Detecting advance fee fraud emails using self-referential pronouns: A preliminary analysis

Rofiat Alli[a], Rebecca Nicolaides[a], Russell Craig[a]*

[a]Portsmouth Business School
University of Portsmouth
Portsmouth, UK

ABSTRACT

We promote awareness of the features of emails that propose advanced fee fraud schemes. These are commonly known as 419 emails (after Section 419 of the Nigerian Penal Code). We outline the structural features of 419 emails and conduct a preliminary study of their distinctive linguistic features, using word frequency counts and DICTION text analysis software. We find that the incidence of first person singular pronouns is seven times greater in 419 emails than non-419 emails. We suggest elements of a future research agenda that can build on our preliminary results to help reduce advanced fee fraud.

*** Corresponding author**
Russell Craig
Portsmouth Business School
University of Portsmouth
Richmond Building, Portland Street
Portsmouth , PO1 3DE
United Kingdom

Primary e-mail: Russell.Craig@port.ac.uk
Secondary e-mail: Russell.Craig@vu.edu.au

Detecting advance fee fraud emails using self-referential pronouns: A preliminary analysis

1. Introduction

Fraud is a misstatement or false representation that is intended to deceive for personal (usually financial) advantage (Action Fraud, 2010). Advance Fee Fraud (AFF) is the most frequently encountered and successful type of fraud in history (Ofulue, 2010; Garrett, 2014). AFF involves a request for advance fees or upfront payments by a dishonest person from a victim, for resources, goods or services that never materialize (Action Fraud, 2010; FBI, 2010; Ultrascan AGI, 2014). AFF schemes commonly employ two or more other kinds of fraud (such as impersonation fraud, identity theft, and/or phishing) (Edelson, 2003).

In this research note, we focus on a specific kind of AFF, known widely as *419 fraud.* This type of fraud is claimed to be "one of the longest running, most successful, omnipresent [and] transnational [frauds] … in history" (Ultrascan AGI, 2014; p.16). *419 fraud* involves:

> … scheme(s) designed by fraudsters purporting to have lucrative but bogus business, humanitarian or philanthropic related deals where the victim is promised large sums of money for no initial investment.… [and]… is persuaded to advance some cash to the scammer for a variety of purposes such as the payment of unanticipated taxes, duty fees and outright bribery. The victim will later discover that there is no fortune to be retrieved or business to profit from, as the scammer disappears with the advance fee he or she had collected (Onyebadi & Park, 2012, p. 182)

Attempts to perpetrate this particular type of fraud are regularly experienced by many members of society, including accountants and their clients. Indeed, typically, government regulatory agencies throughout the world advise persons who are suspicious of any email from a stranger seeking upfront payment of funds to seek advice from "an accountant or financial planner if in doubt" (Australian Competition and Consumer Commission,

Scamwatch website, https://www.scamwatch.gov.au/types-of-scams/unexpected-money/inheritance-scams).

The research note that follows should be particularly beneficial in raising the consciousness and understanding of those with interests in forensic accounting, especially fraud detection. Our aim is to provide a portrait of *419 fraud* and to draw on that portrait to offer some preliminary insight to how *419 fraud* can be detected. We introduce the prospect that the computer aided text analysis software, *DICTION*, can assist in the detection of *419 fraud*. We do not canvass the theoretical underpinnings of fraud. For an understanding of this, see, for example, Choo and Tan (2007) and Schuchter and Levi (2015).

The term *419 fraud* originated in Nigeria and is derived from Section 419 of the Nigerian Penal Code that deals with false pretences (Nykodym & Taylor, 2004; Oriola, 2005; Salu, 2005). The term is used widely in international parlance to describe various fraudulent schemes perpetrated within or outside Nigeria (Adogame, 2009; Chawki, 2009). *419 fraud* has flourished for many decades and has defrauded thousands of "curious, naïve, and/or sympathetic" victims (individuals and companies) of cash and other assets — sometimes resulting in tragic mental health deterioration or even suicide (Glickman, 2005, p. 463). The extent of the monetary losses resulting from *419 fraud* has been estimated by Ultrascan AGI (Advanced Global Investigation). This is an international consulting firm comprising "51 partners managing 3284 experts in 69 countries" which focuses on anti-money laundering and transnational organized crime (http://www.ultrascan-agi.com/public_html/html/about.html). According to Ultrascan AGI, by 2013 over $US 82 billion had been lost to *419* Advanced Fee Fraud, with $US 12.7 billion lost in 2013 ("419 Advance Fee Fraud Statistics 2013", http://www.ultrascan-agi.com/public_html/html/419_statistics.html). A recent case of AFF involved "a lonely [English] beancounter" who was jailed after " he fell for … a classic Nigerian email scam,

and conned £150,000 out of a friend so he could bankroll his fake damsel in distress" (Martin, 2016).

Much legislation has been passed, especially in Nigeria, to curb *419 fraud*, but without success. Media education campaigns have encouraged the adoption of procedures (including recourse to the advice of accountants) to protect proprietary information and to lead to safer and more disciplined use of computers and the Internet. Several scam-baiting sites have also been developed, such as *419eater* (http://www.419eater.com/). Despite such efforts, there has been a strong increase in the perpetrators, victims, and losses from *419 fraud.* Nonetheless, the level of successful convictions of perpetrators has been low (Oriola, 2005). This is largely due to the difficulty in building a case against perpetrators because of technical aspects of the fraud (The Herald, 2017).

419 scammers (known as *419ers*) are adept at evading legal counter measures (Ultrascan AGI, 2014). Attempts to defeat them with traditional legal instruments have been inadequate due to the transnational nature of the fraud (Webster & Drew, 2017). The Internet has facilitated the activities of *419ers* because it offers a wider geographical coverage, access to a large number of potential victims at a low-cost and low-risk, new and readily available scamming opportunities, anonymity and concealment of identity through minimal physical contact (Blommaert & Omoniyi, 2006; Hutchings & Hayes, 2009; Ofulue, 2010; Webster & Drew, 2017). Dion (2010, p. 630) concluded that *419ers* regard the Internet as an "Eldorado" because of the "quasi absence of the rule of law." The failure of previous efforts to curb *419 fraud* prompted Holt and Graves (2007) and Herley (2012) to argue for the development of automated scam filter detection systems based on an understanding of how *419ers* think, and how their victims react.

In a forensic accounting context, the power of linguistic and psycholinguistics analysis techniques to understand the human behaviour reflected in fraud has been studied widely (see

Nicolaides, Trafford, & Craig, 2017). Linguistic analysis techniques have been used to investigate behaviour associated with corporate fraud through study of CEO letters to shareholders, conference calls with financial analysts, and earnings press releases. These investigations have found some significant linguistic indicators of deception (Humphreys *et al.*, 2011; Craig, Mortensen & Iyer, 2013; Purda & Skillicorn, 2015). Nonetheless, with respect to *419 fraud*, linguistic analysis is less developed, despite its strong potential for success (Ofulue, 2010; Carter, 2015). If *419 fraud* can be reduced considerably, it is claimed that this would help curtail undesirable associated activities such as money laundering, organized crime, corruption, and terrorism (Ultrascan AGI, 2014).

Here we conduct a preliminary exploration of how linguistic analysis techniques (including those involving the use of computer assisted text analysis) can be helpful in identifying a distinctive marker of *419 emails*. We are motivated by the promptings of Holt and Graves (2007), Herley (2012) and Lamberger, Dobovšek and Slak (2013) to improve the rate of detection. In particular, we conduct a preliminary investigation of a sample of *419 emails* and *non-419 emails* using a linguistic cue that has been associated with deceptive conduct in written text: a high rate of use of first person singular pronouns. Thus, we address the following specific research question:

> *Is the incidence of first person singular pronouns substantially higher in 419 emails than in non-419 emails?*

The central purpose is simply to draw attention to the importance of linguistic analysis techniques (particularly pronoun use) in helping to identify fraud. In doing so, we contribute to the accounting literature by introducing a wide variety of research on email fraud that has been published in other disciplines, such as in communications, social psychology, criminology, law and ethics. We propose some elements of a future research agenda.

Section 2 reviews literature on the nature of *419 emails*, how *419 fraud* is perpetrated, and outlines some studies of language use that have investigated whether pronouns are indicators of deception. Section 3 describes the research method. Section 4 presents results. Section 5 enters conclusions and offers suggestions for future research.

2. Literature Review

2.1 The nature of 419 fraud

419ers use techniques that distort recipients' rational thought processes, and command a shift in behaviour by them (Freiermuth, 2011a; 2011b). These techniques are implicit in the structure and style of *419 emails*, where every word and sentence has a planned purpose (Ofulue, 2010; Freiermuth, 2011a; 2011b; Carter, 2015).

Scammers often obtain names and addresses of potential victims from trade journals, professional and commercial directories, lists in URLs and newspapers (Glickman, 2005). Some studies suggest the preferred type of victim is an educated citizen (Lamberger *et al.*, 2013; Ultracsan AGI, 2014) while others contend that scammers prefer to target people who are elderly, vulnerable or uneducated (Blommaert & Omoniyi, 2006), or who are greedy and complicit, or greedy and charitable (Freiermuth, 2011b).

Scammers often enlist a high level of technological and cultural competence to reflect the current reality of the countries and persons with whom they interact (Blommaert & Omoniyi, 2006). They often include attachments to enhance the veracity of their emails (Glickman, 2005). They discuss details of the scam offer, and the role of the scammer and victim, in ways that are specifically intended to be compelling to recipients (Carter, 2015). Scammers often assume the names and designations of prominent people, such as government officials, royalty, or wealthy business moguls (Glickman, 2005; Salu, 2005; FBI,

2010). They mix English and African names to persuade recipients to think they are "educated, upscale" individuals in their native country (Lamberger *et al*., 2013, p. 222).

Most *419 emails* contain grammatical errors and structural flaws that seem unbefitting to the identities scammers assume (Blommaert & Omoniyi, 2006). Many contain long convoluted sentences, all uppercase writing, non-standard punctuation, and misspellings (Blommaert & Omoniyi, 2006; Ofulue 2010; Freiermuth, 2011b). Some observers claim this is done deliberately to portray the stereotype that foreigners have of Africans (and Nigerians in particular) of being "childlike, intellectually unsophisticated, innocent in business ways, and corrupt" (Glickman, 2005, p. 464). The effect is that many recipients bask in an illusion of intellectual superiority, rendering them over-confident, and less suspicious (Glickman, 2005). Invariably, scam emails seek to achieve credibility, portray urgency, and implore secrecy (Carter, 2015). An example of a *419 email* (a business proposal) from Ultrascan AGI (2014). is reproduced in Appendix A. Note that the author cheekily uses the name "Dr. Kemy Kazy" – euphony for "kamikaze"? For other examples see Blommaert and Omoniyi (2006) and Onyebadi and Park (2012).

2.2 Structure of a 419 email

419 emails have the following rhetorical sequence at least 30% of the time: opening salutation; personal/professional credentials; offer solicitation; tale; trust statement; historical credentials; offer details; confidentiality plea; urgency statement; invitation for further contact; polite ending; and closing salutation (Freiermuth, 2011b, p. 224). Carter (2015, p.11) observed that this structure helped to "… inform, persuade, convey credibility, demand urgency, and secrecy, and provide re-assurances of legitimacy in ways that address the individualized and localized requirements and concerns of the recipient." *419* scammers have found the above structure to be effective in "building solidarity" with email recipients and distorting their rational decision-making processes (Freiermuth, 2011a, p. 123).

Typically, the introduction to *419 emails* are "narratives of trust" featuring an endearing or religious generic opening salutation (Dear Friend, Beneficiary, Asalam Alaekum (peace be with you)) (Ofulue, 2010). Often, to quell any doubts a recipient might have, there follows a brief mention of how the sender obtained the recipient's address. This is followed by a statement of purpose for writing (Freiermuth, 2011a). Scammers are predisposed to "anticipate a recipient's skepticisms and allay them" (Carter, 2015, p. 2). Ultimately, the recipient is persuaded to believe that s/he is lucky enough to be "the chosen one." Such flattery is intended to lower the recipient's guard and achieve success (Freiermuth, 2011a, b; Carter, 2015).

The body of the email aims to develop credibility, intimacy and trust. Often there is a narrative outlining a tale of misfortune and an invitation for recipients to engage with the scammer to secure some promised future fortune. The cultivation of an air of intimacy is a high priority. Scammers "reveal personal information, and appear vulnerable or flawed," giving "an impression of honesty, transparency and truthfulness" (Carter, 2015, p. 2). This encourages recipients to return the favour and mirror the sender's kind gestures. The email concludes by reiterating that the main service or product in question is a scarce commodity. The recipient is urged to act quickly and is implored to keep in further contact (Freiermuth, 2011a).

2.3 Deception detection in 419 fraud

Victims of *419 fraud* are deceived by misstatements or outright omissions of genuine facts. They form a false illusion of reality and this results in them incurring pecuniary and non-pecuniary losses. Chawki (2009) and Onyebadi and Park (2012) claim that *419ers* frame the discourse of their email deceptively to limit the potential victim's perception of the situation. The vulnerability of some victims is exacerbated by their gullibility and poor ability to detect lies and deceit. Thus, their capacity to differentiate between a scam email and a

legitimate e-mail is limited (Kraut, 1980; Bond & DePaulo, 2006). In a study by Datar, Cole, and Rogers (2014), 163 respondents were presented with four emails (two scam, two legitimate), but only 1.7% correctly identified all four emails, and only 64.5% correctly identified three of the four.

The intermediaries between senders and recipients of fraudulent emails (Internet nodes and networks) afford little control over the activities of scammers, other than mainly through email spam filters. Increasingly, recipients are unable to detect fraudulent emails because the Internet has made it easier for scammers to cut and paste legitimate logos and other data (of known and widely dispersed companies) to their emails (Lamberger *et al.,* 2013). Furthermore, structural and linguistic errors in most *419 emails* have been found to offer ineffective signals of likely deception (Blommaert & Omoniyi, 2006).

Websites of government authorities, professions and businesses contain alerts which highlight the incidence and features of scam emails and other forms of fraudulent correspondence. Most attempts to promote awareness involve warning recipients not to provide confidential information via email, not to pay advance fees in online transactions, and not to engage lightly with any correspondence containing even fleeting mention of Nigeria or Africa (Datar *et al.,* 2014). Despite strong and unambiguous warnings, and regulatory and preventive counter-actions taken locally and internationally, AFF scams have prospered through unsolicited emails in the Internet era (Adogame, 2009; Onyebadi & Park, 2012; Holt & Graves, 2007).

Spam filters have had only limited success in reducing the number of AFF emails that reach addressees (Lamberger *et al*., 2013). Information technology consultants Tiger Technologies claim that AFF spam is one of "the most difficult types for a spam filter to block." Website spam filters are often based on the words in subject lines (Edelson, 2003; Ofulue, 2010), such as "URGENT REPLY NEEDED." Nonetheless, spam filters have been

more successful, on average, than human intuition. This lends support to argument that improving the efficiency of spam filters by incorporating knowledge of linguistic cues and semantic features of fraudulent emails should be given priority over attempts to hone the detective skills of potential victims (Herley, 2012). Here we highlight the potential to develop spam filtering parameters based on the use of first person singular pronouns.

2.4 Pronoun use in 419 emails

Language use is crucial in understanding the psychology of the human mind and the cognitive processes embedded in the language choice of individuals (Pennebaker, Mehl, & Niederhoffer, 2003). Several studies have concluded that deceptive individuals are adept at manipulating language to achieve credibility and success with their fraudulent schemes (Blommaert & Omoniyi, 2006; Onyebadi & Park, 2012; Carter, 2015). Nonetheless, few studies of *419 emails* have delved into the ramifications of the language used (exceptions include Blommaert & Omoniyi, 2006; Ofulue, 2010; Freiermuth, 2011a; Onyebadi & Park, 2012). Holt and Graves (2007) pointed to scammers' use of unique words and phrases such as "risk-free", "urgent", "confidential" and references to monetary activities. However, despite such peculiar word markers occurring consistently in *419 emails,* few studies have identified specific language cues that have strong ability to detect *419 emails.*

Commonly, *419ers* adopt language that appeals to religion, sympathy or pity; flatters personalities; and creates illusions of intimacy, urgency and sincerity (Glickman, 2005; Lamberger *et al.*, 2013). Of particular interest to us is the potential for use of personal pronouns to create a sense of intimacy with recipients, and thereby, help allay their doubts (Carter, 2015). We contend that writers of *419 emails* attempt to create an air of intimacy and familiarity with recipients through extensive use of first person singular pronouns (that is, by means of self-references, such as "I, me, my, mine, myself"). These first person singular pronouns have been associated with deceptive communication (Gupta and Skillicorn (2006);

Newman, Pennebaker, Berry, and Richards (2003)). Their use declares an individual's ownership of a statement and is tantamount to a projection of honesty (Vartapetiance & Gillam, 2012).

Decreased use of self-reference first person *singular* pronouns and increased use of collective first person *plural* pronouns ("we, us, our, ours, ourselves") is reported to be a blame-shifting or disassociation strategy to help deceptive individuals distance themselves from the deceptions in their statements and to shift responsibility from themselves to others (Gupta and Skillicorn, 2006; Craig *et al.*, 2013. See also Pennebaker *et al.*, 2003; Zhou, Burgoon, Nunamaker, & Twitchell, 2004). Because *419ers* are not usually involved in any disassociation or blame-shifting strategies, we do not explore pronoun use related to those strategies.

3. Research Method

Our preliminary study was based on a sample comprising thirty *419 emails* and thirty *non-419 emails.* To reduce potential selection bias, the *419 emails* were chosen from among the first 50 emails in the Ultrascan AGI database of *419 emails*. They were classified according to content, with a view to ensuring fair representation of the major varieties of *419 emails* (Blommaert & Omoniyi, 2006; Ofulue, 2010). Following classification, a close-reading analysis identified and eliminated twenty emails which duplicated content with little or no modification. The resulting sample of thirty *419 emails* fell into the categories listed in the extreme left hand column of Table 1 and comprised 10,880 words (mean = 360 words; range from 68 to 723 words).

The *non-419 emails* were chosen randomly from the email inbox of the first author and several of her colleagues. The fifth column of Table 1 shows the categories of the resultant sample of thirty *non-419 emails*. This sample comprised 6356 words (mean = 227 words; range from 49 to 998 words). We did not select the *non-419 email* sample to match the exact

content categories identified in the sample of *419 emails*. This is consistent with Chang (2014) and findings that any matching the two sets of emails in terms of content does not significantly affect ability to correctly classify an email as deceptive or non-deceptive.

Table 1
Analysis of Email Samples by Category

419 emails				***Non-419 emails***			
Category	*No*	*%*	*Mean Words*	*Category*	*No*	*%*	*Mean Words*
Dormant Accounts	4	13	482	Sales Letter	5	17	329
Job Offers	9	30	328	Offers	4	13	329
Lottery wins	3	10	364	Student Notifications	6	20	194
Settlements & Compensations	4	13	304	Feedback Requests	2	7	152
Foreign Investment	6	20	329	Invitations	2	7	93
Humanitarian & Charity	3	10	417	Membership Notifications	8	27	297
Other	1	4	399	Other	3	10	75
Total	30			Total	30		

We then used two methods to search the samples for linguistic cues that would enable discrimination between *419 emails* and *non-419 emails*. First, we used the search facility of *Microsoft WORD* to tally the number of first-person singular pronouns (I, I'd, I'll, me, mine, my, myself) and first-person plural pronouns (we, we'll, we've, us, our, ours, ourselves) in both groups of emails. To capture the extent to which there is a collective (rather than individual) apportionment of responsibility in the emails, we also calculated "the percentage of all first-person pronouns that are singular," consistent with Chatterjee and Hambrick (2007, p. 364). This more expansive approach offers a stronger view of pronoun use.

Second, we used the computer-assisted text analysis software programme *DICTION 7.0* to identify the potential for first person singular pronouns to provide linguistic cues to

deception. Readers unfamiliar with *DICTION* can find compact summaries of its key features in Hart and Carroll (2013), Amernic, Craig and Tourish (2010), and Murphy (2013). For example, Hart (2001, p. 45), the deviser of *DICTION*, describes *DICTION* as:

> … a dictionary-based package that examines a text for its verbal tone. It deploys some 10,000 search words in 33 word lists. None of these search terms is duplicated in these lists, which allows the user to get an usually rich understanding of a sample text. Lying at the heart of the program are five master variables [CERTAINTY, OPTIMISM, ACTIVITY, REALISM, and COMMONALITY] that are created by combining the subaltern variables. The master variables have been chosen intentionally, the assumption being that, if only five questions could be asked of a given passage, these five would provide the most robust understanding.

In particular, we compared differences between the *DICTION* score for the variable 'SELF-REFERENCE' in the *419 email* sample and *non-419 email* sample, using *DICTION's* "All Norms" dictionary as our normative word-list referent. The variable 'SELF-REFERENCE' is defined as:

> All first-person references, including I, I'd, I'll, I'm, I've, me, mine, my, myself. Self-references are treated as acts of indexing whereby the locus of action appears to reside in the speaker and not in the world at large thereby implicitly acknowledging the speaker's limited vision (Hart & Carroll, 2013, p. 6)

Our expectation was that the focus of the SELF-REFERENCE variable on first-person references will lead to a very high out-of-normal-range score being recorded for it, consistent with our prior contention.[1]

4. Results

Table 2 reports the incidence of first-person singular pronouns and first person plural pronouns in each analysis group. First person singular pronouns occurred 30.42 times per

[1] We note that Onyebadi and Park (2012, p. 181) used DICTION to determine the "main persuasive lexical characteristic" of *419 emails*. Their study does not seem to reflect a strong understanding of DICTION and the interpretations that should be applied defensibly to DICTION scores. In interpreting scores for the five DICTION master variables, they seem to ignore that each of these master variables assumes a normal distribution around a different mean. Thus, the scores for each should be interpreted in the context of each variable's expected normal distribution. Onyebadi and Park's (2012) results show that four of the five master variables are unremarkable, because they fall within the normal range (of ± 1 s.d.) from the mean expected score. The only out-of-range score is for the master variable CERTAINTY, but this is barely out of range. It provides no strong pointer to a distinctive semantic tone. The conclusion Onyebadi and Park (2012, p. 181) draw is that the language of REALISM or "meeting tangible needs in people's everyday lives" is the "main pervasive lexical characteristic in *419 emails*." In our view, this is misleading.

1000 words for the *419 emails*, but only 4.41 times per 1000 words for the *non-419 emails*. Thus, the incidence of first person singular pronouns was almost seven times greater in *419 emails*. The proportion of first person singular pronouns to first person plural pronouns was 63.1% for *419 emails*, but only 17.5% for *non-419 emails*. Thus, the proportion was 3.6 times greater in *419 emails*.

Table 2

First Person Pronoun Use per 1000 words

Group	Total Words	Singular (I, I'd, I'll, me, my, mine, myself)		Plural (We, we've, we'd, us, our, ourselves)		Proportion singular
		n	per 1000	n	per 1000	
419 emails	10880	331	30.42	193	17.74	63.1%
Non-419 emails	6356	28	4.41	132	20.77	17.5%

The last five lines of Appendix B reveal that scores for each of the five *DICTION* master variables (ACTIVITY, OPTIMISM, CERTAINTY, REALISM, COMMONALITY) are all within normal range (±1 standard deviation from the expected mean). With respect to the other 35 *DICTION* variables, thirteen scores (indicated in bold in Appendix B) fell outside the expected normal range. Almost all of these were just outside the normal range. The most pronounced out-of-range score was for the dictionary variable SELF-REFERENCE (31.14). This occurred in the *419 emails*, indicating that use of self-reference words (such as 'I', 'me', 'my', 'myself') was extremely high for this sample. In comparison, the SELF-REFERENCE score for the *non 419 emails* was 0.52, within the normal range.

5. Conclusion

The findings of our preliminary enquiries draw attention to advance fee fraud or *419-type* emails being characterised by the following linguistic features:

1. A much higher proportion of first person singular pronouns than in *non-419 emails*.

2. A much higher proportion (than in *non-419 emails*) of all first person pronouns being singular rather than plural pronouns; and
3. A *DICTION* score for the SELF-REFERENCE variable that falls higher than +1 standard deviation from the expected norm.

The findings lead to the research question being answered in the affirmative. They highlight the potential for overuse of first person singular pronouns to be a linguistic cue to a deceptive or fraudulent *419 email*. However, clearly, there is much more that can be done to improve the identification of *419 emails* by analysing the language they employ. For example, our findings point to the likely usefulness of the following tentative three-point heuristic to identify *419 emails*:

1. There are at least 20 first person singular pronouns per 1000 words.
2. At least 60% of all first person pronouns are singular.
3. *DICTION*'s score for SELF-REFERENCE falls more than + 1 standard deviation from the norm (that is, it has a *DICTION* score higher than 15.10).

Tentative application of this heuristic to our sampled emails indicates a correct rate of identification of *419 emails* and *non-419 emails* of about 80%. There is ample scope to refine this heuristic and to conduct stronger empirical testing of it. This could involve, for example, a much larger data set, and a data set in which the *419 emails* and *non-419* emails are more closely matched in word length. The heuristic should be tested, too, with stronger regard for the type of fraudulent email that it is better at detecting. Our tentative analysis points to the heuristic being particularly strong in detecting requests to apply as a foreign beneficiary for the transmission of funds, and in respect of appeals to send funds on humanitarian grounds. It is less strong in detecting more formal and stylised (but nonetheless fraudulent emails) such as those making job offers.

While the use of first person singular pronouns is highly suggestive of a *419 email*, there are several other language issues that can be explored beneficially. For example, at what point in the email do self-referential pronouns occur? Are they concentrated for rhetorical impact, in the opening and closing sentences? Another beneficial aspect of any forward-looking research agenda would be to examine whether the decision heuristic we propose (or any other classification heuristic, for that matter) is better in detecting scam emails than intuitive judgement exercised by a perceptive reader. Thus, further evaluation should be directed to compare the classification results obtained using the proposed heuristic with those obtained in a laboratory setting by a group of readers who are asked to use intuition to classify a carefully selected sample of emails as fraudulent or non-fraudulent.

References

Action Fraud. (2010). *What is fraud and cyber crime?* Accessible at http://www.actionfraud.police.uk/what-is-fraud

Adogame, A. (2009). The 419 code as business unusual: Youth and the unfolding of the advance fee fraud online discourse. *Asian Journal of Social Science*, *37*(4), 551-573.

Amernic, J., Craig, R., & Tourish, D. (2010). *Measuring and Assessing Tone at the Top Using Annual Report CEO Letters*, Institute of Chartered Accountants in Scotland, Edinburgh.

Blommaert, J., & Omoniyi, T. (2006). Email fraud: Language, technology, and the indexicals of globalisation. *Social Semiotics*, *16*(4), 573-605.

Bond, C.F., & DePaulo, B.M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214-234.

Carter, E. (2015). The anatomy of written scam communications: An empirical analysis. *Crime, Media, Culture*, *11*, 89-103.

Chatterjee A., & Bambrick, D.C. (2007). It's all about me: Narcissistic Chief Executive Officers and their effects on company strategy and performance. *Administrative Science Quarterly*, *52*, 351-386.

Chawki, M. (2009). Nigeria tackles advance free fraud. *Journal of Information Law & Technology*, *1*, unpaginated. Accessible at http://go.warwick.ac.uk/jilt/2009_1/chawki

Chang, A. (2014). Linguistic deception cues in selected narrative disclosures contained in prospectuses of failed and non-failed New Zealand finance companies. Master's thesis, University of Canterbury, New Zealand, Accessible at https://ir.canterbury.ac.nz/handle/10092/8926

Choo, F., & Tan, K. (2007). An "American Dream" theory of corporate executive fraud. *Accounting Forum, 31,* 203–215.

Craig, R., Mortensen, T., & Iyer, S. (2013). Exploring top management language for signals of possible deception: The words of Satyam's chair Ramalinga Raju. *Journal of Business Ethics, 113*(2), 333-347.

Datar, T.D., Cole, K.A., & Rogers, M.K. (2014). Awareness of scam e-mails: an exploratory research study. In *Proceedings of the Conference on Digital Forensics, Security and Law* (pp. 11-34). Association of Digital Forensics, Security and Law. Accessible at https://search.proquest.com/openview/a29198c9976b91fc1d2cf594e1644a7c/1?pq-origsite=gscholar&cbl=60415

Dion, M. (2010). Advance fee fraud letters as Machiavellian/Narcissistic narratives. *International Journal of Cyber Criminology, 4,* 630–642.

Edelson, E. (2003). The 419 scam: information warfare on the spam front and a proposal for local filtering. *Computers and Security*, 22, pp. 392-401.

FBI [Federal Bureau of Investigation]. (2010). *Common fraud schemes*. Accessible at https://www.fbi.gov/scams-safety/fraud

Freiermuth, M. (2011a). Text, lies and electronic bait: An analysis of email fraud and the decisions of the unsuspecting. *Discourse & Communication, 5*(2), 123-145.

Freiermuth, M. (2011b). "This transaction is 100% Risk-Free!" Why do people fall prey to email scams? *International Conference on Language and Communication (LANCOMM)* (pp. 222- 230). Accessible at file:///C:/Users/e5028727/Downloads/LANCOMM2011E-proceeding.pdf

Garrett, E.V. (2014). *Exploring internet users' vulnerability to online dating fraud: Analysis of routine activities theory factors*. Master's thesis. The University of Texas at Dallas. Accessible at http://pqdtopen.proquest.com/doc/1656449717.html?FMT=ABS.

Glickman, H. (2005). The Nigerian "419" advance fee scams: prank or peril? *Canadian Journal of African Studies 39*(3), 460-489.

Gupta, S., & Skillicorn, D. B. (2006). Improving a textual deception detection model. In *Proceedings of the 2006 conference of the Centre for Advanced Studies on Collaborative Research* (p. 29). IBM Corp.

Hart, R.P. (2001) Redeveloping DICTION: Theoretical considerations. In M. West (Ed.), *Theory, Method, and Practice of Computer Content Analysis* (pp. 43–60). New York, NY: Ablex.

Hart, R., & Carroll. C.E. (2013). *DICTION 7.0. User's Manual.* Digitext, Austin, TX.

Herley (2012). Why do Nigerian Scammers Say They are from Nigeria? *World Economic Information Services.* Accessible at http://www.econinfosec.org/archive/weis2012/papers/Herley_WEIS2012.pdf

Holt, T.J., & Graves, D. (2007). A qualitative analysis of advance fee fraud email schemes. *The International Journal of Cyber Criminology*, *1*, 137-154.

Humphreys, S.L., Moffitt, K.C., Burns, M.B., Burgoon, J.K., & Felix, W.F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, *50*, 585-594.

Hutchings, A., & Hayes, H. (2009). Routine activity theory and phishing victimisation: Who gets caught in the net. *Current Issues Criminal Justice*, *20*, 433.

Kraut, R. (1980). Humans as lie detectors. *Journal of communication*, *30*, 209-218.

Lamberger, I., Dobovšek, B., & Slak, B. (2013). Analysis of the fraudulent letters AKA Nigerian letters. In *Proceedings of the Biennial International Conference: Criminal Justice and Security–Contemporary Criminal Justice Practice and Research* (pp. 443-466).

Martin, A.J. (2016). Accountant falls for sexy Nigerian email scammer, gives her £150k he cheated out of pal. *The Register*, 25 October (https://www.theregister.co.uk/2016/10/25/accountant_defrauded_client_to_rescue_nigerian_woman_he_met_online/)

Murphy, A.C. (2013). On "True" Portraits of Letters to Shareholders and the Importance of Phraseological Analysis. *International Journal of Corpus Linguistics*, *18*, 57-81.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003).Lying words: Predicting deception from linguistic styles. *Personality and social psychology bulletin*, *29*(5), 665-675.

Next Web Security. (2016). *419 Scams - Nigerian Advance Fee Fraud: West African Organized Crime Organizations; Detection, Education, Eradication*. Accessible at https://www.nextwebsecurity.com/

Nicolaides, R., Trafford, R., & Craig, R. (2017). 'Helping Auditors Identify Signs of Deception through Psycholinguistics,' *Journal of Financial Crime*, in press.

Nykodym, N., & Taylor, R. (2004). The world's current legislative efforts against cyber crime. *Computer Law & Security Review*, *20*, 390-395.

Ofulue, C. I. (2010). A digital forensic analysis of advance fee fraud (419 scams). *Handbook of Research on Discourse Behavior and Digital Communication: Language Structures and Social Interaction: Language Structures and Social Interaction*, IGI Global: Hershey, PENN, 296.

Onyebadi, U., & Park, J. (2012). 'I'm Sister Maria. Please help me': A lexical study of 4-1-9 international advance fee fraud email communications. *International Communication Gazette*, *74*, 181-199.

Oriola, T.A. (2005). Advance fee fraud on the Internet: Nigeria's regulatory response. *Computer Law & Security Review*, *21*, 237-248.

Pennebaker, J.W., Mehl, M.R., & Niederhoffer, K.G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*, 547-577.

Purda, L., & Skillicorn, D. (2015). Accounting variables, deception, and a bag of words: Assessing the tools of fraud detection. *Contemporary Accounting Research*, *32*, 1193-1223.

Salu, A.O. (2005). Online crimes and advance fee fraud in Nigeria-are available legal remedies adequate? *Journal of Money Laundering Control*, *8*, 159-167.

Schuchter, A., & Levi, M. (2015). Beyond the fraud triangle: Swiss and Austrian elite fraudsters. *Accounting Forum, 39,* 176–187

The Herald (2017). *The life and exploits of Canadian based Nigerian playboy and identity thief: Adekunle Johnson 'Chrome' Omitiran.* Accessible at http://www.herald.ng/life-exploits-canada-based-nigerian-playboy-identity-thief-adekunle-johnson-chrome-omitiran/

Ultrascan AGI (2014). *Examples of 419 E-mail*. Accessible at http://www.ultrascan-agi.com/public_html/html/419_e-mail.html

Ultrascan AGI (2014). *419 Advance Fee Fraud statistics 2013*. Accessible at http://www.ultrascan-agi.com/public_html/html/419_statistics.html

Vartapetiance, A., & Gillam, L. (2012). Deception detection for the tangled web, *Computers and Society*, *42*, 34-47.

Webster, J., & Drew, J. (2017). Experiences of fraud detectives using a victim-focused approach. *International Journal of Police Science & Management*, *9*, 39-53.

Zhou, L., Burgoon, J. K., Nunamaker, J. F., & Twitchell, D. (2004). Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications. *Group Decision and Negotiation*, *13*, 81-106.

APPENDIX A
Example of a 419 email
(Source: Ultrascan AGI, 2014)

Dear friend,
I came to know about you in my private search for a reliable person/company capable to handle a confidential establishment of a sister company abroad on behalf of my family and myself. As a matter of fact, I got your information from the West African Chamber of Commerce and industry, American Export Promotion Council, Eurasia World Trade Journal.

I am Dr. Kemy Kazy, the son to the former Chairman South Delta Petroleum Development Committee and the owner/president of Conja Industries Incorporated. Some Moslem extremist murdered my father immediately he publicly condemned the 11th September terrorist attack on World Trade Center in USA. But before he finally gave up the ghost, he told me that I should try and invest out side African race. Now my life and that of my 86yrs old mother are in danger because the Moslems extremists are still after our lives. They have burnt our family house and our known properties, freeze our known accounts. So what I want from you is to assist us in establishing a sister company in your Country, receiving the fund for the contract in your account overseas so that I can come up to your country and live a normal and happy life. We have gone through so many traumatic experiences. We are urgently waiting for your response.

BELOW IS THE COMPANY INTENTED PROJECT.

This proposal was submitted to you to see if you can be of assistance, represent or help, facilitate the acquisition of a viable factory site for Conja intended project in your country. It should be germane to give you a brief overview of the company's profile and project as follows:-

(i) Brief Company Profile:
Conja Industries Incorporated is involved in the Full trimmings line for the interior decoration sector: tassels, tiebacks, cords, braids, fringes, Fire Extinguisher etc.

(ii) Brief project Description:
To acquire and establish a permanent branch plant in your country.

(iii) Purpose:
To engage in the manufacture of Full trimmings line for the interior and exterior decoration sector/ breakables.

(iv) The Proposal:
To establish contact with a representative in the country who must be ordinarily resident in the area of the company's interest. The ideal person will function as a contact point for the company towards the procurement of a suitable site for the project. The company, as a policy, will pay due compensation upon satisfactory completion of the acquisition.

You will have to consider the advisability of establishing such a project in your country, and a subsequent effort in procuring a factory site viz:

(i) Minimum of three hectares of land, urban or rural area.
(ii) Easily accessible;
(iii) Proximity to power grid
(iv) Proximity to flowing stream or river (optional),

The project will be funded entirely by the family. All materials and information gathered by you in connection to this project are confidential, and considering this letter you agree that no use shall be made of the information beyond the terms of your engagement.

I look forward to working with you on this important project.

Thanking you in anticipation

Yours truly,

Dr. Kemy Kazy.

APPENDIX B
***DICTION* Scores**

Variable	*Low*	*High*	*419 Value*	*Non-419 Value*
Numerical Terms	0.30	15.04	10.77	14.64
Ambivalence	6.49	19.21	**1.89**	9.47
Self-reference	-1.18	15.10	**31.14**	0.52
Tenacity	23.32	39.76	23.43	27.50
Leveling Terms	5.02	12.76	**4.30**	12.40
Collectives	4.04	14.46	**16.65**	6.62
Praise	2.77	9.59	**1.95**	4.03
Satisfaction	0.47	6.09	4.44	5.08
Inspiration	1.56	11.12	6.35	1.77
Blame	0.06	4.16	0.34	1.04
Hardship	1.26	10.48	4.51	**0.42**
Aggression	1.07	9.79	6.01	**0.79**
Accomplishment	4.96	23.78	10.13	18.89
Communication	2.21	11.79	5.33	6.50
Cognition	4.43	14.27	**4.34**	12.75
Passivity	2.10	8.08	7.06	4.13
Spatial Terms	4.17	19.85	**21.87**	11.21
Familiarity	117.87	147.19	**116.95**	**109.40**
Temporal Terms	8.36	21.82	13.35	17.38
Present Concern	7.02	16.66	**16.93**	14.12
Human Interest	18.13	45.49	27.34	43.00
Concreteness	10.70	28.50	27.64	24.79
Past Concern	0.97	6.19	3.55	2.82
Centrality	1.18	7.54	2.99	2.27
Rapport	0.42	4.26	0.48	1.28
Cooperation	0.36	8.44	5.83	8.42
Diversity	0.07	3.81	0.10	2.07
Exclusion	-0.03	4.31	2.82	3.12
Liberation	-0.46	4.72	1.69	2.46
Denial	2.57	10.35	**1.88**	3.46
Motion	0.17	4.35	0.19	2.26
Insistence	9.15	111.15	67.41	40.19
Embellishment	0.16	1.14	0.16	0.35
Variety	0.45	0.53	0.52	0.50
Complexity	4.31	4.91	4.63	4.84
Activity	46.74	55.48	49.71	49.55
Optimism	46.37	52.25	51.51	51.53
Certainty	46.90	51.96	47.01	49.51
Realism	46.10	52.62	51.52	50.04
Commonality	46.86	52.28	49.80	49.13