

How Language Nerds Solve Crimes | Otherwords
<https://www.youtube.com/watch?v=-EVXdhstw14>
Transcript: <https://dontveter.com/ec/crimes.pdf>

Have you seen this man? Probably.

This is one of the most famous police sketches ever made.

And it's supposed to be this guy, Ted Kaczynski, a.k.a. the Unabomber, an American terrorist who mailed homemade explosives across the country for almost 20 years.

Obviously, the sketch artist didn't have much to work with.

And the Unabomber left no fingerprints or DNA.

So how was he finally caught? Linguistics.

I'm Erica Brozovsky, Ph.D., And this is Otherwords.

When the Unabomber sent a 35,000 word manifesto to the newspapers, the FBI had forensic linguists James Fitzgerald and Roger Shuy examine it for language clues.

They found some outdated terminologies like “broad” and “chick” that weren't likely to come from a woman or a young person.

Arcane words like anemic and chimerical suggested a high level of education, and they knew that odd spellings, like “wilfully” and “clew” were briefly popular in the Chicago area in the forties and fifties.

chimera

noun

chi·me·ra [kī-'mir-ə](#)

kə-

[Synonyms of chimera](#)

1

a

capitalized : a fire-breathing she-monster in Greek mythology having a lion's head, a goat's body, and a serpent's tail

b

: an imaginary monster compounded of incongruous parts

2

: an illusion or fabrication of the mind

especially : an unrealizable dream

a fancy, a chimera in my brain, troubles me in my prayer —

John Donne

His utopia was a chimera.

3

: an individual, organ, or part consisting of tissues of diverse genetic constitution

A hybrid created through fusion of a sperm and an egg from different species is a chimera.

In 1995, Linda Patrick was reading the manifesto in the Washington Post and found it disturbingly familiar.

She asked her husband David to take a look and he instantly recognized some of the phrases from letters his brother had written to newspapers.

At the time, the FBI was receiving thousands of tips.

But when they heard that David's brother was a highly educated, middle aged man from the Chicago area, it piqued their interest.

Fitzgerald and Shuy compared the letters David supplied them with the manifesto and concluded that unless it was an elaborate hoax, this had to be their man.

Shortly thereafter, the FBI raided an isolated cabin in the Montana wilderness where they found Ted Kaczynski, along with bomb-making materials and an original copy of the manifesto.

In a way, we are all forensic linguists.

Have you ever reread a friend's text trying to deduce why they're really skipping your party?

Have you ever pored over a warranty before making a big purchase?

Or stared at graffiti on a bathroom wall trying to guess what Jesse could have done to get called out like that?

In all of these cases, you're analyzing linguistic clues like word choice, grammar and punctuation to dig for hidden meaning.

But the term forensic linguistics didn't exist until the 1960s, when Swedish linguist Jan Svartvik reexamined a controversial murder case through a linguistic lens.

In 1949, a Welshman named Timothy Evans confessed to the strangling of his wife and baby daughter in a series of strangely contradictory statements.

Three years after his conviction and hanging it was discovered that Evans' downstairs neighbor, a man named John Christie, was actually a serial killer and the real murderer of Evans' wife and daughter.

Why would Evans confess to something he didn't do?

Since he was functionally illiterate, all his statements had been dictated to police investigators, totaling almost 5000 words.

Svartvik analyzed the text using several metrics, like the frequency of different types of grammatical clauses.

What he found was that whenever the statements veered into fantasy, a different voice seemed to take over.

Maybe Evans changed his speaking style when he lied.

Or maybe the dictating officer took some creative liberties.

This kind of analysis, known as stylometry, assumes that each of us has our own idiolect, a unique way of speaking that can be used like a fingerprint to identify us.

Your idiolect includes obvious things like vocabulary and pronunciation, but also subtle syntactic habits.

Do you prefer active voice or passive voice?

Do you tend to put adverbial clauses before or after the main clause?

You may not notice these habits, but a forensic linguist might.

One of the key methods of stylometry is categorizing words and counting them.

People who use more nouns, articles and prepositions tend to be more rational and analytic, while those who favor verbs, adverbs and pronouns are more dynamic and emotional.

You can also put words into broad thematic categories like family, religion, sensuality and death.

In 2015, social psychologist Ryan L Boyd and linguist James W Pennebaker used this word counting method to analyze Shakespeare's plays and build a psychological profile, much like an FBI agent would for a criminal.

They compared it to the text of a play with disputed authorship called *Double Falsehood*, and found that Shakespeare was the most likely author of the first three acts, with his frequent collaborator John Fletcher writing the last two.

Stylometry has been used to identify modern authors as well.

It's how we know that the Zapatista military leader, Subcomandante Marcos, is actually the respected writer Rafael Sebastián Guillén Vicente.

It's how journalist Joe Klein was outed as the anonymous author of the political novel *Primary Colors* and how whodunit scribe Robert Galbraith was revealed to be a pseudonym for Hogwarts billionaire J.K. Rowling.

But authorship attribution is a lot easier when you have a big body of text to analyze, known as a corpus.

In criminal investigations, it's not every day that the perpetrator sends you a 35,000 word manifesto.

Smooth move, Ted.

Most of the time you're working with much smaller samples.

Take the 1996 murder of six year old pageant queen JonBenet Ramsey.

Even before her body was discovered in the basement of the family home, investigators already thought that there was something hinky about the ransom note left behind.

hinky /hɪŋg'kē/ adjective So peculiar as to arouse suspicion; curiously odd.

Acting [suspiciously](#); [strange](#), [unusual](#); acting in a manner as if having something to hide, or seemingly [crooked](#).

Suspicious (of something).

For one thing, it was about three times longer than an average ransom note, with phrases lifted right out of popular action movies.

“Do not attempt to grow a brain.”

The letter also contained several instances of what linguists call “adversarial stylometry,” when someone deliberately tries to obscure their writing style, like misspelling simple words or claiming to be foreign, which is not an adjective that people often use to describe themselves.

Linguists are also called in to analyze letters sent to the police by supposed serial killers like the Zodiac Killer and the Beltway Sniper.

In 2018, a forensic linguist from the University of Manchester used modern techniques to analyze the letters sent to London police in 1888 by someone claiming to be Jack the Ripper.

Not only did he confirm that the first two letters were written by the same person, but also found a potential connection with a third letter that was long thought to be a hoax written by someone from the Central News Agency to boost newspaper sales.

This might mean that all the letters were hoaxes, that the real killer never wrote to the police at all and never named himself Jack the Ripper.

In such cases, the corpus is so small that they can only make guesses about the type of person who wrote them rather than match it to a specific author.

But that may be changing thanks to technology.

In the era of social media, each one of us has an extensive corpus of posts, tweets and comments that can be mined to build an idiolectic profile.

And artificial intelligence can analyze that corpus with stunning accuracy.

After all, that's exactly what these algorithms were designed to do: find patterns in language.

But the complexity of AI can also limit its usefulness in forensic linguistics.

The most powerful systems are typically unsupervised, meaning their own programmers don't know how they reach their conclusions.

They find patterns in language so obscure that a professional linguist would have trouble understanding it, let alone a jury of laypeople.

And that does not fly in a court of law.

You can't have a computer on the witness stand just say "Trust me, he's your guy."

Perhaps the day will come when A.I. evidence is as accepted and normalized as DNA evidence.

But until then, juries can't convict someone based on the word of an algorithm.

Unless we outsource jury duty to A.I. as well, and I'm terrified to think how many people would be okay with that.

For now, human linguists still have a place in courtrooms and not just to introduce evidence.

They help lawyers fine tune their arguments, parse complicated statements with disputed meanings, monitor police interrogation tactics, and even ensure that non-English speaking defendants are interpreted fairly and accurately.

The fact that we all have our own unique way of speaking is a beautiful thing.

It's at the core of personal expression and contributes a wonderful tapestry of language.

But it also means that no form of expression can be 100% anonymous.

As the famous Miranda warning goes, anything you say can and will be used against you in a court of law.