Predicting Recidivism Among Internet Child Sex Sting Offenders Using Psychological Language Analysis

Michelle Drouin, PhD,[1] Ryan L. Boyd, PhD,[2] and Miriam Greidanus Romaneli[1]

Abstract

In this study, we examined the extent to which computerized linguistic analysis of natural language data from chat transcripts of Internet child sex stings predicted recidivism among 334 convicted offenders. Using the Linguistic Inquiry and Word Count (LIWC) program, we found that reoffenders (including simultaneous and previous offenders) differed significantly from nonreoffenders in measures of clout (a composite measure of social dominance) and percentage of words used in the following linguistic categories: cognitive processes, personal pronoun use, insight, time, and ingestion. In contrast, total word count and percentage of sexual words, two categories that might be assumed to be predictive of recidivism, were not significantly different between these two groups. These analyses help to develop a typology for an Internet sex reoffender as one who is dominant, nonequivocating, and likely to discuss meeting with their target and/or parents' schedules. Moreover, they highlight the importance of examining the functional aspects of language in forensic linguistic analysis, and exemplify the utility of computerized linguistic analyses in the courtroom.

Keywords: language analysis, internet sex offenders, recidivism, sex sting offenders, forensic linguistic analysis

Introduction

CELL PHONE AND INTERNET use has become nearly ubiquitous in the United States and many parts of the developed world.[1] Consequently, many interactions, including illegal activities, occur via electronic mediums. Fortunately, interactions that occur via these technologies produce digital evidence which, according to the National Institute of Justice (NIJ), "is now used to prosecute all types of crimes, not just e-crime."[2] Accordingly, the field of digital forensic science has expanded to include both traditional computer forensics (e.g., the types of material transmitted, to whom, and with what frequency), and also, in accordance with the provisions of The Wiretap Act,[3] content analyses of the digital communication.[4]

A prime example wherein content analyses may be of great benefit is in the realm of online child solicitation sex stings. Recently, two separate studies have shown that the psychological hallmarks of grooming strategies (i.e., the luring language used by sexual predators[5]) are evident in chat transcripts from online child solicitation sting cases,[6,7] and that some dimensions of offenders' language were associated with length of jail time assigned during sentencing.[7] In this study, we build upon the content analyses of previous

work to examine whether offenders' language patterns from online chat transcripts are predictive of recidivism among convicted Internet sex sting offenders.

Online child sex sting offenders

As the Internet gained popularity in American households, the general public and law enforcement agencies (LEA) began to recognize potential dangers lurking for children online. Consequently, the Department of Justice, in conjunction with the National Center for Missing and Exploited Children (NCMEC), has been investigating online child victimization for almost two decades.[8–10] Although the number of children reporting unwanted online sexual solicitation has dropped significantly from 2000 to 2010 (from 19 to 9 percent), concerns remain that Internet and computer-mediated communication technologies may still facilitate child victimization.[9,11] In answer to such concerns, law enforcement agents across the United States regularly conduct online sex stings to apprehend potential child predators before they are able to commit a contact offense with an actual child.[12–15]

With regard to recidivism, numerous studies conducted in various contexts have found consistently that Internet sex offenders have a much lower rate of recidivism than contact

---

[1]Department of Psychology, Indiana University—Purdue University Fort Wayne, Fort Wayne, Indiana.
[2]Department of Psychology, The University of Texas at Austin, Austin, Texas.

offenders,[16] and Internet sex offenders without a history of previous contact offense(s) have a lower rate of recidivism than those who do have a history of such offenses.[17] For example, in a comprehensive meta-analysis of more than 2,500 internet sex offenders, only 4.6 percent were found to have had a sexual reoffense, and only 2 percent had a new contact offense.[16] Thus, research suggests that Internet-only sexual offenders may comprise a distinct category of offenders who are at a relatively low risk of recidivism relative to others.

Importantly, however, most of the offenders in previous studies had been convicted of child pornography possession or distribution; offenders from Internet sex stings comprise only a small portion of the sample. A primary distinction may be made in that child solicitation offenders could be distinct from other Internet-only offenders in that, although they have not committed a contact offense, they are contact-driven (i.e., they have, in most circumstances, made a substantial step toward committing a contact offense by traveling to meet a minor for sexual activity) rather than fantasy-driven.[18] In light of this difference, there is a strong need for more specific studies with Internet sex sting offenders to determine whether there are any markers that predict recidivism in this subsample of online offenders. In other words, disambiguating those factors indicative of a propensity to reoffend is crucial from both a psychological perspective and a legal perspective.

Language as a predictor of recidivism among internet sex offenders

For decades, researchers have been examining the relationships between natural language data and individual personality and dyadic communication characteristics.[19–25] Computerized text analysis software, such as the Linguistic Inquiry and Word Count (LIWC) program,[26–28] has revolutionized this research by allowing researchers to compare natural data sets across contexts using meaningful, validated linguistic categories. Importantly, work with software such as LIWC has found that the very foundations of an individual's psychology, such as emotions, motivations, and attentional processes, are embedded (and quantifiable) within the words people use to communicate. Simply put, much research has been conducted to establish and develop language-based psychological analyses into a reliable and valid mode of assessing psychological and social processes.

As digital evidence has become more popular in the courtroom, language analysis has gained attention from the legal community as a useful forensic tool.[4,29] With regard to Internet sex stings specifically, these analyses may be particularly beneficial as the prosecution of these cases often includes chat transcripts containing extensive natural language data (i.e., interactions between offenders and undercover stings), with word counts numbering in the thousands to tens of thousands. Although it is common for these chat transcripts to be admitted as forensic evidence, in absence of testimony from a forensic linguist to contextualize the language used in the chats, the trier of fact must rely upon their own, likely naïve understanding of language and grooming strategies to make conviction and sentencing decisions.

Recent computerized linguistic analyses from these chat transcripts have revealed important trends that could prove helpful to the trier of fact.[6,7] Black et al.[6] were the first to map these natural language features onto sexual predation indicators, analyzing grooming strategies within transcripts of 44 convicted online sex sting offenders. They found grooming strategies were prevalent throughout their transcripts; however, their frequency and order of appearance did not conform with O'Connell's[30] proposed stage model of online grooming. O'Connell[30] suggested that relationships would be established (first as friendships and then as more exclusive relationships) before offenders engaged in risk assessment and sexual talk (including plans to meet), whereas sex sting offenders engaged in these talks early and throughout the transcripts.[6]

In a more recent study that used the same offender chat database, Drouin et al.[7] found that multiple psychological dimensions of offenders' language (i.e., total word count, sexual words used, and clout) were associated with length of jail time adjudged during sentencing (i.e., offenders with lower rates of word usage in these categories than their undercover agents received less jail time). This suggests that the trier of fact may already be using some linguistic features of these transcripts in their decision making. However, currently, there is no empirical evidence that these language markers are related to increased risk of past, concurrent, or subsequent offenses, which calls into question the validity of relying upon these cues in language for the purpose of forming and rendering sentencing decisions.

Current study

In this study, we sought to extend previous lines of inquiry to determine whether there were natural language patterns that predicted recidivism among online child sex offenders. In line with Drouin et al.,[7] we predicted that those who reoffended would be more overtly predatory, using more *overall words*, more *sexual* words, and displaying higher *clout* than nonreoffenders. Additionally, we focused on more nuanced, functional aspects of language. (See Table 1 for the linguistic categories we included and representative text samples from the chat transcripts.) As shown in Table 1, *cognitive processes* and *insight* are categories that contain words like "think," "know," and "ought." We predicted that reoffenders would use fewer words than nonreoffenders in these categories because they would be goal-oriented, less reflective, and nonequivocating, having already determined their desired course of action and expressing little doubt.[31,32] Additionally, we predicted that nonreoffenders would use increased rates of *first-person singular pronouns*; previous work has repeatedly found that greater use of this class of pronouns is indicative of lower levels of social status and dominance, and decreased self-reflection.[21,33–35] Finally, a manual inspection of more than 200 Perverted Justice Foundation Inc. (PJFI) transcripts showed that many offenders used time words to arrange meeting, verify parents' schedules (corresponding to both the sexual and risk assessment categories of online grooming[6,30]), and otherwise coordinate behaviors toward end goals of contact. Additionally, many undercover agents engaged in talk about food as an engagement tactic when offenders were not forthcoming or dominant with regard to conversational topic. Therefore, we predicted that reoffenders would use higher rates of words in the *time* category and lower rates of words in the *ingestion* category.

TABLE 1. LINGUISTIC INQUIRY AND WORD COUNT LINGUISTIC CATEGORIES, CATEGORY SAMPLE
WORDS/DESCRIPTIONS, AND CHAT TRANSCRIPT EXAMPLES

LIWC category	LIWC category sample words/description	Chat transcript example
Clout	Composite variable: Higher numbers suggest "that the author is speaking from the perspective of high expertise and is confident; low Clout numbers suggest a more tentative, humble, even anxious style." (LIWC, p. 22)	High clout: "We'll have fun, Go for a ride, have a good dinner//Then we'll see where we end up" Low clout: "aww thanks…I will be back I am going to eat ok"
Sexual	Horny, love, incest	"I'm just really horny and hard"
First person singular pronouns	I, me, mine	"I didn't want you to think that I changed my mind because I don't like you because that is not the case."
Cognitive processes	Cause, know, ought	"I don't know. Whatever you want"
Insight	Think, know, consider	"I'm sure. Don't think I'm ready for drastic changes like that."
Time	End, until, season	"Maybe next week if she works nights we can meet//Cool//Friday"
Ingestion	Dish, eat, pizza	"Have any thing good for dinner?… I had cheesy mac and beef."

Note: "//" indicates a line break and/or conversational turn.
LIWC, Linguistic Inquiry and Word Count.

Methods

Data collection and preparation

Chat transcripts. PJFI[a] is a nonprofit organization focused on the apprehension of sexual predators using online sting operation tactics. The PFJI liaisons with local LEA in conducting operations that primarily take place in regional online chatrooms. The PJFI makes transcripts of all chat logs resulting in conviction publicly available[b]—this archive served as the primary source of natural language data. As part of a larger study,[7] data were collected from the PFJI archives in January, 2016. Collected data included complete transcripts of all interactions between stings and offenders[c] and various metadata, such as geographical data and sting demographics (i.e., fictional sting gender, fictional sting age). A total of 590 full transcript collections and associated metadata were collected. For the current study, we included only those individuals who had chatted online with fictional female stings ($n = 538$)[d] and were registered as sex offenders in their respective state, resulting in a final N of 334 transcripts. All offenders in this final sample were male (M age = 33.36, $SD = 10.14$).

Recidivism data. Offenders were searched for by name via the U.S. Department of Justice National Sex Offender Public Registry online portal.[e] In cases where multiple individuals with the same name were found, the registry entry was matched with the conviction information, demographic characteristics (e.g., age and city of residence), and offender photo (if available) to ensure that the correct person was identified. After matching the initial sex sting conviction information to the conviction record, the records were inspected for any prior, concurrent, or subsequent offenses. All offenses were recorded; however, only offenses related to minors (e.g., child pornography, child solicitation, use of computer to harm a minor), sexual offenses (e.g., sexual assault of elderly or disabled person), or unspecified subsequent felonies were counted as recidivism. We included both

contact and noncontact offenses because either would be a significant violation of probation and/or other restrictions placed upon sex offenders, and it also allowed us to cast the widest net possible for the identification of potential reoffenders. Additionally, concurrent or prior offenses reported by PJFI but not included in the sex registries were also recorded. Of the offenders located in the registry, 291 (87 percent) had no prior, concurrent, or subsequent offenses (nonreoffenders), and 43 (13 percent) were categorized as reoffenders. Among reoffenders, 12 (3.6 percent of the total sample) had reoffended after the sex sting conviction (usually child pornography or solicitation offenses; only 2 were contact offenses), 18 (5.4 percent of total sample) had prior offenses, and 9 (2.7 percent of total sample) had simultaneous offenses. Additionally, 4 (1.2 percent of total sample) had multiple offenses (e.g., prior and simultaneous).

Language preparation/analysis. Following data collection, all transcripts were preprocessed in multiple stages. Within each transcript, natural language data were separated by speaker, isolating the words written by offenders from the fictional sting characters. To ensure between-individual consistency, spelling standardization procedures were used to correct common misspellings (e.g., "teh" instead of "the"), netspeak (e.g., "ty" for "thank you"), and other common idiosyncrasies (e.g., elongation of "oh" to "ohhh"). Texts were manually inspected to ensure the general accuracy of the spelling standardization process.

Offender language samples were subsequently analyzed using the LIWC2015 software.[26]

The LIWC2015 software quantifies natural language data into objective measures of psychological processes. Using LIWC2015, language is quantified using a dictionary-based word counting approach, wherein a body of text is quantified along ~80 psychological dimensions as a function of word use. For example, if 1 out of every 10 words in a language sample belongs to the "positive affect" LIWC category, the text will receive a score of 10 percent for positive

Table 2. Descriptive Statistics for LIWC2015 Measures Used in Current Study

LIWC2015 category	M (SD)
Word Count	5,561.49 (7,398.62)
Sexual	1.15 (0.94)
Clout	81.34 (13.09)
First person singular pronouns	7.67 (1.75)
Cognitive processes	12.27 (2.13)
Insight	1.90 (0.72)
Time	5.08 (1.13)
Ingestion	0.43 (0.31)

Note: With the exception of the ''Word Count'' and ''Clout'' categories, which are computed as a sum and internally normalized by the LIWC2015 software (Pennebaker et al.[26]), respectively, all categories reflect the percent of words in a given text that are indicative of each psychological process.

emotionality; this approach has been extensively validated across hundreds of studies (see Tausczik and Pennebaker[25]). LIWC2015 also provides a handful of ''summary'' measures, such as the ''clout'' score, that represent population-normed psychological measures based on the word-counting method just described. These LIWC-based measures can be used in traditional statistical models in the same way as other types of quantified psychological measures—our statistical analyses are reported in the following section. For the current analyses, all texts were of adequate length for inclusion; descriptive statistics for the measures under consideration for the current study are presented in Table 2.

Results

Results from all analytic procedures are presented in Table 3. For all language-based psychological measures included in the current study, we performed independent-samples t tests to investigate differences between reoffenders and nonreoffenders. Both groups met assumptions for homogeneity of variance for all models (Levene's test $ps \geq 0.15$), with the exception of the *ingestion* category. As such, standard t tests were performed for all but this category, wherein we performed a t test accounting for the violated equality of variances assumption. In the current analysis, results for all measures showed statistically significant differences between reoffenders and nonreoffenders with the exception of the

''sexual'' and word count measures. Reoffenders were more likely to exhibit higher clout in their language in addition to a greater use of time words; nonreoffenders tended to use words from the cognitive processes, insight, ingestion, and first person singular pronoun categories at higher rates.

For psychological measures that demonstrated significant between-group differences, we performed follow-up binomial logistic regressions, which were used to model the probability of reoffense as a function of psychological processes measured via language.[f] Log odds ratios were converted to probability scores for ease of interpretation. Estimates of changes in recidivism probabilities were calculated for an increase of 1 standard deviation (*SD*) for each measure.[36] For example, an increase of 1 *SD* in an offender's use of words from the ''time'' language category (i.e., an increase of 1.13; Table 2) corresponded to a 59.57 percent increase in their probability of reoffending.

Discussion

The Department of Justice has long suggested that content analyses of digital evidence may be valuable for many types of criminal cases where online communication takes place.[2,4] At the same time, psychologists have been conducting computerized natural language analyses, showing that language patterns exhibit trait-like psychometric properties.[24,37,38] However, although forensic linguistic researchers have used linguistic analyses in the courtroom for years[39,40] only a handful of known studies have used *computerized* linguistic programs for these analyses.[6,7,19,35] Furthermore, little work has been done with explicitly *psychological* language analyses, which can not only provide objective measures for use in statistical modeling, but interpretable metrics that facilitate valuable psychological insights.[41] Accordingly, forensic computerized linguistic analyses are just beginning to gain attention from the legal community as potential sources of evidentiary support.

Our study extended previous work in this area by examining whether patterns in natural language data are predictive of recidivism within a large sample of Internet sex sting offenders. Previous work has shown that the word count, sexual words, and clout are associated with jail time in sentencing decisions.[7] Of those measures, however, only the clout category was predictive of recidivism in our current analysis; offenders scoring 1 *SD* above the mean were 60 percent more likely to recidivate. This is a notable finding, as it shows that

Table 3. Independent-Samples T Tests and Δ Recidivism Probabilities from A Binomial Logistic Regression Model Using Linguistic Inquiry and Word Count Language Dimensions

LIWC2015 category	T	p	M (SD) Nonreoffenders	M (SD) Reoffenders	Δp (Recidivism)$_{[+1\ SD]}$
Word count	1.12	0.26	5,387.10 (7,182.85)	6,741.65 (8,725.28)	—
Clout	2.04	0.04	80.78 (13.34)	**85.13 (10.64)**	+59.89%
Sexual	−1.14	0.94	1.14 (0.94)	1.24 (0.96)	—
First person singular pronouns	−2.26	0.02	**7.75 (1.75)**	7.11 (1.67)	−40.96%
Cognitive processes	−2.64	<0.01	**12.39 (2.11)**	11.48 (2.10)	−39.31%
Insight	−3.44	<0.001	**1.95 (0.72)**	1.55 (0.61)	−35.16%
Time	2.53	0.01	5.02 (1.11)	**5.48 (1.19)**	+59.57%
Ingestion	73.99	0.001	**0.45 (0.32)**	0.32 (0.21)	−35.71%

Notes: Results from independent-samples t tests and Δ recidivism probabilities from a binomial logistic regression model. For ease of interpretation, mean scores for each group are bolded to indicate significantly higher use of language from a given language-based measure of psychological processes.

the most visible, intuitive language categories that one might assume to be linked to reoffending (i.e., word count and sexual words) are not predictive of recidivism. In contrast, clout, along with some of the other more functional language dimensions (e.g., use of personal pronouns, cognitive processes, and insight), were predictive of recidivism. This is also important to note because these latter language dimensions may not be immediately apparent to the human eye, and current stage models of online grooming[30] do not differentiate between typologies of online offenders in terms of least and most likely to reoffend. These findings, coupled with the higher rates of time category words and lower rates of ingestion related words among reoffenders (as compared to nonreoffenders), allow us to develop a relatively clear picture of Internet sex sting offenders who recidivate. Those Internet sex sting offenders most likely to reoffend are more predatory in their language; they dominate the chat conversations with their fictitious underage targets and show little equivocation (using less ''I think I might'' language, and more ''We are going to'' language). They are not often sidetracked by conversations of what they are eating for dinner but would rather engage in conversations about when they are going to meet up for sexual activity and when the target's parents will be away (i.e., sexual and risk assessment stages).

Limitations and conclusion

Our study does have limitations that need mention. First, all transcripts were drawn from the Perverted Justice online archive. Although the cases were prosecuted in different jurisdictions, it could be that sting protocols that employ different conversational tactics (e.g., avoid talk about food when there is a lull in conversation), may have different results with regard to the ingestion category, specifically. Additionally, we were able to locate and analyze transcripts for only 334/538 (62 percent) of offenders in the sex offender registries; thus, our reported rate of recidivism may be lower or higher than the actual rate as the missing reports may not be random. However, the recidivism rate reported here (3.2 percent) is relatively equal to that found in other studies[17,18]; therefore, we expect that our rates are generalizable.

Overall, our findings align somewhat with Black et al.,[6] who found that sexual and risk assessment stages of online grooming occur early in chats, and with Drouin et al.,[7] who suggested that clout is an important factor to consider in prosecution of these cases. However, they also extend these studies in an important way by showing how these linguistic categories could be used to estimate probability of recidivism. From a practical standpoint, these analyses could be used during sentencing to help direct the trier of fact away from linguistic distractors, like word count and sexual words, and toward the linguistic predictors of recidivism (i.e., use of ''we,'' often checking schedules in attempts to meet up). More importantly, from an empirical standpoint, this study provides a model for how computerized linguistic analysis can be used to estimate and understand recidivism in forensic settings.

Notes

a. Perverted Justice Foundation Incorporated. www.pjfi.org
b. The PJFI transcript archives (www.perverted-justice.com/?con=full) were originally accessed on January 22, 2016.

c. According to the PFJI, the perpetrator initiated interactions in all cases (www.perverted-justice.com/index.php?pg=faq#cat2).
d. Previous analyses (Drouin et al.[7]) showed that the chat transcripts for male perpetrators with male stings were qualitatively and quantitatively different from those with male perpetrators and female stings.
e. The NSOPW links to individual jurisdictions; however, the Department of Justice ''does not guarantee the accuracy, completeness, or timeliness of the information contained in Jurisdiction Websites.'' Moderate rates (62 percent) of offender database matching may be attributable to incomplete Jurisdiction Websites combined with high numbers of incarcerated, deceased, or deported individuals.
f. Note that while we do not present significance tests from the binomial logistic regressions here, the results are parallel to those of the *t* tests.

Acknowledgments

Author Disclosure Statement

No competing financial interests exist.

References

1. Global Media and Intelligence Report. (2015) eMarketer.com. www.emarketer.com/go/2015gmisummary.aspx?publicid=chdZZ%2fSb0yd62ap1QQ8YLuIkBOOkXVghRlRoF36uUds%3d&s=1 (accessed Oct. 16, 2016).
2. National Institute of Justice. (2016) Digital evidence and forensics. www.nij.gov/topics/forensics/evidence/digital/Pages/welcome.aspx (accessed May 25, 2017).
3. Title III of the Omnibus Crime Control and Safe Street Act (The Wiretap Act) of 1968. 18 U.S.C. 2510–2520 (1968).
4. National Institute of Justice. (2007) Digital evidence in the courtroom: a guide for law enforcement and prosecutors. www.nij.gov/topics/courts/Pages/welcome.aspx# (accessed May 25, 2017).
5. Olson LN, Daggs JL, Ellevold BL, et al. Entrapping the innocent: toward a theory of child sexual predators' luring communication. Communication Theory 2007; 17:231–251.
6. Black PJ, Wollis M, Woodworth M, et al. A linguistic analysis of grooming strategies of online child sex offenders: implications for our understanding of predatory sexual behavior in an increasingly computer-mediated world. Child Abuse and Neglect 2015; 44:140–149.
7. Drouin M, Boyd RL, Hancock JT, et al. Linguistic analysis of chat transcripts from child predator undercover sex stings. The Journal of Forensic Psychiatry & Psychology 2017 [Epub ahead of print] http://dx.doi.org/10.1080/14789949.2017.1291707.
8. Finkelhor D, Mitchell KJ, Wolak J. (2000) *Online victimization: a report on the nation's youth. Crimes against Children Research Center.* Washington, DC: National Center for Missing & Exploited Children.

9. Jones LM, Mitchell KJ, Finkelhor D. Trends in youth Internet victimization: findings from three youth Internet safety surveys 2000–2010. Journal of Adolescent Health 2012; 50:179–186.

10. Wolak J, Mitchell KJ, Finkelhor D. *Internet sex crimes against minors: the response of law enforcement. Crimes against Children Research Center 2003*. Washington, DC: National Center for Missing & Exploited Children.

11. Mitchell KJ, Jones LM, Finkelhor D, et al. Understanding the decline in unwanted online sexual solicitations for U.S. youth 2000–2010: findings from three Youth Internet Safety Surveys. Child Abuse and Neglect 2013; 37:1225–1236.

12. Donovan B, Bourne J. (2016) Uber driver, college basketball player among 12 arrested in St. Johns County child sex sting. Action News Jax 2016, April. www.actionnewsjax.com/news/local/uber-driver-college-basketball-player-among-12-arrested-in-st-johns-county-child-sex-sting/212049010 (accessed May 25, 2017).

13. Harris R, Milligan, M. (2014) Principal among 14 arrested in child sex sting. CBS46, 2014, March. www.cbs46.com/story/24864531/14-arrested-in-dekalb-county-in-connection-with-child-pornography (accessed May 25, 2017).

14. Holley P. (2014) Fla. child sex sting nets suspects from Disney and SeaWorld, plus a Christian football coach. Washington Post 2014, March. www.washingtonpost.com/news/morning-mix/wp/2016/04/06/fla-child-sex-sting-nets-suspects-from-disney-seaworld-and-a-christian-football-coach/?utm_term=.3066a7b258af (accessed May 25, 2017).

15. Perverted Justice. (2017) http://perverted-justice.com (accessed May 25, 2017).

16. Seto, MC, Hanson, RK, Babchishin, KM. Contact sexual offending by men with online sexual offenses. Sexual Abuse: A Journal of Research and Treatment 2011; 23:124–145.

17. Goller A, Jones R, Dittmann V, et al. Criminal recidivism of illegal pornography offenders in the overall population—a national cohort study of 4612 offenders in Switzerland. Advances in Applied Sociology 2016; 6:48–56.

18. Seto MC, Wood J, Babchishin KM, et al. Online solicitation offenders are different from child pornography offenders and lower risk contact sexual offenders. Law and Human Behavior 2012; 36:320–330.

19. Bond GD, Lee AY. Language of lies in prison: linguistic classification of prisoners' truthful and deceptive natural language. Applied Cognitive Psychology 2005; 19:313–329.

20. Carey AL, Brucks MS, Küfner AP, et al. Narcissism and the use of personal pronouns revisited. Journal of Personality and Social Psychology 2015;109:e1–e15.21.

21. DeWall CN, Buffardi LE, Bonser I, et al. Narcissism and implicit attention seeking: evidence from linguistic analyses of social networking and online presentation. Personality and Individual Differences 2011; 51:57–62.

22. Kacewicz E, Pennebaker JW, Davis M, et al. Pronoun use reflects standings in social hierarchies. Journal of Language and Social Psychology 2014; 33:125–143.

23. Newman ML, Pennebaker JW, Berry DS, et al. Lying words: predicting deception from linguistic styles. Personality and Social Psychology Bulletin 2003; 29:665–675.

24. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language use: our words, our selves. Annual Review of Psychology 2003; 54:547–577.

25. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and Social Psychology 2010; 29:24–54.

26. Pennebaker JW, Boyd RL, Jordan K, et al. (2015) *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin.

27. Pennebaker JW, Chung CK, Ireland M, et al. (2007) *The development and psychometric properties of LIWC2007*. Austin, TX: LIWC.net.

28. Pennebaker JW, Francis ME, Booth RJ. (2001) *Linguistic Inquiry and Word Count (LIWC): LIWC 2001 manual*. Mahwah, NJ: Erlbaum.

29. Colleluori AJ. (2010) Defending the internet sex sting case. GPSOLO January/February 2010. www.americanbar.org/newsletter/publications/gp_solo_magazine_home/gp_solo_magazine_index/colleluori.html (accessed May 25, 2017).

30. O'Connell R. (2003) A typology of cyber sexploitation and online grooming practices. Preston, England: University of Central Lancashire. http://image.guardian.co.uk/sys-files/Society/documents/2003/07/17/Groomingreport.pdf (accessed May 25, 2017).

31. Boals A, Klein K. Word use in emotional narratives about failed romantic relationships and subsequent mental health. Journal of Language and Social Psychology 2005; 24: 252–268.

32. Joksimovic S, Gasevic D, Kovanovic V, et al. Psychological characteristics in cognitive presence of communities of inquiry: a linguistic analysis of online discussions. Internet and Higher Education 2014; 22:1–10.

33. Arguello J, Butler BS, Joyce E, et al. (2006) Talk to me: foundations for successful individual-group interactions in online communities. In *Proceedings of the CHI'06 conference on human factors in computing systems*. New York: Association for Computing Machinery Press, pp. 959–968.

34. Danescu-Niculescu-Mizil C, West R, Jurafsky D, et al. (2013) No country for old members: user lifecycle and linguistic change in online communities. In *Proceedings of 2013 www Conference*.

35. Pennebaker JW. (2011) *The secret life of pronouns.* New York, NY: Bloomsbury Press.

36. Hosmer DW Jr., Lemeshow S, Sturdivant RX. (2013) *Applied logistic regression*. 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc.

37. Boyd RL, Pennebaker JW. Did Shakespeare write Double Falsehood? Identifying individuals by creating psychological signatures with text analysis. Psychological Science 2015; 26:570–582.

38. Pennebaker JW, King LA. Linguistic styles: language use as an individual difference. Journal of Personality and Social Psychology 1999; 77:1296–1312.

39. Cotterill J. Domestic discord, rocky relationships: semantic prosodies in representations of marital violence in the O.J. Simpson trial. Discourse and Society 2001; 12:291–312.

40. Grant T, Macleod N. Assuming identities online: experimental linguistics applied to the policing of online paedophile activity. Applied Linguistics 2016; 37:50–70.

41. Boyd RL. Psychological text analysis in the digital humanities. In Hai-Jew S, ed. *Data analytics in the digital humanities*. New York City, NY: Springer Science, pp. 161–189.

Address correspondence to:
Dr. Michelle Drouin
Department of Psychology
Indiana University—Purdue University Fort Wayne
2101 E. Coliseum Boulevard
Fort Wayne, IN 46805

E-mail: drouinm@ipfw.edu