University of Central Lancashire UCLan

Central Lancashire Online Knowledge (CLoK)

Title	Public appeals, news interviews and crocodile tears: an argument for multi-channel analysis
Type	Article
URL	https://clok.uclan.ac.uk/23471/
DOI	https://doi.org/10.3366/cor.2015.0075
Date	2015
Citation	Archer, Dawn Elizabeth and Lansley, Cliff (2015) Public appeals, news interviews and crocodile tears: an argument for multi-channel analysis. Corpora, 10 (2). pp. 231-258. ISSN 1749-5032
Creators	Archer, Dawn Elizabeth and Lansley, Cliff

It is advisable to refer to the publisher's version if you intend to cite from the work. https://doi.org/10.3366/cor.2015.0075

For information about Research at UCLan please go to http://www.uclan.ac.uk/research/

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the http://clok.uclan.ac.uk/policies/

Public appeals, news interviews and crocodile tears: an argument for multi-channel analysis

Dawn Archer and Cliff Lansley

Abstract

In this paper, we propose a Six Channel Analysis System (SCAnS) for the (semi-)automatic investigation of potential deception across all communication channels. SCAnS builds on our current system: Six Channel Analysis in Realtime (SCAnR). SCAnR users are trained to code – as Points of Interest (PIns) – relevant occurrences of twenty-seven criteria relating to the six channels, when they appear to point to inconsistencies with respect to the speaker's account (the story they are trying to convey), their apparent baseline and the context. Our experiences to date confirm the view that multi-channel approaches have the potential to lead to higher accuracy rates of deception detection than is possible when using individual methods of detection and/or when focussing on one communication channel independently (Vrij *et al*., 2000: 257), especially when combined with cognitive elicitation strategies. However, we recognise the importance of (in)validating the relevance of the twenty-seven criteria through ongoing research. SCAnS will provide a (semi-)automated means of achieving this. Given our audience, we focus on the usefulness of content-analysis tools like Wmatrix (Rayson, 2008) for this purpose. More broadly, our research has implications for the analysis of data in forensic contexts, across all available channels of communication, and for the coding of (para)linguistic features.

Keywords: Crocodile tears, deception, multi-channel analysis, Points of Interest (PIns), press appeals

1. Introduction

The media use the term 'crocodile tears' to describe people's feigning of grief/anguish over crimes that they are found, ultimately, to have perpetrated. Since 2002, a number of television documentaries have alerted the public to the use of crocodile tears in press appeals. But academic interest in press appeals is scant currently, excepting Vrij and Mann (2001), ten Brinke and Porter (2012) and Wright Whelan *et al*. (2013). We outline these studies in Section 2. In Sections 3 to 3.3, we then outline the manual approach to data-gathering we currently advocate as part of our 'real-time' system to better reading others, and report four trained coders' analyses of a range of press appeals. The Emotional Intelligence Academy (EIA) work with practitioners/professionals who have limited time to make high-stake decisions in their work contexts (including intelligence services, customs, border police, law enforcement, negotiators, military, investors, fraud investigators and sales/procurement). The Six Channel Analysis in Real-time (SCAnR) system is designed to be applied in real time, whilst remaining as effective as other approaches to credibility assessment/deception detection. Simply put, users are encouraged to:

- Ignore behaviour which seems consistent with the account (the story they are trying to convey) of the speaker (S);
- Identify potential Points of Interest (PIns) from the various communication channels, which are suggestive of inconsistency – that is, mismatches within, and especially clusters of PIns across, these channels that cannot be explained by their apparent/emerging baseline and (micro/macro) context; and,
- Pay especially close attention to that person's cross-channel behaviour immediately after a prompt or question (i.e., up to seven seconds) – on the understanding that 'we can reasonably conclude that the behavior is directly associated with the stimulus' (Houston *et al*., 2012: 30).

To aid users in this process, SCAnR's alphanumeric referencing method prioritises twenty-seven research-corroborated criteria involving the voice (V), linguistic content (C), interactional style (IS), facial movements (F), body movements (B, which includes gesturing), and the psychophysiology/ autonomic nervous system (ANS), (see Section 3.1). Users are informed from the outset that they are not to assume these (or other behaviours) map incontrovertibly to deception. In fact, they are made aware that the veracity of particular deception indicators is hotly debated amongst academics. A case in point is avoiding the first person. Some find this particular criterion meaningful when it comes to identifying potential deception (see, for example, DePaulo *et al*., 2003). According to Vrij (2008), however, the relationship between self-references and deception is weak.

Newman *et al*. (2003) and Bond and Lee (2005) go even further, arguing that deceivers tend to use fewer third-person pronouns than truth tellers (but see Section 5.2). The SCAnR approach allows for this by discouraging users from drawing conclusions based on a single criterion or even a single communication (sub)channel and also focusses users on changes (including changes to pronoun use) through an account. In fact, users are cautioned against making *any* veracity conclusions based on PIns alone, even where they cluster across the (sub)channels within a seven-second window. This is because the PIns are understood to be points of interest only, which may or may not point to deception. What they provide users with, then, are areas for 'probing' through focussed questioning, where this is possible, and areas of interest where this is not (see Section 3).

As SCAnR represents a real-time manual approach to credibility assessment/deception detection, we are currently assessing the validity of using additional (semi-)automated means of assessing a person's (cross channel) behaviour. As we report in this paper, our ultimate aim is a Six Channel Analysis System (SCAnS) which provides users with simultaneous but detailed information with respect to the voice, IS, facial movements, body movements and ANS. One motivation for SCAnS is the opportunities it affords to (in)validate the relevance of the twenty-seven indicators used within SCAnR. A second motivation is to augment our understanding of the various cross-channel elements that make up high-stakes deception (and, where relevant, the co-occurrence or clustering of these elements), in case there are important criteria beyond the twenty-seven we currently prioritise. Given the audience of this journal, we will focus primarily on computer assisted analysis of linguistic content in this paper by assessing the usefulness of Wmatrix3 (Rayson, 2008) in highlighting potential deception indicators. This work thus builds on that of McQuaid *et al*. (2015: 21), who recently sought to 'enhance our understanding of verbal elements' of seventy-eight press appeals using the same content-analysis tool. Our analysis is described in Section 5.1 and the McQuaid *et al*. study, in Section 5.2.

2. Three previous investigations involving press appeals

Vrij and Mann's (2001) study involved fifty-two uniformed police officers from the Netherlands who viewed video clips of five UK press conferences where relatives were appealing for help with respect to missing or murdered kin; the relatives were found to be guilty of the crime at a later date. The police officers knew nothing of the cases, for Vrij and Mann's aim was to determine whether these 'professional lie-catchers' performed better than lay people in detecting deceit. Vrij and Mann found that the police officers did no better than chance (demonstrating accuracy rates of 50 percent, generally speaking) – although three officers scored an accuracy rate of 80 percent. When asked 'which cues had prompted' the fifty-two 'police officers to make their decision', the officers mentioned twenty-three different cues in total (Vrij and Mann, 2001: 129). However, between twenty-two and twenty-seven officers mentioned the same three 'cues' at least once: fake emotions, real emotions and gaze aversion. On further analysis, Vrij and Mann (2001: 129) found that 'the more often the officers indicated . . . the target person showed real emotions, the less accurate they were'. They also 'found a gender effect . . . with males being more accurate than females' (2001: 130). In line with previous research, Vrij and Mann (2001: 124) suggest that this may be because 'men are usually more suspicious than women', and that 'women are better than men in decoding information that someone wants to convey'. This particular investigation – in using deceptive pleaders only (i.e., those named as the perpetrators at a later date) – thus benefitted the men's approach: namely, looking for evidence of what the pleaders were trying to conceal as opposed to what they were trying to convey. A third finding worthy of comment is that, although the officers' experience with interviewing suspects did not map to their actual accuracy in detecting deceit, it was significantly correlated with confidence – that is, the more experienced officers assumed incorrectly that they were more effective deception detectors than they were (Vrij and Mann, 2001: 128–9).

For their investigation of crocodile tears in press appeals, ten Brinke and Porter (2012: 469) coded 74,731 frames of televised footage relating to seventy-eight pleaders, thirty-five of whom were deceptive pleaders. The coding scheme adopted expands on both Porter and ten Brinke's (2008) revision of Ekman *et al*.'s (2002 [1976]) Facial Action Coding System (FACS) and Pennebaker *et al*.'s (2001)

Linguistic Inquiry and Word Count (LIWC) software. The scheme can, thus, capture features such as emotional facial expressions, facial movements (blinks, gaze aversions, *etc*.), facial manipulations (instances when participants touched, scratched or covered their faces), illustrators (arm or hand gestures used to supplement speech), verbal cues (length of plea in words, speech rate, percentage of words that were speech hesitations, *etc*.), pronouns, words signalling tentativeness (e.g., *maybe* and *guess*), and emotion/affect terms (*joy*, *grief* and *hate*). Ten Brinke and Porter favoured their own facial coding procedure over FACS, because of the apparent ease with which it can be 'translated into practical recommendations for relevant professionals' (2012: 472): nonetheless, they have the ability 'to isolate particular facial areas of interest for future, intensive FACS coding' (2012: 472). The coding was undertaken by trained coders and 'involve[d] classifying the emotional expression in each 1/30-s frame of video in the upper and lower regions independently'. The authors explain that they use LIWC to capture verbal cues, but do not explain the process used to capture body language. They do, however, stress that coders' inter-rater reliability was high when it came to coding both verbal cues and body language. When discussing their findings, ten Brinke and Porter focussed on particular cues which, thanks to previous research, are already considered to be potentially indicative of deception and they used them to test five hypotheses. The hypotheses all assume that deceptive pleaders will somehow behave differently from genuine pleaders when it comes to their producing convincing sadness and distress expressions, speech rate, hesitations, pronoun use, use of affective terms, blink rate, *etc*.

Wright Whelan *et al*. (2013) used Canadian, New Zealand, UK and US video footage of thirty-two people making televised appeals for help in finding out what happened to a missing or murdered relative. Sixteen were later found to have perpetrated the crime concerned and were convicted accordingly. The authors were particularly interested in determining the occurrence of seven verbal cues previously related to honesty: expressions of positive emotion toward the relative; avoidance of 'brutal language' in favour of euphemisms such as 'taken from us'; expressions of concern (for the relative); expressions of hope (for a positive outcome); expressions of pain or grief; pleas for help (in finding the relative/perpetrator) or for the relative to return; and implicit references to (societal) 'norms' and their violation (see Wright Whelan, 2012). They were also interested in any occurrences of deception cues which, for them, were more typical of high-stakes situations. These cues included equivocation (markers), speech errors, phrase repetition, fillers, avoidance of the first-person, extraneous information outside the context of the incident, and use of 'brutal language' or detail about the relative. All verbal cues were found manually by one coder, after initial testing of inter-rater reliability (using a quarter of the datasets) was found to be high. Wright Whelan *et al.* (2013) also went on to investigate three pre-determined non-verbal behaviours: gaze aversion, head shaking and shrugging. As with the verbal cues, one coder was responsible for finding and coding these nonverbal cues in the thirty-two video clips (once high interrater reliability had been established for a random sample of 25 percent of the appeals). In a third step to their investigation, the authors considered in detail six cues that had been found to help them to discriminate between liars and truth-tellers significantly in investigations 1 and 2, namely, 'expressions of positive emotion towards the relative as a dichotomous variable, expressions of hope, references to norms of emotion or behaviour, equivocation, gaze aversion and head shaking' (2012: 10). They concluded that these particular behaviours – and especially equivocation, speech errors, gaze aversion and head shaking – may prove to be a useful focus when distinguishing deceptive pleaders from truthful pleaders in this particular high-stakes context. We pick up on some of these 'honesty' and 'deceit' features in our findings sections, when discussing the linguistic content of our video clips (see Sections 3.3 and 5.1).

3. SCAnR

SCAnR users are advised to ignore any behaviours that are consistent with the account, (emerging) baseline and context. Our reasoning for this is that, generally speaking:

- People are more likely to detect truths correctly (see, for example, Vrij's, 2008, 'truth bias' and, in particular, his assessment that 67 percent of people can correctly evaluate truths whilst only 44 percent correctly evaluate lies);

- People will demonstrate more cognitive load and/or might experience (and thus leak more indicators of) emotions when lying than when telling the truth – in particular, fear, excitement/*duping delight* and guilt (Ekman, 2004); and,
- A statement derived from memory of an actual experience is believed to differ in both content and quality from a statement based on invention or fantasy (Undeutsch, 1967).

Rather than attempting to note everything in realtime, then, users are deliberately 'freed up' to prioritise those occasions when an individual demonstrates inconsistencies across communication channels. These PIns can be coded manually, using the SCAnR coding scheme. Following a question, SCAnR users are encouraged to pay close attention to that individual's (cross-channel) behaviour for seven seconds following it (see Houston *et al*., 2012: 30). Users are asked to be particularly mindful of any communication cues which seem to be inconsistent with S's account (i.e., the story S is trying to convey), their emerging idiosyncratic baseline (within the situational context), and all relevant micro/macro contexts. To help in this process, users trained in the SCAnR approach are orientated towards twenty seven criteria (which are outlined below). These research-corroborated criteria relate to all six communication (sub)channels. Researchers have suggested that a multi-cue approach (such as this) can be bolstered by having prior knowledge of (i.e., a baseline for) a target's truthful behaviour. However, this possibility tends to be limited/non-existent in high-stake contexts such as an airport or in televised press appeals where people plead to the watching public for help. For this reason, we advocate that SCAnR users pay close attention to change(s) in communicative behaviour, during a given interaction, and what immediately follows each change or cluster of changes within and across the communication channels. Such changes are not, at this point, to be taken as evidence of deception. Rather, as explained above, they are to be viewed as PIns to be probed, using a forensic questioning approach.[1] Where such targeted probing is not possible, they remain as points of interest only. Users are counselled to resist veracity judgments based on passive behaviour only, as what makes an expert adept at lie detection is not only the passive observation of S's demeanor, but also the active prompting of diagnostic information (Levine *et al*., 2014).[2]

3.1 Coding scheme for PIns

The psychophysiological/ANS channel (P) accounts for seven of the twenty seven PIns, five of which capture physiological signals that users can sometimes see and hear without technical aids: that is, changes in skin colour, perspiration (P3), blood pressure on visible veins (P4), breath (P5), dryness of the mouth (P6) and pupil size (P7). P1 and P2 relate to changes which usually require technology to be detected (e.g., heart rate and galvonomic monitoring; see Section 6).

Similar to ten Brinke and Porter (2012), we draw on insights from FACS-related research for the first of the five Face codes (F1). Specifically, our (FACS-trained) SCAnR coders catalogue FACS anomalies with seventeen key FACS codes. F2 marks a durational misfit (Ekman and Frank, 1993), F3 marks evidence of asymmetry (unless indicative of contempt), and F4, evidence of asynchrony between muscle movement across the face for 'felt' emotion(s unless subtle). Finally, F5 marks onset/offset profiles which do not display the smooth onset/offset patterns characteristic of felt emotions (Ekman and Frank, 1993). As with all features, we required inter-annotator agreement between three of the four coders for an individual inconsistency to be verified as a PIn in the SCAnR analysis proper (for which, see Section 3.3).

The Body channel (B) captures features shown to be of value in emotion and veracity judgments (Vrij *et al*., 1996): specifically, micro gestures or gestural slips indicative of 'leakage' (B1); evidence of change(s) in illustrator behaviour (B2) and/or manipulators (B3); evidence of (muscle) tension in the body (B4); and changes in eye behaviour (blinks, eye gaze/movement/closure, *etc*.). Codes B1 to B5 are based on a more detailed system of Action Descriptors (ADs) developed by the team at EIA[3] (cf. ten Brinke and Porter, who simplified FACS Action Units/Descriptors or AUs to aid realtime annotation and multi-rater coding comparisons).

The SCAnR coding adopted for the Voice (V) enables real-time analysis of changes to pitch (V1), volume (V2) and tone (V3), but coders can also note sound lengthening, backchannels, stressed syllables, utterance trail offs, *etc*. (Rockwell *et al*., 1997).

Interactional Style (IS) is our label for phenomena such as fillers, parroting, evasion strategies (including equivocation markers), response latency, emphatic statements, repetition, qualifiers, pronoun usage (e.g., use of third person/avoidance of first person, or *vice versa*), (de)personalisation, distancing devices, *etc*. (following Jurafsky *et al*., 2009). More specifically, I1 marks changes to the rhythm (or 'flow') of the interaction because of features such as (filled) pauses, stutters, disfluencies, response latency, and so on. I2 marks evidence of evasiveness/ambiguity/equivocation (Wright Whelan et al., 2013). I3 encompasses influencing or impression management strategies. For example, the use of religious belief/values/character references, credibility labels, or a proof/evidence frame (Houston *et al*., 2012), representational frames relating to the Other, inappropriate politeness, repetition, *etc*.

The Content channel (C) contains four PIns. C1 captures changes in tense or inappropriate tense usage (such as when someone pleads for the return of a loved one, but refers to them in the past tense). C2 captures distancing language, following DePaulo *et al*.'s (2003) observation that deceivers will sometimes use linguistic constructions (e.g., fewer self-references, more tentative words) which serve to distance them from the subject(s) of their speech (see also Hancock and Woodworth, 2013). Here, SCAnR users might consider pronouns, tentativeness features, subject/noun changes, emotional terms/affective language, inappropriate concern, qualifiers, minimisers and other epistemic modality markers, *etc*. (Bond and Lee, 2005; ten Brinke and Porter, 2012; and Newman *et al*., 2003). The third Content criteria, C3, makes use of an adapted version of Criteria-Based Content Analysis (CBCA).[4] Although CBCA is primarily used to assist (European) courts in evaluating the credibility of children's (transcribed) narratives of sexual abuse, it has been used to evaluate adult accounts relating to issues other than sexual abuse (Porter and Yuille 1996; and Vrij *et al*., 2000). When drawing on CBCA criteria, Vrij *et al*. (2000) used a restricted set in combination with Reality Monitoring criteria. The SCAnR method, in contrast, has been to amend CBCA criteria so that users might record, as a PIn, occasions when the content of the story that S conveys: (*i*) lacks coherence, (*ii*) lacks unstructured, spontaneous reproduction, (*iii*) includes inappropriate detail, especially relative to the core of the story and what we know about memory (the account may also be void of related associations and unusual/superfluous details), (*iv*) exhibits contextual vagueness (as opposed to being characterised by contextual embedding), (*v*) is devoid of descriptions of interactions (including [recalled] verbatim conversations), (*vi*) is devoid of admissions of poor memory recall/spontaneous correction of memory errors (without prompting) and self-deprecation, and (*vii*) is devoid of accounts of mental states (self and other).[5] The final Content criteria, C4, is used when the SCAnR user recognises a verbal slip as a PIn (Ekman, 2004: 40).

3.2 Press appeals coded using SCAnR

To help readers better understand the SCAnR method, we include a SCAnR analysis of four televised clips representative of recent UK cases involving press appeals in Section 3.3:

- Ten-year-old friends, Holly Wells and Jessica Chapman, went missing from their village (Soham) in Cambridge on 4 August 2002. Ian Huntley, a school caretaker in the same village, was the last person to see them alive and gave a number of television interviews to this effect. On 16 August, Ian Huntley and his then girlfriend, Maxine Carr, gave witness statements to the police. After items of 'major importance' were recovered from their home, they were subsequently arrested and charged on suspicion of murder. That same afternoon (17 August) the girls' bodies were recovered from a ditch near a local RAF base. Huntley was found guilty of murder on 17 December 2003. Carr was found guilty of assisting an offender by falsely claiming they had been together at the time of the murders –when, in fact, she was away visiting relatives.

- Mick and Mairead Philpott made what the press described at the time as an emotional news conference on 16 May 2012, following a fire at their home in Derby which killed six children. Two weeks later, the Philpotts were charged with their murder. On 2 April 2013, they were found guilty of manslaughter (along with their friend and neighbour, Paul Mosley).
- Twelve-year-old Tia Sharp disappeared from the home of her grandmother, Christine Sharp. Christine's partner, Stuart Hazell, told police that Tia left the house on 3 August 2012 to travel to Croydon (5 miles away) but did not return. On 7 August, Tia's uncle, David Sharp, made a televised plea for her safe return. On 10 August, following the discovery of a body in the loft of Christine and Stuart's home, police immediately launched a search for and then arrested Hazell on suspicion of murder. Hazell pleaded not guilty when he appeared in court on 8 March 2013. He subsequently changed his plea to guilty six days into his trial and was sentenced to life imprisonment the following day.

We opted for material relating to these particular cases for several reasons. First, the four televised clips (described below) represent high stakes situations: the loss or disappearance of children. This is important as when lies concern more serious matters, researchers hypothesise that liars will be more emotionally invested and aroused, leading to more pronounced cues to deception (Buckley, 2012). If this is so, we might hypothesise, in turn, that coders should find more PIns when analysing deceptive pleaders than when analysing innocent pleaders. Four SCAnR coders (who are also experienced/ certified FACS coders) completed the coding, so that we could achieve inter-annotator agreement for any PIn mentioned in Section 3.3.

Our second reason for choosing these clips is to ensure that that we have 'ground truth' in respect to who perpetrated the actual crime (from a legal perspective). For example:

- We now know that Huntley was being deceptive when interviewed. The press interview we draw upon takes place outside his home; at this point, he had already murdered Holly and Jessica, and had disposed of their bodies. Coders assessed a clip totalling 170 words over 2,100 frames (over 49 seconds) of video.
- Philpott represents a deceptive pleader, who is now known to have started the fire. At the time of the appeal, he was attempting to conceal any involvement. That appeal totals 372 words and is 2 minutes 37 seconds in duration. Philpott's wife, Mairead, sat with him throughout, facing the cameras, but did not speak.
- Sharp represents an innocent pleader. This appeal was given by a police officer (DCI Nick Scola), as well as Sharp. It is 2 minutes 2 seconds in duration, and totals 299 words. Sharp was responsible for 90 of the 299 words. Two other family members sat with Sharp, facing the cameras, but did not speak.
- Hazell confessed and was judged to be involved and this means that he was lying in the ITV interview we draw upon in this paper. In fact, as he gave this interview in the home of Tia's grandmother, he would have known that Tia's body lay in the loft, wrapped in polythene. The Hazell clip that coders were asked to analyse consists of 685 words, and is 3 minutes 39 seconds long.

These cases are well known in Britain – and were known by the coders. This is not problematic, in this instance, as the coders' analyses are not included in this paper in order to demonstrate the predictive potential of SCAnR when it comes to assessing credibility or detecting deception. In fact, we gave careful instructions to our coders not to conclude, given that (*i*) there is no opportunity to probe any PIns which may occur, and (*ii*) we have legal ground truth anyway. Rather, we asked coders to pay special attention to – and, if present, to code – any cross-channel inconsistencies, regardless of the outcome of the cases. In other words, it did not matter whether the individual being coded was an honest or deceptive pleader or an honest or deceptive interviewee.[6] Our third reason for choosing these particular clips is that readers will be able to access some of the above relatively easily using sites such as YouTube. This enables readers to verify the coders' analyses, if they so desire. Finally, the inclusion of interviews, in addition to the press appeals, demonstrates that the SCAnR model can cope with monologues and also dialogues (albeit ones that are quite formal in structure). Note, however, that

SCAnR users are informed of the importance of allowing for the priming effects of the questions of others (Vrij, 2008), when coding their data – especially where this appears to shape recipients' utterances to some degree.

3.3 PIns identified by the coders

The first finding worthy of note is the number of PIns identified by the four coders which averaged 11.25 when it came to the deceptive accounts (within a range from 9 to 15) and only 1.5 for Sharp (the truthful account) over similar time periods (within a range of 1 to 2). The low number of PIns with respect to Sharp (the innocent pleader) means that, overall, coders judged his behaviour to be consistent with appearing on television to make an appeal for the return of a missing loved one. In other words, they expected that all pleaders/interviewees would exhibit (and, hence, allowed for a level of) nervousness, given the context. This said, three of the four identified one or two anxiety indicators as PIns.[7] When asked, the three coders stated that they coded lip-licks/swallows as PIns, even though they may well be a result of context, because they acted as a marker for engaging in more detailed, post-event analysis to eliminate/confirm any possible deception hypothesis (see Section 5). The low frequency of PIns, within the Sharp appeal, appears to be further corroborated by the occurrence of Wright Whelan's (2012) honesty features (e.g., plea for safe return of relative, expression of positive emotion toward the relative, plea for help in finding the relative, expressions of hope for positive outcome). Caution is needed, however, as some of Wright Whelan's honesty features also appeared in our other datasets: they were used by Philpott, for example (see Section 5.1). Future research would thus need to determine whether honest pleaders use more of these features more frequently than deceptive pleaders and perhaps in ways that cluster (as in the case of Sharp's appeal). A larger dataset would also enable the application of statistical measures of significance (see Sections 5 to 5.2).

In contrast to Sharp, all three deceptive interviewees demonstrated a number of behavioural inconsistencies. For Huntley and Hazell, in particular, these inconsistencies tended to occur immediately following each of the interviewers' questions, thereby validating a focus on behaviour within a seven-second window following a question stimulus (Houston *et al*., 2012). Huntley and Hazell were found to demonstrate facial expressions that were inconsistent with their accounts, emerging baselines and/or contexts. Like Philpott, they also exhibited prolonged, asymmetrical facial expressions (FACS codes AU1 and AU4; SCAnR code F1/2/3); for example, inner brows raised and squeezed together as a pose, involving no corresponding, synchronous action from the lip corner depressors (AU15), through an account that was descriptive rather than sad (F1/4).

With respect to the body channel, Hazell used the micro-gesture head shake 'no' (twice): the first, when he affirmed that Tia 'left the house at 10.30' and, the second, when stating that she 'walked out of my house'. Ground truth now affords us the knowledge that both statements constituted lies in this case, as Tia's body lay in the loft, above the heads of Hazell and the interviewer. Huntley and Hazell were also found to display tension in the body which, for the coders, was 'out of proportion with the context/account' (allowing for the fact that the context – interviews –will produce some level of anxiety in the interviewee). This tension co-occurred with firm hand/arm grasping actions. The coders further noted that Hazell's illustrators were suppressed (and, thus, demonstrated a change from previous behaviour) during recall of the events on the day following Tia's murder. The illustrators were not synchronised with his speech through this phase. There were also significant increases in blink rate in response to questions about what Tia said to Hazell as she left and about the chronology of events. Given Mann *et al*. (2002) found that deceivers blink less, it is worth reiterating that the four coders were asked to not conclude in these particular cases, because of not having the opportunity to probe PIns (if found). However, they were empowered to suggest alternative hypotheses for any PIns which did arise, regardless of ground truth, given that miscarriages of justice do occur. In this case, for example, the coders commented that 'heavy' control of body movement could be indicative of contextual nervousness or fear (television interview). They also noted that nervousness/fear can be experienced by both the guilty and the innocent (cf. fear of being caught *versus* fear of the truth being disbelieved). This empowerment to hypothesise is crucial, in our view, given that 'people typically have incorrect

beliefs about cues to deception' (Vrij, 2008: 125). Indeed, the general public tend to associate lying with cues that actually demonstrate nervousness, and fail, thus, to realise which cues are related to deception (and sometimes only weakly). Thanks to phenomena such as confirmation bias and beliefs perseverance, they also 'tend to seek information that confirms rather than disconfirms their beliefs' (Vrij, 2008: 131) – however unhelpful – with the result that people who lie (but do not demonstrate nervousness) often go undetected, and people who are truthful (but who demonstrate nervousness for whatever reason) are falsely labelled as liars. Moreover, because of a lack of adequate feedback in most contexts (especially feedback which is 'given frequently, reliably, and immediately'), people do not tend to 'discover that their views are inaccurate' (Vrij, 2008: 132). This, in effect, allows them to perpetuate their false beliefs – be it by finding examples which appear to support them or by disregarding any evidence that would serve to undermine them.

In terms of the voice channel, the coders noted that, for example, Hazell's volume decreased when he claimed Tia 'walked past me . . . to go out'. And Philpott was found to have inconsistencies with high pitch and high volume: these were noted during portrayals of 'sad' moments in his account. With respect to content, statement credibility issues were highlighted as Pins for Hazell in particular. For example, coders believed Hazell to have used inappropriate detail (C3.3) to describe activities, as opposed to providing contextual detail (C3.4) around the core of the event under discussion (i.e., Tia's disappearance). These were over-structured in a linear form (C3.2) and populated with script memory: 'she wanted a sausage roll, well, cos she's always eating sausage rolls'; 'she doesn't take her washing up out'; 'she uses it [phone] as it's charging so there's no charge going through'. These examples contain an additional linguistic feature, which was used by Philpott as well as Hazell: third-person reference to children in (apparent) preference to using their personal names. This particular feature represented a PIn (C2) for the coders because of (potentially) pointing to 'distancing language' (DePaulo *et al*., 2003). We will return to the use of distancing by deceptive pleaders and interviewers in Sections 5.1 and 5.2.

4. SCAnS

We have a comprehensive Six Channel Analysis System (SCAnS) to allow users to revisit and further explore PIns identified in realtime at three sublevels, using SCAnR, through:

- Transcribed accounts of an interaction (=text);
- Voice recordings of the interlocutors (=audio); and,
- Visual recordings of the participants involved (=video).

The creation of transcriptions, as part of this multi-tiered system (akin to ELAN),[8] cannot be done effectively in realtime as yet. Transcriptions are, nonetheless, worthwhile, in our view: they afford SCAnS users the opportunity to engage in computer-assisted breakdowns of linguistic content. Jurafsky *et al*. (2009) have had some success in automatically coding for backchannels, appreciations and collaborative completions, for example. Automated linguistic analysis, in turn, opens up the possibility of searching for and counting the occurrences of specific features, and (where the size of the datasets allow) applying statistical measures to them (examples can be seen in Sections 5.1 and 5.2). As a rule, such automated Corpus Linguistic (CL) techniques work better if the necessary time is given to processing (preferably large) bodies of text. Yet, the SCAnR method is designed to be undertaken in realtime on individual datasets which, as this paper has established, can be very small. Currently, it is very difficult to carry out meaningful and accurate automated linguistic analyses in realtime. If or when it becomes possible for users to have access to such results realtime (as opposed to post-event), we will, therefore, need to consider whether simultaneous data overload leads to a deterioration of performance in users' realtime assessments across the six channels (as opposed to a better reading of people).

The issue of data overload is especially important to us, since most of our real-world users need to garner cross-channel information in real time. Where CL techniques will probably be most useful, then, is in allowing for further scrutiny of their datasets shortly thereafter (or, more realistically, at a later

date), should this be deemed necessary. This highlights an issue between the more applied (professional) fields and academic disciplines, in our view: the SCAnR method encourages users to pass over consistencies and, instead, focus on inconsistencies. Users are not undertaking the comprehensive analyses that are typical of academic disciplines such as forensic linguistics and CL. Within applied, professional contexts, however, decisions on next questions and action(s) have to be made in the moment with no permission or opportunity for recorded playback and analysis. As well as being a luxury in such working environments, time also has financial implications: hence, professionals across various fields have to be convinced that a second (and sometimes third, fourth and fifth) examination of the data will prove fruitful (by, for example, pointing to something the initial investigation did not, as opposed to validating something they had 'already identified'). Professionals also have to consider the effect of over-analysis: in an airport context a passenger spends around ninety minutes in the terminal and only three to twelve minutes engaged with airport staff who are responsible for gauging the possibility of mal-intent and deception. Any more than this creates delays and customer dissatisfaction.

In the remainder of this paper, we begin to address real-time and post-event analysis by assessing the value of incorporating tools like Wmatrix3 (Rayson, 2008) within SCAnS. Specifically, we will use Wmatrix3 to analyse the datasets described in Section 3.1, before comparing our results with McQuaid *et al.* (2015). Whereas we are using Wmatrix3 to compare two deceptive interviewees, one deceptive pleader and one honest pleader, McQuaid and colleagues used the same tool to compare the language of thirty-five deceptive pleaders and forty-three innocent pleaders (using data provided by ten Brinke and Porter, 2012; see Section 2). Like ten Brinke and Porter, McQuaid *et al.* (2015) focussed on both the pleas as a whole, and also the direct part of the pleas only, on the assumption that a lie should be easier to detect in the latter, given the deceptive pleader 'is either asking for information that they already know, or speaking directly to an individual whom they know is dead (in the case of addressing the missing family member) or non-existent (in the case of addressing an unknown perpetrator)' (2015: 10). However, they later asserted that there are probably more merits in analysing the whole plea, as opposed to restricting analysis to the direct plea.

5. Automated capture of the content of interactions usingWmatrix3

Wmatrix3 is a corpus analysis and comparison tool which annotates running text automatically according to various pre-determined categories: 232 semantic categories and 273 part-of-speech (POS) categories. The tool enables users to glean the frequencies with which particular lexical, semantic-field and POS tokens appear within a given text or texts, and to determine the statistical significance of words, semantic fields or POS, as well as any statistically relevant collocates they may have. Statistical significance (or keyness) is achieved, in this case, by comparing a target text with a normative dataset: words, semantic fields or POS are deemed to be statistically significant when they achieve a Log-Likelihood (LL) value of 6.63 or above (the cut off for 99 percent confidence of significance). In our case, we compared our datasets against the BNC Sampler (spoken), which provided us with 237 keywords, eleven key POS and nineteen key semantic fields (see Section 5.1).

The belief that SCAnS might benefit from incorporating automated linguistic analysis tools like Wmatrix3 stems from two inter-related hypotheses:

- That such tools are able to capture patterns of language use that are typically not easy for human coders to recognise spontaneously or consistently (McQuaid *et al.*, 2015); and,
- That the patterns they uncover point to the 'aboutness' of a given text or texts (Phillips, 1989), as well as to cultural/ideological belief systems (Sinclair, 2004) and, perhaps more relevantly for our purposes, individuals' emotional and cognitive experiences (Pennebaker, 2011), including providing potential cues as to whether or not someone is lying (Pennebaker *et al*., 2003; and Vrij *et al*., 2010).

Although Paul Rayson, the creator of Wmatrix3, does not claim that his tool is able to tap into psychological processes or to detect potential deception, it has been used in this way. We have already mentioned McQuaid *et al.* (2015) in this respect. In addition, Hancock *et al.* (2013) have explored the language of psychopaths using Wmatrix3. For McQuaid *et al.* (2015: 2), in particular, language content is 'potentially more diagnostic of deception' than 'nonverbal/speech-related cues' when it comes to high-stakes lies, and Wmatrix offers an effective means of identifying these content cues. In fact, the authors argue that this particular tool is more effective than both:

(1) Current veracity tools, which rely on human coders (see also Bond and Lee, 2005); and,
(2) Other automated content analysis tools like LIWC, because of 'taking into consideration the context in which [a] word is used' (McQuaid *et al*, 2015: 6).

Although Wmatrix may be more contextually sensitive than LIWC, it is not always as context-sensitive as it might be – not least because it normally assigns the first semtag in a string to a word, when that word has more than one semantic meaning.[9] In addition, the accuracy levels quoted by McQuaid and colleagues (i.e., POS tagging achieving 96–97 percent accuracy and semantic tagging typically achieving 92 percent accuracy) relate to modern general English only. In which case, Wmatrix3 results always need to be verified by the user (as opposed to being trusted unreservedly). With this caveat in mind, Section 5.1 summarises our own Wmatrix3 findings; and Section 5.2 provides a comparison with McQuaid *et al*.'s (2015) findings.

5.1 Wmatrix3 findings

Using LL 6.63 as the cut-off point, Wmatrix3 found:

- Thirty-seven keywords (LL 6.78 to LL 35.42) and four key semantic fields (LL 8.40 to LL 17.51) for Huntley (no POS categories with LL values of 6.63 or higher were found).
- Eighty keywords (LL 6.92 to LL 63.90), seven key semantic fields (LL 7.02 to LL 15.76) and four key POS (LL 7.68 to LL 23.41) for Philpott.
- Nineteen keywords (LL 6.73 to LL 74.94), two key semantic fields (LL 6.91 to LL 10.33) and three key POS (LL 8.54 to LL 24.27) for Sharp.
- 101 keywords (LL 6.88 to LL 220.35), six key semantic fields (LL 7.48 to LL 39.48) and four key POS (LL 8.26 to LL 29.40) for Hazell.

The first thing to note is how much statistically significant informationWmatrix3 provides (which may pose an overload issue for some SCAnS users). As we are dealing with very small datasets in each case, individual LL values highlighted by Wmatrix3 need to be treated with a healthy level of caution. In our case, this has led us to prioritise the key semantic-field and key POS findings (where they do not represent mistaggings), in our analyses, and draw in the keywords where they seem to support the findings of these key semantic fields/POS and/or have very high LL values. In this way, we are still 'reliably identify[ing] all terms representing or relating to a certain semantic category (e.g., terms relating to [S]'s level of knowledge)': which, as McQuaid *et al*. (2015: 4) highlight, is something that human coders often struggle to do systematically and/or reliably. Table 1, then, captures the key semantic fields/POS/words that we discuss explicitly as part of our analyses of the four datasets.

Dataset	Keywords	Key semantic fields	Key POS
Huntley	*There's* (LL 33.70); *seemed* (LL 18.85); *they're*, *I've*, *glimmer*, *St Andrews*, *Miss Carr* (LL 17.51); *cheerful*, *chatty* (LL 14.74); *happy* (LL 14.04); *on the doorstep* (LL 13.69); *absolutely* (LL 13.40); *came across* (LL 12.51); *run away* (LL 11.23); *teach* (LL 10.46); *not at all* (LL 9.91); *stood* (LL 9.38); *direction* (LL 9.20); *library* (LL 8.68);	'light' (LL 17.51); 'happy' (LL 13.58); 'seem' (LL 9.77); 'location and direction' (LL 8.40)	(None found)

	walked (LL 8.12); *just* (LL 7.91); girls (LL 6.78)		
Philpott	*just* (LL 31.95); *please* (LL 25.56); *fire brigade* (LL 24.31); *save* (LL 19.07); *police* (LL 17.18); *organs*, *his*, *grieve*, *disrupting*, *been*, *ambulances*, *Duwayne*, *Daniel Stephenson*, *Butler* (LL 15.97); *gratitude*, *God know*s (LL 13.20); *tried* (LL 12.53); *pass away*, *donate*, *beg* (LL 12.16); *firemen* (LL 10.57); *leave alone* (LL 8.78); *ambulance* (LL 8.26); *Joe* (LL 7.59); *suffering*, *pain* (LL 7.51); *alone* (LL 7.36); *children* (LL 7.19); *cope* (LL 7.10); *nurses* (LL 7.03); *helping*, *doctors* (LL 6.92)	'trying hard' (LL 15.76); 'medicines and medical treatment' (LL 12.22); 'polite' (LL 12.21); 'success and failure' (LL 8.93); 'people' (LL 8.70); 'temperature: hot/on fire' (LL 8.53); 'sad' (LL 7.02)	Possessive pronouns (LL 23.41); obj. person. pronoun *us* (LL 17.31); plural common nouns (LL 15.50); 'been' (LL 7.68)
Sharp	*you're*, *Tia* (LL 37.47); *come home* (LL 21.91); *urge* (LL 12.46); *want* (LL 12.15); *public* (LL 7.76); *phone* (LL 7.45); *trouble* (LL 7.29); *knows* (LL 7.16); *any* (LL 6.73)	'wanted' (LL 10.33); 'law and order' (LL 6.91)	Base form of verb (LL 24.27); 'being' (LL 10.92); 'doing' (LL 8.54)
Hazell	*she's* (LL 220.35); *n't* (LL 81.56); *she* (LL 50.09); *charge* (LL 46.41); *Tia* (LL 29.38); *phone* (24.68); *know* (21.22); *her* (19.54); *you're*, *sausage roll*(*s*), *res-responsible*, *deep down*, *accusations*, *Tia's*, (LL 14.69); toast (LL 14.59); *clock* (LL 14.21); *God's*, *charge up* (LL 11.92); *I'd* (LL 9.29); *angel* (LL 8.96); *loving*, *golden* (LL 8.67), *kitchen* (LL 8.35); *a little while* (LL 7.82); *my* (LL 7.50); *on her own* (LL 7.36); *charging* (LL 7.23)	'Religion and the supernatural' (LL 39.48); 'Moving, coming and going' (LL 22.18); 'Knowledgeable' (LL 14.88); 'Telecommunications' (LL 9.54); 'Pronouns' (LL 8.21); 'Food' (LL 7.48)	Past tense of verb (LL 29.40); third person *he*, *she* (LL 19.30); –*s* form of verb (LL 11.49); possessive pronouns (LL 9.22); base form of verb (LL 8.26)

Table 1: Keyness results for the four datasets

As Table 1 reveals, the 'key' language features of the two pleaders, Sharp and Philpott, are markedly different. A possible reason for this, beyond innocence and guilt, is that Sharp was appealing for help in finding a missing loved one, whilst Philpott was talking about the loss of loved ones in a house fire. Hence, 'wanted' is a key semantic field in Sharp's speech, alongside keywords such as *come home*, *urge*, *want*, *public* and *knows*. We would argue that Sharp's keyness terms (across the word, semantic field and POS levels) appear to support the coders' decision to code very few PIns (none of which were linguistic in nature). Indeed, they very much point to Wright Whelan's (2012) honesty features of pleading for the safe return of, or help from, the public in finding a relative, expressing positive emotions toward them – by, in this case, reassuring Tia to 'come home' and that she was 'not in any trouble', *etc*. Simply put, Sharp's speech seems to 'fit' (i.e., be consistent with) the context.

Philpott also makes use of one of Wright Whelan's (2012) honesty features: the avoidance of 'brutal language' in favour of the euphemism, 'passing away', with reference to his son, *Duwayne* (the last child to die as a result of the house fire). A particular honesty feature – pleading for help in finding the perpetrator(s) – is noticeable for its absence, however, given that the police were still looking for the perpetrator(s) at this time (at least officially). Keywords to capture our attention, because of having a pleading component, were *beg*, *leave alone*, *please* and *grieve*. Philpott used them as part of a request to the general public:

> There's one thing I would request is please please leave my family alone. If you've got any questions or anything at all, please don't come through me or my family, please go to the police, because what's happening at the moment, you are disrupting what these officers are trying to do. So please I beg you leave us alone and let us try and grieve in peace and quiet. That's all I ask. Thank you.

Although the SCAnR coders did not raise this direct plea as a PIn, it contains elements which might be captured by two of our SCAnR criteria:

- Representational frames relating to Other(s) – in this case, the 'disruptive' press/public and the 'hindered' officers (I3);
- Repetition/inappropriate politeness being used as a possible influencing technique (I3); and,
- The use of (overly strong) emotional terms, given what is being discussed at the time (*beg*; C2).

Another example of potential impression management (and, hence, I3) within the Philpott transcript relates to Philpott's mention of Duwayne. Specifically, Philpott signalled his and his wife's intention to donate Duwayne's 'organs to save another child' (so that good might come from tragedy), and asserted that helping another child to live helped to 'take a bit of the pain away' for them both. A possible way of making use of Wmatrix3 within SCAnS, in which case, is in raising criteria which (like the above) 'fit' one or more of the twenty-seven PIns, regardless of whether they have been identified by a human coder (see Section 6).

Apart from the above, the bulk of Philpott's appeal functioned more like a thank you speech to the professionals/neighbours who had tried valiantly to rescue the children from the burning house: hence, Wmatrix3 assigning statistically key semantic fields such as 'trying hard', capturing words such as *tried*, *effort*, *trying* and *try*; 'medicines and medical treatment', capturing *ambulance*(*s*), *doctors* and *nurses*; 'polite' (LL 12.21), capturing *thank* and *gratitude*; and 'temperature: hot/on fire' (LL 8.53), capturing *fire brigade* and *firemen*. Some of these words also function as keywords in their own right (see Table 1). Additional keywords to support the aforementioned 'aboutness' hypothesis include *Daniel Stephenson*, the *Butler* brothers and *Joe* (the neighbours who tried to save the children).

Interestingly, this referencing of the neighbours using their personal names stands in stark contrast to how Philpott referenced the children: only Duwayne was mentioned by name (and, thus, constitutes a keyword for Philpott). This may be an example of psychological distancing on Philpott's part, thus supporting the coders' hypothesis that Philpott used such distancing techniques (e.g., C2). In this case, Wmatrix3 might be seen as a useful means of (linguistically) verifying the (in)validity of some PIns.

Huntley and Hazell's datasets capture their contributions in televised interviews. It is possible, then, that some of the keyness results – and, in particular, their key semantic fields – have been primed by the questions of the respective interviewer. Thus, as the last known people to see the children alive, they were asked, and hence provided details of, the movements of the children, such that 'location and direction' was a key semantic field for Huntley, and 'moving, coming and going' was a key semantic field for Hazell. As Section 3.3 reveals, the SCAnR coders felt able to identify statement credibility issues as PIns, particularly for Hazell. They included Hazell using (what, for the coders, constituted) inappropriate detail/relying on script memory to describe activities, as opposed to providing detail around the core of the episode under discussion (i.e., Tia's disappearance): 'she wanted a sausage roll, well, cos she's always eating sausage rolls'; 'she doesn't take her washing up out'; 'she uses it [phone] as it's charging so there's no charge going through'. This particular finding is not easy to discern and, hence, (in)validate using the keyness facilitiy in Wmatrix3 – unless we opt to focus on specific keywords like *charge*, *charging*, *phone*, etc.: although *eating* and *sausage roll*(*s*) were part of the key semantic field, 'food', and *phone*, the reason for the key semantic field, 'telecommunications'. Given the inconspicuous nature of these key words/semantic fields in light of the macro topic (Tia's disappearance), it is debatable whether Wmatrix users would opt to focus on them without being specifically prompted to do so for some reason (such as SCAnR coders prioritising them as PIns potentially pointing to inappropriate detail/reliance upon script memory). There is strong overlap between some of the coders' PIns and Hazell's key POS results, however. *She* and *she's* constituted key POS and keywords for Hazell, for example. *She* is also part of Hazell's key semantic field of 'pronouns' (along with *I*, *it*, *you*, *her*, *my*, *they*, *we*, *me*, *that*, *its*, *your*, *whatever*, *anything*, *something* and *them*). Such results corroborate the coders' belief that Hazell used the third person significantly. They hypothesised further that this may have been a potential distancing technique. Third-person reference has been linked to distancing language by, for example, DePaulo *et al*. (2003). McQuaid *et al.* (2015) suggest, further, that deceptive pleaders, in particular, will tend to make more use of specific third-person references, statistically speaking, when compared to honest pleaders – that is, *they*, *somebody*

and *anybody* (we outline McQuaid *et al*.'s study in Section 5.2). The authors believe that 'a guilty individual [may] subconsciously distance him or herself from the victim or the crime' in this way 'due to [their] having guilty knowledge' (McQuaid *et al.*, 2015: 9) and so they advocate that researchers be especially alert to words indicating a level (and especially lack) of knowledge on S's part. Hazell is not a pleader, of course: he is an interviewee. Nonetheless, 'knowledgeable' is a key semantic field for him (Philpott and Sharp, in contrast, have *God knows* and *know* as keywords, see Table 1). Hazell used *know* thirteen times in his interview. In contrast to McQuaid *et al*.'s (2015) finding that dishonest pleaders avoid temporal words (when compared to honest pleaders), some of Hazell's uses of *know* relate to Sharp convincing the interviewer that he is 'sure' about specific times: 'Tia's come down the stairs urrhhh mmm round about uh half past ten eleven something like that maybe. I know cos I know she was on about going up to Croydon [. . .]'. Other uses of *know* include seeking clarification of the interviewer's understanding of (and, potentially, support for) a proposition made by Hazell: '[. . .] she's she's a happy go lucky golden angel do you know what I mean she's she's she's perfect [. . .]'. *Angel*, in turn, explains why 'religion and the supernatural' is a key semantic field for Hazel (see Table 1).

Know is also used by Hazell to claim limited knowledge: '[. . .] she walked out the front door and that is all I know [. . .]'. Interestingly, this particular example occurs at one of the points at which Hazell provides extraneous information outside the context of the incident – in this case around charging the phone (i.e., the type of behaviour which constituted a PIn for the coders: see above and also Section 3.3). Such 'extraneous information' has also been identified as a potential deception indicator by Wright Whelan *et al*. (2013), alongside speech disfluencies and repetition – that is, the features which Hazell also exhibits in the above turns.

Given that Hazell seems to contrast significantly with McQuaid *et al*.'s (2015) findings, let us turn our attention, now, to this study.

5.2 Comparison with McQuaid *et al.* (2015)

Prior to undertaking their study, McQuaid *et al*. (2015: 8) hypothesised that they would find:

> [. . .] differences in the language of genuine and deceptive pleaders that may be indicative of factors previously linked to deceptive behaviour (e.g., psychological distancing and verbal leakage). Deceptive pleaders were expected to use language that distanced them from the truth (potentially in a non-conscious effort to minimize the difficulty of deceiving others), whereas genuine pleaders were expected to use language that personalized the situation for them. Certain aspects of language were expected to be leaked more often by deceptive pleaders (such as terms relating to an individual's level of knowledge) due to the cognitively demanding task of maintaining a lie, particularly an extremely high-stakes lie.

They tested these hypotheses by comparing the language of deceptive pleaders with the language of honest pleaders and *vice versa*, whilst paying particular attention to certain semantic categories within Wmatrix3: level of knowledge (e.g., 'dunno', 'insight' and 'anybody's guess'), words describing personality characteristics (e.g., *friendly* and *generous*), and physical attributes or appearance of the missing person (e.g., *beautiful*), self and other references, and the use of discourse markers. As with our study, they were particularly interested in statistically significant items with a LL value of 6.63 or above.

Although McQuaid and colleagues highlight that the 'knowledgeable' semantic field may be a useful semantic field to check for, given that it was used by both deceptive and honest pleaders, the semantic field was not key for either group. One reason for this may be to do with their decision to compare the former with the latter (and *vice versa*) such that any statistical results highlight the differences between the two groups and, in effect, background any similarities they may share. A second reason may relate to theWmatrix program itself not being able to pick up some instances as being indicative of knowledge (or lack of); for example, statements like 'whoever's got her', which 'imply that the pleader does not know who has taken the victim' (McQuaid *et al.*, 2015: 11).

McQuaid *et al*.'s (2015) decision to compare the language of the deceptive and honest pleaders does reveal some interesting statistical differences, however. Features that are said to characterise the language of the forty-three honest pleaders, when compared to the dishonest pleaders, included the use of more temporal words (in particular, *days* and *weeks*). This was taken to mean that honest pleaders were placing their experiences within a more detailed temporal (and, hence, arguably more concrete) context, as a means of 'placing the disappearance', murder or 'investigation in a concrete, and arguably more real, temporal context' (2015: 10). McQuaid *et al*. (2015: 10) hypothesise that deceptive pleaders probably used 'fewer tangible time words to avoid providing temporal details that' are false and/or 'they may eventually lose track of' (but see Hazell: Table 1).[10] Honest pleaders were also found to use *we* more than their deceptive counterparts (but in the entire pleas only). McQuaid and colleagues believe that, in this case, *we* was indicative of more personalised self-referencing as a family unit (rather than as an individual), and that it afforded honest pleaders a means of steering the public's attention towards 'the pleaders and their feelings' in the hope of enhancing 'the public's desire to assist' them. As previously mentioned (see Section 5.1), the thirty-five deceptive pleaders had a (statistical) preference for third- as opposed to first-person reference: *they* (LL 9.26) and singular indefinite pronouns such as *anyone* and *somebody* (LL 14.47). They also used exclusivisers/particularisers, such as *just* (LL 13.98), more than honest pleaders. This statistically significant use of *just* (compared to honest pleaders) is seen by McQuaid *et al.* (2015) to serve a number of inter-related purposes:

- To make the words that follow sound more salient;
- A conscious technique to present the deceptive pleader as more distraught (in the hopes of masking a lie);
- A distancing technique (see also Wright Whelan *et al.*'s 2013 inclusion of *just* in their equivocation category); and,
- A 'stalling for time' mechanism (be it conscious or unconscious).

McQuaid *et al.* (2015) is an excellent example of one means of combining CL techniques and forensic linguistic knowledge in order to tease out the differences between the language of, in their case, innocent and deceptive pleaders. But, as the authors themselves would emphasise, these predictable differences always need to be (in)validated by the context. *Just* constitutes a keyword for two of our deceptive interlocutors: Huntley and Philpott (see Table 1 for LL values). Huntley's discourse markers cluster at the point at which he lied about Jessica and Holly having left 'in the direction of the library' following a brief discussion with him, and, thus, act as distancing devices (along with the verb *seem*):

> They just came across, and asked how Miss Carr was, as she used to teach them at St. Andrews. Erm, I just said she weren't very good, as she hadn't got the job. And they just said please tell her that we're very sorry. And off they walked in the direction of the erm library over there. They seemed fine, very cheerful, happy, chatty. I didn't see anything untoward. Nobody hanging around. Just seemed like normal, happy kids.

Philpott's uses of *just* also co-occur (in this case, with *actually* and *can't believe*):

> Erm, I've actually been down to my our- our home and what we saw, we- we just can't believe it. We grew up in a community that's been had a lot of problems with violence and- and God knows what else, and to see this community to- to come together like this is just- it's just too overwhelming.

Philpott's usage of *just* seems to have less to do with distancing, and more to do with making his assertions more salient. The lesson here, then, is that even when deceptive pleaders or deceptive interviewees use a feature known to correlate with deception, we still need to check how specific individuals are using these features in context. This explains our position that a PIn does not point to deception, but, rather to a feature that is worthy of further probing/investigation.

6. Concluding comments and future work

As we have demonstrated in Section 3.3, the prioritisation of PIns has the capacity to provide information worthy of further probing without creating data overload (cf. users having to attend to all available cross-channel information). In addition, users benefit from the specification of particular PIns known to be correlated with possible deception. Nonetheless, it is important to engage in ongoing work that (in)validates the legitimacy of such PIns, as potential deception indicators. Such work, in addition, may point to PIns not yet included within SCAnR. Sections 5 to 5.2 are the start of such a process, from a (corpus) linguistic perspective. We are also exploring ways of incorporating other automated approaches within SCAnS in order to support the manual analyses of, for example, the voice, IS and ANS (see Section 5). Access to voice recordings would mean that, where relevant or necessary, users could capture automatically derived acoustic representations relating to pitch (FO measures), volume (amplitude measures) and pauses. The latter, in particular, is important when considering two features often used to determine deception: S's speech rate and articulation rate (ten Brinke and Porter, 2012) because speech rate and articulation rate can only be accurately measured post utterance (and, preferably, post speech, after S has finished speaking). Recent technological innovations and advances mean that it should also be possible to incorporate tools within SCAnS, that provide ANS information beyond the physiological signals which can be humanly seen or heard. We are currently experimenting with:

- Thermal cameras to help users to 'see' local blood circulation;
- Pixel movement/colour changes (amplified by a factor of a hundred) so that users can pick up skin colour changes and movement/micro-expressions that the naked eyemay not (including a pulse);
- 'Kinect'-type technology – used in computer gaming – to pick up the slightest body movement; and,
- Remote pupilometry, which is making advances into indexes of cognitive activity/load.

The above bring challenges, as well as potential benefits. The effect of overt technology can create behavioural changes which contaminate evaluations, as we know from the criticisms around the field of polygraphy. The ethics and practicalities of using covert equipment complicates real-world analysis. In addition, if and when it becomes possible for users to have access to the results of such tools during real-time (as opposed to post-event) analysis, we will need to consider whether simultaneous data overload leads to a deterioration of performance in users' real-time assessments across the six channels. A related issue is achieving 'buy in' from professionals (working in high-stakes contexts) who have limited time and capacity to capture simultaneous, multi-modal data from a human, especially when they are working alone (Buckley, 2012). They may also be restricted (by ethics, protocols, legislation, *etc*.) from using covert or overt technology. This includes police, military personnel, security professionals, negotiators, medics, purchasers, social workers, and so on. Such analysts need, therefore, an effective way of maximising their performance in short, real-time situations. Current CL techniques require too much processing time to be effective in such cases (as well as transcripts on which to base their results –which, at present, are extremely difficult to generate successfully in real-time). If processing time could be improved sufficiently, CL assumptions around the use of random sampling and representativeness would be difficult to address, nonetheless, especially where analysis as soon as possible after the event is the aim, because we are seeking to help practitioners better understand what specific individuals do in context as efficiently and quickly as possible.

Another (and possibly more important) issue for the user is having access to technologies which are as supportive and user friendly as possible (rather than providing lots of data to be analysed). One possible route may be for SCAnS users to consider only semantic fields/POS found to be key and to ignore everything else – in which case, users would need to consider what type of normative corpus to use, in order to achieve the best keyness results (as well as whether such a normative corpus already existed, or would need to be created). A second route may be for SCAnS users to have a more targeted approach similar to McQuaid *et al.* (2015) by, for example, prioritising specific Wmatrix semtags (or similar) given what we know about deception currently, and looking for evidence of statistical significance

therein (through, for example, concordance results). Once again, however, users would need to think carefully about their normative corpus. This approach would also take time and, thus, appears to preclude real-time analysis completely. As Archer (2014) has shown, it is possible to use Wmatrix semtags to trace pragmatic phenomena but without prioritising statistical significance. This might be extended, in the case of credibility/deception studies, to include specific words, POS, semantic fields, *etc*. (whilst being careful to check the results of such targeting, to ensure that features such as *just* are used to achieve deception in context; see Section 5.2). Once again, then, time will prove to be a factor.

There appears to be much more mileage in CL techniques playing a useful role in helping us (and other researchers) to determine the validity of language features previously identified as potential deception indicators. Indeed, as Sections 5.1 and 5.2 reveal, Wmatrix3 is able to (in)validate PIns. It also has the potential to point to things not yet included in our (or others') list of potential deception indictors. A benefit of the SCAnS approach, which applies to the CL discipline more generally, would be to provide a means of linking linguistic analysis back to the visual and auditory communication channels, since SCAnS and SCAnR are designed on the premise that we have to pay attention to all six of the communication channels if we are to truly understand communication.

Acknowledgments

We would like to thank our coders for their SCAnR analyses of the Hazell, Huntley, Philpott and Sharp datasets, the two anonymous reviewers, for their insightful comments in respect to an earlier version of the paper, and the guest editor, Sam Larner, for his help throughout. All remaining infelicities are ours.

Endnotes

1 Our experience suggests that professionals operating within high-stake contexts share our view of the folly of making veracity judgments on video subjects where PIns cannot be tested by skilful questioning, and also prefer the safer option of identifying signals 'to be researched further . . . by subtle and skilled questioning' (Pearse and Lansley, 2010).
2 The design and practice of such prompting is outside the scope of this paper.
3 The EIA team have catalogued/illustrated each action/signal with a still photograph or video, and assigned it a specific AD reference number. The photographs and videos are not to be taken too literally, as they are designed to exemplify, not typify. There will be slight variations from person to person due to their anatomy and flexibility.
4 CBCA assumes that 'a statement derived from memory of an actual experience differs in content and quality from a statement based on invention or fantasy' (Vrij, 2008: 209).
5 This aspect of SCAnR may prove to be the most contentious, given that CBCA was developed to assess witness credibility, not whether a person was telling the truth or being deceptive, and has previously struggled to distinguish short lies (non-experienced elements) within otherwise truthful stories (i.e., experienced events; see Vrij and Mann, 2001). In our defence, we point to the fact that we are not the only researchers to use CBCA in deception detection research (see, for example, Colwell, 2007; and Vrij *et al*., 2004, 2007), and the fact that our adapted version of CBCA is but one component of the SCAnR method.
6 It is possible that the coders were primed to notice more PIns in the deceptive cases, and less in the case of Sharp, because of having ground truth. However, as part of their training, coders are asked to code cases for which they do not have ground truth, but we do (without concluding), and a similar pattern emerges to that reported in Section 3.3 (i.e., more PIns in the deceptive cases). They also code cases, like the above, where ground truth is known but discuss the possibility of potential miscarriages of justice (especially where those involved are judged guilty in law but continue to claim their innocence).
7 Lip-licks/swallows are suggestive of physiological resources being channelled towards the sympathetic nervous system (the automatic fight/flight response towards a threat) and away from the non-essential digestive process.
8 ELAN is a professional tool for the creation of complex annotations, with video and audio in mind (for details, see: https://tla.mpi.nl/tools/tla-tools/elan/).
9 McQuaid *et al.* (2015) note at least one example of *just* being mis-assigned by Wmatrix3, for example (see Section 5.1).
10 Deceptive pleaders' lack of temporal terms was also explained to be 'consistent with ten Brinke and Porter's (2012) finding that more tentative words were used by deceptive than genuine pleaders' (McQuaid *et al.*, 2015: 10).

References

Archer, D. 2014. 'Exploring verbal aggression in English historical texts using USAS: the possibilities, the problems and potential solutions' in I. Taavitsainen, A.H. Jucker and J. Tuominen (eds) *Diachronic Corpus Pragmatics*, pp. 277–301. Amsterdam and Philadelphia: John Benjamins.

Bond, G.D. and A.Y. Lee 2005. 'Language of lies in prison: linguistic classification of prisoners' truthful and deceptive natural language', *Applied Cognitive Psychology* 19, pp. 313–29.

ten Brinke, L. and S. Porter 2012. 'Cry me a river: identifying the behavioral consequences of extremely high-stakes interpersonal deception', *Law and Human Behavior* 36(6), pp. 469–77.

Buckley, J.P. 2012. 'Detection of deception researchers need to collaborate with experienced practitioners', *Journal of Applied Research in Memory and Cognition* 1, pp. 126–7.

Colwell, K. 2007. 'Assessment Criteria Indicative of Deception (ACID): an integrated system of investigative interviewing and detecting deception', *Journal of Investigative Psychology and Offender Profiling* 4(3), pp. 167–80.

DePaulo, B.M., J.J. Lindsay, B.E. Malone, L. Muhlenbruck, K. Charlton and H. Cooper. 2003. 'Cues to deception', *Psychological Bulletin* 129, pp. 74–118.

Ekman, P. 2004. *Telling Lies: Clues to Deceit in the Marketplace, Politics and Marriage*. New York: W.W. Norton.

Ekman, P. and M. Frank. 1993. Not all Smiles are Created Equal: The Differences between Enjoyment and Non-Enjoyment Smiles. San Francisco: University of California.

Ekman, P., W.V. Friesen and J.C. Hagar. 2002 [1976]. *Facial Action Coding System: The Manual on CD-ROM*. Salt Lake City, Utah: Network Information Research.

Hancock, J. and M. Woodworth. 2013. 'An "eye" for an "I": the challenges and opportunities for spotting credibility in a digital world' in B.S. Cooper, D. Griesel and M. Ternes (eds) *Applied Issues in Investigative Interviewing, Eyewitness Memory, and Credibility Assessment*, pp. 325–40. Springer: New York.

Hancock, J.T., M.T. Woodworth and S. Porter. 2013. 'Hungry like the wolf: a word-pattern analysis of the language of psychopaths', *Legal and Criminological Psychology* 18(1), pp. 102–14.

Houston, P., M. Floyd and S. Carnicero. 2012. *Spy the Lie*. New York: St Martin's Press.

Jurafsky, D., R. Ranganath and D. McFarland. 2009. 'Extracting social meaning: identifying interactional style in spoken conversation', *Proceedings of the 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies*, pp. 638–46.

Levine, T.R., J.P. Blair and D.D. Clare. 2014. 'Diagnostic utility: experimental demonstrations and replications of powerful question effects in high stakes deception detection', *Human Communication*
Research 40(2), pp. 262–89.

Mann, S., A. Vrij and R. Bull. 2002. 'Suspects, lies, and video tape: an analysis of authentic high-stakes liars', *Law and Human Behavior* 26(3), pp. 365–76.

McQuaid, S.M., M. Woodworth, E.L. Hutton, S. Porter and L. ten Brinke. 2014. 'Automated insights: verbal cues to deception in real-life high stakes lies', *Psychology, Crime and Law*, pp. 1–31.

Newman, M.L., J.W. Pennebaker, D.S. Berry and J.M. Richards. 2003. 'Lying words: predicting deception from linguistic styles', Personality and Social Psychology Bulletin 29(5), pp. 665–75.

Pearse, J. and C. Lansley. 2010. 'Reading others', *Training Journal*. October 2010, pp. 62–5. Available online at: www.TrainingJournal.com

Pennebaker, J.W. 2011. *The Secret Life of Pronouns: What our Words Say about Us*. New York: Bloomsbury Press.

Pennebaker, J.W., M.E. Francis and R.J. Booth. 2001. *Linguistic Inquiry and Word Count (LIWC): LIWC2001*. Erlbaum Publishers: Mahwah, New Jersey.

Pennebaker, J.W., M.R. Mehl and K.G. Niederhoffer. 2003. 'Psychological aspects of natural language use: our words, our selves', *Annual Review of Psychology* 54, pp. 547–77.

Phillips, M. 1989. *Lexical Structure of Text*. Birmingham: University of Birmingham.

Porter, S. and L. ten Brinke. 2008. 'Reading between the lies: identifying concealed and falsified emotions in universal facial expressions', *Psychological Science* 19, pp. 508–14.

Porter, S. and J.C. Yuille. 1996. 'The language of deceit: an investigation of the verbal clues to deception in the interrogation context', *Law and Human Behavior* 20, pp. 443–59.
Rayson, P. 2008. 'From key words to key semantic domains', *International Journal of Corpus Linguistics* 13(4), pp. 519–49.
Rockwell, P., D.B. Buller and J.K. Burgoon. 1997. 'The voice of deceit: refining and expanding vocal cues to deception', *Communication Research Reports* 14(4), pp. 451–9.
Sinclair, J.M. 2004. *Trust the Text: Language, Corpus and Discourse*. London: Routledge.
Undeutsch, U. 1967. 'Beurteilung der Glaubhaftigkeit von Aussagen' in U. Undeutsch (ed.) *Handbuch der Psychologie Vol. 11: Forensische Psychologie*, pp. 26–181. Göttingen, Germany: Hogrefe.
Vrij, A. 2008. *Detecting Lies and Deceit: Psychology of Lying and its Implications for Professional Practice*. Second edition. Chichester: Wiley.
Vrij, A. and S. Mann. 2001. 'Who killed my relative? Police officers' ability to detect real-life high-stakes lies', *Psychology, Crime and Law* 7(1–4), pp. 119–32.
Vrij, A., W. Kneller and S. Mann. 2000. 'The effect of informing liars about Criteria-Based Content Analysis on their ability to deceive CBCA raters', *Legal and Criminological Psychology* 5, pp. 57–70.
Vrij, A., G.R. Semin and R. Bull. 1996. 'Insight into behavior displayed during deception', *Human Communication Research* 22(4), pp. 544–62.
Vrij, A., L. Akehurst, S. Soukara and R. Bull. 2004. 'Let me inform you how to tell a convincing story: CBCA and reality monitoring scores as a function of age, coaching, and deception', *Canadian Journal of Behavioural Science* 36(2), pp. 113–26.
Vrij, A., K. Edward, K.P. Roberts and R. Bull. 2000. 'Detecting deceit via analysis of verbal and nonverbal behavior', *Journal of Nonverbal Behavior* 24(4), pp. 239–63.
Vrij, A., P.A. Granhag and S.B. Porter 2010. 'Pitfalls and opportunities in nonverbal and verbal lie detection', *Psychological Science in the Public Interest* 11(3), pp. 89–121.
Vrij, A., S. Mann, S. Kristen and R.P. Fisher. 2007. 'Cues to deception and ability to detect lies as a function of police interview styles', *Law and Human Behavior* 31(5), pp. 499–518.
Wright Whelan, C. 2012. *High Stakes Lies: Identifying and Using Cues to Deception and Honesty in Appeals for Missing and Murdered Relatives*. PhD thesis. University of Liverpool, UK.
Wright Whelan, C.W., G.F. Wagstaff and J.M. Wheatcroft. 2013. 'High stakes lies: verbal and nonverbal cues to deception in public appeals for help with missing or murdered relatives', *Psychiatry, Psychology and Law* 21(4), pp. 523–37.