I was watching television tonight (as of 1982), and between 7pm and 8pm I heard a loud noise. My mother was in the kitchen and I heard her say I was on the phone. I was called for me. My mother said I was on the phone and I walked into the kitchen. I was called for me. I was just after 7pm.

The Language Detective

Forensic Linguistics First Certificate Course

Unit 1

www.thetext.co.uk
Aim of Unit 1: becoming a 'language detective'

The aim of this unit is to make you aware of some of the more important forensic phenomena in language. We have devised two ways of doing this. Firstly, learning about text measurement, and, secondly, learning to transcribe texts accurately. Measuring text is an interesting (and actually quite simple) way to find out about a text’s origins. At the same time, learning to transcribe accurately will increase your powers of observation, and hence your ability to detect what may be wrong with a text.

As you enter this first phase of forensic linguistics you need to be aware of two important considerations: (1) forensic linguistics is not psychology - we do not analyse people’s motives or behaviour, and (2) because we all use language every day it is very easy to take it for granted. This is why transcribing and measuring are so valuable - they teach us to examine and analyse language in ways we perhaps never thought about before. They set us on the path to becoming language detectives.

With the Forensic Linguistics Institute, you are learning with the world’s leading experts in this field. This is the world’s only educational institution where the teachers are full-time practitioners, routinely accustomed to handling criminal work and giving evidence in court.

John Olsson, Course Director
Contents

A prior word on 'forensics'
What is 'forensic linguistics'?
What do forensic linguists do?
The origins and development of forensic linguistics
Authorship
Linguistic register
Notes on the Derek Bentley statement
Basics of text measurement
Exercise
Assignment 1
Corpus and Collocations

The Language Detective

The Derek Bentley statement
Collocations in the statement
Transcribing forensic texts
Transcription
Assignment 2
Unit conclusion
Email address for assignments
Bibliography
Other Forensic Linguistics Institute courses
Notes
Accreditation
A prior word on 'forensics'

The word forensic comes from the ancient Roman word *forum*, a kind of Roman civic centre where legislative and judicial matters were decided, a sort of public debating chamber where citizens could discuss their problems and air their grievances.

In modern times, the word forensic is most closely associated with the word 'science', as in the phrase 'forensic science'. Partly because of its close connection with the word 'science' and partly because of the way the media has reported some of the more spectacular results achieved in criminal detection, the word forensic has come to imply something like 'special science' or 'criminal science'. This has given it a hallowed status, implying that its results are unassailable or invincible in some way. Whatever the validity of these observations relative to the so-called 'hard' sciences, the application of the word forensic to linguistics should not be taken to mean unassailable or invincible. We should not invest our deliberations with the aura of a DNA investigation.

This is because a good deal of forensic linguistic work is observation and analysis, as opposed to the measurement of probability. This does not mean that we cannot do plenty of good, worthwhile language science. It just means that we must be careful in our conclusions - but most people would take that to be a principle of good science anyway. Having sounded a note of caution with regard to the word forensic and its collocates, let us now consider what forensic linguistics is.

What is Forensic Linguistics?

Forensic linguistics is the application of linguistic expertise to legal, judicial and criminal contexts - forums (or fora) if you prefer. So, the true meaning of forensic linguistics is much closer to the origins of the word forum. We use our knowledge of language in a forum publicum (public forum) where matters of public interest, to do with the law, are debated. In this three unit course our main attention will be on the application of forensic linguistics to the criminal context, in other words linguistics and criminal investigation.
**What do forensic linguists do?**

A forensic linguist is a linguist first and foremost, a student of linguistics, of the science of language. This individual then applies linguistic expertise to a specific criminal, legal or judicial context. He or she may be asked by a police department, a prosecutor, or a judge, to identify the author of a text, give an opinion as to the genuineness of a suicide text, or assess the threat level of a ransom demand. Forensic linguists also have private clients: for example, an individual may receive anonymous hate mail. In such cases the forensic linguist can apply linguistic knowledge to the question: courts have relied on forensic linguists to carry out such identification processes in the past.

**The origins and development of forensic linguistics**

The first known use of the phrase forensic linguistics occurred in a paper by Professor Jan Svartvik: The Evans statements: a case for forensic linguistics? In this 1968 paper Jan Svartvik, an eminent Swedish linguist then doing pioneering work on language corpora, analyzed the police statements of Timothy John Evans and found that there were two distinct authorship styles in the texts. Between 1968 and the 1980's there was little public activity in the field, but in the late 1980's and early 1990's a number of significant miscarriages of justice came to the fore in the United Kingdom and forensic linguistics underwent something of a revival, with such well known cases as that of Derek Bentley, the Birmingham Six, etc. Since that time forensic linguistics has branched out into a number of fields which we will consider in some detail in the following sections.

**Authorship**

One of the main areas in forensic linguistics is authorship. Linguists and other professionals, including psychologists and statisticians, have long been interested in establishing the characteristics of individual style. The earliest known authorship controversy was that of the bible. In 1711 a German priest, HB Witter, suggested that the multiplicity of biblical names for the deity pointed to the possibility of multiple authorship of the bible. Others reached the same conclusion: Jean Astruc, a French doctor, and JG Eichhorn, a German professor. The other significant authorship controversy focused on the authorship of the Shakespeare plays. In the late 18th century, James Wilmot contended that Bacon was the real hand, as did James Corson Cowell in 1805 before a philosophical society meeting at Ipswich, England. In the mid nineteenth century Augustus de Morgan adopted
a scientific approach to the question of authorship. He proposed word length average (over long texts) as an authorship indicator. TC Mendenhall, of the Ohio State University, made similar proposals later in the century. By contrast Yule, a Cambridge statistician in the late 1930’s, saw mean sentence length as a useful marker.

A number of other researchers have claimed to have discovered techniques which give reliable and robust authorship markers. However, to date nobody has been able to demonstrate complete reliability of method. The reasons for this are complex. They relate to such issues as individual variation and the tension between language as a social property and language as an individual attribute. For any individual to be understood by any other individual in the same language community, it is necessary for language to be a common property of that community. This simple fact, along with such diverse socially impelling factors as standardisation in education and media globalisation, mitigates strongly against language being an individual property. However, there are indications of a degree of individualism from one author to another and, as Heike Hanlein demonstrated in her 1998 study, we are all familiar with a certain sense of déjà vu when reading the text of an author familiar to us.

The other main problem in identifying viable authorship markers is that style seems to be text type dependent. In other words, our style depends on the type of text we are writing (formal academic, personal letter, chatty email, etc.), as well as who our intended audience is (close friend, as opposed to senior aged colleague whom we greatly respect, or - by contrast - someone whom we deem to be lower in the social pecking order than ourselves, etc.). Many writers who have proposed style markers have proved too insistent on the reliability of their results, and have failed to present their statistics in a way that courts can understand. We will be looking at quantifying some aspects of text measurement in this course, but we will not be ‘doing’ statistics as such.

Linguistic Register

Linguistic register is a term used by linguists to describe the social components of an item of language. Who is saying/writing it? Who is it addressed to? What is the topic? How formal is it? Is its origin as a spoken or written text, or was it produced by some other means? This last question relates to what linguists term mode.

Mode is the study of language form and production. It looks at such questions as whether a text may have been dictated or spoken. It helps us to question language more closely, and
Forensic Linguistics First Certificate Course

not to take the form in which we may find it at face value. Very often we are presented with a language encounter which may seem spontaneous, but which in fact was scripted. For example, an interview between a news anchor and a reporter, speaking ‘live’ from a crime scene, the White House, Downing Street, etc. We are told the encounter is live, and we assume that the interview is off-the-cuff and unstructured. This is to help us believe that the media organization carrying out the interview has absolutely no political agenda, when - in fact - that may not be the case. This is just an everyday example.

Notes on the Derek Bentley statement

This is a disputed police statement, namely the statement allegedly given by Derek Bentley (hanged for being at the scene of a shooting in which a policeman was killed by his accomplice, Christopher Craig - but later pardoned). The point about this statement was that it was supposed to have been dictated by Derek Bentley (i.e. not written or spoken). It occurred to me early on that transcription from dictation was likely to be very different from ordinary written language, despite superficial similarities, and similarly that dictation and speech would also be different from each other. I was able to pinpoint two different styles in the text, shown here as DB1 and DB2.

The Language Detective

DB1

I have known Craig since I went to school. We were stopped by our parents going out together, but we still continued going out with each other - I mean we have not gone out together until tonight. I was watching television tonight (2 November 1952) and between 8pm and 9pm Craig called for me. My mother answered the door and I heard her say I was out. I had been out earlier to the pictures and got home just after 9pm. A little later Norman Paskley and Frank Fasey called. I did not answer the door or speak to them. My mother told me that they had called and I then ran out after them.

DB2

There was a little iron gate at the side. Chris then jumped over and I followed. Chris then climbed up to the drainpipe and I followed. Up to then Chris had not said anything. We both got out on to the flat roof at the top. Then someone in a garden on the opposite side shone a torch up towards us. Chris said: ‘It’s a copper, hide behind here’. We hid behind a shelter arrangement on the roof. We were there waiting for about ten minutes. I did not know he was going to use the gun.

Study the above examples from Derek Bentley’s statement - there is an Assignment relating to them on Page 15.
Forensic Linguistics First Certificate Course

Basics of text measurement:

**Sentence Length Average** - We measure sentence length to help us determine the origin or provenance of a text, or differences between one section of a text and another. Simply count the number of words between sentence-enders (periods/full stops, question marks or exclamation marks).

**Word Length Average** - We measure word length by first removing all punctuation marks, then counting the total number of characters divided by the total number of words.

**Exercise**
Transcribe the example text in bold type below into your word processor and calculate (i) the number of words in the text, (ii) the average sentence length, and (iii) the average word length. The answers are on the next page.

**Example Text**
Here is a text consisting of a number of words and sentences. We are using it as an example. You can easily count the number of words and sentences in it.

Sample data from the previous page:
31 words; 3 sentences; 121 characters.
Average Sentence Length: \( \frac{31}{3} = 10.3 \) words per sentence
Average Word Length: \( \frac{121}{31} = 3.90 \) letters per word

**TIP**
Give average sentence length to one decimal place, and average word length to two decimal places.

**Assignment 1.** Now go back to the earlier examples, DB1 and DB2 (see Page 13). Carry out the same tests on these two excerpts. Remember to calculate sentence length average to one decimal place and word length average to two decimal places (where possible). Comment on the style differences between the two sections. What do you think these differences mean in terms of the text’s origin? Comment on any other aspect of the text which may relate to these questions (300-500 words).

See Page 21 of this volume for the email address to send your assignments to.
Some technical terms:

**Corpus**
A corpus of language (plural corpora) is a body of language collected for research purposes. We can have general corpora and case specific corpora where we examine a relatively small body of language to research facts specific to a case. Linguists sometimes also build specialist corpora of a particular genre or text type.

**Collocation**
A collocate is a word which occurs near, or next to, another word. There are many regularly occurring collocates in the language, for example 'street' often occurs near 'life'. Thus 'street' is a collocate of 'life'.

**Exercise**
In the following text, which is the Bentley statement in full, see if you can observe any collocates. (Note: click on the text to get a downloadable copy.)

---

The Derek Bentley statement
I have known Craig since I went to school. We were stopped by our parents going out together, but we still continued going out with each other - I mean we have not gone out together until tonight. I was watching television tonight (2nd November 1952) and between 8pm and 9pm Craig called for me. My Mother answered the door and I heard her say I was out. I had been out earlier to the pictures and got home just after 7pm. A little later Norman Parsley and Frank Paoey called. I did not answer the door or speak to them. My Mother told me that they had called. I ran out after them. I walked up the road with them to the park shop where I saw Craig standing. We all talked together and then Norman Parsley and Frank Paoey left. Chris Craig and I then caught a taxi to Croydon. We got off at West Croydon and then walked down the road where the toilets are - I think it is Tamworth Road. When we came to the place where you found me, Chris looked in the window. There was a little iron gate at the side. Chris disappeared around the corner and I followed up to there. Chris had not said anything. We both got out and I held the little iron gate. Then someone in a garden on the opposite side shouted a torch towards us. Chris said: 'It's a copper, hide behind here.' We hid behind a hedge which was on the roof. We were there waiting for about ten minutes. I did not know he was going to use the gun. A plain clothes man climbed up the drainpipe and got up to the roof. The man said: 'I am a police officer - the place is surrounded.' He caught hold of me and as we walked away Chris fired. There was nobody else there at the time. The policeman and I went round a corner by a door. A little later, the door opened and a policeman in uniform came out. Chris fired again then and this policeman fell down. I could see he was hurt as a lot of blood came from his forehead just above his nose. The policeman dragged him round the corner behind the brickwork entrance to the door. I remember I shouted something, but I forget what it was. I could not see Chris when I shouted to him - he was behind a wall. I heard some more police officers behind the door and the policeman with me said, 'I don't think he has many more bullets left.' Chris shouted, 'Oh yes, I have' and he fired again. I think I heard him fire three times altogether. The policeman then pulled me down the stairs and I did not see anyone. I knew we were not going to break into the place. I did not know what we were going to get - just anything that was going. I did not have a gun and I did not know Chris had any until he shot. I now know that the policeman in uniform is dead. I should have mentioned that after the plain clothes policeman got up the drainpipe and arrested me, another policeman in uniform followed and I heard someone call him Mac. He was with us when the other policeman was killed.
Collocations found in the Bentley statement: door and policeman

A little later the door opened and a policeman in uniform came out. Chris fired again then and this policeman fell down. I could see he was hurt as a lot of blood came from his forehead just above his nose.

The policeman dragged him round the corner behind the brickwork entrance to the door. I remember I shouted something but I forget what it was. I could not see Chris when I shouted to him - he was behind a wall. I heard some more policemen behind the door and the policeman with me said, 'I don't think he has many more bullets left.'


There is a software program available on the website called Notebag. You can perform a number of analyses on texts with Notebag. A full help file and instructions are included with the program, which is available here.

Transcribing Forensic Texts

Forensic texts can come in various shapes and sizes. They can be handwritten (as you will see in the next example), or typewritten or word-processed. Even if they are word-processed you will still need to make some kind of a copy. Copying from one computer program to another, or from one file location to another can still result in errors. Scanning a text as an image and then transforming it into an optically recognisable document also has pitfalls. The most difficult of all types of transcription is probably the transcribing of handwritten documents. However, any form of transcription is precise and exacting work. To be a good language detective you need to carry out the transcription task with great accuracy and care.

Transcription

Earlier, we began to acquire our skill set for forensic linguistics by learning a little about measuring text. In this section we will continue to advance our knowledge about text measurement and at the same time we will acquire another very important skill, that of transcribing text accurately. Recall that a forensic text is evidence. What good is
Assignment 2:
Transcribe the Susan Smith confession. Then measure it for average word length and average sentence length. Comment on any internal differences between one part of the confession and another. In your view what could have caused those differences? Consider, for example, the circumstances under which the confession was produced.

[Count each icon as a single word]
Bibliography


De Morgan, S. 1882. Memoir of Augustus de Morgan (By his wife Sophia de Morgan, with selections from his letters). London.


Other courses:

Please visit our website to view our other courses. These include a Diploma course, a Forensic Transcription course, a variety of micro-courses. By the beginning of 2009 we also plan to have a Forensic Phonetics course.

Accreditation

The Forensic Linguistics Institute is accredited as a learning provider by the Open and Distance Learning Quality Council of the United Kingdom.
Notes:

Right-click anywhere on this page and then click on 'Notes'. Then click on 'Add New Note'. You can add any notes you wish to this document and save them. If you wish you can do the same on any other page in this book.

If you have reached this page, you have probably finished this Unit. Unit 2 looks at genuine vs. simulated text, and Unit 3 considers the question of authorship. We hope you are enjoying the course so far. We look forward to your continued studies with us and to receiving your assignments. We welcome your suggestions for any improvements we can make to future courses.

www.thetext.co.uk