.

Heydon, G. 2005. *The Language of Police Interviewing: A Critical Analysis*. Basingstoke: Palgrave.

Internet Watch Foundation, 2013. *Annual Report* https://www.iwf.org.uk/report/2013-annual-report

MacLeod, N. 2016. "I thought I'd be safe there": Pre-empting blame in the talk of women reporting rape. *Journal of Pragmatics* 96, 96-109.

MacLeod, N. and Grant, T. 2012. Whose tweet?: authorship analysis of micro-blogs and other short form messages. *Electronic Proceedings of the International Association of Forensic Linguists' 10th Biennial Conference*, Aston University, Birmingham, UK, July 2011. www.linguisticaforense.ufsc.br/tiki-index.php?page=IAFL+2011

MacLeod, N. and Grant, T. 2017. 'go on cam but dnt be dirty': linguistic levels of identity assumption in undercover online operations against child sex abusers. *Language and Law/ Linguagem e Direito* 4:2, 157-175

MacLeod, N. and Haworth, K. 2016. Developing a linguistically Informed approach to police interviewing. In R. Lawson, R. & D. Sayers (eds) Sociolinguistic Research: Application and Impact. Abingdon: Routledge.

Martellozzo, E. 2012. *Online Child Sexual Abuse: Grooming, Policing and Child Protection in a Multi-Media World*, London: Routledge.

Nini, A. 2015. *Authorship Profiling in a Forensic Context*. PhD dissertation. Aston University, Birmingham, UK.

Pavlenko, A. 2008. Non-native speakers of English and the Miranda warnings. *TESOL Quarterly*, 42: 1, 1-30

Rock, F. 2012 The caution in England and Wales. In L. Solan and P. Tiersma (eds) *The Oxford Handbook of Language and Law*. Oxford: Oxford Press, 312-325.

Searle, J.,1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Tiersma, P. 2001. Textualising the law. *International Journal of Speech, Language and the Law* 8:2, 73-92.

Wright, D. 2013. Stylistic variation within genre conventions in the Enron email corpus: developing a text-sensitive methodology for authorship research. *International Journal of Speech, Language and the Law* 20: 1, 45-75.

Wright, D. 2017. Using word n-grams to identify authors and idiolects: a corpus approach to a forensic linguistic problem. *International Journal of Corpus Linguistics* 22:2, 212-241.