Language Properties of Phone Scammers:
Cyberdefense at the Level of the Human
Aug. 4, June 2016
Judith Tabron, Ph.D.
Hofstra University
Overview:

Forensic linguistics is the study of language as evidence for the law. It is a relatively new field and has not previously been applied to cybersecurity. Linguistic analysis uncovers several features of language interaction in a limited data set (recorded IRS phone scammers) that begin to answer how forensic linguistics could assist in cybersecurity defense.

Current phone security technology records all calls and in some cases analyzes them to ascertain the identity of the caller. Forensic linguistics could be used to identify the crime, not the person, and therefore has a broader - and cheaper - applicability.
Judith Tabron

Studying languages
- InterNIC
- VMS

World literatures in English
- Global popular culture
- Fiction writing
- Screenwriting

People management
- Project management
- Multimedia development

Cybersecurity

Unix, Web stuff, sysadmin
- Windows - DOS to NT

Forensic linguistics

Teaching
My mother was a sysadmin for DARPA.
Cybersecurity and Forensic Linguistics: What’s out there for securing the human?
Phoneprinting: Pindrop

- Won a patent and $35 million in investor funding last year
- At that time already used at 2 of the 4 biggest banks
- Builds a profile not just of your voice but of your voice on your phone line
Phoneprinting: Pindrop
Phoneprinting: Pindrop
Phoneprinting: Pindrop

Identifies the Caller
And the Caller’s location
Uptivity: “Stress detection”

Formerly CallCopy

- Now available to companies “of all sizes”
- Sound based
- Keyword spotting
- Accuracy “greater than 80 to 90 percent”
- Real time
Uptivity: “Stress detection”

“Talk analysis can identify patterns within calls, such as long hold times or periods of silence, as well as the frequency of an agent cutting off a caller.” (underlining mine)
IRS scam phone call features both gaps and interruptions
Competitors in this space

- **Verint** (acquired Victrio)
  Creates a voiceprint for on-the-fly user identification
  - Identifies known criminals by voiceprint

- **TrustId**
  Verifies ANI before the phone call even starts - “pre-call authentication”

- **HP Autonomy Inc.**
  Collects unstructured data for adding to CRMs, etc.
Wanted to improve phone security:

- Cheaper
- Fewer privacy issues
- Focus on a different goal
- Fail closed
Forensic Linguistics

Applied linguistics to understand language evidence (according to legal standards)

- Threat analysis
- Author attribution
- Legal language disputes
- Conviction exoneration (Innocence Project)

...not deception detection (for me)
Forensic Linguistics

“I’m sorry.” “Hey, it’s okay.”

“The dog ate my dessert.”

eat, ate

eat, /eɪt/
Forensic Linguistics

Roger Shuy and the “devil strip” case
-creating a profile to help locate the author of a ransom note
Forensic Linguistics

Do you ever want to see your precious little girl again? Put $10,000 cash in a diaper bag. Put it in the green trash kan on the devil strip at corner 18th and Carlson. Don’t bring anybody along. No kops!! Come alone! I’ll be watching you all the time. Anyone with you, deal is off and dautter is dead!!!
Do you ever want to see your precious little girl again? Put $10,000 cash in a diaper bag. Put it in the green trash kan on the devil strip at corner 18th and Carlson. Don’t bring anybody along. No kops!! Come alone! I’ll be watching you all the time. Anyone with you, [the] deal is off and [your] dautter is dead!!!
Do you ever want to see your precious little girl again? Put $10,000 cash in a diaper bag. Put it in the green trash kan on the devil strip at corner 18th and Carlson. Don’t bring anybody along. No kops!! Come alone! I’ll be watching you all the time. Anyone with you, deal is off and dautter is dead!!!
Training police to emulate the language of a potential victim

G-clipping in Pilgrim data

Dr. Nicci MacLeod, Aston University, UK
Forensic Linguistics

Robert Leonard’s work identifying the author of a threat letter that preceded a murder Forensic Files episode 30 (season 11) “A Tight Leash”

The author did not use negative contractions
Jim Fitzgerald of the FBI did similar profiling work -- the Unabomer case and with threat letters in Australia.
Forensic Linguistics

Tammy Gales’ work on FBI’s threat database - What are the features of threats that are carried out versus threats that are not not carried out?

- Corpus analysis
Forensic Linguistics

Tim Grant in the UK - work identifying the author of SMS messages in a murder case

● “TXT 4N6” article in various venues
Forensic Linguistics

My internship with Robert Leonard and Eric Freedman of the Hofstra Law School, Distinguished Professor of Constitutional Rights

- Improving instructions for juries considering the death penalty
Forensic Linguistics

.... Not deception detection.

.... And not a “unique fingerprint”.
Forensic Linguistics

Can we identify features of social engineering phone interactions such that we could teach our community to recognize when they are occurring?

“Can you reset my password?”
Forensic Linguistics

“I’m sorry.” “Hey, it’s okay.”

“The dog ate my dessert.”

Sociolinguistics

Semantics

Pragmatics

Morphology
eat, ate

Phonology
/eit/

Orthography

Discourse analysis
Forensic Linguistics

- Not automated
- Strengthens the human link, the weakest link in cybersecurity defense

“Can you reset my password?”
Forensic Linguistics

- Not automated
- Strengthens the human link, the weakest link in cybersecurity defense

“Can you reset my password?”

Identifies the Crime
Forensic Linguistics

IRS Scam phone calls

“This January, the Treasury Inspector General for Tax Administration (TIGTA) announced they have received reports of roughly 896,000 contacts since October 2013 and have become aware of over 5,000 victims who have collectively paid over $26.5 million as a result of the scam.”

“Police are coming to arrest you”
Forensic Linguistics

- Americans are uncertain about tax penalties
- And we tend to blame victims of fraud
- Fraud is underreported and underprosecuted

“Police are coming to arrest you”
Forensic Linguistics

Qualitative study of IRS scam phone calls posted on YouTube

Transcribed and analyzed for discourse features

“I’m sorry.” “Hey, it’s okay.”

“I’m sorry.” “Pass the bread.”
Polar tag questions, okay?
### Forensic Linguistics

<table>
<thead>
<tr>
<th>Time</th>
<th>Speaker</th>
<th>Soundmark</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>3:25</td>
<td>50</td>
<td></td>
<td>All right ma’am you would be getting this tax pay order (.) there, OK?</td>
</tr>
<tr>
<td>3:30</td>
<td>51</td>
<td>OK</td>
<td></td>
</tr>
<tr>
<td>3:31</td>
<td>52</td>
<td></td>
<td>So, ma’am but first we have to go to that store before going to that store ma’am, because the store people they won’t be, they don-, they just only accept the cash, so you have to go to your bank first, need to withdraw the money, and then you have to go to the store, to purchase this tax pay vouchers. All right?</td>
</tr>
</tbody>
</table>
Forensic Linguistics

Polar tag questions:

- Prefer a response
- Prefer a positive response
- Prefer a response that accepts the terms of the previous statement

7 to 33 occurrences on the part of the scammers in my samples
Forensic Linguistics

Topic control & Question deferral

“Can I just --”

“No.”
Speaker B informational questions and Speaker C responses

who do I send them to
what’s your address
why don’t you just give me, um, your address
so can you
can you let me know what I do should I just come home and call you
who to send this to would you
where do I mail this
do I mail it to the IRS
Are you ready?
are you ready? [2nd occurrence]
do you want me to put you on hold…
which Winn Dixie do you want me to go to?
What’s your name?
Are you gonna hold?
deflected to future
deflected to future
deflected to future
deflected with “anger”
deflected with “anger”
deflected to future
deflected to future
deflected to future
responds with a question
responds with a question
deflected to future
deflected to future
ignored
Answered ("Yes")
Violations of narrative structure
Forensic Linguistics

Violations of narrative structure
William Labov, 1960s, ff

1. Abstract: What is the story about?
2. Orientation: Who, when where, how?
3. Complicating action: Delays resolution, adds suspense...
4. Evaluation: Who's the hero? Who's the bad guy?
5. Result / Resolution: What happened in the end?
6. Coda "And that's why I can never go back there again."
Forensic Linguistics

Violations of narrative structure

“You owe the IRS.
There’s nothing you can do about it.
Police are coming to arrest you.

…..”
Forensic Linguistics

Violations of narrative structure

May be the toughest to realize in the moment

Tends to require after-the-fact analysis

… but I’m giving you the information
Forensic Linguistics

*Polar tag questions & question deferral:*
Can you strengthen your organizations by educating them about these linguistic peculiarities?
Forensic Linguistics

The Institute for Forensic Linguistics, Threat Assessment, and Strategic Analysis at Hofstra University

Robert Leonard
Tammy Gales
Forensic Linguistics

Judith Tabron

judith.tabron@gmail.com

@judithtabron