INVESTIGATION · SCIENTIFIC AND TECHNICAL INVESTIGATION

The Contribution of Forensic Linguistics to Investigative Knowledge

Forensic linguistics applies the study of language to investigations and legal proceedings. This article explains its main contributions, including authorship analysis, the examination of disputed statements and interviews, voice comparison, language analysis in asylum cases and the meaning of contested words, along with the evidence for its reliability and its limits.

ModQu research article · 1,476 words · 18 references (APA 7th edition)

Language is evidence. Threatening letters, ransom notes, disputed confessions, emails, text messages, social media posts, recorded calls and police interview transcripts all consist of words, and the way those words are chosen, arranged and spoken can reveal who produced them, whether they were recorded accurately and what they meant. Forensic linguistics is the application of linguistic knowledge to such questions in investigations and legal proceedings. This article describes its main contributions, the methods behind them and the limits investigators should keep in mind.

Origins

The term "forensic linguistics" was used by Svartvik (1968) in an analysis of statements attributed to Timothy Evans, who had been hanged in 1950 for the murder of his daughter. Svartvik found that the incriminating parts of the statements differed in style from the rest, suggesting they were not in Evans's own words. Evans was later posthumously pardoned after it emerged that his neighbour, John Christie, was a serial killer. In the decades since, linguists have been involved in many cases concerning disputed police records, authorship and meaning (Coulthard & Johnson, 2007; Gibbons, 2003).

Disputed statements and interviews

Before interviews were routinely recorded, police statements were often written by officers and signed by suspects. Linguists examined such statements for signs that words had been added or altered. Coulthard (1994) showed in several British appeals that statements attributed to suspects contained features more typical of police register, such as particular word orders and phrases, than of ordinary speech, and that some passages were too similar to other documents to have been produced independently. Such analysis contributed to convictions being reconsidered.

Linguists also analyse interviews themselves. Shuy (1998) examined recorded interrogations and undercover conversations and showed how questions, interruptions and topic control could lead suspects to appear to agree with propositions they had not actually accepted. Eades (2010) has shown how differences in communication norms, for

example among Aboriginal Australians, can lead to misunderstanding in interviews and courtrooms, including apparent agreement that reflects politeness rather than assent. These analyses inform both the evaluation of evidence and the design of fairer interviews.

Authorship analysis

Authorship analysis asks who wrote a text, or whether two texts were written by the same person. It rests on the idea that each writer has habitual linguistic choices, sometimes called an idiolect, that distinguish their writing from others'. Nini (2023) sets out a theory of linguistic individuality to explain why such differences exist and how they can be measured.

Methods range from qualitative comparison of distinctive features, such as unusual spellings, punctuation habits or phrases, to statistical analysis of frequencies of common words and character sequences. Mosteller and Wallace (1963) used the frequencies of function words such as "upon" and "while" to attribute disputed Federalist Papers to James Madison, a landmark in statistical stylometry. Juola (2006) and Stamatatos (2009) review the many computational methods since developed, which can achieve high accuracy in controlled conditions with enough text.

Forensic texts, however, are often short, such as a few text messages or a single threatening letter, and the pool of possible authors may be uncertain. Grant (2007) argued that forensic authorship analysis should express conclusions in terms of the strength of evidence, using methods whose error rates have been tested, rather than claiming identification. Authorship analysis is usually best used to prioritise suspects or to test whether a known person's writing is consistent with a disputed text, rather than as sole proof.

Linguists have also contributed to online investigations. Grant and MacLeod (2020) describe how analysis of language helps undercover officers take over the online identities of offenders convincingly, and how language can reveal identity in online communities.

Voice comparison

Forensic speaker comparison examines whether a voice in a recording, such as a threatening call, belongs to a known person. Analysts combine auditory analysis of accent, voice quality and speech habits with acoustic measurements. Rose (2002) explains why speaker identification is difficult: a person's voice varies with emotion, health, recording conditions and context, and different speakers can sound similar. Morrison (2009) argued that forensic voice comparison should adopt the likelihood ratio framework used in DNA evidence, expressing how much more probable the evidence is if the known person is the speaker than if someone else is, and should be tested empirically.

Language and origin

In asylum procedures, authorities sometimes use language analysis to assess whether an applicant's speech is consistent with the region or community they claim to come from. Linguists have warned that such analysis is complex, because people's language reflects their life histories, migrations and contacts, not only their birthplace. Guidelines produced by a

group of linguists recommend that analysis be carried out by qualified linguists, that conclusions be stated with appropriate caution and that analysts address speech communities rather than national borders (Language and National Origin Group, 2004).

Meaning and interpretation

Many legal disputes turn on what words mean. Did a statement amount to a threat, a confession, an agreement or a warning? Was a consumer contract clear? Solan and Tiersma (2005) show how the meaning of words in criminal cases, for example in consent to searches, requests for a lawyer and threats, often depends on context and on assumptions that courts do not always make explicit. Linguists can help investigators and courts understand how ordinary speakers would interpret a message.

Deception and language

Some researchers have studied whether the language of lies differs from the language of truthful accounts. Newman et al. (2003) found statistical differences on average, for example fewer first-person references and more negative emotion words in deceptive texts. Such differences are small and inconsistent across settings, and they do not allow reliable judgements about whether an individual is lying. They are research findings, not investigative tools for classifying statements.

Contributions of forensic linguistics

What each area can offer an investigation, and its main caution

	What it offers	Main caution
Disputed statements	Tests whether recorded words are the speaker's own	Depends on comparison material
Interview analysis	Shows how questions shaped answers	Interpretation of intent is limited
Authorship analysis	Prioritises or tests possible authors	Short texts and open suspect pools reduce reliability
Voice comparison	Assesses whether a known person is the speaker	Voices vary; methods need validation
Language and origin	Assesses consistency with a claimed community	Life histories complicate conclusions
Meaning of words	Explains how ordinary speakers would understand a message	Context can support several readings

ModQu analysis based on Coulthard & Johnson (2007), Grant (2007) and Morrison (2009) ModQu.com

Figure 1. The main areas in which linguists contribute to investigations, what each can offer, and the main caution for each. Diagram by ModQu.

Working with a linguist

Investigators get the most from linguistic analysis when they involve a linguist early and brief them carefully. Useful steps include preserving the original texts and recordings in their native format, since retyping or converting them can remove features such as spacing, punctuation and metadata; collecting comparison material from each candidate that is similar in type, date and topic to the disputed text; and recording how each sample was obtained. To reduce the risk of bias, the linguist should not be told which candidate investigators suspect, and ideally should analyse the disputed text before seeing the comparison samples. Investigators should also ask for the analysis to be written so that a non-specialist can follow the reasoning, with examples of the features relied on. Where texts are in more than one language, or in a dialect or online style the linguist does not know well, a specialist in that variety may be needed. Linguistic analysis takes time, so investigators should agree the questions, scope and timetable at the outset.

Reliability and admissibility

As with other forensic fields, the reliability of linguistic evidence depends on the method. Courts applying standards such as that in *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993) ask whether a method has been tested, has known error rates and is accepted in the field. Methods based on statistical analysis of large comparison corpora and tested on known cases are easier to defend than impressionistic judgements by a single expert. Investigators instructing linguists should ask what method will be used, whether it has been validated on similar material, how much comparison text is needed, and how conclusions will be expressed.

A short example

The following hypothetical case illustrates authorship analysis in an internal investigation. A company receives a series of anonymous emails making false allegations about a manager. Three employees have been mentioned as possible authors. A linguist compares the emails with samples of each employee's work emails. She finds that the anonymous emails share several uncommon features with one employee's writing, including a distinctive abbreviation and a habit of ending sentences with double punctuation, but that one of the other employees also uses some of these features. Her report states that the evidence moderately supports one employee as the author rather than the others, explains the comparison method and its limits, and recommends seeking other evidence. Email server logs then show that the anonymous messages were sent from a device registered to that employee. The linguistic analysis helped prioritise the enquiry; the logs provided the proof.

Conclusion

Forensic linguistics contributes to investigations by testing disputed statements and interviews, analysing authorship, comparing voices, assessing language in asylum cases and clarifying the meaning of contested words. Its history includes important corrections of miscarriages of justice. Its reliability varies with the method and the amount of material, and conclusions are strongest when based on tested methods and expressed as degrees of support. Used in this way, linguistic analysis can add valuable knowledge to investigations without claiming more certainty than language allows.

References

Coulthard, M., & Johnson, A. (2007). *An introduction to forensic linguistics: Language in evidence*. Routledge.

Coulthard, M. (1994). On the use of corpora in the analysis of forensic texts. *Forensic Linguistics, 1*(1), 27–43.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

Eades, D. (2010). *Sociolinguistics and the legal process*. Multilingual Matters.

Gibbons, J. (2003). *Forensic linguistics: An introduction to language in the justice system*. Blackwell.

Grant, T., & MacLeod, N. (2020). *Language and online identities: The undercover policing of internet sexual crime*. Cambridge University Press.

Grant, T. (2007). Quantifying evidence in forensic authorship analysis. *International Journal of Speech, Language and the Law, 14*(1), 1–25. https://doi.org/10.1558/ijsll.v14i1.1

Juola, P. (2006). Authorship attribution. *Foundations and Trends in Information Retrieval, 1*(3), 233–334. https://doi.org/10.1561/1500000005

Language and National Origin Group. (2004). Guidelines for the use of language analysis in relation to questions of national origin in refugee cases. *International Journal of Speech, Language and the Law, 11*(2), 261–266. https://doi.org/10.1558/sll.2004.11.2.261

Morrison, G. S. (2009). Forensic voice comparison and the paradigm shift. *Science & Justice, 49*(4), 298–308. https://doi.org/10.1016/j.scijus.2009.09.002

Mosteller, F., & Wallace, D. L. (1963). Inference in an authorship problem. *Journal of the American Statistical Association, 58*(302), 275–309. https://doi.org/10.1080/01621459.1963.10500849

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin, 29*(5), 665–675. https://doi.org/10.1177/0146167203029005010

Nini, A. (2023). *A theory of linguistic individuality for authorship analysis*. Cambridge University Press. https://doi.org/10.1017/9781108974851

Rose, P. (2002). *Forensic speaker identification*. Taylor & Francis.

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

Solan, L. M., & Tiersma, P. M. (2005). *Speaking of crime: The language of criminal justice*. University of Chicago Press.

Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology, 60*(3), 538–556. https://doi.org/10.1002/asi.21001

Svartvik, J. (1968). *The Evans statements: A case for forensic linguistics*. University of Gothenburg.