Arts & Humanities Research Council

Forensic linguistics

What is it?

Linguistics is the scientific study of language as a human activity. It deals with both the structure of language and the ways in which it functions in different settings. There are many fields in linguistics, and one increasingly prominent area of applied research is forensic linguistics. This field has developed from a research-based understanding of language.

Forensic linguistics involves the application of scientific knowledge to language in the context of criminal and civil law. Forensic linguists have an interest in understanding the language of the written law, its complexity and its origin, as well as the use of language in forensic procedures. They also study the judicial process from point of arrest, and through the interview, charge, trial and sentencing stages. For example, linguists are interested in the language of police interviews with witnesses and suspects, and in the language of lawyers and witnesses in cross-examination.

Forensic linguistics and the justice system

Linguistic research of this sort, which is publicly funded and conducted at universities, has made significant contributions to our understanding of justice. For example, it has led to the development of new ways of communicating rights to suspects in police custody and has challenged the norms of persuasive language in courts.

Forensic linguists have also been called upon to give expert testimony in a variety of criminal cases across the UK and abroad. Science is becoming increasingly important in relation to the law, and forensic linguistics is one area where research is leading to advances that are increasingly used to solve crimes. A recent example is the high-profile murder case of Danielle Jones, where linguistics was used to decode text messages, resulting in a conviction.

Forensic linguists also apply their scientific knowledge and techniques to private disputes between parties, which may at a later stage result in legal action. Recent examples range from trademark disputes between producers and distributors of whiskey to cases of plagiarism at universities, as well as several slander cases.

Typical types of linguistic evidence include

- Author analysis: determining who wrote a particular text by comparing it to known writing samples of a suspect in order to try to link texts by the same author. Evidential texts have varied considerably in length and type in recent cases, ranging from text messages and threat letters to long terrorist conspiracy documents. Authorship analysis includes detection of plagiarism.
- Discourse analysis: analysing the structure of mostly written conversations, to help determine issues such as whether a suspect is agreeing to engage in a criminal conspiracy. Other areas of application include the analysis of police questions in interviews and the management of courtroom exchanges. Research in both of these areas has resulted in procedural changes in other countries and to some extent in the UK. An example of this kind of change can be seen in the 'Plainer English' version of the Californian Criminal Code.
- Linguistic competence: for example, did a suspect understand the police caution? Judgements of linguistic competence need to be subtle and determined by experts, as a (non English-speaking) defendant may have a sufficient command of English for casual daily interactions but may be disadvantaged in more formal and stressful contexts.
- Trademark disputes: often concerning questions of when a trademarked term begins to be used generically. In such a situation, legal 'trademark death' may be said to occur. This leaves

companies in a dilemma, for example if the generic verb to search the internet becomes 'to google', then Google will suffer 'trademark death' and lose its legal protection

Forensic linguists can intervene in the legal process at the investigative stage, at the trial stage or at the appeal stage. Perhaps because it is much easier to eliminate someone as a suspect with a reasonable degree of confidence than it is to prove him or her guilty beyond a reasonable doubt, it is not surprising that most linguistic expertise in criminal cases has been for the defence. Nonetheless, linguistics can be useful to public prosecutions as well, especially during the investigative stages.

Forensic linguists typically work with corpora when providing linguistic evidence. Corpus studies involve the systematic analysis of language through the collection and study of large bodies – or corpora – of texts or oral testimony (transcripts). Linguists often build corpora themselves, for example with witness statements and police statements, and check their data against a reference corpus. Forensic linguists use these collections of a particular type of writing or other linguistic data as a reference point for their more detailed analyses, where individual cases of linguistic usage can be compared with that of the larger group or corpora of linguistic use and characteristics.

Future developments

Forensic linguistics is currently improving the reliability and validity of its methods. This will have important implications for admitting evidence in court.

A very hot topic in forensic linguistics is investigative interviewing. Linguistic research will almost certainly influence police guidance and training on interview techniques. Other current research topics with potentially important societal impact include courtroom linguistics, (including how vulnerable witnesses might be assisted in a trial), communication between legal professionals and laypeople, and the study of the language used in prisons and offender treatment programmes.

Forensic linguistics timeline

1968 The term 'forensic linguistics' is used for the first time by the linguistics professor Jan Svartvik in his book "The Evans Statements: A Case For Forensic Linguistics". He presented an analysis of statements given by Timothy John Evans, suspected of murdering his wife and baby, to police officers at Notting Hill Police Station in 1953, and demonstrated that the disputed parts of the statement had a grammatical style measurably different from the style of uncontested parts.

1988 The Federal Criminal Police Office (BKA), Germany, organises a two-day conference in forensic linguistics. The BKA has had a linguistics unit since the 1980s. In contrast, the UK police do not have such a unit and rely on university-based forensic linguistics.

1989 Conferences on forensic handwriting analysis, at which there are also presentations in forensic linguistics. Organised by the late Lothar Michel at Mannheim University, Germany.

1990 and 1991 Conferences in York on forensic phonetics and linguistics, organised by Peter French.

1990, 1991, 1992 and beyond GAL, the German Applied Linguistics Association, has a working group on forensic linguistics in its programme every year.

March 1992 The First British Seminar on Forensic Linguistics is held at the University of Birmingham, and is attended by delegates from Australia, Brazil, Ireland, Holland, Greece, Ukraine and Germany as well as the UK. There is a consensus that an international association is needed. It is from this seminar that the International Association of Forensic Linguistics was born.

November 1992 A one-day seminar (Second British Seminar on Forensic Linguistics) is held in Birmingham.

1993 International Association of Forensic Linguistics (IAFL) founded.

1994 Launch of the journal Forensic Linguistics. The journal changed its name to The International Journal of Speech, Language and the Law in 2003 to reflect a broadening of academic coverage and readership.

1999 First MA course in forensic linguistics introduced at Cardiff University.

2008 Establishment of the Centre for Forensic Linguistics at Birmingham's Aston University, to cope with the increasing demand for forensic linguistic skills.

September 2008 The Council for the Registration of Forensic Practitioners (CRFP) opens up its register for forensic linguists.

An example of author analysis and use of corpus evidence

Forensic linguistics contributed significantly to the posthumous overturning of Derek Bentley's conviction for murder in 1998. Nineteen-year-old Bentley was hanged in 1952 for his part in the murder of a policeman. The fatal shot had been fired by Bentley's sixteen-year-old friend, Christopher Craig, when Bentley was already in police custody. Bentley was convicted as an accessory to murder, partly on the basis of his statement to police. Two police officers swore at the original trial that the police statement was a verbatim account by Bentley of a spoken monologue.

Linguist Malcolm Coulthard from the University of Birmingham examined the statement when the case was reopened, and found a number of features which indicated police co-authorship. This suggested that at least part of the statement resulted from questions and answers, as Bentley claimed, rather than a "verbatim record of dictated monologue", as claimed by the police. One such feature was the frequent use of the word "then". This might not seem remarkable given that Bentley was reporting a series of sequential events, and that one of the obvious requirements of a witness statement is accuracy about time. However, a glance at a series of other witness statements suggested that Bentley's usage was atypical.

Feeling that the use of that word could be expected to be higher than average in witness statements, Coulthard compiled two corpora, one of witness statements and one of police statements. The results were striking: The word "then" occurred once every 930 words in the former, but once every 78 words in the latter. This can be compared to Bentley's statement, where "then" occurred once every 53 words, and to the Corpus of Spoken English, where it occurred once every 500 words. Thus, Bentley's usage of "then" grouped his statement firmly with those produced by police officers.

This led Coulthard to focus more on the use of the word "then". He noted the frequent post-positioning of temporal (time-related) "then" after the grammatical subject ("I then" rather than "then I"), which occurred seven times in the 582-word text. The Corpus of Spoken English showed "then I" to be ten times as frequent as "I then", the latter occurring only once every 165,000 words. That structure did not occur at all in Coulthard's corpus of witness statements, but appeared once every 119 words in the corpus of police statements. These features, combined with others, contributed to a successful argument that the Bentley "confession" was, in part, the written work of police officers, and not simply a word-for-word transcript of Bentley's spoken statement as the police alleged.