

## **Linguistic Profiling**

**from Roger W. Shuy, *The Language of Murder Cases* Oxford U Press  
(in press)**

We don't know when linguistic profiling was first practiced, but its history goes back at least as far as the Old Testament of the Bible, where Judges 12 describes the aftermath of a battle in which the Gileadites had soundly defeated and killed 42,000 Ephramites, after which they occupied and controlled the shore of the Jordan River. After the battle ended, some of the surviving Ephramite soldiers pretended to be civilians, and requested permission from the victorious Gileadites to let him cross the river. Suspecting that they were surviving Ephramite soldiers, the Gileadites devised a linguistic profile test, asking them to pronounce the word, *shibboleth*. The /sh/ consonant cluster was not in the Ephiamite's language, and therefore they pronounced the word with an /s/ sound instead. The linguistic test worked, and the Gileadite soldiers slew them on the spot.

Another historically recorded case of linguistic profiling is reported to have taken place during the Revolt of the Sicilian Vespers that broke out in Palermo on Easter of 1282. The records are not totally clear and they may be less than totally factual, but they report that the Sicilians rebelled against the army of the French king, Charles of Anjou, who had been ruling the Kingdom of Sicily for 16 years. While Sicilians were celebrating on the night of the Vespers, some French officials joined them. A few of the Frenchmen then made sexual advances to local Sicilian women, after which the Sicilians grew angry and began to riot. Thousands of French inhabitants were slaughtered as the local Sicilians broke into French homes, businesses, and even into the Franciscan and Dominican convents. In order to distinguish the local Sicilians from the unwelcome French invaders, it is said

that the rebels required them as to pronounce the word *ciciri* (garbanzo beans). Since no Frenchman could pronounce the “c” with the proper Sicilian /ch/ sound, the Sicilians killed some three thousand French men and women, and after a few weeks King Charles lost control of the country (Runciman 1958, 212).

Today, linguistic profiling is based entirely on more tools of linguistics than the consonant sounds used by the Gileadites with the word, *shibboleth*, and by the Sicilians with the word, *ciciri*. Linguistic profilers now rely on research from dialect geography, lexicography, sociolinguistics, historical linguistics, and psycholinguistics, all of which rely on the linguistic tools of phonology, grammar, semantics, and pragmatics. Although few linguists specialize in all of these areas, a good linguist will know a lot about some of them and is trained well enough in the others to be able to provide helpful information to law enforcement agencies. For the past few decades, linguistic profiling has become a useful way to help law enforcement agencies narrow down their lists of suspects and, in some cases, even use it to confront suspects with these linguistic profiles as a way of convincing them to admit their crimes.

Linguistic profiling is not the same thing as behavioral profiling. It is believed that the idea of behavioral profiling originated in the Behavioral Science Laboratory of the Federal Bureau of Investigation, where specialists in psychology and criminology worked together to assess the behavioral characteristics of as yet unidentified criminals. Necessarily, the profilers base their conclusions on information and theory on research in cognitive psychology, on their personal experience with previous crimes, and on their subjective interpretations based on whatever facts about a crime are available, including the type of crime and whatever physical evidence exists.

Turvey (1999) calls it an artful process that can produce confirmation bias, because it begins by searching for proof to confirm a theory, rather than beginning with evidence data. Even though behavioral profilers have little or no evidence data to begin with, their technique is believed to be valuable in some investigations and it has achieved considerable status in the area of criminal investigations.

When people or businesses receive hate mail or threat messages, the FBI and non-government forensic psychologists provide their behavioral profiling services. Until recent years, however, behavioral profiling had not taken full advantage of the analysis and assistance that linguistic profiling can provide.

To do linguistic profiling effectively, it helps greatly if the linguist's training and competence includes a strong component of language variation and change. In fact, the linguistic profiler would do well to have specialized in sociolinguistics, because it is the variability of language that offers many clues to the possible identity of an anonymous writer or speaker. Other language clues also may be evident, relating to such things as the person's political beliefs, social standing, economic status, ethnicity, and attitudes. The sole focus of linguistic profiling is to point out sociolinguistic features about suspects that their own language suggests.

The important legal concepts of intentionality are premeditation are clearly apparent in such threat messages, for otherwise the writers wouldn't have sent them in the first place. This transparency makes it possible for linguistic profiling analysis to serve as a touchstone upon which the issues of intentionality and predisposition are less than clear in the 13 murder cases described in chapters 5 through 10.

First I describe the role of linguistic profiling in one of the most famous murder events of the 1990s, the case of the Unabomber, ultimately identified by the FBI as Ted Kaczynski.

## **The Unabomber Case**

Between 1978 and 1995, sixteen mailed packages containing home-made bombs killed three people and seriously injured 23 more. Up to 1986 the bombs sent in intricately whittled wooden boxes had been mailed or planted in various locations in the country. Two were sent to professors in Illinois, two to professors in California, one to Yale, one to a professor in Tennessee, one to a professor at MIT, two to airline executives, and one to a California computer store. During the following seven years, no more such bombs were sent. But the bombings resumed between 1993 and 1995 with bombs mailed to a California geneticist, a New Jersey advertising executive, and a California lumber lobbyist. Most of the bomb packages contained short notes or letters, parts or all of which survived the explosion in one shape or another.

In the fall of 1994, FBI Special Agent Sharon Smith took my graduate seminar in Linguistics and Law while she was working on her PhD in psychology at Georgetown University. At that point I had been consulting with security officers of various corporations and occasionally with the FBI about various threat messages, and it occurred to Agent Smith that her FBI colleagues on the Unabomber Task Force might benefit from what I had called “linguistic profiling.” At that time, the FBI already had an ongoing program using psychologically based criminal profiling, comparing the behavioral characteristics of current crime settings and types with past

crimes of a similar nature, trying to determine behavioral characteristics that might help them discover or narrow down lists of suspects. As mentioned above, the major difference between a psychological criminal profile and a linguistic profile is that the former makes predictions based on comparisons of the apparent behavioral characteristics of the unknown perpetrator of the current crime with the known behavioral characteristics of past similar crimes. In contrast, the findings of linguistic profiling are based entirely on currently existing and known spoken or written language evidence. Rather than relying on inferences about the suspect by comparing the bombings with past similar crimes, the Unabomber case provided some written texts that provided considerable physical language evidence-- the messages he sent with the bombs, the letters he wrote to magazines and newspapers, and ultimately his long manifesto.

A linguistic profile does not claim to identify specific authors. Instead, it describes how the suspect's language matches social, economic, education level, and other information that has been identified in previous sociolinguistic research. Its purpose is to help law enforcement narrow down its list of existing suspects or to suggest directions for locating new ones (Campbell and DeNevi 2004).

In May 1995, before the Unabomber had sent his manifesto to the newspapers, Special Agent Smith approached me about the possibility of providing the FBI with a linguistic profile of the bomb messages and letters written by the then unknown Unabomber. After I agreed to do this, a group of agents met with me at my home in Washington DC to explain the case and leave with me the brief notes the Unabomber had sent with his bombs along with some letters the Unabomber had written to several magazines and

newspapers as he tried to convince them to publish his as yet unfinished manifesto.

On June 28, 1995, about six weeks after my first meeting with the FBI agents, both the Washington Post and the New York Times received a 56 page, typewritten manuscript in the mail, demanding that unless they published it, the author would “start building our next bomb,” which was very clear evidence of intentionality and predisposition. The writer called for a worldwide revolution against the effects of modern society’s “industrial-technological system.” Immediately, the newspapers turned it over to the FBI. Earlier, this anonymous author had also written similar letters to Penthouse magazine and Scientific American, saying that he intended to plant only one more bomb if they would publish his paper in serial form or as a book. At around the same time, the San Francisco Chronicle received an anonymous letter in which the author threatened to blow up an airliner at the Los Angeles airport. These notes, letters and manifesto constituted the language evidence available for producing linguistic profile of the suspect.

Shortly after the FBI Task Force received the manifesto, they also sent it to a large number of college professors who specialized in the history of science, hoping that the language used in the Manifesto would jog their minds to recall a particular student who might have written it. Because searching for a former student who may have written term papers about these ideas was probably not the most efficient way to identify the criminal, the professors’ comments were not considered very helpful. On the other hand the lengthier language evidence that the FBI now possessed made it possible to examine the Manifesto for possible language clues to the writer-- a linguistic profile.

Many sociolinguistic clues for identifying characteristics of the writer were evident in these writings. For example, the author wrote about going out in the "sierra" (not used as a proper noun) after a hard day's work, little realizing that such an innocuous expression could help the FBI identify northern California as one location in his past life. On the other hand, the fact that his writings did not include other common western terms, such as "ranch," "fork" (for a branch of a stream), "range," or "mesa," led to the suspicion that his western use of "sierra" did not mark him as a native Westerner. His use of the Northern dialect variant, "rearing children," as opposed to the Midland dialect's "raising children," offered one of the clues that he had possibly grown up in the Northern dialect area ranging from New England to the Upper Midwest.

Another clue to the Unabomber's background location was found in some of his spellings of common words that matched those being used by the Chicago Tribune during the 1940s and 1950s. During this period the Tribune's editor was spearheading a movement toward what he considered more standardized spellings of English, such as changes in the spelling of words with doubled consonants, such as "willfully," to a single consonant, "wilfully," and the change of "clue" to "clew." I was aware of this because I lived in the Chicago area in the 1950s. It is possible, if not likely, that a literate Chicago area schoolboy might well adopt some of the Tribune's spellings as his own, which is one of the reasons why my linguistic profile hypothesized that the Unabomber grew up in the Chicago area, which was eventually shown to be accurate.

These unconventional spellings also suggested a clue to the writer's age. If the Unabomber had grown up during the period of the Tribune's spelling reforms, he would have to have been around fifty years old at the

time his mail bombs were sent, which was verified after Ted Kaczynski was captured. Other clues about his age were that Kaczynski used terms that date a writer as having grown up in the 1960s, such as his use of the expression, "Holy Robots," borrowed from Batman's assistant, Robin, in the popular 1960s Batman television program. Other 1960s fad terms found in the Unabomber's writings included "playing footsy," and "working stiff."

The Unabomber's gender references indicated that he was either ignorant of or isolated from the inclusive gender references that were beginning to be expected during the time of his writing. He used "he" and "him" where other writers were avoiding the masculine pronoun in references to both sexes. His sociological terms also gave clues to his age, such as his frequent use of "other directed" and his many references to individual "drives," which suggested his acquaintance with the sociology of the sixties, particularly that of David Riesman.

The Unabomber's lexicon also gave clues to his religious background, as he commonly used expressions such as, "unclean thoughts," "time of troubles," "the human soul," "cradle to grave," "impersonal demon," "sin" (many times) and "God's will." He also told a parable of a weak neighbor and a strong neighbor, using language that is very similar to Biblical style: "If he lets the strong man survive and only forces him to give the land back, he is a fool, because when the strong man gets it he will take again all the land for himself." His possibly Catholic background came through in his lament that modern society postpones having children, consistent with that church's ban on birth control. He argues in favor of spanking children and, in references to the sex drive, frequently mentions "sublimation" as the answer--more language suggestions that he may have grown up in a Catholic context.



These language clues provided a stark contrast with the FBI's behavioral profiles. In 1980 after the first bombs were placed and exploded, the FBI's profile opined that the murderer was 18 to 22 years old, with some undergraduate education, possibly a frustrated, withdrawn college student with low self-worth, who may like to torture animals and set fires. In 1985, the FBI profile added ten or so years to the bomber's age and opined that he may be an unmarried loner working as a technician and living near the crime scenes. In 1986 a revised profile said he was probably a male in his late 30s or early 40s, with 2 to 4 years of college, who changed technical jobs frequently, had a rich fantasy life, and was excessively neat and clean.

By 1993, the Unabomber had begun to write letters to magazines and newspapers under the pseudonym of Freedom Club (or FC), offering them first rights to publish his still uncompleted manifesto (which he referred to as a book or a serialized essay). These letters contained language that pointed to a much more educated bomber than the FBI originally thought. Now the revised FBI profile said he was a loner who was highly intelligent, technically competent, and a voracious reader in the areas of science, history, psychology, and the social sciences. It is difficult to know what this profile meant when it added that he had high self-esteem but low self-concept.

In 1995, at about the time the FBI had received my linguistic profile, the older versions of the FBI behavioral profile had changed a bit. More in keeping with my linguistic profile, the FBI now suggested his age as between 40 and 50, but it held fast to his education being that of a high school graduate with some college and trade school training. My linguistic profile apparently had convinced them about the bomber's place of origin, however, for that version of the FBI's profile said the Unabomber lived,

worked, attended school, or made extensive visits to the Chicago area. The profile continued, however that he was a loner with low self-esteem but at the same time had a strong sense of superiority and had spent some time on a college campus.

The letters the Unabomber sent to magazines suggested that he was a much more highly educated person (a fact which became more apparent after The New York Times and Washington Post later published his Manifesto). In those letters he used a somewhat learned vocabulary, including words such as "surrogate," "oversocialization," and "tautology", along with complex grammar (including the subjunctive), and wrote a rather lucid style most of the time. Whatever one might think of his radical ideas, his sequential organization of these ideas was usually logical, and he gave evidence of having read enough about such fields as history, archaeology and sociology to feel that he had the right to discount the contributions that scholars in these fields had made to the "well-being of the human race." On the other hand, the references he cited usually were not up-to-date, his uses of punctuation and spelling would not be acceptable in the humanities or social sciences, and he shifted back and forth from the scholarly to the casual styles in less than the expected scholarly way. He was clearly a well-educated man whose writing would need a considerable amount of editorial help if it would ever become publishable. His writing style would not pass muster in the humanities or social sciences, but might, with help, get by in a hard science. This suggested that his academic background was likely in the hard sciences (it turned out to be mathematics).

He took a dim view of college professors, whom he called "university intellectuals," noting in one letter, "people with advanced degrees aren't as smart as they think they are." It was apparent that if Kaczynski was himself

a college professor, he certainly did not think much of his peers and he had no empathy for the profession. The fact that his Manifesto used references to books and articles without indicating dates or publishers suggested that he was no longer associated with the university life and that he had little access to academic libraries. In fact, there were no such libraries in or near where he then lived in the very rural area of Lincoln, Montana.

The Unabomber's writing, therefore, provided some reasonably useful clues about his education, age, occupation, religion, and geographical background, along with a broad sense of his personal life. As it turns out, he had indeed been a college professor at one time, was Catholic, grew up in Chicago, was about fifty years old, once lived in Northern California, and since he later chose to live in rural Montana, he no longer had access to his much needed, up-to-date academic library resources.

Linguistic profiles make no claim to exact identification of suspects. They simply can't do this, and my linguistic profile of the Unabomber was no exception. It had absolutely nothing to do with Ted Kaczynski's ultimate capture. The credit for this goes to his brother David and David's wife, who noticed that the topics and ideas in the manifesto were similar to those in the letters that Ted had written to them, after which David heroically identified his brother and pointed out exactly where he was currently living. Not many murder cases are advantaged by having a close relative available to compare the ideas of two sets of written texts in this way. Technically perhaps, the information produced by David and his wife was not a linguistic profile, but it certainly did the much-needed job.

Even though linguistic profiling turned out to have nothing much to do with Ted Kaczynski's capture, this case still illustrates one way that it's possible for linguistic profiling to assist law enforcement. Here, the task was

to compare the note fragments, letters to the media, and manifesto in order to try to discover clues that might help the FBI locate an identifiably lone anonymous writer among 300,000 million American potential suspects.

The more common use of linguistic profiling, however, is when several suspected authors of messages are already on law enforcement's radar screen. Comparison of evidence documents can be a useful way to determine which of the suspects is the more feasible one to pursue. In this sense, linguistic profiling is not the same as conventional authorship analysis, which attempts to identify authors of anonymous documents. The current conventional approaches to authorship identification are (1) stylistic comparisons of known and unknown writers that focuses on language features such as lexicon, morphology, usage, and punctuation (McMenamin 1993; 2002), and (2) statistically oriented approaches that focus primarily on syntax (Chaski, 2001). Since the Unabomber case had evidence from only one unknown writer, conventional authorship analysis comparisons were neither relevant nor useful. The product was simply a linguistic profile, with no effort to identify a single, specific murderer, but rather to discover clues about the general characteristics of such a writer so that law enforcement might be better informed about what type of person they were looking for.

### **Aftermath**

Thanks to his brother, the FBI located Ted Kaczynski at his Lincoln Montana cabin and arrested him. He was later tried for murder, found guilty, and sentenced to life in prison. I include this case here because his writings clearly indicated his intentionality and predisposition to kill people, and there was no question that he wrote these things voluntarily.

As an ironic side note, a month or so after I submitted my linguistic profile to the FBI, I retired from teaching at Georgetown University and moved to Missoula, Montana, only about sixty miles from where Ted Kaczynski had been living in a simple cabin in the woods. It may have looked like I really knew where he was. I didn't.

### **Gary Indiana Women's Medical Clinic**

In some cases in which there is evidence to compare threat messages with the writings of known or possible suspects, linguistic profiles can also offer a range of possible clues, including the writer's ethnicity. A murder threat that took place in Gary Indiana provides an example of how clues to ethnicity played a significant role in helping to resolve the case. Over a period of three weeks, an anonymous writer had sent three carefully typewritten bomb threat messages to the director of a women's medical health clinic.

Women's medical clinics that perform abortions sometimes become targets of fierce opposition, but this clinic did not do abortions, removing this as the threat writer's possible motive. Baffled by the case, the Gary police called in the FBI to help them, after which the special agent in charge asked me for whatever assistance I could provide. Since there were no comparison documents to analyze, the conventional authorship analysis approach was not relevant. Instead, this case called for a linguistic profile of the apparently single author.

As I examined the three long and rambling threat messages, the language used in them made it clear that the writer provided an undeniable intention and predisposition to kill everyone in the building when it was to be bombed. Further examination made it evident that the writer's language

competence was consistent with that of a native speaker of American English in most ways, but it was also oddly inconsistent in others. For one thing, I noticed that the writer used the adjective "proper" in expressions such as "she did not give a *proper* exam," and "she was not given *proper* care." Since this adjective is used far more by British speakers than by Americans, its presence here suggested that the author had a British English background, as opposed to most American English, where "proper" is less likely to be used in this context. That the writer had a British background also was consistent with the spelling of the noun "device" as "devise."

In addition, the bomb threat letters contained syntax which sometimes used subject-object-verb sequences, as in "she will finally the seriousness of the problem recognize," where the expected English syntactic pattern of subject-verb-object would have been, "she will finally recognize the seriousness of the problem." The bomb threat messages also used more repetition for emphasis than would most American writers, and they often deleted articles that American writers would use habitually, as in the writer's missing "the" in "you know time is soon," and the missing article "a" in "you can be transferred to better position" and "I will not give warning." In addition, the writer sometimes omitted prepositions and pronouns where English syntax expects them, as in "if I address it [to] her," and "you should change [your] place of employment." These linguistic features are all consistent with a writer who could have been educated in a former British colony, such as Pakistan, where the remnants of Hindi-Urdu subject-object-verb syntax, article deletion, and pronoun deletion might remain along with tell tale elements of British English.

Although research on gender differences is fairly recent and is by no means conclusive, there are some findings that are believed to characterize

female language, such as frequent tentativeness and hedging, indirectness, the use of politeness markers and formal reference terms such as "Ms." and "Mr.", frequent use of "so" intensifiers with emotional import such as "I'm so happy," certain adverbs such as "surely," certain characteristically female adjectives such as "lovely," "charming," and "delightful," and a tendency to miniaturize, as in "our nice little home," "the little black dress," or "the child's tiny little hands." In addition, female writers are said to focus more on the feelings of the people being mentioned, while males are said to be less concerned about recipients' feelings.

This bomb threat writer also used many tentative and hedged expressions such as "it seems like," "she seemed confused," and "I suppose I should have." The writer even justified writing the letters by explaining that this was done "on the advice" of others, a justification that male threat givers would not be as likely to feel needed. The writer also relied heavily on expressions of feeling, such as "I felt completed," "I deeply regret," "my hopes were crushed," and "I was so upset." The writer politely thanked the reader for taking the time to read the letter and apologized for being a bother, while also saying "please" many times, and helpfully but oddly explained that he or she no desire to hurt anyone. The writer also used the "so" emotional intensifier in expressions such as, "I was so upset," and "there is so much danger in your clinic" and also used empathetic expressions such as, "you are surely in harms way," and "how sad for you." Since it is difficult to imagine that most male threat writers would use such language, I hypothesized that the author appeared to be female.

The FBI had already interviewed all ten of the female employees at that clinic, thinking perhaps that the threats may have been the result of some internal conflict, but their interviews yielded nothing suspicious.

Because at this point the FBI had no identifiable suspects in the case and they had no comparison documents, I suggested that they start by going back to the ten employees and asking them and the director to write narratives telling everything they did, saw, felt, and thought from the time they got up in the morning until they went to bed on the day that the most recent bomb threat appeared. In contrast with the FBI's previous interviews, I was more interested in how the employees said things than in the content of what they might say. I wanted to look for linguistic clues that might match those of the language of the writer of the bomb threats. Thus, the original linguistic profile was transformed into a type of authorship identification analysis.

One of these writing samples stood out strongly as using language that was very similar to the writing found in the bomb threats. Although the FBI knew the identities of each of the writers, I asked to be intentionally blinded to this information. The FBI agents were surprised (as was I) when my analysis demonstrated a very good match between the language of the threat messages and the writing sample composed by the director of the clinic, the very person who originally reported the matter to the police. She had grown up in Pakistan where she spoke both Urdu and English and went to British schools, leaving the above-mentioned clear ethnic traces of this in her three threat messages. What had started out to be a task of linguistic profiling ended up fulfilling the task of authorship identification.

Not being linguists, the law enforcement agencies did not notice the Hindi-Urdu, British English, and female style clues in the three bomb threat letters, nor did they associate them with the language of the clinic director. Linguists trained in ethnic variability can do this. There could be no question about the intentionality and predisposition of the writer, who wrote these threats voluntarily.



## **Aftermath**

The law enforcement officers could not even guess at any motive that the director might have had to send these bomb threat messages to herself. When the FBI's supervising agent showed the director how my linguistic profile matched her own writing sample, she confessed immediately and also revealed her motive for writing those threats. She tearfully explained that her husband had been studying medicine at a California medical school, where recently he had failed his exams. His wife, the physician who was Director of the Gary clinic, admitted that she wanted to move back to California to help her husband study to retake the exams and finish his medical training. Apparently she could think of no other excuse to do this than to invent a bomb threat scenario that she felt would be an adequate excuse to close the clinic and allow her to move back to her husband. Since no physical harm occurred, no criminal charges were brought against her.

Like the Unabomber case, the Gary bomb threats provided clear evidence of the suspect's intentionality and predisposition to kill people. In most linguistic profiling cases, there is usually no question about the intention, predisposition, or voluntariness of the sender of the anonymous messages. The language evidence is very different in solicitation to murder cases, when the targets are still suspects and the government's task is to determine intentionality and predisposition that is elicited voluntarily. Even when murder suspects are in custody where the primary task is to elicit a confession, it is the language used by the suspects, police interviewers, and lawyers that frames the issues of intentionality, predisposition, and voluntariness.

The rest of this book deals with 13 murder cases in which law enforcement provides known suspects and defendants in whose cases the

legal terms, intentionality, predisposition, and voluntariness were not nearly as clearly represented as they were in these threat message cases.

