Forensic Linguistics

Robert A. Leonard

Forensic Linguistics[1]

Sometime in 2001, Brian Hummert discovered a letter on his car's windshield that began, "Here is the proof that your wife is a slut." The letter writer said he had engaged in a "one niter" with Charlene Hummert years before, and she had ruined things for him with his girlfriend. In chillingly precise detail, he described Charlene's recent movements and activities. He related that she had bought sex paraphernalia through the mail. He had followed her through a local York, Pennsylvania shopping mall, and not only did he know that she had had a "glamour photograph" taken at a certain gallery, but he had also obtained a copy of that photograph—Charlene holding a red rose—and included it with the letter. The writer even knew about the Hummerts' home surveillance camera and the code—7805—of their security system. The letter stated: "the time is now right for payback."

In March of 2004, Charlene Hummert was found strangled to death, her body dumped in the back of her SUV and abandoned in a supermarket parking lot. Police discovered a blurry surveillance video of a suspect entering the supermarket. The autopsy showed the cause of death to be ligature strangulation and a search of the Hummert home yielded, among other evidence, a red dog

[1] The English language has two main meanings for the word linguist: one, a speaker adept at a foreign language (indeed, in many agencies, such as the FBI, a "linguist" has this meaning), and two, a scientist who studies human language as a set of real-world phenomena. Academic, scientific linguists belong to the second group (although many are also adept at foreign languages).

Forensic linguistics is the application of the science of linguistic investigation to issues of law. Forensic linguistics augments legal analysis by applying rigorous, scientifically accepted principles of analysis to legal evidence like contracts, letters, wills, confessions, and recorded speech.

Linguists—as all scientists—seek to explain the nonrandom distribution of data. Just as bullets do not randomly issue from firearms nor chemical concentrations randomly spread throughout a human body, words are not randomly found to issue from the keyboards and mouths of speakers of English or any other language. Words adhere to patterns; these patterns are the subjects of systematic observation by scientific linguists.

As in all other sciences, linguistics solves problems by constructing competing hypotheses and then testing which hypothesis better explains the nonrandom distribution of the data. For example, Galileo demonstrated that while the hypothesis that the Sun revolves around the Earth explained much of the data (it certainly looks like it does)—the competing hypothesis, the Copernican heliocentric model that states the Earth revolves around the Sun, explained more of the nonrandom distribution of the data (for example, the observed, nonrandom orbits of the planets), and explained the totality of the data better, and was therefore the superior hypothesis.

R.A. Leonard (✉)
Institute for Forensic Linguistics, Threat Assessment and Strategic Analysis, Hofstra University, Hempstead, NY 11549, USA
e-mail: cllral@hofstra.edu

leash that matched the markings on Charlene's neck. Her pants were on backwards, suggesting she had been dressed after her murder. Someone apparently then dragged her across her own driveway into her car. Physical evidence was overwhelming that she had been killed at home. At this juncture, a letter was received by the lead detective and the press from a self-confessed serial killer. It read, "This is the fifth woman I killed. I'm getting good at it." It was signed, "John."

Modern Forensic Linguistics Is the Application of the Science of Linguistics to Issues of the Law

Linguistics is the scientific study of language. With hundreds of professional peer-reviewed journals, it is a well-established science, recognized by the American Academy of Sciences, and regularly granted research funds by the National Science Foundation. In virtually any major university or college, a student can specialize in linguistics and many major universities grant a Ph.D. degree in linguistics. Forensic linguistics applies linguistic science to legal cases, such as this murder, and is recognized by the courts. NOTE: See Coulthard, 2004, Coulthard & Johnson, 2010, Grant, 2013, Leonard, 2006, Leonard, 2012, McMenamin, 2002, McMenamin, 2004, Shuy, 1993, Shuy, 1998, Shuy, 2006, Shuy, 2014, Solan & Tiersma, 2004, Tiersma & Solan, 2002

This chapter focuses on two case studies—first, the murder of Mrs. Hummert, and second, the kidnapping of a little girl—that exemplify investigatory strengths of forensic linguistics.[2]

Among other things, forensic linguistics narrows the suspect pool of possible authors, discerns demographic information from language evidence, and then, given samples from subjects, helps identify or disallow possible authors. In this case, the Pennsylvania State Police Major Case Team wanted to know, "What can you tell us about whoever wrote these letters?" Knowing almost nothing about the murder, my colleague Dr. Benji Wald and I analyzed the letters.

Authorship analysis seeks to answer questions such as who wrote a bomb threat, a ransom note, a threat letter, a blog post, or an e-mail. For example, in a case the prosecutor referred to as the "Facebook Catfishing Murders of East Tennessee" I was asked to seek the identity of the "CIA agent Chris" who sanctioned the assassination of a young couple.

Linguistic demographic profiling and authorship analysis are on a continuum, incrementally narrowing down the suspect pool. One links the questioned (Q) documents to ever smaller groups.

An illustrative example of a profiling case is one analyzed by my research partner, Dr. Roger Shuy, the founder of forensic linguistics in the USA. Investigators in the Midwest gave him a ransom note and asked essentially the same question the police asked me in the Hummert case: What can you tell us about the writer of the document? The ransom note given to Dr. Shuy was "scrawled in pencil" and left on the doorstep of the parents, who contacted the authorities, who then came to Dr. Shuy. Here is the note, transcribed[3]:

---

[2]Forensic linguists help investigators, litigators, triers of fact, and threat assessment and threat management professionals *extract maximum intelligence from language evidence such as letters, e-mails, notes and texts*, often involving cases such as the following, types other than those discussed in this chapter, e.g.:

- Overt and implied threats of violence
- Stalking
- Assessment and interdiction of internal and external threats
- Identification of extortionists and threateners
- Fraud detection, investigation, and deterrence
- Internal and external illegal information leaks
- False information spread
- Corporate espionage detection and interdiction

---

[3]Dr. Shuy, Georgetown Distinguished Research Professor of Linguistics, Emeritus, wrote in 2001 about this case, and linguistic demographic profiling in general: "It is believed that the idea of psychological profiling originated in the Behavioral Science Laboratory of the FBI, where specialists in psychology and criminology worked together to assess the characteristics that would point to a specific type of perpetrator of a recent crime. Several non-governmental groups now offer their psychological profiling services to private industry after hate mail or threat messages are received. Until recently, however, such profiling has not included the analysis of linguistic clues about the geographical origins, socioeconomic status, race, age, gender, and even occupation of the writers. In short, the resource of knowledge about dialect geography, lexicography, and sociolinguistics has been largely overlooked for this task. Threat letters and ransom notes can be a rich source of forensic information. The problem is that most law enforcement officers and prosecutors are unfamiliar with linguistic variation in English speech and writing that can give them the most help." (Shuy 2001, p. 2).

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!!!

Certain features jump out. Probably the most noticeable are the obvious misspellings:

- the substitution of *k* for *c*: kan for can, kops for cops
- the spelling of daughter as dautter

Further, the cadence and structure of the last sentence is awkward, and omits some possible words:

- It reads "Anyone with you, deal is off and dautter is dead!!!"
- It could read: "*If* anyone *is* with you, *the* deal is off and *your* dautter is dead!!!"

There are any number of possible explanations for the misspellings and the odd last sentence. Three of them are:

1. The writer is a native speaker of another language that only uses *k* for the K sound in *cat* and *cops*. (But note that *corner*, *Carlson*, and *come* are spelled not with *k*, but correctly, with *c*.) If the writer is a nonnative speaker of English, this can also explain the poorly done last sentence.
2. The writer is an English speaker, but only partially literate.
3. The writer is well educated and is consciously *pretending* not to know how to spell or write a standard sentence—that is, the misspellings and poor last sentence are an attempt at *disinformation*.

It is of course normal for analysts to have several possible explanations for the patterns they see. As in other sciences, we treat these as *competing hypotheses*, and, as in other sciences, the question becomes: which is the *superior hypothesis* that can best "*explain the nonrandom distribution of the data*"?

One obvious measure of superiority is that a hypothesis can explain all, or more, of the data, rather than only some of it. So while Hypothesis 1 can explain some of the data—the last sentence, and *kops* and *kan*—it cannot account for *corner, Carlson*, and *come.*

Hypothesis 2 can explain the last sentence, and all the misspellings. It can also explain the variation between the *kops* and *kan* misspellings and the correct spellings of *corner*, *come*, and *Carlson*, since a semiliterate writer might only sometimes misspell simple words, and thus might spell *corner* and *come* correctly, and might also know the street name *Carlson* by sight, from street signs. But there are patterns in the data that Hypothesis 2 cannot explain, for example: *watching*, *diaper*, and *precious* are spelled correctly, and, even more important—because it is systematic and not a single word like *precious*, the spelling of which perhaps an uneducated writer could look up—the *punctuation* in the *entire* ransom note is quite standard, and fully literate. It also makes sense to the forensic linguistic analyst that punctuation might be unnoticed by someone attempting to "dumb down," and not be foremost in a person's mind as a giveaway to educational level. Further, while it is possible to attempt to dumb down one's writing, it is less likely, and far more difficult, to "dumb up," to coin a phrase—indeed, that is rare; dumbing down is common.

Note also how the use of *precious* in the lead sentence "Do you ever want to see your precious little girl again?" conveys a tone that is mocking and cruel. This further supports Hypothesis 3—the word *precious* is totally unnecessary to the mere functionality of the note—"Do you ever want to see your ___ little girl again?"—would work as well; or the whole sentence could just be omitted. The mocking, cruel use of the word suggests someone with a reasonable command of the language—certainly someone who could spell *cops* and *can*. So Hypothesis 3 is superior—it can explain all the patterns in the entirety of the data. We are left with the conclusion that the writer is well educated but is trying not to seem so.

What else can the ransom note tell us about the author? There is a further clue that, especially if one is a fluent reader, is easy to miss the first few times one reads the note. An important skill

all fluent readers have is being able to unconsciously ignore terms that do not immediately make sense, and continue along to get the gist of the meaning even without that one piece. In the present case, it is the term *devil strip* in this sentence:

> Put it in the green trash kan on
> the devil strip at corner 18th and Carlson.

What is the *devil strip*? It is not a very common term, and it turned out to be an extremely important piece of evidence for Dr. Shuy as he sought to learn what there was to know about the ransom note's author. *Devil strip* is a term for the strip of grass in between a sidewalk and the curb. The reason it was important is that it is a term used in and around Akron, Ohio—and ONLY in that area. Outside the Akron area the term is relatively unknown.[4] I once gave an address in Columbus, Ohio where I discussed this case. A District Attorney told me she was standing behind two uniformed police officers, one from Akron and the other from nearby Columbus. When I mentioned *devil strip* the officer from Akron turned to his friend and said, "But that's what everybody calls it, no?" It was not.

This is even a better clue to the geographical speech community of the writer than it seems at first. The reader should consider what word he or she uses for the strip of grass in between a sidewalk and the curb. If you have one—and most of the hundreds of speakers I have asked do not—it is highly unlikely you realize your term is only regional, and not simply *the* term for it, like *cat* is the normal, generic, universal name for that animal. By using *cat*, you would not think you were giving away much information on where you were from, and you would be right.

Some respondents suggest they call such a grass strip a median, or some other such term. In New Haven, I discovered, it is called a *planter strip*. In Nassau County, NY, my father, a govern-

ment official, called it a "county strip" and explained that the county controlled it. I never thought it might have another name until I came across Roger Shuy's devil strip case. Thus, a term like this is unlikely to be consciously manipulated to deceive, so we may take the information that it gives at face value; the ransom note author quite likely thought he was using a generic term, as generic and universal as "trash kan." Perhaps in the future, when criminals routinely assume their words will be analyzed by trained forensic linguists, they might plant disinformation clues that are that subtle. But that time has not yet come.

So, Shuy looked at the note and asked the police if they had, on their suspect list, a "well-educated person from Akron." They did, and were no doubt amazed at the rapid, precise, detailed Sherlock Holmesian feedback. Shuy explained his rationale, and armed with this, the police presented this analysis to that suspect, and he confessed. This is an excellent point in our discussion to stress that in authorship cases, forensic *linguistics cannot identify a particular individual as a writer or speaker*. But contrary to popular lore, even DNA—considered the gold standard of forensic tests—cannot identify a particular individual. DNA can exclude a suspect, it can narrow a suspect pool, but its practitioners state that it does not identify individuals.[5]

4 A sign in Akron, Ohio.

---

[5] As Dr. Daniel Drell of the Human Genome project states: "If the sample profiles don't match, the person did not contribute the DNA at the crime scene. If the patterns match, the suspect may have contributed the evidence sample…there is a chance [–exceedingly slim–] that someone else has the same DNA profile." Analogous to DNA matches, Dr. Drell explains the value of gathering "linking evidence in a chain":

Assume that type O blood is found at the crime scene. Type O occurs in about 45% of Americans. …If, in addition to being type O, the suspect is a blond, and blond hair is found at the crime scene, you now have two bits of evidence to suggest who really did it. … If you find that the crime scene has footprints from a pair of Nike Air Jordans (with a distinctive tread design) and the suspect, in addition to being type O and blond, is also wearing Air Jordans with the same tread design, you are much closer to linking the suspect with the crime scene. In this way, by accumulating bits of linking evidence in a chain…you can argue that your suspect really is the right person (Drell, n.d.).

A similar case holds for linguistic evidence: by accumulating matches of linguistic features as links in a chain, in addition to nonlinguistic evidence, a much stronger inference may be made that a suspect is the correct person.

Similarly, forensic linguistics can exclude suspects and narrow the suspect pool. For example, if a note is written in Mandarin, and all good intelligence says that a suspect does not know Mandarin, she is excluded. On the other hand, if the note presents consistent features of, e.g., New York dialect, that narrows the suspect pool.

In court cases, it is solely the task of the trier of fact—the jury, or in a bench trial a judge—to determine if the expert's analysis of the distribution of linguistic patterns actually should cause them to conclude that a particular suspect wrote a particular document, and, further, whether that helps them decide that he is guilty or not guilty. There may be other persons in the world who could have generated the same or similar patterns as we find in whatever document is of interest.

Turning back to demographic profiling, recall that throughout the ransom note the punctuation was fully literate, and this suggested a well-educated writer. Punctuation, and orthography in general (i.e., spelling, spacing, and other aspects of transmitting a language into writing), can also indicate possible demographic features.

Readers familiar with Spanish will recognize the inverted question mark placed at the beginning of a question, as in *¿Quieres ir?* ("Do you want to go?"). This is so different from English that it is unlikely a Spanish speaker writing English would unconsciously write "¿Do you want to go?" But there are other, subtler differences.

Spanish (and other languages as well) does not capitalize the names of months, days of the week, nationalities, and languages, as does English, as in "*El español es una idioma bonita*" ("Spanish is a beautiful language"). Consider:

> I'm always watching when she walks to spanish class.

Not capitalizing *Spanish* could be merely a mistake, or it might be one indication of a Spanish speaker as opposed to an English speaker. In an actual threat case in California, we noticed:

> I challenge that you have the right to have her to yourself. I have known her since a very long time myself, perhaps even longer than you.

Consider *since*. This use of *since* suggests a nonnative English speaker. English has *for a long time* and *since 2013* but not *since a long time*. Perhaps this was a direct translation of French *depuis longtemps* (literally "since a long time"). Was there a French speaker in the suspect pool? There was, and as it turned out, patterns in his known documents matched several other patterns in the threat as well.

Concerning French vs. English speakers, consider the following, from an Internet page on French and English:

> The French punctuation requires a space before double signed punctuations marks such as:
>
> this one :
> this one ;
> this one ?
> this one !
> and this one %

Is this writer likely a speaker of English or French? (That there are spaces before the punctuation marks in the example do not indicate either a French or English speaker, because they are illustrations of correct French.) Let us assume, for non-linguistic case-related reasons—perhaps the suspect pool only has two such speakers—that there are only these two choices: English, or French (of course, from only the data given here, the writer may be a speaker of a language other than English or French).

> There are two indications the person is French rather than an English speaker: the use of *the* and *-s*:
> **The** French punctuation requires a space before double signed punctuation**s** marks such as…
> Standard English would be "French punctuation…" French would be "La [The] ponctuation française…".

As for *punctuation**s***, that, of course, is not English. A French speaker might use it because the word "ponctuations" is a commonly used plural noun that by itself means "punctuation marks," or she perhaps mistakenly thinks it should agree with mark**s**. In any event, that *-s* points more towards a French speaker than an English one.

Indeed, the person who contributed this example to the website signed in as a French woman:

Agnès E.; Senior Member; location: France; native language: French of France [others are from Belgium and elsewhere].

A further example, from a phishing e-mail, reveals someone doing a poor job of masquerading as the English-speaking iTunes team. There is more than one mistake, but just consider the first line, which of course has an un-English space before the final "!":

Verify your iTunes account !
Dear customer,
We have received your iTunes account is used for fraud. Your account will be suspended until you confirm that you are the original user account.

– To confirm that you are the original user of this account: Click here

"We have received your iTunes account is used for fraud" does not specify whatever they claim to have received that led them to believe that your account is "used for fraud"—*being used fraudulently* would be a decent English phrase. My professional advice: Don't click there.

So we have seen how seemingly small details can provide intelligence that can prove useful in the investigation of cases. Let us return to the Hummert murder, to analyze the Stalker and Serial Killer letters. Here are the letters as we received them, followed by the retyped texts of the letters:

On the surface, the letters seem very different.

Here is the proof that your wife is a slut. Do what you will with it. Sorry it took so long. I only come occasionally back to the area on business. Merry Xmas. I will send you several copies of this so you get the information in case the slut intercepts one.

Before I tell you how I got it, I want to tell you a little about myself. I played in a band back it the late seventies/early eighties. I had a one niter with your wife. She was a fine piece of ass that I enjoyed several times that night. Rumor had it that she occasionally took several guys at once and she sucked cock really well. I would have loved to have found out. A couple of days later she made sure my fiancée found out. She dumped me and then had an abortion. We have since patched things up and gotten married, but she can't have any children. I blame your wife for that. The time is now right for payback. I hope to see your wife miserable the next time I am in the area.

I ran into your wife back in September at Gabriel Brothers. I almost didn't recognize her with her dyed hair. I have been following her around hoping she would mess up. On October 6, I followed your wife over to Capitol City Mall. She was dressed up more the usual for a Saturday of shopping. She went into the Picture People. This was around 10 AM. A couple of weeks later I went in and got copies of the pictures enclosed. On the negative holder she had written that the photo was a gift. There was no indication of which one she had printed up.

I ask you who was it for? Also she does not have her wedding ring on. Why not? A red rose is a symbol of love. For who? I don't think you know about these. Do you? Also she has purchased a lot of sexy bras and panties. Have you seen them or the red nightie? Were they brought for your enjoyment? You may also want to ask her about her Spencer Gift purchases. Do you love lubes with her? So you see once a slut always a slut.

Stalker Letter

Here is the proof that your wife is a slut. Do what you will with it. Sorry it took so long. I only come occasionally back to the area on business. Merry Xmas. I will send you several copies of this so you get the information in case the slut intercepts one. Before I tell you how I got it, I want to tell you a little about myself. I played in a band back in the late seventies/early eighties. I had a one niter with your wife. She was a fine piece of ass that I enjoyed several times that night. Rumor had it that she occasionally took several guys at once and she sucked cock really well. I would have loved to have found out. A couple of days later she made sure my fiancée found out. She dumped me and then had an abortion. We have since patched things up and gotten married, but she can't have any children. I blame your wife for that. The time is now right for payback. I hope to see your wife miserable the next time I am in the area.

I ran into your wife back in September at Gabriel Brothers. I almost didn't recognize her with her dyed hair. I have been following her around hoping she would mess up. On October 6, I followed your wife over to Capitol City Mall. She was dressed up more the usual for a Saturday of shopping. She went into the Picture People. This was around 10 a.m. A couple of weeks later I went in and got copies of the pictures enclosed. On the negative holder she had written that the photo was a gift. There was no indication of which one she had printed up.

I ask you who was it for? Also she does not have her wedding ring on. Why not? A red rose is a symbol of love. For who? I don't think you know about these. Do you? Also she has purchased a lot of sexy bras and panties. Have you seen them or the red nightie? Were they brought for your enjoyment? You may also want to ask her about her Spencer Gift purchases. Do you love lubes with her? So you see once a slut always a slut.

Serial Killer Letter

I killed Charlene Hummert, not her husband. We had an affair for the past nine months. She wanted to break it off. So I broke her neck! I wrote letters to her husband and to Det. Loper [the lead detective].

I used a white nylon rope to kill her they won't find me I am leaving. I am writing because of Easter. I am sorry I killed her.
They won't find the cell phone she used to call me, it is in the river and not under my name.
I carried her into the kitchen and then dragged her outside to her car. This is the fifth woman I killed. I am getting good at it.
Cops have no idea how easy it is to pin husband when they only look there.
She knew about pictures on PC. She told story to set up husband for the Divorce. Ha Ha
ByeBye for now
John

What can an analyst find in these letters that responds to the question from the police? Remember, as we saw in the *devil strip* ransom note, the writer(s) may well have attempted to disguise their language patterns. Thus, the most important advice here is not to take anything in the document at face value—neither the content, nor the language used in the document. It is safe to assume that if writers do not sign their real names at the bottom of a document then they likely do not want to be identified. One must always **beware of disinformation**. I have used these two letters for training purposes many times, and it is common for people to start analyzing the different psychological underpinnings of the authors. But none of the content in these letters can be assumed to be true.

As we have seen, disinformation on demographic features such as education level or dialect may be revealed in their *inconsistent patterns*. That is, unless one is a trained linguist—and even then—**it is difficult to assume a false linguistic identity**. It is difficult to alter all systems of language together and to the same degree. There are simply too many systems and details to keep track of. We saw this in the devil strip case, where the reader dumbed down his spelling, but not his punctuation.

These two letters have many surface differences. The writer of the stalker letter gives evidence of being more educated than the author of the serial killer letter. There are ungrammatical and other peculiarities in the second letter. But we must remember the circumstances of the writing of both letters. What were their immediate goals? The stalker letter intended to embarrass,

reveal secrets, and cause havoc. The serial killer letter intended to do something else, even though, like the first letter, it tells a story of Charlene Hummert's alleged infidelities. It begins, "I killed Charlene Hummert, not her husband," and ends, "Cops have no idea how easy it is to pin husband when they only look there…She knew about pictures on PC. She told story to set up husband for the Divorce." It is clearly stating that the police are wrong to suspect the person who, by this stage of the investigation was their prime suspect, and, although Dr. Wald and I did not know it, was already in custody. That was Charlene's husband, Brian.

Without going into explanations of the more technical terms, here is a nonexhaustive list of some linguistic investigative features that we might use in such a case to demographically profile and also to compare the two letters for possible common authorship:

- choice of words or syntax that may indicate dialect, or underlying native language;
- grammar, e.g., clause embedding, preposition usage, discourse markers, "that" complementizer deletion;
- patterns of usage and nonstandard punctuation;
- management of narrative time structures, and how departures from the narrative sequence (flashbacks, flash-forwards, asides) are handled;
- word choice;
- mechanics of register type, e.g., letter, ransom note, detective novel; formality level;
- style mechanics, e.g., parallel structures.

As noted before, it is important always to be aware of possible disinformation, including what I term *masking* and *masquerade*. Masking is simply attempting to mask one's own normal usage, as we saw in the dumbing down of the author of the devil strip note. To masquerade is more focused, disguising one's normal language patterns in an attempt to assume a particular false identity. This is attempting to write in the voice or style of someone else. We commonly see this, for example, in analyzing the circumstances of changed wills: someone attempts to mimic the

language of a terminally ill person and writes instructions to change the will to leave everything to them or a crony.

A variant of masquerade is what we see in the serial killer letter, and to analyze this type of document we use what I call *template analysis*. It essentially means discerning what template the author is using to assume a false identity—here "serial killer"—purposefully putting that to one side, and paying special attention to what remains. Doing so here gives us, for example, a reference to a computer as a *PC*. Especially in 2004, who would refer to a computer as a PC? Perhaps someone who worked with computers, rather than a member of the general public. And, indeed, Brian Hummert, the chief suspect, was a computer technician for the State Police.

Analysis showed that for all the surface differences, the letters bore a remarkable similarity. They were of course both in English, and they were in a quite similar dialect—there were no examples, for instance, of obvious nonstandard dialect, surprisingly for the apparently poorly composed serial killer letter, written in clumsy handwriting and dumbed down by leaving out some words. They both demonstrated the author's ability to structure a narrative with competence, as evidenced by a seamless execution of time shifts. If we follow the time sequences in the letters, we find they effortlessly flash back, and flash forward, and step out of the narrative time flow to add information—all in a way that reads naturally and is unobvious. Like the excellent punctuation of the devil strip letter, this belies the apparent unsophistication of the second letter.

Our analysis revealed information that the investigators already had suspected but did not have scientific evidence to support, namely that both letters had been written by the same person, in this case, Brian Hummert. And although the similarities just described may have narrowed down the suspect pool of likely letter writers, and did not reveal any meaningful inconsistencies between the letters, still this was not sufficient to obtain a search warrant. The police sought an examination of all the available known documents of the chief suspect, yet the similarities thus far described did not necessarily establish that a

single person probably authored both. But in the letters we also noticed an odd pattern of repetition: *found out* and *found out*; *break* and *broke*:

Stalker Letter

> Rumor had it that she occasionally took several guys at once and she sucked cock really well. I would have loved to have **found out**. A couple of days later she made sure my fiancée **found out**. She dumped me and then had an abortion.

Serial-Killer Letter

> I killed Charlene Hummert, not her husband. We had an affair for the past nine months. She wanted to **break** it off. So I **broke** her neck! I wrote letters to her husband and to Det. Loper.

This device consists of repeating the same verb in two consecutive sentences in a passage but changing the context of use in such a way as to express irony and cruel humor. In the first letter the writer repeats the verb but shifts the subject from *I* (the writer) to *she* (the victim), and in the second letter from *she* (the victim) to *I* (the writer). He shifts the complement of *find out* from "hypothetical sex acts" to "having had an affair," and he shifts the complement of *break/ broke* from *the affair* to *her neck.*

This is quite a precise rhetorical device, and in both letters is highly similar in structure and effect. There has existed a vast scholarship on rhetorical devices for many centuries. I researched whether this device, which we termed "ironic repetition," was a common device. It was of course possible that it was, and that just by chance two different writers might have chosen to use it. Unlikely, but I have seen writers accused of plagiarism when all they actually did was use the same device used by earlier works. A good example of this is the *chiasmus* device of President John F. Kennedy's famous "…ask not what your country can do for you; ask what you can do for your country." Its similarity to others' earlier speeches led to accusations of plagiarism leveled at JFK.

I discovered that the ironic repetition device was not common. I could not find a description

that matched it, and I eventually asked the curator of the massive online encyclopedia of rhetoric, the "Silva Rhetoricae" (rhetoric.byu.edu) hosted by Brigham Young University; he was unfamiliar with it, did not have a name for it, and said that it might best be classified within a general category of repetitive devices called *ploce*.

I explain the rhetorical device in detail here because it illustrates an important, if ultimately obvious, principle: the rarer the features, the more indicative they are of either a particular language variety (e.g., narrowing the suspect pool because someone uses a technical term of a group of specialists who are the only users of the term) or of a particular writer (the "ironic repetition" in this case). How likely was it that this uncommon rhetorical device just happened to be chosen by two different authors, both of whom supposedly had affairs with, and wrote letters about, Mrs. Hummert, and on either side of the time of her murder? Not very likely. It was certainly not unreasonable to think that a single person might have written both letters. Put another way, which was the superior hypothesis? Random chance, or single author?

A judge granted a search warrant, and we compared samples from Mr. Hummert's known writing to the writing of the questioned documents—the stalker and serial killer letters. The samples were workplace e-mails, handwritten writings from Hummert's workplace, and Hummert's legal notes and complaints, etc. while he was in custody in prison. Unknown to us, he had already been arrested and incarcerated. We didn't expect to find rhetorical devices like ironic repetition in such writings, and we did not. But we did find something else quite noteworthy.

All the documents shared not just a tendency, but a categorical skewing in the patterning of contracted verbs. Certain verbs can contract in various types of speech and writing, so that *I am, she is,* and *you did not* can contract to *I'm, she's,* and *you didn't*. But in the Brian Hummert documents, while negative verbs were sometimes contracted, positive verbs never were. That is, *did not* sometimes became *didn't* but *I am* never contracted to *I'm.* As the chart below shows, not one of the 74 positive verbs in the known writings

was contracted, and not one of the 23 positive verbs in the stalker and serial killer letters was contracted.

Known Hummert	Contracted	Non-contracted
Negative	15 (e.g., *didn't*)	25 (e.g., *did not*)
Positive	**0** (e.g., *I'm*)	74 (e.g., *I am*)

Questioned—Stalker and serial killer letters	Contracted	Non-contracted
Negative	6 (e.g., *didn't*)	2 (e.g., *did not*)
Positive	**0** (e.g., *I'm*)	23 (e.g., *I am*)

Such an extreme skewing suggested a personal idiosyncrasy. Although we had never seen this precise skewing before, to gauge this pattern's actual uniqueness we needed to compare it to a base rate from a reference data set. I studied the contraction patterns in letters to the editor to the local York, Pennsylvania, newspaper, on the assumption that they were the most local and least edited and publicly obtainable writings. Indeed, there was no indication that writers in York had a pronounced tendency to avoid positive contractions. Had that been the case—were it a regional tendency—we might unwittingly only be narrowing the suspect pool down to "writers in York, PA." We derived large reference databases through Google Web and Google Scholar and also found that this contraction/noncontraction skewed pattern shared by both the known and questioned writings was not one that matched up to any reference databases we could access or construct.

Below are examples from two transcribed e-mails written by Hummert, similar to slides I walked the jury through. Note that the one negative (boldface) is contracted—"do not" becomes "don't." Positives (italics), on the other hand, are never contracted—for example, "I am" never becomes "I'm."

> **From**: Hummert, Brian D
> **Sent**: Friday, September 24, 2004 2:54 PM
> **To**: [Redacted]
> **Subject**: Dental Appointment
> [Redacted],
> I have a Dental appointment on Monday morning. I **don't** know if *I will* come in before I go or not. *I will* fill out a leave slip when *I am* in after the appointment.
> Brian

From: Hummert, Brian D
Sent: Thursday, May 06, 2004 8:11 AM
To: [Redacted]
Subject: How goes.
[Redacted]
How goes it over there?
PFA second phase is in and working well after a few minor glitches. *I am* currently working on the Premium time reporting system. *It is* using SAP extract files to replace CMIC files and half the data is missing or wrong. What fun. I guess *I will* be over here until they clear me. This is an in house vacation. No on call, no CLEAN.
Brian

Notice the same pattern in the stalker letter, and the serial killer letter. Negatives (boldfaced) are sometimes contracted—e.g., in one case, "did not" becomes "didn't"; in another, "does not" is not contracted. Notably, positives (italics) are never contracted.

> *Here is* the proof that your *wife is* a slut. Do what you will with it. Sorry it took so long. I only come occasionally back to the area on business. Merry Xmas. *I will* send you several copies of this so you get the information in case the slut intercepts one. Before I tell you how I got it, I want to tell you a little about myself. I played in a band back in the late seventies/early eighties. I had a one niter with your wife. She was a fine piece of ass that I enjoyed several times that night. Rumor had it that she occasionally took several guys at once and she sucked cock really well. *I would* have loved to have found out. A couple of days later she made sure my fiancée found out. She dumped me and then had an abortion. *We have* since patched things up and gotten married, but she **can't** have any children. I blame your wife for that. The time is now right for payback. I hope to see your wife miserable the next time *I am* in the area.
> I ran into your wife back in September at Gabriel Brothers. I almost **didn't** recognize her with her dyed hair. *I have been* following her around hoping *she would* mess up. On October 6, I followed your wife over to Capitol City Mall. She was dressed up more the usual for a Saturday of shopping. She went into the Picture People. This was around 10 AM. A couple of weeks later I went in and got copies of the pictures enclosed. On the negative holder *she had* written that the photo was a gift. There was no indication of which one she had printed up.
> I ask you who was it for? Also she **does not** have her wedding ring on. Why not? A red rose is a symbol of love. For who? I **don't** think you know about these. Do you? Also *she has* purchased a lot of sexy bras and panties. Have you seen them or the red nightie? Were they brought for your enjoyment?

You may also want to ask her about her Spencer Gift purchases. Do you love lubes with her? So you see once a slut always a slut.
I killed Charlene Hummert, not her husband. We had an affair for the past nine months. She wanted to break it off. So I broke her neck! I wrote letters to her husband and to Det. Loper.
I used a white nylon rope to kill her they **won't** find me *I am* leaving. *I am* writing because of Easter. *I am* sorry I killed her.
They **won't** find the cell phone she used to call me, *it is* in the river and not under my name.
I carried her into the kitchen and then dragged her outside to her car. This is the fifth woman I killed. *I am* getting good at it.
Cops have no idea how easy it is to pin husband when they only look there.
She knew about pictures on PC. She told story to set up husband for the Divorce. Ha Ha
ByeBye for now
John

Experts from several other forensic disciplines also testified. Hummert was ultimately found guilty of first-degree murder, and was sentenced to life in prison without parole.

Forensic Linguistics' Foundation Is Linguistic Theory Derived from Language Use in the Real World

It may appear at first that forensic linguists could do what they need to do without ever leaving the office, and on one level that is true. But the training and experience of linguists like Dr. Shuy, Dr. Wald, or myself, is that of sociolinguistic field researchers, collecting and analyzing the actual language that speakers and writers use, and analyzing how, out in the real world, language systematically varies in different interactional contexts. Many forensic linguists were trained in linguistic variation and analysis by initiating and recording interviews on street corners, in living rooms, in the East African savannah, on Swahili sailing *dhows*, and in bars, in language communities all over the world—from Harlem and Detroit and London to Bangkok and Mombasa—and who then used this firsthand data to build data banks and construct theories of how language works in the real world. (This is in contrast to the far more numerous

purely theoretical grammarian linguists, for whom the intuitions of native speakers about sentence acceptability serve as the primary data.)

From experience with the kinds of language that real speakers actually use to communicate, we promulgate theoretical constructs that can explain linguistic variation; the demographic profiling and authorship cases revolve around understanding such linguistic variation. In such cases we search for constellations of features. Just as a dialect can be described by a collocation of concurrent linguistic features, a linguist can conduct detailed, multi-level linguistic analyses on the language in written documents or speech samples, and weigh whether or not the evidence supports the hypothesis that the linguistic patterns in that document can best be explained as instances of the linguistic patterns found in the various subjects' known samples.

The growth of the field of *corpus linguistics* enhances our ability to utilize large, more targeted databases for reference and analysis. Current corpora include the 450-million-word Corpus of Contemporary American English (COCA), the 100-million-word British National Corpus (BNC), and the 1.9-billion-word Corpus of Global Web-Based English (GloWbE). For example, in a recent case, we explored the rarity of certain words followed by a comma, used to begin a sentence. For example, "Secondarily, if payment is not made …". Using *secondarily* was a feature found in the questioned document, and in the known writings of just one subject. The other subjects used *secondly* and *second.* The corpora show, clearly, that this *secondarily* usage is quite rare, compared to its alternatives; it was an important feature in the analysis.

		ALL GloWbE	US GloWbE (per million)	GB GloWbE (per million)	COCA (per million)	BNC (per million)
1	SECOND	13117	11.07	5.48	15.30	13.15
2	SECONDLY	11998	4.16	6.67	3.17	16.05
3	**SECONDARILY**	**43**	0.05	0.02	0.05	0

In the investigation of the 2009 Coleman homicides, the CTAD, a corpus originated by SSA Jim Fitzgerald of the FBI's Behavioral Analysis Unit, proved invaluable. As I wrote:

> Death threat letters and emails sent prior to the murder, and spray-painted words on the walls and a victim at the murder scene all began…with the same obscenity. I had been hired by the FBI Behavioral Analysis Unit (BAU) some years before to analyze and advise on their Communicated Threat Assessment Database (CTAD), a "computerized database/software program designed to be the primary repository for all communicated threats and other criminally oriented communications" within the FBI (Fitzgerald, 2007, p. 6). I thought a CTAD search of that usage would be useful, and asked that one be done. Corpus linguistic analysis showed the use of these obscene words to begin a threat to be extremely rare in CTAD's database, and thus they were a noteworthy pattern linking the utterances together. I also found other language features that linked together the threats and spray-painted words. (Leonard, 2017)

Conclusion

Forensic linguistics may be seen as an intelligence gathering methodology. It is applicable to a wide range of cases and situations. As we have discussed in the kidnapping and murder cases above, forensic linguists help investigators and triers of fact to extract maximum intelligence from language evidence such as letters, e-mails, notes, texts, wills, confessions, and recorded speech. This scientific language analysis can assist in criminal investigations, threat assessment, counterterrorism, fraud detection, company-internal sabotage, and many other areas of forensic interest that involve the use of language.

This introduction has discussed some methodologies that have been accepted in courts in the USA, the UK, and many other countries.

Acceptance in the courts is growing (in the past several years, I, for instance, have been qualified to testify in courts in 12 US states and five federal district courts. I have also testified as a linguistic expert before World Bank ICSID Tribunals in Washington, DC, and Paris). The field is growing, with undergraduate and graduate training becoming increasingly more available and with more scientific journals dedicated to the field.

Linguistics may stand alone in the forensic sciences in that after forensic linguists present their analysis, non-linguists often indicate that the analysis is obvious and self-evidently true—even if, before the analysis was presented, they could not predict what it would show. Lay users certainly know the structure of their language, but it is largely an unconscious knowledge; scientific research in linguistics seeks to make those structures explicit, and training in linguistics teaches linguists what to expect when they analyze language evidence. To illustrate this, a useful analogy might be to medical experts who read X-ray films. Although we untrained viewers can certainly see the X-ray films, we can't tell what their significance is; the trained medical experts can. Similarly, linguistic experts describe and define the underlying structure of written and spoken language. Both sets of experts can do their jobs because they are trained and skilled in what to look for as they assess the meanings and implications discoverable in their observations.

References

Coulthard, M. (2004). Author identification, idiolect, and linguistic uniqueness. *Applied Linguistics*, *25*, 431–447.

Coulthard, M., & Johnson, A. (2010). *The Routledge handbook of forensic linguistics*. Routledge.

Drell, Dan. (n.d.). https://web.archive.org/web/20130701183932/www.ornl.gov/sci/techresources/Human_Genome/elsi/forensics.shtml.

Fitzgerald, J. (2007, February). The FBI's communicated threat assessment database; history, design, and implementation. *FBI Law Enforcement Bulletin, 76(2)*, 6–9.

Grant, T. (2013). TXT 4N6: method, consistency, and distinctiveness in the analysis of sms text messages. *Journal of Law and Policy*, *21*, 467–494.

Leonard, R. (2006). Forensic linguistics: applying the scientific principles of language analysis to issues of the law. *International Journal of the Humanities*, *3*, 65–70.

Leonard, R. (2012). Retslingvistik. Sprogforskeren som efterforsker og ekspertvidne' ['Forensic linguistics: the linguist as investigator and expert witness'] (translated into Danish by Nicoline Munck Vinther). *Mål og Mæle*, *35*(3), 11–17.

Leonard, R., Ford, J., & Christensen, T. (2017). Forensic Linguistics: Applying the science of linguistics to issues of the law. *Hofstra Law Review, 45*, 881.

Leonard, R. (2017). In press. *"Communicating linguistic theory and analyses to judge and jury in the highly adversarial US justice system: theatrical cross examinations vs. the facts" in Legal Discourse and Communication, Centre for Research in Language and Law*. Naples, Italy: Cambridge Scholars Publishing.

McMenamin, G. (2002). *Forensic linguistics: Advances in forensic stylistics*. Boca Raton, FL: CRC Press.

McMenamin, G. (2004). Disputed authorship in US law. *International Journal of Speech Language and the Law*, *11*, 73–82.

Shuy, R. (1993). *Language crimes: The use and abuse of language evidence in the courtroom*. Cambridge, MA: Blackwell.

Shuy, R. (1998). *The language of confession, interrogation and deception*. London: Sage.

Shuy, R. (2001). Discourse analysis in the legal context. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The handbook of discourse analysis*. Cambridge, MA: Blackwell Publishing.

Shuy, R. (2006). *Linguistics in the courtroom: A practical guide*. Oxford University Press.

Shuy, R. (2014). *The language of murder cases*. Oxford University Press.

Solan, L., & Tiersma, P. (2004). Author identification in American courts. *Applied Linguistics*, *25*, 448–465.

Tiersma, P., & Solan, L. (2002). The linguist on the witness stand: Forensic linguistics in American courts. *Language*, *78*, 221–239.