Forensic Linguistics

John Olsson

Forensic Linguistics:
Second Edition

Also available from Continuum

Language and the Law, Sanford Schane

Forensic Linguistics: Second Edition

John Olsson

Contents

This page intentionally left blank

Introduction to the second edition

It is a pleasure to be able to write an introduction for the second edition of this book, especially as I did not expect the first edition to be as successful as it seems to have been: three reprints in three years, translation into Arabic, its adoption on courses around the world, its presence in many major university and city libraries.

In presenting a second edition of *Forensic Linguistics* my intention has been to try to make it a better book, more informative and descriptive, more useful to students coming to grips with these questions for the first time, and more useful to practitioners.

I also have nearly six years more experience as a forensic linguist than I had when I started writing the first edition. This experience includes my involvement in a number of major court cases, including murder, witness intimidation, suspicious death, copyright infringement, narcotics investigations, examination of mobile (cell) phone texts, and many others. I therefore hope that my experiences will help me to produce a book which will improve on the first edition – the only possible reason for writing a second edition.

In undertaking this edition I have tried to keep as close to the original format as possible. Having agonized long and hard about what to keep in from the first edition and what to put in that was not there, in the end I adopted a simple policy: if I thought it was useful and still relevant I kept it. If not out it went. The addition of new material was carried out on the same kind of basis – if it added something useful it went in, if not it did not, no matter how attractive it might seem.

Who should use this book

This book is intended for students of forensic linguistics at undergraduate and postgraduate levels. It can be used on its own or in conjunction with other titles in the field. As indicated above, the book is also intended for non-specialists,

such as law enforcement officers, legal professionals at all levels – from the newly qualified solicitor to the senior barrister, the magistrate to the judge. Psychologists and sociologists will also find the title of interest, particularly those who already have an interest in language. I have also received many emails from forensic chemists, fingerprint experts and even handwriting analysts who have been keen to learn about forensic linguistics. I believe that experts in all forensic disciplines may find this book useful. The Forensic Linguistics Institute also runs several distance-learning courses on the internet, and these are well attended by linguists and non-linguists alike from around the world.

About the author

John Olsson spent a number of years in the 1970s as an interpreter for the Metropolitan Police in London, after which he studied psychology at degree level. A stint in business was followed by two postgraduate degrees in linguistics at Bangor, Wales and Birmingham, England. His M.Phil thesis was a quantitative and qualitative analysis of the process of dictation, as applied to the Derek Bentley text (see Text Appendix). He has attended court many times as an expert witness, and is frequently consulted by solicitors and law enforcement agencies requiring opinions on forensic texts. In all he has prepared nearly 200 reports on linguistic and phonetic cases for the courts.

What is Forensic Linguistics? 1

Chapter Outline

Text types

One way to answer the question 'What is Forensic Linguistics?' is by considering the kinds of text forensic linguists are sometimes asked to examine. Literally any text or item of spoken language has the potential of being a forensic text. If a text is somehow implicated in a legal or criminal context then it is a forensic text. A parking ticket could become a forensic text, a will, a letter, a book, an essay, a contract, a health department letter, a thesis – almost anything. In practice, however, forensic linguists have mostly confined their attention to a small number of text types, some of which are included in the forensic text appendix in this book.

Thus, the appendix contains, among others, a forged will, a statement alleging sexual assault, a letter from a pipe bomber attempting to rationalize his crimes, a complaint about the practice of verballing,[1] a young mother claiming that she was carjacked and her children kidnapped, the same woman appealing at a press conference for the safe return of her children, and, finally, the same woman admitting to having killed her children.

We have a stalker text from John Hinckley – the man who attempted to kill President Reagan – addressed to a film star; there is the well-known police statement wrongly attributed to Derek Bentley, as well as several other defendant and witness statements; a number of emergency calls to a fire department – several of which are hoax and the ransom note from Carlos the Jackal demanding safe passage out of Austria for himself and the OPEC delegates he and his associates had kidnapped. There are also several other ransom demands, including that for the Enigma machine stolen from Bletchley Park in England. There are the Gilfoyle 'confession' and 'suicide' letters, as well as examples of smear, hate, trick, terror and incitement mail; and there are various confessions; a pair of plagiarized texts; an excerpt from the Unabomber's text; and the apparently forced confessions of the three westerners in the Saudi terror case. There is the ransom note in the JonBénet Ramsey murder case, as well as the ransom note in the Lindbergh kidnap and murder; several death row final statements; a number of the seventeenth-century Salem confessions; a 'suicide' note faked for literary purposes (this is not strictly speaking a forensic text, but is of interest to forensic linguists for comparison purposes). There are several attested suicide notes including that of author Virginia Woolf and several elderly patients in hospital who were unwilling to burden their relatives with their illnesses, and miscellaneous other texts. Many of these texts are disputed, quite a few are attested. There are *no* invented examples. The list of texts has been expanded considerably from the first edition, and I have also now included examples of internet scams and hoaxes, email threats, email communications which include computer viruses, etc. I have also included, on a special website, a number of examples of disputed spoken language, and some of the case material from voice identification cases. I have also removed a number of texts which I felt were not of real use to the student.

I hope that the breadth of text I have made available will enable the reader to see that the work of the forensic linguist is very varied. In the first edition I wrote that those interested in developing their knowledge of Forensic Linguistics were coming into the discipline at a very interesting time. Since then, in the six years since I first wrote those words, Forensic Linguistics has matured

considerably. Law enforcement agencies have long since seen the importance of Forensic Linguistics and inquiries from solicitors for analysis of statements, police notebooks and even of courtroom interaction, continue to grow. A new development has been in the field of security, a major growth area since the tragic events of 11 September 2001 and the tragic bombings in Madrid and London. The intervening years have also seen a phenomenal growth in organized crime, notably in people-trafficking and narcotics, and my own work as well as that of other forensic linguists has reflected these trends.

Applied linguistics

Another way to answer the question 'What is Forensic Linguistics' is to consider it as the application of linguistics to legal questions and issues. That is also a useful starting point, but like all answers it is imperfect and serves only to stimulate more questions. For example, what does *'the application of linguistics'* mean?

When Forensic Linguistics is referred to as an *application of linguistics* or, more concisely, an *applied* linguistic science, the word *applied* is not necessarily being used in the same sense as, for example, in the phrase *applied* statistics, where what is being applied is a theory underpinning a particular science to the practice of that science. Forensic Linguistics is, rather, the application of linguistic knowledge to a particular social setting, namely the legal *forum* (from which the word *forensic* is derived). In its broadest sense we may say that Forensic Linguistics is the interface between language, crime and law, where *law* includes law enforcement, judicial matters, legislation, disputes or proceedings in law, and even disputes which only potentially involve some infraction of the law or some necessity to seek a legal remedy.

Given the centrality of language to life in general and the law in particular, it is perhaps somewhat surprising that Forensic Linguistics is a relative newcomer to the arena, whereas other disciplines, such as fingerprint identification and shoeprint analysis, are much older, having a well-established presence in judicial processes.

The application of linguistic methods to legal questions is only one sense in which Forensic Linguistics is an application of a science, in that various linguistic theories may be applied to the analysis of the language samples in an inquiry. Thus, the forensic linguist may quote observations from research undertaken in fields as diverse as language and memory studies, Conversation

Analysis, Discourse Analysis, theory of grammar, Cognitive Linguistics, Speech Act Theory, etc. The reason for this reliance on a broad spectrum of linguistic fields is understandable: the data the linguist receives for analysis may require that something is said about how the average person remembers language, how conversations are constructed, the kinds of moves speakers or writers make in the course of a conversation or a written text, or they may need to explain to a court some aspects of phrase or sentence structure.

In summary, we can say that the forensic linguist applies linguistic knowledge and techniques to the language implicated in (i) legal cases or proceedings or (ii) private disputes between parties which may at a later stage result in legal action of some kind being taken.

A brief history of Forensic Linguistics

Like almost all sciences it is not possible to say that Forensic Linguistics began at a specific moment in time. Authorship identification is one area of Forensic Linguistics, but questions of authorship have exercised minds since the times of the ancient Greek playwrights who not infrequently accused each other of plagiarism. Since at least the eighteenth century scholars and amateurs alike have pondered over the authorship of some of the world's most famous texts, including sacred texts and the plays of Shakespeare. However, these applications differ from the forensic context, because it is only with Forensic Linguistics that the linguist is obliged to defend an opinion in a public forum, a court of law, and not just on questions of authorship, but on many other issues as well.

The actual phrase *Forensic Linguistics* was not used until 1968 when a linguistics professor by the name of Jan Svartvik recorded its first mention in a now famous analysis of statements by Timothy John Evans (see 'Authorship and the birth of forensic linguistics', Chapter 2).

For a long period in English law a set of rules had been established regarding the interrogation of witnesses, in particular how statements were to be taken from them. These prescriptions were known simply as Judges' Rules which laid down that suspects were to dictate their narrative to police officers, that police officers were not to interrupt suspects, and that questions were not to be asked of the suspect at the statement stage except for minor clarifications.

In practice this almost never happened. Typically, a police officer would ask a series of questions, take down notes and then write or type the suspect's

statement, not in the words of the suspect, but in a form and pattern which police custom had long dictated. Thus, police statements contained phrases like 'I then observed', etc. This type of phrasing is not at all typical of how people speak, but rather reflects a way of phrasing which has come to be known as 'police register', itself an area of study within Forensic Linguistics.

The learned judges who formulated the rules for statement taking were not aware that dictating a statement and transcribing it verbatim are difficult – perhaps even impossible – tasks for the average speaker. Learning to dictate a narrative in a coherent, sequential, articulate form is extremely difficult, but the person taking the statement has an even harder task if the speaker is not skilled at pacing his/her delivery. Usually, people do not deliver their statements in a coherent, ordered fashion: they speak too fast, they omit important details, they speculate aloud, they backtrack, and so on. In effect, the Judges' Rules were unworkable. This was why police officers had their own way of taking and regrettably in some cases making statements. It was simply impossible to follow the prescriptions of the Judges' Rules.

This was why in the early days of Forensic Linguistics, at least in the United Kingdom, many cases involved questioning the authenticity of police statements. The first example of expert evidence being given from the witness box on this matter was at a murder trial at the Old Bailey in 1989.

Notable cases included appealing against the convictions of Derek Bentley (posthumously pardoned), the Birmingham Six, the Guildford Four, the Bridgewater Three, and so on. In the United States forensic work began slightly differently, but also concerned the rights of individuals with regard to the interrogation process. In 1963 Ernesto Miranda was convicted of armed robbery, but appealed on the grounds that he did not understand his right to remain silent or to have an attorney present at the time of questioning. The Court of Appeal overturned his conviction in 1966.

In the US there were many *Miranda* cases, as they came to be known. On the face of it the provision of Miranda is a simple one: police officers are obliged to advise arrestees that they need not speak unless they wish to, that they are entitled to have a lawyer present, and that anything they say can be used against them in court. However, many issues arose, as discussed by Professor Roger Shuy: (i) a confession must be voluntary, (ii) questioning should not be coercive, (iii) arrestees must be asked whether they understand their rights, etc.

With regard to the first point Shuy pointed out that an arrestee is hardly in a position to agree *voluntarily* to being questioned. Effectively, the very nature of questioning (as pointed out by the US Supreme Court) is coercive. Shuy

(1997: 180) gives a good example of the issue of coercion in an interrogation process. He describes how a suspect, having declined to speak following the reading of his Miranda rights, was escorted in the back of a police car to the police station by two officers, who then began to talk to each other about the possibility of the murder weapon in the case (a shotgun) being accidentally stumbled upon by children at a nearby school. The suspect immediately waived his rights and led officers to the location of the weapon. The suspect, a man by the name of Innis, was convicted of murder and his lawyers appealed. The issue before the appeal court was whether the suspect had been coerced into making the confession. This in turn caused lawyers and judges to consider the meaning of the word *interrogation.* The Rhode Island Supreme Court concluded that interrogation need not involve the asking of a question, and that in this case subtle coercion had occurred and that this was 'the functional equivalent of interrogation'. In the US Supreme Court it was thus appreciated that 'interrogation need not be in the form of a question . . . [and] may involve the use of psychological ploys'. However, it was also realized that the conversation between the officers was probably more in the nature of casual remarks than a deliberate ploy.

Shuy raised many important points about Miranda, and vigorously questioned many of its assumptions of simplicity. He cites one case where a 15-year-old boy from Houston, Texas was read his rights and ultimately signed a confession of murder. After analysing tape-recorded interviews between the attorney and the child Shuy concluded that though the boy often said he understood what he was being asked it was clear that his level of comprehension was extremely low. His school confirmed that his comprehension ability was no more than that of an 8-year-old. Thus in this and other cases Shuy explores the most basic premises of Miranda, and – by extension – similar legal provisions. He does not take even the 'simplest' word or concept for granted: what does 'voluntarily' mean, does 'understand' mean 'I say I understand' or 'I actually understand'?

The work of Roger Shuy, and other US linguists, has encompassed many areas of civil and criminal practice, but right from the beginning, the law itself was, as it were, subject to questioning: what does this law mean? How do different people perform when asked if they 'understand' their rights? There is a very readable review of early Forensic Linguistics in the United States, written by Judith Levi (Levi, 1994). In her account Levi recalls a case in which she was asked to analyse a 'bad news about your social benefits' letter written by the Illinois Department of Public Aid to recipients of child benefit payments whom they had categorized as 'non-cooperative'.

One of the tasks Levi undertook was to determine whether the vocabulary selections made by the drafters of the letter had used technical and bureaucratic language in place of ordinary, everyday language. Also included in the analysis were pragmatic questions such as inferencing (Were inferences made by recipients of the letter justified by the facts of the case? Did the writers of the letter 'provide incomplete information which could lead ... to the making of misleading inferences'? Was the reader forced 'to infer information that should have been made explicit'?). The result of Levi's work was a stunning success for the recipients of the letters. Most of the recipients of the benefit were single mothers who had suffered real hardship as a result of the State's actions.

The judge in the case awarded $20,000,000 to the beneficiaries and ordered the State to rescind its classification of 'non-cooperation' until it had complied fully with the court's consent order. Finally, the State was ordered to re-write the letter in terms comprehensible to the beneficiaries (Levi, 1994: 18). An important point noted by Levi is the comment by a linguist acting in another case, namely that the legal system is 'linguistically naïve and vulnerable' (Levi, 1994: 22). This point is referred to in the next section.

Another early application of Forensic Linguistics in the United States related to the status of trademarks as words or phrases in the language. A major case involved a dispute surrounding an aspect of the brand name 'McDonald's', owners of the multinational fast food chain. In this case the linguists were Genine Lentine and Roger Shuy (as reported in Levi, 1994: 5). Quality Inns International announced their intention of opening a chain of economy hotels to be called 'McSleep'. 'McDonald's' claimed that the attachment of the 'Mc' prefix to many unprotected nouns, such as 'Fries' in 'McFries' 'Nuggets' in 'McNuggets', etc., barred Quality Inns from use of the 'Mc' prefix. In this case the plaintiff was not just claiming implicit ownership of a name, but over a morphological principle, namely the attachment of a particular prefix to any noun.

It appears that the claim was inherently one of a 'formula for combination' (Levi, 1994: 5) and it was this formula for which protection was being invoked. 'McDonald's' also claimed that they had originated the process of attaching unprotected words to the 'Mc' prefix and had run advertising campaigns which illustrated this. In their evidence Lentine and Shuy showed that the 'Mc' prefix had had previous commercial applications, and that as 'McDonald's' had not objected to any of these they had no grounds for doing so in the present instance. Despite the overwhelming evidence presented by Lentine and Shuy, judgement was for the plaintiffs and Quality International Inns were unable to launch their chain of motels under the 'McSleep' banner.

In Australia linguists began meeting in the 1980s to talk about the application of linguistics and sociolinguistics to legal issues. They were concerned with the rights of individuals in the legal process, in particular difficulties faced by Aboriginal suspects when being questioned by police. They quickly realized that even such phrases as 'the same language' are open to question. An important instance of this is the dialect spoken by many Aboriginal people, known colloquially as 'Aboriginal English', wrongly thought by many white Australians to be a defective form of the English spoken by whites. It is in fact a dialect in its own right. Thus, when being questioned by police, Aboriginal people bring their own understanding and use of 'English' to the process, something which is not always appreciated by speakers of the dominant version of English, i.e. 'white English'. More than this they bring their own interactional, culturally based, styles to the interview. An individual's own interactional style, if perceived to be at variance to that of the dominant culture, might compel responses to questions in particular, non-confrontational ways which could lead to a false assumption on the part of a questioner that the suspect was being evasive or, worse still, that an admission of guilt was being made.

Other Australian research focused on how Aboriginal witnesses and defendants understood the legal processes involved in land claim hearings and examined the impact of cross-cultural differences between white settlers and Aboriginal people on the presentation and even outcome of cases. In this context Gibbons (1994: 198) observes 'the . . . system . . . around interrogation in the courtroom is alien to Aboriginal culture'. Gibbons is the author of two major books on Forensic Linguistics, *Language and the Law* (Longman, 1994) and *Forensic Linguistics: An Introduction to Language in the Justice System* (Blackwell, 2003). In these books he not only summarizes some of his own considerable experience as a Forensic Linguistics expert in the court system, but also details much of the history of the development of Forensic Linguistics.

Surprising as it may seem, the thread which connects many of these different forensic activities is *authorship*. Essentially, in considering the question of whether an individual dictated a statement, or whether a statement was in the words of its alleged speaker, analysts were actually asking the question 'Who was the author of the statement attributed to X?' This applied to the statements of Derek Bentley, Timothy Evans, the Birmingham Six, the Bridgewater Three in the United Kingdom (and many others), as well as to Australian Aboriginal defendants who claimed that police had 'verballed' them (i.e. altered what they had said).

Conditions of authorship in police interviews

In the case of US defendants whose Miranda rights were being investigated, there was a slightly different kind of authorship nexus of questions which included: 'Did the putative authors of statements (such as the 15-year-old Houston boy, or Mr Innis) make their statements voluntarily, knowingly and in full possession of their rights?' In other words, the issue here relates to the conditions of authorship: a series of questions put by police, for which answers are required, structures and even distorts a narrative of events; answers which appear vague, ambiguous or reluctantly given may slant a narrative in a way which is disadvantageous to the defendant and, as pertinently, to the apparent truth of the narrative. In any case, the assymetric nature of the relationship between authority figures (the police) and the defendant – who may be (i) illiterate (ii) a speaker of another language than the language of interview (iii) young/disabled/culturally disadvantaged, etc., can result in a text (such as a record of interview, video or audio recording or written statement) which is considerably at variance with what the suspect would have said had he/she been given the opportunity to make a statement in a non-coercive or less threatening environment.

Authorship and disadvantage

In a broad sense to be an author is to possess the language you are using. It is the use of language to produce a text over which you as the author have control, and whose course you are free to direct. Disadvantaged speakers are scarcely in control of the authorship process when 'giving' a statement to powerful authority figures. We can realistically challenge many texts produced under conditions of duress even where the duress may not have arisen through the intention of questioning police officers, court officials, or any other authority figure within the justice system. If a suspect's way of using language is at some remove from that of the officials with whom the suspect is dealing then the potential for distortion of the authorship process is clearly exacerbated, probably in proportion to the differences of perspective, interactional styles and cultural norms between the institutional, authority figures on the one hand and the suspect on the other. This is not to suggest any malice or intrinsic lack of fairness or

justice on the part of officials: they work within institutional structures which are not always the most conducive for taking individual circumstances into account.

The *meanings* of a word

In Germany, an early case involved an alleged slander by a tenant in an apartment complex of a fellow tenant (Kniffka, 1981). The issue at stake was whether the word *concubine* was an insult. Linguists advised that for some speakers the word might be amusing, for others a way of addressing each other as a joke, while yet others might find it insulting under some circumstances: it was not possible to say that a given word or phrase, on its own, was an insult, or constituted verbal injury. Rather, the relationship between speaker and hearer, the context of the situation, the speaker's education level – all needed to be taken into account. A word does not have a single, universally agreed, meaning within a speech community. Other issues which emerged in the early days of Forensic Linguistics in Germany involved authorship attribution, and the development of methodologies for doing so. An early case, reported by Kniffka, concerned the theses of twin sisters whose previous academic performance was, according to university authorities, at a much lower level than the work they submitted for their final examination. Kniffka argued that an authorship attribution in the case was not possible because the language used was essentially the meta-language of the thesis subject and that it was not easy to attribute such language to any given individual. He suggested the university authorities subject the students to a written examination on their theses to test their knowledge, rather than relying on subjective comparisons with their previous, known, work.

In the years since Forensic Linguistics began to establish itself as a discipline its scope has grown considerably. From its beginnings as a means of questioning witness and defendant statements, linguists have been called on to give evidence in many different types of case, including authorship attribution in terrorist cases, product contamination cases and suspicious deaths; the interpretation of meaning in legal and other documents, the analysis of mobile (cell) phone text messages to establish a time of death. The list continues to grow. In the next section, an important area within Forensic Linguistics will be considered: Forensic Phonetics, the analysis of speech through auditory and acoustic means and its application in the legal and criminal arena.

Forensic phonetics

Forensic Phonetics deals with questions of speaker identification, resolution of disputed content of recordings, the process of setting up voice line-ups and ear line-ups and related topics. It has a more established presence in the legal forum than Forensic Linguistics and its progress has been assisted by recent advances in acoustic engineering. Phoneticians are able to analyse the distinctive speech characteristics of a speaker relative to other candidate speakers in an inquiry much more easily than as little as 20 years ago. An important ethic within Forensic Phonetics is that no means exists which can infallibly identify an unknown speaker in a legal case (such as a hoax or bomb threat caller to an emergency service). Rather, like all branches of science Forensic Phonetics examines a set of phenomena, in this case aspects of recorded speech, and offers opinions based on the observations arising from the analysis.

Among the earliest British forensic phoneticians were John Baldwin, Stanley Ellis and Peter French, while in Germany Hermann Künzel was also active. Künzel (with Eysholdt) considered many aspects of speech production with reference to social situations, including the influence of alcohol on speech (Künzel and Eysholdt, 1992). Kniffka's (1990) collection contains accounts of some of the early forensic phonetic cases – see especially Ellis's and Baldwin's contributions to that collection. The earliest recorded voice identification testimony in the UK was in 1965, given by Stanley Ellis at Winchester Magistrates' Court. This book contains a separate chapter on forensic phonetics.

The meaning of legal words

What happens when words we use every day are also used as technical words by the law? The short answer is that life can become very difficult for those perceived to be on the wrong side of the law. The difficulty can be compounded when judges instruct juries using technical legal words which are also common, everyday words.

Stratman and Dahl highlighted some of these difficulties in their work on the language of temporary restraining orders, and the difficulties ordinary speakers may have in comprehending them. They cite a case where a man served with a temporary restraining order drove to his partner's apartment and slipped a letter under her door in order to elicit from her what the

problems in their relationship were and how they could address them. The court argued that he had violated the restraining order's injunction not to 'molest, interfere with or menace' his partner (Stratman and Dahl, 1996: 212). It was clear that the drafters of this particular law had a different conception of the words *molest, interfere with* and *menace*, since it is highly likely that most people would not ordinarily consider placing a letter under someone's door to be an act of molestation or menace on its own. In court judges often refer to dictionaries for the meaning of words which occur in legislation. However, this approach has been criticized. Generally speaking, linguists view dictionaries as imprecise and limited. Meanings are probably best not taken from a dictionary, but from experimentation and observation of how words are used. It is generally agreed that words have a core meaning and a number of 'fuzzy' meanings (see Goddard, 1996: 254). While the core meaning is probably well understood by 'most' people, it is as the word approaches the boundaries of its semantic envelope that difficulties arise. We would probably all agree that dogs, cats and hamsters are pets. But what if a prospective tenant in a block of apartments which allows pets were to bring a chicken or a crocodile and claim such an animal as a pet? How does this kind of meaning difficulty equate with interpretations of words in our previous example, such as *molest* and *menace*?

When common words are also technical terms

Thus, while the word *pet* encompasses a range of familiar domestic animals in the minds of most of us which may or may not exclude such creatures as a chicken or a crocodile, words like *molest* and *menace* have status as legal terminology. Though legal drafters are nowadays obliged to use words in a meaning as close to ordinary language as possible, words like these do present special problems, since they may have been used in legal language for hundreds of years in a more or less fixed fashion, yet in ordinary language their meaning will probably have changed considerably. Linguists have proposed a number of ways of dealing with this kind of difficulty, including carrying out semantic surveys. However, this approach has not found universal favour among linguists. For example, Solan notes: 'People cannot explain what, for instance, makes a snake a snake, a game, a vehicle, etc. Generic categorization

is a matter of induction and intuition, which we are rarely able to describe' (cited in Goddard, 1996: 259). Goddard notes that it would be absurd for forensic linguists to promote themselves as experts in the meaning of legal words, because this is really the province of judges. Using surveys, for example, to determine the meaning of a word, could produce contradictory results. He suggests that if linguists confine themselves to non-legal words, this may make more sense, but in any case, he points out that semantics is still a relatively underdeveloped area of linguistics and that there is still considerable disagreement among semanticians as to methodology. Corpus Linguistics has allowed the semantic survey approach to flourish because in the technological age it is easy to collect many samples of a word in its ordinary usage. However, an important competing process is that of semantic reduction (the 'reductive paraphrase' – Goddard, 1996: 269), which puts into practice Plato's dictum that a definition must use words which are simpler than the word which is being defined.

Judges and juries

Another area of potential conflict with regard to word-meaning is in the directions given by judges to juries. It has often been pointed out that such directions are full of legal terminology, some of which may be present in the language as ordinary, everyday words. How are jurors, by definition 'ordinary citizens', to understand whether a word is being used as a technical term or as an ordinary word, let alone understand the legal terminology? Would all the jurors in a case agree as to the meaning of a particular word? In recent years, in England and Wales at least, judges have received recommendations to illustrate their jury directions with visual presentations, to avoid giving directions about the law, and to keep reminding juries throughout the jury-direction phase what the issues in the case are. In the US in some jurisdictions, judges are now being trained in how to talk to juries. These are very welcome developments.

Linguists in court

When it comes to linguists giving evidence in court, it is clear that lawyers and linguists have different goals. The job of the lawyer is to convince or persuade

the jury that the defendant is guilty or innocent. The job of the linguist is to present an opinion and to explain that opinion. The lawyer may interrupt the expert witness, use rhetoric, 'spin' or even guile, and may choose to ignore anything the expert witness says. It is safe to say that it is not necessarily the case that the lawyer is intent on discovering or promoting the 'truth'. The linguist, on the other hand, is a witness and must observe the oath or affirmation which the court requires all witnesses to take, no matter what lawyerly stratagems may be used to suppress or distort that evidence. Lawyers can also play on the notion of 'cooperative'. The linguist will usually attempt to be cooperative, but linguist and lawyer may conflict about what *cooperative* means in practice in a given instance. This, again, will be due to the differing discourse practices of lawyers and linguists. The lawyer will in all probability bring a folk-semantic meaning to the idea of *cooperation*: 'Why aren't you cooperating with the court? After all, it's a *simple* question.' The linguist on the other hand, mindful that the lawyer is attempting to direct the discourse away from the evidence, struggles with the lawyer's notion of *cooperative*: to the linguist co-operation here means that lawyer and linguist *cooperate* to uncover the truth. However, I would not wish to suggest all lawyers use such stratagems – we should bear in mind that they have a difficult task to do and are judged on results. Nevertheless, lawyers and linguists do bring a different 'world view' to the courtroom.

Legal standards of expert evidence across the world

Finally in this section, it is important to consider some aspects of the different methods of admitting expert witnesses into courts, in particular linguists. In the US each state has its own rules of evidence, some of which will be applicable only to district courts, and some to higher courts. There are also Federal Rules of Evidence and these differ in kind from the evidence rules of lower courts. The rules governing expert evidence are complex and not always understood. They require that scientific evidence meets certain standards. Increasingly, the 'Daubert' standard is what is insisted upon. This requires, among other things, that witnesses demonstrate the known error rate attached to their opinion. This of course implies that the linguist must present quantifiable data. However, in linguistics it is not always possible to present quantifiable data, and it may

indeed be misleading to do so. Some courts have interpreted 'Daubert' more flexibly than this, and it is an ongoing debate in legal and linguistic circles, with some insisting that any authorship attribution analysis must be backed up by the use of inferential statistics, which is the only way to demonstrate a known error rate in a particular case. However, contrary to popular belief there is in reality no such thing as a 'linguistic fingerprint' and it is not always possible to quantify a view that a particular individual is the author of a questioned text in a case.

In other countries it is sufficient that the method on which the expert bases an opinion should be acceptable to the scientific community, and that the expert should be qualified to give it. Both Canada and Sri Lanka, for example, follow this method of accrediting a witness and accepting an opinion.

In the near and medium term future it is likely that the question of how linguists verify their opinions will be given a great deal of attention. Some have argued that linguists have inhabited the ivory tower of academia for too long. For this reason moving into the rough-and-tumble arena of the law, where they are required with great rigour to justify what they do, say and believe, has been a culture shock for many. Understanding of this culture is critical: some would argue that it is not productive to describe the law as alien or hostile to the linguistic viewpoint. The law is blind: it has no favourites, nor should it. It is surely necessary for linguists to accept this culture and adapt to it, while remaining true to their discipline.

In this section an attempt was made to illustrate some of the issues linguists face when interacting with the legal system. It is now seen as imperative among linguists that both they and legal professionals work towards a better understanding of each other's perspective. If linguists claim that lawyers are ignorant of linguistics, then it is up to linguists to ensure that this situation does not continue. Lawyers can equally claim that linguists are ignorant of the law and it is certainly up to linguists to ensure that this gap in their knowledge is addressed as a matter of some priority.

It will also be important for linguists, in this age of international courts, to understand the discourse practices of international law, and to familiarize themselves with the customs and mores of other countries' legal systems, as Forensic Linguistics moves into a new millennium and an uncertain terrain in a world of organized crime, international terrorism and human rights abuses in many countries. It is likely in the future that increasing numbers of those seeking to enter the field of Forensic Linguistics will have additional qualifications in areas such as the law and mathematics and statistics, and

gain greater understanding of scientific techniques, methods and presentation. Many universities are already equipping their undergraduates with some of this information. With a broad but accurate insight into the law and an appreciation of how science is 'done' in other fields than their own, forensic linguists of the future will have greater means at their disposal than the founders of the discipline.

Previous authorship studies

2

Chapter Outline

Biblical *authorship* disputes

Authorship controversies do not seem to have existed prior to the eighteenth century, quite possibly because up until then books had not been so widely distributed, and the Cartesian concept of individuality and the individual ownership of ideas had yet to receive general acceptance. The earliest known controversy related to the authorship of the Bible, and was voiced by a German priest, H. B. Witter, who in 1711 pointed out that the different names for divinity in the Pentateuch could indicate that several authors had contributed to it. Jean Astruc, a French medical doctor arrived at a similar conclusion later in the century, while J. G. Eichhorn, a Lutheran professor at the University of Jena, reached this view about 100 years after Witter (Eichhorn, 1812). Interest among scholars in biblical authorship questions intensified with the rise of Darwinism and related theories and continues to the present day.

Shakespearian authorship disputes

Some time after Witter but before Eichhorn, according to Olivia Hill (online[2]) the earliest reference to the Shakespeare controversy occurred. In 1785 the Rev James Wilmot wrote that Bacon was the real author of the Shakespeare plays, but did not publish his ideas as he had no appetite for contention. For this reason the controversy did not really surface until James Corton Cowell first presented it to his local philosophical society in Ipswich, England in 1805. However, it was not until almost half a century later that the interest in authorship questions began to be pursued scientifically.

Early scientific and statistical approaches

Among those who expressed an early interest was Augustus De Morgan, the first professor of mathematics at University College, London. In 1851 he wrote a letter to his friend the Rev Heald at Cambridge, suggesting that Heald, who had expressed an interest in the question of the authorship of biblical letters attributed to St Paul, should

> ... count a large number of words in Herodotus, say all of the first book, and count all the letters; divide the second number by the first, giving the average number of letters per word in that book. Do the same in the second book. I should expect a very close approximation. If Book 1 gave 5.624 letters per word, it would not surprise me if Book 2 gave 5.619. But I should not wonder if the same result applied to two books of Thuycidides gave, say 5.713 and 5.728. That is to say I should expect the slight difference between one writer and the other to be well maintained against each other, and very well agreeing with themselves. If this fact were established, then, if St Paul's Epistles which began with ... gave 5.428 and the Hebrews gave 5.516 for example [implying a difference of just below 2 per cent as evidence of a difference in authorship], I should feel quite sure that the Greek of the Hebrews (passing no verdict on whether or not Paul wrote in Hebrew and others translated) was not from the pen of Paul ...
> (De Morgan, 1882: 215–6)

Although De Morgan appears to have been the first to express these ideas, it is not known whether he studied the topic in any greater depth than this. Then,

30 years later, a professor of physics and mechanics at Ohio State University, T. C. Mendenhall, published his thoughts on authorship of text, primarily using word length average as a marker (Mendenhall, 1887). He published a second article some years later (Mendenhall, 1901), but does not appear to have taken any further interest in the subject of statistical measurement of authorship.

Later statistical studies

Udney Yule – the well-known Cambridge theoretical statistician – had been interested in the statistics of what we now term genetic engineering and the work of Mendel in particular before becoming interested in authorship questions. He saw *mean sentence length* (Yule, 1938, 1944) as a viable marker. However, neither his approach, nor that of any of the other scientists of the time, resulted in anything remotely resembling the ability to 'fingerprint' authors, although each of the values suggested – e.g. average word length, average sentence length, etc. – has proved useful in different kinds of language testing. Other statisticians have since shown great interest in authorship attribution, including the analysis by Mosteller and Wallace (1964) of the Federalist papers, and Kenny (1982). However, one of the disadvantages with pure statisticians approaching this issue is that generally they do not bring any knowledge of language with them to the problem. If Yule, for example, had been a linguist, he would have been aware there are significant differences between sentence length in speech from that of writing. To be more precise, it is probably the case that mean sentence length reduces in proportion to formality. Sentence length, just like any other measure, is therefore dependent on genre and register as much (if not more) than it is dependent on the individual writer/speaker.

Text metrics and statistical techniques used in authorship identification

In addition to measures described in the previous section, authorship measures in use by a number of analysts have included word length average, average number of syllables per word, article (determiner) frequency, type-token ratio, punctuation (both in terms of overall density and in terms of syntactic boundaries) and the measurement of hapaxlegomena (unique words

in a text). For a full account of metrics used and statistical techniques see Holmes (1985), Forsyth and Holmes (1996) and Chaski (2005).

Statistical approaches have included factor analysis, Bayesian statistics (Mosteller and Wallace, 1964), Poisson distributions (Särndal, 1967), multivariate analysis (Burrows, 1987), discriminant function analysis of function words (Baayen *et al.*, 1996). More recently Chaski has used discriminant function analysis of syntactic boundaries and structures as well as a range of other techniques (Chaski, 1998, 2001b, 2005).

Authorship and the birth of forensic linguistics

Almost simultaneous with the rise in computational hardware – but conceptually quite separate – the discipline of forensic linguistics came into being with the publication by Svartvik of his classic study into the police statements in the Evans case (Svartvik, 1968) some 15 years after Timothy John Evans had been hanged. John Christie had murdered a succession of women at his home in 10 Rillington Place, west London, and had duped Evans into believing he, Evans, was partly responsible for the death of his own wife and child which had actually taken place at Christie's hands. Evans had gone to his uncle's home in Merthyr Tydfil, south Wales, but after a short period he had handed himself into the police and confessed to having killed his wife. Two statements were given at Merthyr Tydfil and two at Notting Hill police station, London, where Evans was taken on the day following his arrest. In his analysis Svartvik demonstrated the presence of two very different registers in the Evans' statements. However, this case was not an authorship study as such: its importance lies in Svartvik's pioneering technique in analysing textual alteration, and his name for the new science, *forensic linguistics.*

Some years later Coulthard (1994) undertook a groundbreaking linguistic study of the alleged statement of Derek Bentley who, like Evans, had been hanged in the 1950s. Coulthard's analysis of this statement was a major factor in Bentley's posthumous pardon, but whereas Svartvik's investigation was primarily a statistical study, Coulthard's is more of a general descriptive analysis. Coulthard refers to his approach as *forensic discourse analysis,* but it can also be seen as a combination of insights from different linguistic fields including speech act theory, corpus linguistics, register, and even psycholinguistics.

Cusum Plot of short text by author
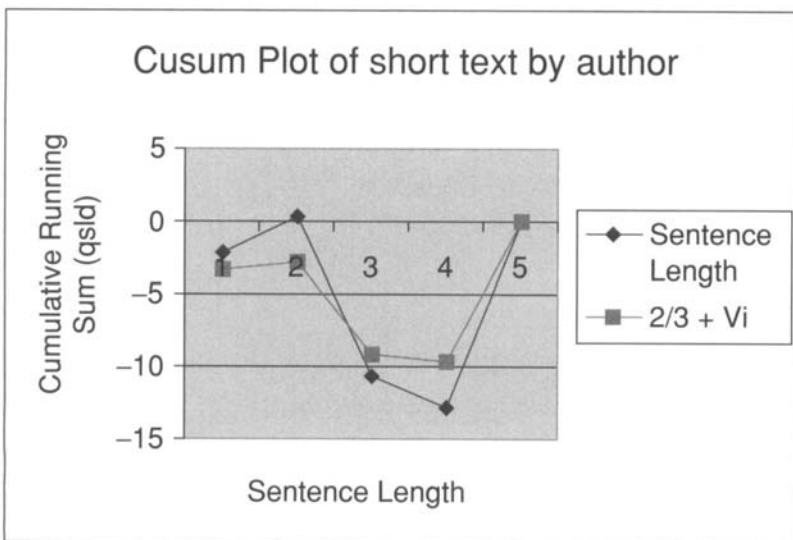
Chart 2.1 A Cusum Plot of an earlier version of the preceding three paragraphs

Morton: Cusum analysis

In marked contrast, Morton has proposed the analysis of text using the Cusum (abbreviation for cumulative sum) method (Morton and Michaelson, 1990). Morton's method is interesting because Morton himself admits to having no idea why Cusum (or Qsum) works. Although a biblical scholar of some reputation, Morton makes no claims to any linguistic knowledge. However, he does make high claims for his method, the chief of which is that Cusum always works, even on short texts. According to Morton speakers' habits do not vary significantly, and there is no difference between speech and writing in terms of an individual's habits.

The assumption behind Cusum analysis is that in using language each speaker exhibits a set of unique habits, and that these habits form statistically identifiable patterns in the text. One such habit, according to Morton and Michaelson (*ibid.*), is the number of two- and three-letter words in a sentence. Another is vowel-initial words. To carry out a Cusum test on habits such as the use of vowel-initial words and 2–3 letter words, you would first identify all occurrences of each type of word in the text, and then plot their distribution in each sentence. You would compare the Cusum distribution for these habits with the average sentence length for the text. These two sets of values should track each other. Certainly this does seem, from Chart 2.1, to be the
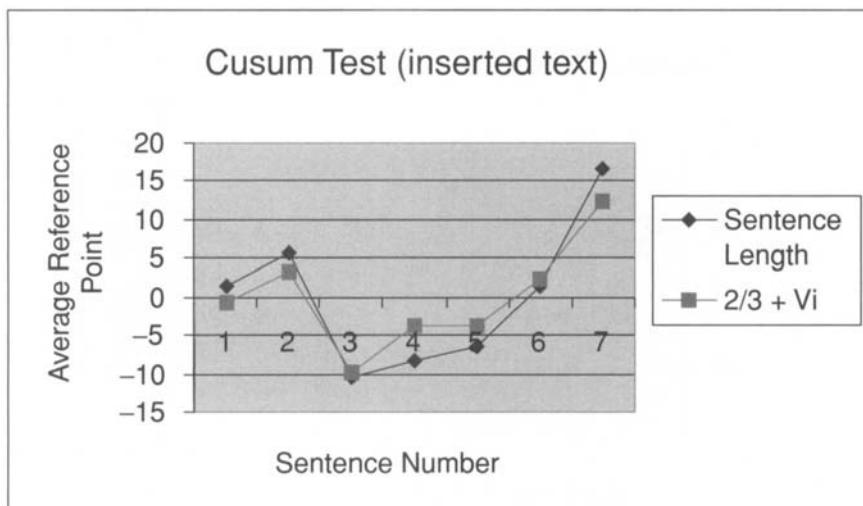
Chart 2.2 A Cusum Plot of the same text as the previous chart, but with an inserted section

case. According to Morton and Michaelson (1990) and Farringdon (1996), any altered section of the text would exhibit a different pattern from the rest of the text.

In fact this is exactly what Chart 2.2 seems, at first, to show: text from another author was inserted after sentence 2. Looking at Chart 2.2, we do seem to see some discrepancy between the values of the two reference points between sentences 3–5, which does indeed appear to be larger than the differences seen elsewhere in the two plots. However, the key question is this: just exactly what constitutes a discrepancy? How do we judge when we have a good fit? Clearly, interpretation of such results will tend to be subjective. Schils and de Haan (1995) expressed other concerns about Cusum, namely that there is 'considerable intra-author variation', which would preclude a reliable basis for inter-author discrimination, while Sanford, Aked *et al.* (1994) found that the Cusum technique was 'based on assumptions which are at best of limited reliability, and are most likely completely false'. Canter and Chester (1997) conducted a detailed evaluation of a revised Cusum technique and found that 'the weighted Cusum technique does not reliably discriminate between single- and multiple-authored texts'. Methods which claim to be infallible, and those which blatantly run counter to linguistic knowledge and experience should be treated with scepticism. The Cusum technique falls into both of these categories. In the next section I am going to take a brief look at statistics – at least a more

conventional use of statistics in forensic linguistics than that discussed in this section.

Computational linguistics and forensic authorship attribution

From the mid 1980s onwards, powerful computers became more widely available and the science of *computational linguistics* was developed. The field is really a partnership between computational studies and linguistics.

In computational linguistics the goal of authorship research is essentially that of finding authorship algorithms. For a few computational linguists, thankfully not all, it is axiomatic that individual style can be attributed to the level of precision we associate with the science of fingerprinting. However, the main difficulty with computational approaches is the amount of data computational linguists are used to working with: for the computational linguist a text of 5,000 words is a short text, usually within a corpus of ten or more texts. The forensic linguist, however, may have to work with no more than two or three texts of perhaps no more than a few hundred words in length. It is for this reason that I prefer nowadays to see forensic authorship work in terms of author comparison rather than attribution. Computational linguists are becoming more aware of forensic requirements, but the gap is still wide. Moreover, forensic linguists do not, in general, see the goal of automatic authorship attribution as a viable forensic aim. In fact, in the context of forensic work, any automatic procedure must be considered positively dangerous. In the end the forensic linguist must deliver an opinion, and the court must make the final decision. Even DNA scientists do not rely entirely on automatic procedures, but use their judgement and offer courts an opinion.

Chapter conclusion

In this chapter I have given a very brief outline of some current authorship studies, including a description of methodology where appropriate. However, even a brief visit to the internet, or to your university or faculty library, will show you that there are literally thousands of references to this topic, and that my choice is of necessity very limited, perhaps narrowly reflecting those areas

I have found to be of interest in my own day-to-day forensic researches. One thing, however, is certain and that is that the debate as to what constitutes an appropriate approach will continue. Part of this debate centres around the question of whether forensic authorship work is a science or an art. I believe it is a science, but not just a quantitatively based one. The linguist observes and describes, then makes matches or states that the compared items are not a match.

Forensic authorship is only in its infancy, and is still capable of being shaped and influenced in major ways. There is, therefore, an exciting opportunity awaiting today's students to be involved in the genesis of a whole new approach to linguistic evidence.

Individuals and language use

3

In this brief chapter I will consider some of the arguments relating to the notion of the linguistic fingerprint, or individual style. First, I will consider the prerequisites for style, mainly to do with acquisition, education and culture. Then I will look at style itself, what it is, its relationship to genre, and its status as a psychological construct. Finally, I will consider the question of individual style from a practical point of view, as applied to forensic linguistics.

What is the linguistic fingerprint?

The *linguistic fingerprint* is a notion put forward by some scholars that each human being uses language differently, and that this difference between people can be observed just as easily and surely as a fingerprint. According to this view, the linguistic fingerprint is the collection of markers which stamps a speaker/writer as unique. However, although the linguistic fingerprint is a powerful concept with great attractions for law enforcement and other agencies worldwide, there is so far little hard evidence to support the notion.

Despite this, there are nevertheless innumerable references to a 'linguistic fingerprint'. These are claims that language style is a result of unconscious forces and that these form some kind of 'linguistic fingerprint'. For example, Pennebaker (1999) claims that using 'new behavior/language sampling technology for field research', he and his group will 'reconstruct a person's "linguistic fingerprint" from his or her daily interactions and relate it to a variety of self-reported personality characteristics, situational variables and physiological markers (e.g., blood pressure, cortisol, testosterone)...' See also, Bax (online) who refers to an undergraduate course which would 'describe aspects of their [Johnson, Boswell, etc.] language use (e.g. grammatical variation, idiosyncratic spelling systems, vocabulary) in an attempt to discover what could be called their "linguistic fingerprint"'.

As the reader will see from the above, the notion that individuals have a *linguistic fingerprint* is now widely disseminated, not only in paper journals and other publications, but extensively across the internet, and not only by academia, but also by a number of commercial companies anxious to cash in on the current widespread interest in the forensic sciences, often claiming to be able to offer miracle solutions to 'problems' of author identification.

The difficulty I have with the above claims is simple: nobody has yet demonstrated the existence of such a thing as a linguistic fingerprint: how then can people write about it in this unexamined, regurgitated way, as though it were an established fact of forensic life?

Perhaps it is this word 'forensic' that is responsible. The very fact that it collocates so regularly with words like *expert* and *science* means that it cannot but raise expectations. In our minds we associate it with the ability to single out the perpetrator from the crowd to a high degree of precision, and so when we put *forensic* next to *linguistics* as in the title of this book we are effectively saying *forensic linguistics* is a genuine science just like *forensic chemistry, forensic toxicology*, and so on. Of course, insofar as a *science* is a field of endeavour in which we seek to obtain reliable, even predictable results, by the application of a methodology, then forensic linguistics is a science. However, we should avoid giving the impression that there are easy, wholly automatic ways of identifying an author out of a crowd of authors. In the next several sections I will demonstrate why, in my view, the notion of linguistic fingerprint is essentially flawed. However, this chapter is not the complete story – in the next chapter I will look at a number of arguments surrounding the question of author variation, both within an author and across several authors.

Language acquisition

Language is an acquired property, not an inherited one. It is not like an individual's fingerprints or DNA. Moreover, language is acquired, not all at once, but in stages. Individuals continue acquiring language throughout their lives, although the core development is generally complete by 30 months or so (Radford, 1990). The reverse of acquisition and development is atrophy: language is not only acquired, it is also lost. At no time in an individual's life is language 'fixed'. In that case, if language is not a fixed property, how is style to be measured? Aside from the variation exhibited by the individual, how often does an individual's use of language change, and to what extent? What makes it change? Of course this is aside from diachronic change in language generally (see Lightfoot, 1979).

Universal education and linguistic homogeneity

Several hundred years ago we may have been able to distinguish most individuals from each other through their use of language, because in those days dialectal variation was significantly greater than now. There was little standardization.[3] Those who were able to write frequently invented their own spelling and grammar rules. However, with the advent of universal education, cheap books, rail and road networks, air travel, and finally mass paper and electronic media, those who speak the same language have started to use language more like each other in the way they speak and write than at any previous time. For example, aside from minor lexical differences which are not always easily observed, and a few orthographic variations, there is frequently little to distinguish English language native speaker writers from each other: we do not always know if the writer is Caribbean, Canadian, South African, Australian, British, etc. unless perhaps specific cultural issues are being dealt with.

Moreover, frequent linguistic intercourse across cultural and national boundaries has an homogenizing effect on language: we become like each other in the way we use language in order to communicate better. Nowadays, we are all of us virtually using a language product which is being handed down by the combined forces of dictionary makers, pulp fiction writers, schools, universities, the media, and – not least – the political establishment (each with

its fair share of prescriptivists). If the decline in the number of languages being spoken across the world is anything to go by, we may observe that linguistic uniformity is in direct proportion to the degree of mass human intercourse. As Crystal (online) notes: 'In the 19th century, there were more than 1,000 Indian languages in Brazil, many spoken in small, isolated villages in the rain forest; today there are a mere 200, most of which have never been written down or recorded.' In Africa, the United Nations estimate that more than half of the 1,400 indigenous languages are either in decline or under threat, largely due to the influence of English and other colonial languages, while other major languages, such as Russian and Chinese have probably contributed to similar linguistic reduction in their own countries. I believe that standardization within a language, such as English, and the decline in the number of languages, due to the power of the world's major languages (English, Chinese, Spanish, Arabic, Russian, French, German, Portuguese) is part of the same process of linguistic homogenization facing the world at this time, and may contribute to *reducing* linguistic individuality. There is little room, then, for believing that humans use language idiosyncratically. Finally, it should be said that the difficulties in distinguishing one author/speaker from another are increasing, not diminishing. Technology is not solving the problem of fingerprinting authors, it is making the task more difficult: as the availability of information explodes exponentially, writers have access to more and more information, and in many cases are free to plagiarize from each other, almost with impunity.[4]

Section conclusion

In this section I have tried to outline my reasons against the notion of style being an individual property. The reasons I have put forward relate to acquisition, culture and education, as well as internationalization and globalization. In the next section I will discuss style in more detail, what it is, and its possible role in authorship questions.

Towards a theory of style

Style must be one of the most widely abused words in forensic and literary linguistics. To some it is a way of using language, a set of mannerisms which fulfil a certain formal or defined communicative need. To others it is an

individual's marked way of using language habitually, something which is special to that individual, also called *idiolect*, a term we take to mean the language of an individual as opposed to the language of a society, i.e. *sociolect* Bilgrami (1993: 66–73). Properly speaking idiolect refers to speech, but has come to include all forms of language. Idiolect and sociolect are – as I have argued in previous sections – not mutually exclusive: your idiolect is 'part of' your sociolect. Your sociolect changes as your idiolect changes. There is, in fact, a mutual dependency between the two.

There are two diametrically opposed views on style: (i) it is, as suggested above, a collection of markers consciously chosen by a speaker, a set of markers which can be observed and measured, or (ii) it is a set of unconscious habits not normally observed by the author, but which – once uncovered by the linguist – can also be observed and measured.

These two opposing views place the linguist in something of a difficulty: if on the one hand the set of markers used by the speaker can be observed, then the speaker can alter them and anyone else can imitate them. Thus, if a style can be imitated by anyone its potential for remaining unique is surely reduced.

On the other hand if the speaker's habits are unconscious then we are in difficulty because we do not know whether the set of 'unconscious' markers differs from author to author, or whether it is a fixed set across all authors. This is similar to the habits/choices debate. According to Morton and Michaelson, 1990 (see previous chapter), language idiosyncrasy comes about because we each have a set of (unconscious) habits, while according to others – including Ephratt – linguistic individuality exists because we as writers/speakers exercise linguistic choice.

If the set of markers differs from author to author, then it is impossible to know in advance what to look for, and – since each individual (according to this view) would be different – the researcher would be able to decide arbitrarily what to test for. This could produce subjective results. Conversely, if the set of style markers were fixed across all authors, this would imply some kind of predetermined relationship between language and the individual. This would lead us to wonder how such a relationship could be acquired, and to ask what there is about *style* to enable it to be fixed.

There are several important points to be noted about style markers. First, to measure unconscious style markers meaningfully, you need a great deal of text – such as a full-length novel, or hundreds of short texts. On the other hand, the fact that we can observe certain style markers tells us that they are

open to imitation – unless we are able to demonstrate that there is some kind of systematic or structural link between them. I discuss this in the chapter on authorship, where I cite several actual forensic cases.

Genre and style

Genre is a type of text arising from a specific set of communicative requirements. It can be taken in the narrow sense to refer to the work of a community of language users to whom text contributions are a kind of badge of membership (Swales, 1990) or, in the broad sense, a work of a particular type which has certain recognizable characteristics, e.g. crime genres, romance genres, the private letter genre, etc.

Whichever definition of genre we use, though, it is impossible to avoid the realization that there is a conflict between the notions of style and genre, which comes about because if the style of an individual is a property which is separable from the style of all other individuals, then no matter what genre the speaker writes in style will be constant. On the other hand, each genre has its own demands, one may almost say its own 'style': and if so, surely the style of the individual will not remain intact when it comes into contact with the 'style' of the genre? In that case it is logical to doubt whether style is a property of the individual. It may in reality be a property of the genre. That is the essence of the conflict between style and genre. However, we can usefully talk about *generic style* and *personal style* – as long as we realize the descriptive limitations of these terms.

We may consider a kind of compromise to the above, namely that style may be partly conscious and partly unconscious – some features could be within our control, while others may not be; similarly it may be the case that yet other aspects of style relate to or stem from the genre rather than the individual. In other words that in reality *style* – the way an individual uses language – is a complex, fluid interaction of the individual, language and society. In that case, it is certainly not a fixed property, and so cannot be used as a means of identification. This returns us, in a sense, to square one.

Whichever view of *style* we take it seems certain that style is not a primitive; it is a multifaceted construct which appears to be beyond the complete control of the individual and is at least in part dependent on genre and context. It is important to realize that language is not a personal possession. We use it in common with all other speakers of the same language as ourselves. It is,

essentially, the property of its community of its users, or rather of the communities of its users. It is a social property, not an individual one.

The 'linguistic fingerprint': a practical approach

At the practical level in forensic linguistics, there is probably little need in most authorship investigations to think in terms of unique style. It is simply not necessary. Rather, when conducting an investigation, the emphasis should probably be upon the relative differences between the candidate authors and how we can classify their texts. Hence, insofar as it is necessary to think of style at all, it may be better to focus on *distinctive* rather than *unique* style. Distinctive style is the style of a set of texts within an inquiry where there are several sets of texts. As I have suggested above, it is a *relative* term rather than an absolute one. As an investigative goal it is more useful than attempts to prove each individual in the world has a unique way of using language.

Chapter conclusion

The intention of this chapter has been to pose background questions relating to the existence of a unique language style in each indivividual. The concept of uniqueness and the individual is dear to the human psyche. We are indoctrinated from an early age in the belief that we are each unique. This is meant also to imply that we are all 'special', 'wonderful', etc. In fact our belief in our uniqueness is part of the western creed. It is reinforced by such observations as the uniqueness of the individual's fingerprints, the individual's DNA, and the apparent uniqueness of other kinds of biometric data. Somehow we have the same expectation when it comes to language: we expect that the language of the individual should be unique, we want to believe that we each use language uniquely, that it is part of our individuality. It is surely no accident that the third most frequent collocate of the word *unique* is in fact *style*, according to Cobuild English Collocations (1995).

However, the evidence to support the belief that individuals possess a unique style of language needs careful examination, and we should be aware of the pitfalls, especially in an age of mass standardization in education. As both

language and education become more regulated in industrial societies across the globe, individuality in language use may be more and more difficult to detect. As I have suggested above, the most practical option for the forensic linguist is to look for ways of distinguishing individuals from each other rather than attempting to establish the uniqueness of each individual. In the next chapter I will discuss author variation – within authors and between authors. In the chapter after that I will consider aspects of markedness theory applied to authorship comparison.

Variation $\boxed{4}$

In this chapter I will be discussing the notion of *intra*-author variation – how and why an author's texts vary from each other. *Intra*-author variation is to be distinguished from *inter*-author variation.

> **Intra-*author variation*:** the ways in which one author's texts differ from each other.
>
> **Inter-*author variation*:** the ways in which different authors vary from each other.

We should begin by considering a number of different causes of *intra*-author variation.

Genre as a cause of variation: Just because two texts are by one author does not mean they will have less variation than two texts by different authors. Charles Dickens wrote in several different genres, including rags-to-riches stories, crime mysteries and children's stories. When we measure texts in these different genres we see that they exhibit considerable variation, even though they are by the same author.

Text type: Personal letters are likely to involve a less formal lexicon and will contain more interrelationship bonding strategies than, say a term paper or academic article, even if both are written by the same person. On the other hand, several writers working in the same genre and text type are likely to output very similar (at least superficially) types of text. A good example is the *Federalist Papers*, where three leading politicians in eighteenth-century America, James Madison, Alexander Hamilton and John Jay all wrote newspaper essays on impending constitutional changes. On first reading any of the papers, it would be difficult for even experienced readers to observe differences between Madison, Hamilton or Jay. This is because they all essentially had the same message, were publishing in very similar journals, and were consciously using the devices of the literary political essay of the time. All of the articles were published under the single pen-name, *Publius*, and each of the writers referred to previous papers in the series as though there were only one author of all of them. All of these factors contributed to low inter-author variation.

Fiction vs *non-fiction*: Some fiction writers, for example, are also journalists. These media are totally different from each other and will cause *intra*-author variation because of the different demands of each medium.

Private vs *public texts*: A politician writing to a friend or family member will undoubtedly use a different lexicon from that used in a political speech. The letter is a private text, the speech is public.

Time lapse as a cause of variation: The greater the time lapse between two works, the greater the likely variation. This is partly to do with our susceptibility to language changes around us, partly because language changes a great deal more than we realize in a relatively short space of time (e.g. 10 or 20 years) and partly because of the human ageing process, as well as changes in our health, circumstances, etc. (see below).

Disguise as a source of variation: A writer may choose to publish anonymously, and so wish to disguise output to prevent recognition. This can apply as much in the forensic context as in literature, politics, journalism, etc.

Change in circumstances: A writer's change in health, domicile, domestic status, bereavement, etc. can cause variation.

Sociometric parameters: Factors such as register, relationship with addressee, gender/s of writer and addressee, age, educational level, social class, social circumstances, context – to the extent that any of these variables differ across two texts, they will contribute to the degree of variation found between those texts.

Authorship markers

A wide variety of measures are used by analysts as authorship markers. Among many possibilities are function words. Certainly, in long texts they can be indicative of authorship. However, even in long texts – which are not the subject of this book, given that few forensic texts *are* long – there can be pitfalls in using function words as an authorship marker. There are very good reasons for this which I will explain below, using the word *the* as an example of a function word.

The word *the*

The word *the* is the most commonly used word in the English language. On average *the* occurs once in every 20 words that any of us, as native speakers of English, speak or write. However, the frequency of *the* varies with text type: for example, in emails and other forms of personal correspondence, *the* is much rarer than it is in, say, news reports. As a raw indicator of authorship across text types, therefore, it is completely without merit. However, if we control for genre and text type, it does – under certain circumstances – prove useful as an authorship marker in long texts.

Not surprisingly many authorship scholars, few of whom work in a forensic setting, have tested the usefulness of *the* for attribution purposes. Perhaps the classic paper on this subject is Burrows (1987) who, in a well-designed multivariate[5] study, measured just five function words in Jane Austen's novels: *and*, *of*, *the*, *a(n)* and *to*. He found that Jane Austen was very stable in her measurements of these words and could be distinguished from other writers using just these words.

I decided to test a range of authors for their consistency of use with regard to *the*. Sixteen novels by four novelists were measured. The following results were obtained:

Table 4.1 Measurement of *the* frequencies across a range of novels

Occurrence	Length	Text	Author	Frequency
4,105	119,606	Sense	Austen	0.034
4,330	121,864	Pr & Prej	Austen	0.036
6,209	159,922	MansfdPR	Austen	0.039
9,597	157,909	OlTwst	Dickens	0.061
12,794	255,667	BarnRge	Dickens	0.050
1,573	28,619	XmasCarl	Dickens	0.055

(cont.)

Table 4.1 (*cont.*)

Occurrence	Length	Text	Author	Frequency
13,704	359,381	DavCop	Dickens	0.038
14,969	355,568	BleakH	Dickens	0.042
6,806	131,284	PrBlEyes	Hardy	0.052
7,786	142,504	Return	Hardy	0.055
7,046	114,509	TrmptMjr	Hardy	0.062
6,775	116,425	Mayor	Hardy	0.058
8,723	207,814	Mill	Eliot	0.042
12,588	318,397	MMarch	Eliot	0.040
3,655	71,474	SilasM	Eliot	0.051
10,040	215,255	AdamB	Eliot	0.047

The column to look at for instant information is the far right column, where frequency of 'the' in each of the texts is given. Thus in Jane Austen's *Sense and Sensibility* we see the number '0.034'. This simply means that in that novel 3.4% of all of the words was the word 'the'. By contrast, if we look further down the table we see that in Thomas Hardy's *The Trumpet Major*, the 'the' frequency is '0.062', 6.2%. This is much higher than that of any of Jane Austen's novels. From the above table, averages and sample standard deviations were calculated as follows:

Table 4.2 Mean frequencies and deviations for *the* by author

Author	Mean frequency of the	Deviation
Austen	0.036	0.002
Hardy	0.057	0.004
Eliot	0.045	0.005
Dickens	0.049	0.009

As can be seen from the above tables, Jane Austen does indeed exhibit a very stable measurement for *the*. The other authors, however, seem less stable in this regard. I wondered why this should be so. In order to see if I could discover the reason, I devised another test, this time not related to function words. I calculated the 30 most frequent lexical words in each text, and then compared each text with every other text to see whether authors would use more or less the same high frequency words in one novel as they would in another. I particularly wanted to test to what extent an author's high-frequency word list would vary

across his/her novels. In other words, how much variation did authors show? The results were striking. Consider the table below:

Table 4.3 Mean variation rates for four authors (HFL – high-frequency lexicon)

Austen	0.19
Hardy	0.26
Eliot	0.32
Dickens	0.38

Of the four authors, we see that Jane Austen again exhibits the least variation (or most stability). More interestingly, the four authors rank in the same order for variation as they do with deviation of *the*. Austen shows the least *the* deviation, and the least high-frequency word variation; Hardy is the next highest with regard to both measures, then Eliot and then Dickens. I found that using just the HFL (high-frequency lexicon) I could discriminate the works of the four authors from each other (i.e. 'attribute' authorship correctly). Austen could be attributed 100% of the time, Hardy 94.4% of the time, Eliot 94.1% of the time and Dickens 89.1% of the time. Why should some authors be attributable more consistently than others? There are several answers. First, Jane Austen, aside from her foray into Gothic writing and her unfinished novel, *Sanditon*, was what may be termed a single-genre author. This would tend to narrow the range of the high-frequency lexis. Dickens, by contrast, wrote in a range of genres, as did both Hardy and Eliot, though to a lesser extent. This is one explanation for the varying degrees of attributability of each author. There is, however, a more tangible reason: time. If we look across a range of novels by each of the authors we see that the average time period between their novels varies in precisely the same order as their stability with the use of *the*, their degree of variation of high-frequency lexis and hence their degree of attributability, as the table below shows:

Table 4.4 Years between each author's book and his/her every other book

Text	Author	Score	Years
Pr&Prej	Austen	24.5	1.5
Sense	Austen	24.5	2.5
MansfdPk	Austen	24	2
			(cont.)

Table 4.4 (cont.)

Text	Author	Score	Years
FarMadCr	Hardy	22.6	8.4
Mayor	Hardy	22.6	9
Return	Hardy	22.4	7.3
TrmptMjr	Hardy	21.9	7.3
PrBlEyes	Hardy	21.6	6
Mill	Eliot	21.6	9
MMarch	Eliot	20	10
SilasM	Eliot	19.6	6
AdamB	Eliot	19.6	6.4
BleakH	Dickens	19.5	6.75
BarnRge	Dickens	19.5	9.75
DavCop	Dickens	18.3	7.5
OlTwst	Dickens	17.8	9

Summarized, the above table yields the following average time lapse between each novelist's works:

Table 4.5 Mean time lapse between each author's works

Author	Mean Time lapse (years)
Austen	2
Hardy	7.6
Eliot	7.85
Dickens	8.25

Again we see the same order as before, with Austen exhibiting the lowest mean time lapse between her novels, then Hardy, Eliot and Dickens (in that order). Could this be the main reason for the relative ease with which some authors can be attributed and, conversely, the reason for the difficulty of attributing other authors? For example, it is very difficult to find ways of distinguishing, by automatic means, some Hardy works from some Eliot works.

Variation is caused by a number of factors such as topic or genre, text type, changes in an author's circumstances, etc. However, other factors aside, the most important – at least for long texts – seems to be time. In short texts, however, different considerations apply. As mentioned before, if we consider

a complete text as an item of language behaviour, it seems reasonable that a very short text will only give us a brief glimpse into behaviour, while a long text will reveal much more. Measurement of function words, which seems to be a successful indicator of individual variation (especially when we control for time) shows that with short texts there is little stability.

For example, in one well-known case the issue was whether an alleged suicide note had been dictated by the deceased's husband or the deceased herself. In the following table *the* frequencies for each of the texts are given:

Table 4.6 *the* density for the Gilfoyle texts

Text	the *frequency*	Text length	the *density*
E1	10	141	0.071
E2	5	232	0.022
E3	0	147	0.000
E4	12	418	0.029
P1	4	251	0.016
P2	11	554	0.020
P3	6	233	0.026
Q1	8	408	0.020
Q2	18	610	0.030

From this table we can see that there is little stability for *the* across these short texts. The text marked 'E2', for example, despite being 147 words in length, records no uses of *the*. This is not as rare as one might think and is also found with other common function words. Short texts are notoriously fickle when it comes to stability in function word densities. As an authorship test for short texts, therefore, I would suggest that the measurement of function words is completely unsuitable: they are simply too susceptible to variation.

Variation: general principles

The preceding sections are intended to illustrate the following general principles relating to *intra*-author variation:

1. All authors exhibit variation.
2. The general causes of variation are known: genre, text type, time lapse, change in author circumstances, etc.
3. The degree of variation is probably predictable.

4. Variation in short texts can be extreme – rendering them unsuitable for many types of measurement.
5. Variation between authors, inter-author variation, is really *intra*-author variation seen through a different lens. Two authors working in the same genre, producing the same text type, with similar sociometric parameters (type/gender of addressee, social class, educational level, age, social circumstances, context) will in all probability have low inter-author variation.
6. In short text analysis we usually find high *intra*-author variation and low inter-author variation if the texts are of the same type. This is known as the *short text stability problem* and contributes to the difficulty of comparing authorship of unknown texts in the forensic setting.

Chapter conclusion

This chapter will hopefully have given the reader some insight into the question of textual variation, its causes, the fact that we can measure it and that we can predict it to some extent. It should also be clear from this chapter that if there are ways of confidently attributing authorship, then we cannot rely on the surface structures of language. Rather, to extend several well-established linguistic terms: we analyse syntagmatically, but in a paradigm. That is to say, what we analyse is on the surface, but our representation of it is not. This will become clearer in the next chapter when we discuss different forms of linguistic markedness within the framework of authorship comparison methods.

Authorship comparison

5

Plan of the chapter

In this chapter the outline of an authorship comparison method is presented. It is not sufficient simply to find a set of similarities or differences between two text styles. We need to understand the nature of those points of similarity and difference and the linguistic levels at which they can occur, including phonology, syntax, morphology, idiom and orthography. We should think of language as a system which occurs, or is representable, in layers. It is probably useful to talk in terms of some layers being deeper than others. However, they are not really 'deeper' in any structural sense. The use of the word 'deeper' here is simply a metaphor. All the layers are connected, or as a Washington DC literacy project leader has expressed it: 'The layers of language are interwoven'

(Moats, 2001). Thus, for example, we may think of the lexical layer being separate from the syntactic (or 'grammatical') layer. However, items in the lexicon are syntactically tagged: we know that certain nouns are count nouns and others are not, and we know this because the lexicon is 'licensed' by the syntax. Thus we cannot preface a non-count noun with 'a/an', e.g. 'a sugar', 'a money'. The lexical entry for a non-count noun only licenses determiners like 'some', 'any', 'more', etc. Nor would we usually have 'some sugars' if we mean packets of sugar. The plural '–s' morpheme is, like the use of certain determiners, not licensed in this usage.

Hence we see that items in the lexicon have to 'work with' the grammar of the language as well as its morphology (i.e. the morphemes of the language, such as plural inflection –s; past tense inflection –ed; prefixes pre–, inter–; suffixes –less, –ful, etc.). Neurolinguists studying language in patients suffering from aphasia (e.g. stroke victims), for example, have found that difficulties with word finding and naming are related not only to a semantic and phonological layer but also a syntactic layer. When therapies that are designed to access syntax are used, patients show a greater improvement in word naming and in speech output (see Herbert and Best, 2005).

Language layers at the 'other' end of the spectrum relate to context, social appropriacy, and so on. In using language we are constantly working at all layers of language: searching for the appropriate word in the lexicon, placing it in a well-formed syntactic structure, assessing its meaning, determining its contextual and social appropriacy, building human relationships through language and paralanguage, etc. Many of these layers are easily accessible to observation: we can comment on how other people use language, why someone used a particular phrase, how one person's choice of vocabulary will appear different from another's, how tactful some people are when speaking, etc. We know that there are such things as a formal and an informal lexicon, that personal letters are different in particular ways from academic texts which in turn differ from novels or recipes. Effectively, we know that a particular text type will have a particular generic style. We can also 'recognize' the way a friend will use language – we appear to know their 'style' of writing. However, style recognition in this sense is dependent on many factors, and is probably as much a process of conscious recognition of behavioural patterns, expected topics, minor spelling, punctuation and layout quirks, as it is to do with a writer's actual way of using language. As a way of differentiating authors in a formal *forensic* context, it is limited, subjective and wholly without merit. You can prove this simply by trying to imitate the style of another writer – especially

one whose language appears unusual in terms, say, of spelling, punctuation, etc. As you will see if you try the experiment, it is very easy to do. Thus, when comparing authorship of texts, we need to go beyond descriptive qualities. We need to look at linguistic structures.

At the 'deepest' layer we have syntax, the structure of phrases. The deeper the layer, the less accessible it is to the writer's awareness and also the harder to imitate. This point was first made in the forensic context by Chaski in 1997, who stated that the production of '. . . syntactic structures . . . would appear to be very difficult to control' (Chaski, 1998: 19). Shakespearian scholars have also long been aware of this problem. For example, Lancashire cites Horton (1987) as follows: 'Unfortunately the textual features that stand out . . . usually reflect a writer's conscious stylistic decisions and are thus open to imitation, deliberate or otherwise. Tests of authorship that are founded on subconscious habits are a desirable goal in most (if not all) applications' (1987: 9). Hence the syntactic layer is usually more useful for forensic purposes, if not significant, than what we see at the 'surface' layers – i.e. spelling,[6] document layout and so on. Later in the chapter you will see a suggested order-of-importance table which outlines the hierarchy of similarities and differences in more detail.

Information Panel 5.1: Summary of authorship comparison method

- Check **Known, Questioned** and **Unknown** texts (**K** and **Q**, **K** and **U**, etc.) for compatibility and comparability.
- Examine texts for points of **similarity and difference**. Do not be one-sided, include **counter-examples**.
- Ascertain the **linguistic** layers or level/s of these points of similarity and difference.
- Look for **relationships** between the different layers.
- Using a **corpus**, test the extent to which these features exist in the general population.
- Determine their relative **significance** in terms of an order of importance table.
- **Quantify** the differences, using a statistical approach.
- Apply **a scale of judgement** reflecting the likelihood or unlikelihood of authorship.

So, let us assume you have analysed a set of texts and found what you think are similarities and differences. You will still need to do two further things: test the importance of the differences by looking at a corpus or, sometimes

(and with due care), the internet. The second thing you need to do is to calculate the relative importance of these differences on an order-of-importance table. Having done all this there is still more to do: how significant are your findings? You can do one of two things. If you have enough data, you can feed them into a statistical package, or consult someone who is qualified in statistics if you do not have this knowledge yourself. Finally, having made your calculations, decide what level of confidence you can personally invest in your results. This is not the technical term 'confidence level' that statisticians use, but your own personal opinion. It is the very last step in the chain of activities. All of this is not as intimidating as it sounds. I will take you through the steps one by one, using actual case examples. First, however, there are some preliminary matters to discuss.

Authorship – attribution, comparison or identification?

The phrase 'authorship attribution' is much more common than the phrase 'authorship comparison' as any internet research will testify. However, for forensic purposes I prefer *comparison* to *attribution* for a number of reasons. Because forensic texts are typically short they represent only small samples of human behaviour. Even if we find that several short texts by one author exhibit similar features we cannot be sure that these features are not random. There is always the possibility that some quirk of individual variation is acting on the features we have found. Therefore I do not believe we can *actually* attribute short texts, because – out of all the possible population in the world – we cannot say for certain that *x* or *y* wrote a particular text. This would be the same as saying that humans possess an idiolect – a unique way of using the language. Some writers have referred to the 'linguistic fingerprint'. However, there is no evidence for that viewpoint. All we can say is that, given *x* and given *y*, *x* or *y* is the more likely candidate of the two. Therefore what we are doing is not attributing but comparing. As for long texts, even then we cannot speak of attribution with any safety – the same applies, even though the behaviour sample is much larger. Thus, when using the term 'authorship attribution' in the context of any type of text, short or long, we cannot use it in a technical sense. This is why I suggest that the most appropriate technical term for authorship testing is *authorship comparison*. Some linguists use *authorship identification*. My reason

for disliking this term is that, again, if we have two possible authors as candidates, we are not really undertaking an identification exercise. To say that our writer is more likely to be *x* or more likely to be *y* is not the same as saying that our writer is *x* or *y*. We are not actually identifying *x* or *y*. We are proposing a greater likelihood of *x* than *y*, or the other way around. If we were absolutely certain that *x* was the writer, for example, then we could say we are identifying him or her, but it is a basic premise of any scientific method that results are not given as certainties, but as probabilities or classifications. It is very rare in authorship work that we can *identify* a writer. Also, note that identification here suggests being able to pick out one individual author from all other authors. This takes us back to the 'linguistic fingerprint' claim. This is another reason why I think the term *authorship identification* should not be used in forensic work.

It is also important to point out that in some instances we may have only one candidate suspect author. In that case we cannot make a comparison. What we can do, however, is to say that a feature set is consistent with that of the suspect, but – again – this is not attribution and it is not identification, it is simply a matching process of features. It stands to reason that a matching process of this nature is a very different kind of process than one in which we are able to make comparisons across several authors. Some would argue that the whole process is somewhat weakened if there is only one candidate.

These issues aside, one of the key problems with authorship comparison is that some of the methods are not methods at all, and they have no theoretical underpinning. They are simply *discursi* of feature sets. This is especially true of some stylistic approaches where the analyst is free to cite whichever features s/he finds unusual, which seem significant. However, it is important to have a reason for what we find significant, and that reason needs to be underpinned by a theory of individual language use. For this purpose I propose that we consider the theory of markedness. This is not a new idea, and in fact, to put this in its proper context we need to begin by considering an outline of what markedness is in linguistics and how it arose.

What is markedness?

Markedness originated in structuralist and generative grammar as a way of explaining constraints on grammatical rules. It is a standard linguistic concept and has been so for more than 50 years. Both Jakobson and Chomsky sought

to find linguistic universals[7] – rules of structure which applied in all languages. In generative grammar the way the child acquires language is explained by saying that there are two settings for any given rule – a marked setting and an unmarked setting. The unmarked setting is preferred because it requires the least informational input, the marked setting is used to provide contrast. For the most part the child chooses the unmarked setting. This can lead to errors if a child generalizes an ill-formed structure from a pattern, for example 'seed' instead of 'saw' on the basis that past tense verbs in English are most often formed by adding the morpheme '–ed'. This is why it is quite useless to correct very young children: they are choosing the incorrect setting because as yet they do not have sufficient information to choose the correct one – in this case the marked one.

The point about markedness here is that it is a version of some linguistic sign or structure which stands out because it is unusual in some way or 'non-standard'. These are two different concepts, despite their superficial similarity. By unusual I mean a lack of conformity to a common pattern. In the present case, apart from the kind of child language error described above, I mean a lack of conformity to a general pattern, such that if you took a survey of how a thousand people in a community pronounced the word 'pretty' and found only one person who said 'preety', then this person's pronunciation would be 'marked' in the sense that it would be unusual.[8] That is one way in which I mean 'marked'. The other way is slightly different. By 'non-standard' I mean the usage of a word or term in a way which does not fit that of a prescriptive dictionary or academy (and hence a generally accepted) definition. Thus, in a given class at a particular school it may just happen that five out of ten children spell 'receive' as 'recieve'. This could be because their teacher can't spell, or because they are not yet old enough to have learned how to spell *receive* 'correctly'. However, although their linguistic usage in this instance is not unusual in the environment they all share, it is nevertheless 'non-standard'. It simply does not fit the dictionary requirement for that particular word. We will now turn to an example of how the concept of markedness has been applied to authorship problems.

Chaski's application of markedness

The first linguist to consider markedness in terms of authorship systematically was Carole Chaski, whose statistical analysis of syntax in authorship has met

the Daubert challenge in the US court system. She uses Head Driven Phrase Structure Grammar (HDPSG) as a model. Despite its complex-sounding name, HDPSG is in fact a very simple grammar. Like all generative grammars, it is based on the minimum number of rules, in this case four. One of them is known as the specifier-head rule. The specifier-head rule for English states that a specifier should precede its head. A simple example of this is the word order of 'determiner + noun' in English. We do not routinely find, in English, phrases like 'dog the', because this, quite simply, would violate the specifier-head rule. If we did find a writer frequently producing such formulations this would certainly be a marked syntactic form. However, many languages allow the form 'head + specifier', for example Swedish, where the definite article 'the' is attached to the end of the noun, e.g. bil = car, bil*en* = the car. For a formal discussion of syntactic markedness in discourse, see – for example – Joshi *et al.* (1998: 275, 289) and Davison (1984).

Chaski's method of authorship comparison involves the use of a software program called *Alias* which implements a patent-pending proprietary way of calculating markedness across the heads. The user checks the tags of the speech parts but does no more than that. Instances of markedness can relate to the specifier-head rule, the head-complement rule, the head-modifier rule and so on. They can apply to noun phrases, verb phrases, preposition phrases, adjective phrases and so on. Once a text has been marked up for syntactic markedness, the results are computed statistically. In addition to syntactic markedness, punctuation in conjunction with syntax can also be used to evaluate authorship. This involves measuring how an individual uses punctuation, not just how much or what types of punctuation are used. Punctuation can occur at word level, 'bring me food, wine and . . .', at phrase level, 'they had several bottles of wine, a number of whiskies, glasses of brandy . . .', at clause level, etc. Punctuation also occurs at word-internal level, e.g. as apostrophes or dashes. Given the power of both of these methods, it is somewhat surprising that this approach has not received more widespread recognition in the forensic linguistic community. Chaski should be credited with having brought forensic authorship comparison (as opposed to long text authorship 'attribution') into the scientific arena, and out of the darkness of literary criticism, canonical literary corpus construction and discourse analysis models of authorship identification. The lack of recognition to date may have something to do with the fact that many modern linguistics departments work only in pragmatist paradigms, and simply ignore traditional and theoretical grammar.

In this chapter I will consider variations on the study of markedness with a view to looking for systematic relationships between different language levels or layers. I will also cite an example of using internet and corpus data to test for markedness. In this chapter I will not be going into the detail of statistical analysis, but will confine myself to a broad discussion of the comparison methods I use.

Markedness and authorship: a case history

I will just preface this section by repeating that my approach is one of authorship comparison not authorship attribution. If I am given two candidatures for authorship the answer I will give, assuming I can give an answer, is that it is more likely to be *x* than *y*, or *y* than *x*, but this is *if and only if* I can reasonably assume that these are the only two likely candidates. That is all: authorship comparison cannot go beyond that. If a third viable candidate, *z*, is hiding in the wings, unknown to any of us, then my comparison falls down. However, as a forensic linguist I do not get to choose the candidates. I don't even want to know their names – labels such as A, B, C or *X*, *Y*, *Z*, etc. will do just fine. I can advise on general principles of candidate selection, but I cannot do the choosing. This is why I can only *compare x* and *y* because it is mostly unknowable if there is a *z* too. Of course I am not obliged to choose either: I can say, and sometimes do say, that there seems little likelihood that the author is either *x* or *y*. We need to keep in mind that the candidates we have may not be the only possible candidates, and we need to keep in mind the need to avoid stating a likely comparison when the linguistic evidence simply does not offer support for any of the candidates that we know about.

Having made that short preface, I would now like to illustrate an example of my own application of markedness to authorship by means of a case I undertook some years ago. Bear in mind that space restrictions will mean I can only cite a few examples from within an authorship comparison. Hence, this is not the entire analysis – only a small part.

Information Panel 5.2: Comparison assumptions

> When comparing known with questioned (abbreviated 'K' with 'Q') texts do not treat Q as a single corpus since you cannot know in advance, if ever, whether all of Q is of a single authorship.

Background to case

This case concerned a number of malicious and threatening letters. The recipients reported receipt of these letters to the police. It rapidly emerged that all of the recipients had something in common. They were witnesses in a rape case. They didn't actually witness the alleged rape, but they claimed to be in a position to testify as to the location of the main suspect on a particular day. Other individuals, who were not witnesses, knew of several visits by the victim, to the suspect's house, or knew of other information which was damaging to the suspect, or perceived by him to be damaging. The letters all contained claims, supposedly made by a 'friend' of the suspect, that he, the suspect did not rape the victim.

At about the same time the rape case suspect himself wrote several letters to a number of the witnesses and others, effectively making the same claims as the questioned author/s, and in very similar language. I will now give some examples of the language of both text groups, the known (K) and the questioned (U). There are two known examples (K1 and K2) and two unknown examples (U1 and U2). In these examples the rape suspect is 'Joe' and the victim is 'Mary'.

Text Sample 5.1: Witness intimidation case: Text K1 (Known 1)
. . . I was arrested and charged with the rape of Mary. I could not believe this as I am and was impotent.

Text Sample 5.2: Witness intimidation case: Text K1 (Known 2)
. . . when I was arrested, I was firstly charged with raping Mary, I stated to the Police that I could not rape anyone, as I was impotent, and had been since 1996.

Text Sample 5.3: Witness intimidation case: Text U1 (Unknown 1)
The good fortune is Joe was able through solicitors, police Forensics and medical experts to prove 100% that he could not and did not rape or have unlawful sex, through his being Impotent and is and has been for five years or more.

Text Sample 5.4: Witness intimidation case: Text U2 (Unknown 2)
Neither would have known that Joe, was to enter hospital to have tests to Proving that he is and was impotent.

Application of 'markedness'

The reader will have observed a number of references to 'impotent' in both K and U text excerpts. These are abstracted into the table below:

Table 5.1 References to impotence in K and U texts

Text	Projecting clause	Projected clause
K1	I could not believe this	as I am and was impotent
K2	I could not rape anyone	as I was impotent and had been since 1996
U1	Joe was able . . . to prove . . . that he could not and did not rape . . .	through his being Impotent and is and has been for five years or more
U2	Neither would know that Joe	that he is and was impotent

What I find curious about these examples is the tense inversion: the projecting clauses are in the past tense, but the thematic verb of the projected clause is in the present tense. This is a marked structure since (i) it is more usual to cite events in their chronological order, i.e. earlier event before later event and (ii) the projecting clauses themselves are in the past tense, and it would be more usual that the event nearer the projecting tense in time is mentioned first. The only exception is K2 'as I was impotent and had been since 1996' which is no more unusual than '~as I had been impotent since 1996 and was still impotent~'. This example aside, some readers may consider this to be an issue of historical present tense usage. It is certainly true that the narrator in each case may be simply emphasizing the current 'impotence' of the subject. However, this is perhaps a narrative matter rather than a syntactic or semantic matter. Second, the positioning of the present tense within two past tense events in this way does not suggest the historical present tense, at least not in this context.

I suggest, rather, that it is a marked form of some kind. It is not syntactic, since there is no phrase structure rule that would specify a chronological sequence in this context. Similarly, it is not a matter of 'non-standardness', since, again, there is no generally accepted standard that we could turn to and say, 'This clause violates a known standard'.[9] What we appear to have is simply an uncommon or unusual formulation rather than a non-standard one. To test this we need to consider some form of language frequency measurement, in such instruments as language corpora or, failing that, a suitable internet search engine. The corpus I used, the BNC corpus, gave the following results for these configurations:

Table 5.2 Corpus results for phrases like 'am and was'

String	Frequency
was and is	1
is and was	0
am and was	1
was and am	0

This result is not surprising for a corpus result, because corpora are often geared to higher register texts than the examples we see here. Such texts generally reflect a higher percentage of nominalized constructions than informal texts. This can be fairly easily tested if we measure frequencies for determiners such as 'the' and 'a/an'. The percentage of these words in higher register texts, such as academic papers, novels and news media articles is greater than it is in less formal texts, such as letters, emails and other similar texts. The same corpus, the BNC corpus, for example, found only 267 instances of 'you and me' and no more than 564 instances of 'you and I' – and this is on a base of 100 million words. This clearly suggests that the corpus is much more formal in its register stock than, for example, the internet would be. Hence, since the corpus had nothing to offer in the present instance I turned to the internet.

We should be cautious about the use of search engines for a number of reasons. Because of the way search engines work we cannot rely on the exact frequencies given. Most will say, for example, '*about* 1,360,000 results'. The 'about' here is significant. The search engine server decides how much time to allot to the search, depending on its current workload. It then reports what it found in the given time on the servers that were most readily available to it. This is why internet searches vary in frequency. However, we can certainly get a very good idea of frequencies, even if we cannot depend on a precise number. Bear in mind, though, that the results you get have to be very significant to be used at all, e.g. 10,000 vs 50 is a lot more valid than 10,000 vs 9,000.

Where you cannot use a corpus but have to use the internet, for example if you are searching for a stop word, try to triangulate your results using different constructions.

One search engine gives the following counts for various versions of *pronoun + be (present) + be (past)*, and *pronoun + be (past) + be (present)*:

Table 5.3 Internet frequencies of subject pronoun + past and present forms of to be

'I was and am'	155,000
'I am and was'	20,600
'he was and is'	174,000
'he is and was'	21,300
'she was and is'	55,100
'she is and was'	727
'we were and are'	50,300
'we are and were'	904
'they were and are'	122,000
'they are and were'	31,800

From these measurements we see that the past-present formulation is much more frequent than the present-past formulation. There appears to be a strong tendency for the past tense verb to precede the present tense verb. Even if we cannot be sure of the precision of search engine frequencies, the search still seems valid, given that the search has been widened to test other pronouns than the ones we are interested in. Additionally, I suggest that intuition tells us that the past-present sequence is more typical than the present-past sequence, and also that if we are projecting ('I thought', 'he said', 'I believed', etc.) in the past tense then it would be more usual to refer to the past or earlier event adjacent to the present or later event.

In both the known and questioned texts, the kind of disordered sequence I have referred to above occurred on a number of occasions, including such examples as 'the proof and evidence was given to the police', 'in the past years and months', etc. In each case the sequence was disordered: thus, 'proof and evidence' defies the expectation that 'evidence' comes before 'proof': we have to collect the one to establish the other. Similarly in 'the past years and months' we are in a disordered sequence because 'months' go to make up 'years'. The reader can verify this information using both internet (Google^{TM}) and corpus searches (e.g. BNC lists ['months and years': 30 – 'years and months': 1]).

This kind of disordering also occurs at the syntactic level in the present case. Thus we have a syntactic break caused by punctuation (a comma) in these examples:

- The Police officer dealing with the case, stated that he did not believe me.
- ... no proof of sexual activity having taken place, between Mary and myself.
- I and my partner, have over the past years and months, studied the case.

Here we have punctuation breaks occurring at syntactic boundaries, but not at boundaries where they are licensed to occur. Thus, in the first example the noun phrase 'The police officer dealing with the case' is separated from its complement, the verb phrase 'stated that ...' by a comma. In the next example the verb phrase takes an indirect object 'between Mary and myself', yet the object, an intrinsic part of the verb phrase is separated from it. In the last example the noun phrase 'I and my partner' (itself a disordered sequence by convention) is connected to part of the verb phrase 'studied the case' by separating the auxiliary plus adverbial clause of time from the rest of the sentence in commas. Each

of these three examples clearly illustrates how we can use syntactic punctuation (see Chaski, 2005: 5) as part of an analysis.

So far, therefore, in analysing these texts we seem to have found, at the (informal) semantic level, a disordering of sentence components, to do with tense and time, and other kinds of units, while at the syntactic level we have a syntactic break where we do not expect it. I suggest these two phenomena are really to do with ordering and arranging structures so that they either become disordered sequentially, or disordered structurally. They appear to be symptoms or indications of the same overall type. I suggest that this kind of disordering or re-arranging is something most of us do, but on a, typically, very small scale. What struck me about all the texts in this inquiry was the scale and typology of the de-structuring.

I then looked at the lexical level: here too something interesting appeared to be happening. A number of words were made into two words by the intro-duction of a space between them – for example, in both K and Q we find 'inter course' rather than 'intercourse':

KNOWN: At that interview the Police withdrew the charge of rape, and substituted it to one of unlawful sexual inter course.
QUESTIONED: Joe was again interviewed by the Police, the rape charge was dropped and changed to one of unlawful sexual inter course.

The break which occurs here is based on the notion that 'inter course' is two words. It is a disordering of the structure of the lexis. For the relevant meaning, the lexis does not store 'inter course' but 'intercourse'. The lexis probably stores, but for quite a different reason, a collection of prefixes in English, e.g. 'inter' as in 'international', and of course it stores 'course'. Hence two storages – 'inter' and 'course' exist, but are not relevant in the present context. What we essentially have, I am proposing, is a disorder of the structure of the lexis, seen in both K and Q, other examples of which include: 'photo copied', 'back fired', 'other wise' and 'some where'. As with the punctuation breaks in the sentence, we are seeing here a break – a word space – where one is not expected or licensed by the grammar.

There were also many spelling similarities across K and Q, e.g. 'intensions', 'coarse' instead of [course], a number of identical or near identical phrases, e.g. 'for a so-called rape' vs 'for the so-called rape', etc. However, the spelling errors require a different kind of categorization than the other categories in terms of markedness and I will not be dealing with this topic here.

The examples I have given here are simply a small sample of the similarities I found across the texts. In each case I prepare a 'match table', an example of which I give below:

Table 5.4 Sample table of matches across known and unknown texts

Unknown	Matches	Known
is and has been [impotent]	**Temporal order**	I am and was impotent
	is and has been / am and was	
	Lexeme	
	impotent	
	Relevance	
	The less temporally relevant item is cited first.	
Result: Match		Assessment: seems an unusual combination of common features

As noted before, some types of match are more important than others. Thus, exact matches of long clauses or whole sentences would have more value than, for example, misspelling matches. Document layout and style would have the least value, even less than orthographic matches. The point about matching characteristics is that each claimed match needs to be based on the concept of markedness. For example, if a candidate and the known author both spell 'receive' as 'recieve' this may indicate markedness, but it is not likely to be significant markedness. About 2 per cent of the population spell 'receive' incorrectly. It is not that unusual. When we look for markedness, we need to determine – as noted before – whether we are dealing with structures which are unusual or structures which are unconventional. These are two types of markedness.

Information Panel 5.3: Reliability of results

> This is a reminder about testing. The only way to ensure your result is reliable is to test it. You need to test the actual method on a number of text sets before you apply it to a forensic comparison, and you need to test the results you obtain by measuring them against corpora and search engines. In the absence of either of these, consider an experiment based on a questionnaire. Discuss your method with other colleagues and/or students to get feedback and criticism.

With syntactic markedness we are on surer ground than studying features at a more superficial level, such as document layout: we know, as native speakers, that in English the form 'dog the' is marked because it violates a simple, easily observable, grammatical rule. Moving through the layers, with semantic or idiomatic markedness we are on less sure ground simply because these are not at the same layer of linguistic organisation as syntactic structures (see Chaski, 2001b: 40). In the present case, I believe that both known and questioned writers share the same tendencies, namely to disorder chronology, syntax and idiom. Problems or symptoms of very similar types pervade both sets of texts. The next question is calculating the similarities and differences.

To do this, and following Chaski (1998, 2001b, 2005) I compute the number of marked constructions for each type of phenomenon, syntactic and otherwise. It is also important to list counter-examples. Recall, that it is not the linguist's task to *prove* anything. This is why, in the table below, there is a column labelled 'Unmarked'. The task is one of discovery not demonstration. The table below illustrates how you might list your calculations. Of course, if you are in a position to computerize or otherwise automate parts of the calculation process, so much the better.

Table 5.5 Suggested layout of calculations

Phenomenon	Marked	Unmarked
Syntax	K:	K:
	U:	U:
Punctuation	K:	K:
	U:	U:
Semantics and/or use of idiom	K:	K:
	U:	U:
Orthography	K:	K:
	U:	U:
Spelling	K:	K:
	U:	U:
Document layout	K:	K:
	U:	U:

Judging your results

How do you judge your results? If you have adequately tested your results and your method and you are able to quantify them statistically, you can

repose a high level of confidence in them. 'Confidence' here is being used in a non-technical sense. It is not the same thing as a statistical 'confidence level'. 'Confidence level' in statistics means the error-rate you assign as the decision point for claiming an authorship match or, rather, a valid authorship comparison. One tool I use to assist in reaching a judgement is an order-of-importance table. This gives hierarchical order to markedness:

Table 5.6 Order-of-importance table of types of style features

Feature	No.	Comment	Assessment
Types of grammatical structure	1	See Chaski (1998, 2001b, 2005). A rare, unusual or erroneous structure across known and questioned. Knowledge of generative grammar helpful.	Significance depends on number and type of matches.
Punctuation	2	Unusual punctuation habits or punctuation errors can also be significant. Consider combining punctuation analysis with syntax.	Can be significant, but not necessarily on its own.
Idiom	3	Curious phrasing can be significant. Research its markedness wherever possible.	Depends on number and type of matches.
Spelling	4	Spelling errors, especially rare spelling errors. Look for dialectal spellings, such as US vs UK spellings in English.	As above, but research degree of rarity.
Document layout	5	This can be of interest if the layout style is unusual. Beware, however, of cultural styles of layout.	Caution required, e.g. US vs UK vs European styles.

Important notes: (i) when considering similarities and differences: avoid taking isolated instances of phenomena into account, (ii) consider the possibility of dialectal variation to account for (some) differences.

What this table says is that if we find two texts, supposedly independently produced, with a significant number of syntactic matches, then this is likely to be of greater significance than, say, matches relating to document layout.

Finally, your last step is to apply a scale of judgements. See the Appendix at the end of this book for the scale I use, adapted from the original kindly provided by my colleague, Dr Kate Storey-Whyte. It is a very clear scale and will help you to determine your level of confidence in your work. It is important not to overstate your confidence in your findings. This is an ethical question. It is just as important as being completely neutral and impartial, being rigorous in your method, using only the most tried and tested statistical procedures, and practising the utmost confidentiality in your work.

Further authorship testing

In the previous section I described a particular kind of inquiry, one in which the nature of the markedness was highly visible. The known author appeared to have a number of quite distinguishing features which were able to be categorized and, when compared with the questioned documents, seemed to provide strong and measurable similarities.

However, in many inquiries we are dealing with writing which, superficially at least, appears to be very bland. The example below will illustrate:

Text Sample 5.5: First email in child pornography case (questioned text)
I have spent some time this evening at your site and have not been able to access any material of the type that was used to advertise it. It was a bit like memory lane. Some of the material was lifted straight from 35 mm films and videos that I watched 20 years ago. Seriously, though, it really was nothing like the preview and I have no wish to subscribe to this type of site. Please would you give me access to the type of material promised or cancel the charge to my card.
Incidentally, I had no trouble with the foto archive but what is one supposed to do with New CamsVideo and Erotic Show sections? I was unable to get anything from them.

In this particular case the above email was found on the home computer of a businessman who denied authorship. He explained that a friend of his, a plumber, since deceased, had written the email and even used his credit card to complete the purchase of the site membership.

There was some difficulty obtaining letters written by the deceased who, according to his family, had never been 'much of a letter writer' and had never owned a computer. However, some correspondence was obtained, and the following is a typical example:

Text Sample 5.6: Known example of plumber's correspondence
Dear Mr Brown, Please could you contact me as soon as possible as to my situation at the moment and why I am being keep here at the prison. And when will I be released as not knowing is worse of all. I hope to see you soon.

The above example shows a number of basic prescriptive errors, including 'being keep' instead of 'being kept' and the termination of a sentence before 'and'. Additionally, the plumber confuses the superlative with the comparative adjective, i.e. 'worse' instead of 'worst'. He thus appears to exhibit difficulties of three different types: grammar, punctuation and lexis. On the surface, at least, his writing is not easily comparable to that of the email found on the businessman's computer. The businessman on the other hand, demonstrates a somewhat greater competence, as can be seen from the following text:

Text Sample 5.7: Known example of businessman's correspondence
On Tuesday 1 May 2001, my wife and I went out to dinner with my daughter and her partner. When I returned home, at about 01.30 there were two messages on the answerphone. One was from the central station, that monitors the alarm at Greenlake Street, Smithville, informing me that the telephone connection to the premises had failed. The second was from my PA, Mrs Jones who is an alternate key holder at Greenlake Street, to the effect that she had been informed of the line failure, had attended the premises but found no indication of anything amiss. I returned the call to the monitoring station and learned that the probable cause was a line fault and that the incident had been "policed" (log 101 timed at 22.35) and that the police had attended and reported the premises to be secure.

Analysis

I began by looking for signs of markedness in the different known texts. At first sight, the businessman's texts seemed to be very formal, perhaps even pretentious in parts. However, at this stage this is just an impression and needs to be tested. In his survey of markedness as a linguistic topic, Battistella (1996) notes that several linguists, including Greenberg (1963), Zwicky (1978) and Moravcsik and Wirth (1986) all concur that relative frequency is a sign of markedness vs unmarkedness. Some forms of markedness are purely stylistic, e.g. the use of excessive repetition. As such, they are hard to measure and can only be used for an impressionistic account of authorship. In the present instance I decided to take an internet and corpus survey of the phrases found in the known and questioned texts, and test them for frequency. All of the phrases were surveyed on Google and samples of these were rechecked using the British National Corpus (BNC). Further checking of some of the samples was done using Cobuild.

The best way to begin your search for instances of markedness is to print out a copy of the text and underline as many phrases as possible. Aim to do approximately one phrase of three to five words in length for every ten words of text. Some phrases, because of narrowness of context, will simply not be suitable, especially those containing proper nouns and personalized or coded references. Do not bias your sample by just selecting what appears unusual or marked.

Once you have selected phrases, try to look for alternative ways of saying or writing those phrases. Change only one feature at a time – but keep to the text's register, i.e. if the text is formal, businesslike, your changes should reflect this. Here is an example from the businessman's known text:

Table 5.7

Textual phrase	Alternatives
the probable cause was	the cause was probably
	probably caused by
	probably because

In the above table we see an excerpt from the businessman's letter, the phrase '... the probable cause was ... ' As a noun 'cause' is fairly rare when compared with its presence as a verb. Additionally, the adjective 'probable'[10] is fairly rare when compared with its adverbial form, 'probably'. This is attested on Google, Cobuild and BNC. Thus, what we find from the phrase 'the probable cause was', is that it contains two marked items, relative to the items in the text. Of the two alternatives 'probably caused by' is by far the most frequent of the three possibilities given here.

The next step is to check frequencies for each selected phrase and the alternatives you devise. Do not choose only those examples which seem marked. Compile a table something like this:

Table 5.8 Examples of known text markedness

Phrase	Search engine (Global)	Search engine (UK)	BNC	Markedness
the one which	1,150,000	401,000	244	
the one that	26,000,000	1,150,000	1030	Unmarked
the cause was probably	505	1,220	0	
the probable cause was	10,800	373	1	Marked
probably caused by	1,070,000	117,000	1	
of anything amiss	566	351	1	Marked
of anything wrong	28,500	10,500	1	

Note: The author's text choices are indicated here in **bold** type.

I then looked at the known text of the plumber and found that, typically, his marked structures were to do with prescriptive data, largely spelling, but also grammatical, e.g. 'being keep' instead of 'being kept', 'worse of all' instead of 'worst of all'. These types of error are also marked and can be quantified using corpora and search engines (though compilers of corpora tend to 'tidy' up text and thus suppress errors). I found no examples of marked phrases, as the table below shows:

Table 5.9 Phrases from the plumber's texts

Phrase	Search engine (Global)	Search engine (UK)	BNC	Markedness
contact me	66,400,000	1,530,000	180	Unmarked
get in touch with me	1,020,000	115,000	1	
as to my situation	282,000	7	0	Unmarked
about my situation	175,000	9,130	1	
on returning home	155,000	44,600	1	Unmarked
on coming home	49,200	3,290	1	

The questioned text also contained a number of marked constructions, as these examples show:

Table 5.10 Examples of questioned text markedness

Phrase	Search engine (Global)	Search engine (UK)	BNC	Markedness
material of the type	93,500	2,790	0	Marked
type of material	1,180,000	148,000	1	
it really was nothing like	24,600	2	0	Marked
it was nothing like	170,000	14,600	1	
I have no wish to subscribe	3	2	205 ('no wish to')	Marked
I do not wish to subscribe	18,100	1,910	255 ('do not wish to')	

Note: The author's text choices are indicated here in **bold** type.

The next step was to compile a comparative table of types of markedness:

Table 5.11 Type and density of markedness

	Lexical	Grammatical	Categorical	Competence	Density
K (Businessman)	'probable', 'cause', 'amiss', 'to the effect that'	'reported the premises to be' [construction], 'anything amiss', [word order]	Nominalization: 'cause', 'effect'		0.05
K (Plumber)				'being keep' 'worse of all'	0.04
Questioned	'access' vs 'get', 'obtain'	'not able to access', 'that was used' [passive]; 'material of the type'	Nominalization: 'wish'		0.04

As can be seen from the above table, the density of marked constructions and unmarked constructions found in the texts is roughly the same across both known authors and the questioned author. However, the second known author, the plumber, exhibits a completely different type of marked construction when compared with that of the businessman. The businessman's text is, with respect to the type of markedness shown, as well as its register, much more like the questioned text. On the other hand, the plumber's issues of markedness relate to competence in use of the language. In my report I concluded that it was highly likely that of the two candidates, the questioned email originated from the businessman, not the plumber. This evidence was submitted to the court, at which point the businessman pleaded guilty to downloading and possessing child pornographic images. This case is listed on the forensic linguistics case files of the Wikipedia.

Theoretical considerations

I will summarize this chapter by saying something about theoretical bases for a discussion on authorship.

Style: Style can be generic, e.g. the style of a particular genre, or it can be personal, e.g. the 'personal style' of an author. However, description of superficial style is simply not adequate for forensic purposes.

Markedness: Markedness occurs at many levels, formally and informally. What I have described here are simply some aspects of markedness as they can be applied to a forensic setting. We need to link different types of markedness to try to understand the underlying system of markedness for a particular author or set of texts. We should also consider that, in some sense, all texts are marked. When we analyse even the blandest texts we will find signs of markedness. *Note*: do not even consider 'invented' texts for analysis.

Idiolect: We cannot assume that, just because we can differentiate the authorship of different writers, individual writers have an idiolect. The way in which we acquire use of language socially, and continue to modify it through education, peer usage and the media, all mitigate against the idea of idiolect.

Stylistics: The use of descriptive stylistics for forensic purposes is simply not acceptable in today's scientific climate. In the early days of forensic linguistics a number of linguists, including myself, studied or used such methods. However, in recent years courts have rightly insisted on more formal approaches which should wherever possible be quantifiable, either in terms of probability or classificatory statistics.

Chapter conclusion

Note that the methods I use are described somewhat informally here. In practice the process is somewhat more rigorous, involving a more detailed analysis than space permits. I have reserved discussion of statistical matters in authorship for a separate work as, I believe, it requires an entirely different treatment.

Evidence in court

6

In this chapter I offer a brief outline of how courts in a number of countries view expert evidence.

Legal requirements across the globe

The United States: In the USA until fairly recently, evidence of a scientific or technical nature had to pass the Frye test. This was a fairly simple test to administer: providing a method had acceptance from the scientific community it could be held to be valid in a court of law. In recent years, however, this standard was shown to be flawed after it was demonstrated that the use of hypnosis to re-activate memory could result in false memory syndrome: although the use of hypnosis had 'acceptance from the scientific community' it still produced a false result. Hence the standard had to be revised, and this occurred in the Daubert vs Merrell Dow Pharmaceuticals case, in which a group of mothers asserted that their ingestion of Benedictin, a prescription drug, had caused their children to be born with defects.

An expert witness, acting for the respondent (the pharmaceutical company) claimed that there was nothing in the professional literature to suggest that Benedictin produced birth defects. The literature produced by the petitioners (the parents) did not show, according to the expert, that the scientific methods used by the defence's experts to validate their claim that Benedictin did cause birth defects, had received general acceptance (as per the Frye test).

Because the petitioners' evidence failed the Frye test, their evidence was ruled inadmissible. Later, the Court of Appeal overturned the Federal Court's ruling. However, the Supreme Court of the United States vacated the judgment, ruling that there were other criteria than 'general acceptability' for the admission of expert testimony, including:

Knowledge and stature: The expert must have sufficient knowledge of the subject area ('knowledge, skill, experience, training or education'). Additionally, the expert must have stature in the academic or other peer community.

Testing: The technique must be empirically tested. It must be falsifiable and refutable.

Peer review: The technique must have been subjected to peer review and publication.

Scientific method: The expert must be able to demonstrate the known error rate.

'Straightforwardness': The technique must be able to be explained with sufficient clarity and simplicity so that a court and jury can understand its plain meaning.

Collectively these criteria have come to be known as the 'Daubert' criteria, or just 'Daubert'. Note that, in essence, Daubert is nothing more than an interpretation by a particular court of existing rules of evidence, in this case Rule 702, 'Testimony by Experts', which states that:

> If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.

In the United States it is now becoming common for all federal and some state courts to demand evidence which passes the 'Daubert test', and includes the proviso that error rates must be known if a statistical analysis is undertaken.

Not all forensic disciplines can pass a strict form of the Daubert test. These include ballistics, forensic phonetics, forensic anthropology, child abuse accommodation syndrome, etc. On the other hand, polygraph (lie-detector) analysis,

previously inadmissible under the Frye test, is now acceptable to many courts under Daubert.

In the US case law exerts a strong influence, and courts have some latitude when implementing Daubert, including which criteria to apply in any given instance. See Appendix 4, Notes on Daubert.

Australia: In Australia, the situation is somewhat different from the USA. Prior to a revision of the Australian expert witness guidelines in 1998, the task of the expert witness had in fact been described in 1997 by Gordon Samuels, a former president of the Australian Court of Appeal as one of furnishing 'basic scientific or technical data', and 'to present inferences and conclusions from the facts which the judge or jury, for lack of specialised knowledge, cannot draw themselves'.

The emphasis in Australian law has now been placed on the requirement to be relevant and reliable (the relevance and reliability rule). Expert witnesses are not required to provide the kinds of scientific evidence required under a strict interpretation of Daubert, rather they are obliged to qualify their findings (where they believe there may be incomplete or inaccurate evidence), and they must give reasons for their opinions. This implies, of course, that Australian law does not exclude experts from giving their opinions. However, as is the case in several countries, lawyers are reluctant to accept linguistic evidence because many believe themselves to have adequate linguistic knowledge to form their own opinions. Many do not see the study of language as a separate, technical subject.

England and Wales: In the early 1990s Lord Woolf was appointed to oversee a wide-ranging reform of the civil court system in England and Wales. There was growing concern among legal professionals that the law was becoming inaccessible to members of the public. One of Woolf's main concerns was the contribution of expert evidence to the quality of justice. Woolf contended that the use of experts could contribute significantly to a case's costs and therefore deter individuals from seeking legal redress. Amongst other reforms, Lord Woolf has proposed the appointment of a single expert for both parties – a controversial concept and one which is anathema to many solicitors and barristers who have grown up with the traditional adversarial system. In those cases where the parties are unable to agree on an expert, Woolf proposed that the court itself should make the appointment. In essence, Woolf's aim was to give courts wide-ranging discretionary powers in the matter of experts: '... the calling of expert evidence should be under the complete control

of the court'. According to Woolf, the use of experts should be 'economical' and the issues between experts – where there is more than one expert (which, given Woolf's adherence to the notion of single experts is far from certain) – should be narrowed 'as early as possible'. Woolf has not gone so far as to propose scientific standards of evidence, such as were put forward in Daubert. It is worth noting that Lord Woolf's reforms were made relative to the civil justice system. However, there is no suggestion that these principles, or at least some version of them, will not also apply to criminal cases in the future. Many Commonwealth countries have based their legal system on that of England and Wales, and so similarities can be found with regard to expert witness evidence almost across the globe.

Canada: In Canada the determination as to whether an expert witness should be called was formulated as early as 1931,[11] with the memorable juridical phrase:

> . . . the subject matter of the inquiry must be such that ordinary people are unlikely to form a correct judgment about it, if unassisted by persons with special knowledge . . .

More recently, this observation was added to with the injunction that expert evidence must be both helpful to the court in its role as 'trier of fact' and relevant.[12] Several important guidelines for experts were formulated in Perricone vs Baldassarra:

> Expert evidence presented to the court should be, and should be seen to be, the independent product of the expert uninfluenced as to form or content by the exigencies of litigation. An expert should provide independent assistance to the court by objective unbiased opinion in relation to matters within his or her expertise. An expert witness should never assume a role of advocate. An expert should state the facts or assumptions on which the opinion is based and should not omit to consider material facts which detract from that opinion. An expert should make it clear when a particular question or issue falls outside of the expert's expertise. If an expert's opinion is not properly researched because insufficient data is available, this must be stated with an indication that the opinion is no more than a provisional one.

Crucially, Canadian courts demand that experts are unbiased, and may not act as advocates: expert evidence may be ignored insofar as it attempts to bypass this rule of impartiality. As regards Daubert, the Canadian expert

witness newsletter *Economica*[13] notes that scientific criteria can be applied to subject areas which do not constitute pure science 'with some modifications'. So although, for example, some techniques do not have a known error rate, yet the fact that they may have been subjected to 'peer review and publication' will constitute an acceptable criterion. Canadian courts have wide latitude as to what constitutes acceptable scientific practice.

France, Germany (inquisitorial system): In France,[14] Germany and other countries which follow an inquisitorial legal system (as opposed to an ad-versarial system, such as the USA and the UK) experts are appointed by the court, not by the parties to a case. The investigation is carried out by a judge or panel of judges. Lawyers acting for each side submit questions the judge may ask the witness. The rather restrictive rules of evidence found in Anglo-Saxon systems do not apply: in the inquisitorial system evidence is much more freely evaluated. Several other countries use an inquisitorial system. These are mostly Western European countries.

Other countries: In the Philippines, the rules of evidence with reference to expert witnesses state that 'The opinion of a witness regarding a question of science, art or trade, when he is skilled therein, may be received in evidence.'

7 Non-authorship cases

A wide range of case types will cross the linguist's desk in even a short period of time. These range from authorship cases to textual influence, plagiarism to witness intimidation, content of threat communications, hate mail, blackmail, text provenance, poison mail, authorship of witness and other statements, etc. Having previously looked at cases which primarily involved authorship comparisons, I will now cite some other types of case and consider the linguistic issues at stake in those cases. These cases relate to the following linguistic issues: contemporaneity of text, power and disinformation, memory for language, linguistic influence on witnesses, linguistic provenance of written text. For a list of recent cases, which include work on many other types of linguistic issues, please see Appendix 5, 'Selection of author's recent cases'.

Contested diary entries

In a well-known city in England a 40-year-old man had been arrested on suspicion of acts of criminal damage. These included setting fire to sheds, slashing car tyres, burning a van, throwing paint over another vehicle and disconnecting the electricity supply of an apartment block for elderly people. Police seized a diary from the suspect's home and found many entries relating to the crimes. The suspect, we will call him Mr Smith, claimed that he had made the entries in his diary only after he fell under suspicion, reasoning to himself that if he was going to be accused of crimes he had not committed he might as well write about them. Smith said he had often chatted to people who lived in the areas where the criminal damage had taken place and found out about the incidents in that way. Although the police did not believe him, there was nevertheless a linguistic issue here: were the diaries written contemporaneously with the events as they occurred, or had they – as the suspect claimed – been entered some months afterwards? The suspect claimed that the diary he had used had been given to him the previous year by his father, and that he had only begun to use it once he had decided to make the diary entries relating to the incidents.

Asked how it was that the diary contained so many references to matters which had nothing to do with the criminal incidents, Smith claimed that he had copied those from his old diary and then destroyed that diary. He claimed that the entries in the new diary were exact copies of the old diary entries. However, a claim of this nature cannot be taken too literally.

Diary entries do not encompass a single text type, and have the potential to include an almost limitless number of topics. There are thus no simple techniques for dealing with issues of contemporaneity in diary writing: diarists are free to use a range of styles in a diary, to move from informal to formal registers or vice versa, to alter addressivity from, for example a 'Dear Diary' mode to a form of third-party address, etc.

Method of analysis

The diary entries were first divided into two types: (a) incriminating and (b) non-incriminating. The first type details events which appear to incriminate Mr Smith as the perpetrator of certain criminal actions detailed in the diary, such as the slashing of car tyres, the burning down of sheds, and the setting alight of motor vehicles. The second type consists of descriptions of everyday life and do not appear to relate to the perpetration of criminal acts. They include references to his family relationships, financial worries, talk about a car purchase, visits to and from friends, minor health problems, etc. Another way of considering the question of typology for the diary entries is to consider them as 'original' entries and 'added' entries. On some dates the entry contains both 'original' and allegedly 'added' material, just as – by the classification given in (a) and (b) above – those same dates contain 'incriminating' and 'non-incriminating' entries.

I began by examining the diary for possible stylistic features. Ultimately, what I was interested in was seeing whether there were any appreciable differences of style between the incriminating and non-incriminating entries. Examination of the diary revealed a number of regular stylistic features of potential interest:

(1)   The omission of subject pronouns, e.g. 'Met George' in place of 'I met George'. This is of interest because it appears to occur somewhat regularly in the 2003 diary.

(2)   The use of symbols known as emoticons, such as a smiling face or a frowning face, e.g. ☺ and ☹. These appear to have been used to add emphasis to a viewpoint or reaction to a situation.

(3)   The addition of a drawing to an entry: this can be in the form of a map, a diagram, or a heart, presumably to indicate affection.

(4)   Underlining, apparently to add emphasis to a word or phrase.

(5)   The use of parentheses, apparently to expand a point, or to de-emphasize non-essential material.
(6)   The use of quotation marks to highlight a word or phrase.

For the purposes of this discussion authorship style may be described as a collection of identifying features which occur more or less regularly in a given writer's text. An author acquires a particular style as a result of early language influences, education, social contacts and exposure to language in various media, such as newspapers, television, films, books and – in many cases – as a result of use of the internet. An author's style can change in some respects over time, probably mostly in terms of the words added to that individual's lexicon. These language changes which occur in an adult's life are probably mostly attributable to media and peer influences. Other changes can come about through illness, changes in marital status, moving location or switching careers, bereavement, trauma, imprisonment, or any other major alteration in an individual's life. Authorship style is also dependent on topic, register, relationship to the addressee and other contextual factors. Thus, for example, a formal letter by a writer to an employer will exhibit different characteristics from an informal note to a friend. Finally, as with any kind of text analysis, we need to be aware of the possibility of variation.

As a result of the above observations it seems that if Smith added the material relating to acts of criminality at a later date rather than having written them contemporaneously with the other, non-incriminating, events recorded in his diary, then given that (a) they would have been written at a different time and (b) that they were different in topic from the other entries, it follows that there is a strong possibility that the types of diary entry (i) incriminating and (ii) non-incriminating would be observable through analysis of style features. Note that this inquiry differed from an authorship inquiry 'proper' in that there was no dispute as to who the writer of the diaries was.

The first step was to calculate the frequency of features (listed above) across samples of both incriminating and non-incriminating texts. It seemed that if distribution across the two sample populations was similar it could, in conjunction with other tests, add support to the likelihood that the incriminating entries had been written as an integral part of the diary, and not added – as claimed by Mr Smith – at a later date. On the other hand a major difference in the styles of the incriminating vs non-incriminating sections could lend support to Mr Smith's claim that he had indeed added the entries later. The features were counted and the following chart summarizes those findings.
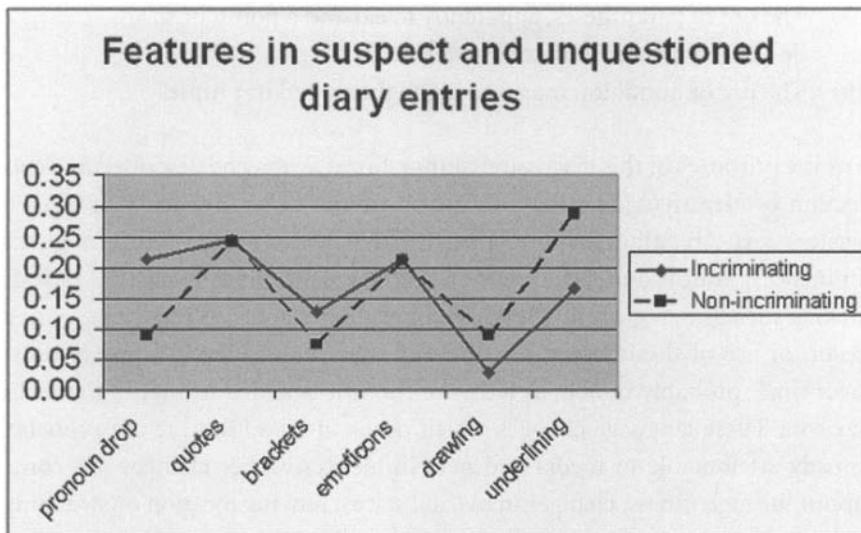
Chart 7.1 Comparison of features across known and questioned diary entries

The frequencies of the style features seem to correspond across both known and questioned entries. It does not seem to be the case that there are any exceptional differences. Two possibilities present themselves. Mr Smith may have been wrong in saying that he had copied the earlier entries 'exactly' as they had been in his old diary – what could have happened was that he may have just adapted his style to the new entries. It rather depended on whether he had made the 're-entries' contemporaneously with the 'new entries'. However, the known and questioned entries sometimes 'interleave' with each other. Here is an example, typical of a number of entries:

Text Sample 7.1: Diary excerpt showing 'interleaving' of entry types
Went to the doctor's. I haven't got a hernia after all ☺ Deb was a real "Gav-Gav" to me today but, I still love her & dream of the day I can buy her a lovely new 'velocipede' like this one (arrow to drawing). J so far, I have £18.96. "Stalked the night"; Browns Walk ☺ Phoned dear Mary? Told her how much I miss her ☺"

Note that Smith often used the phrase 'stalked the night' and that this is a phrase which occurred on dates when incidents of criminal damage had taken place in the areas named, e.g. here 'Browns Walk'. We see that the first part of the above entry relates to his family life (or fantasies about his family life). He then writes about a 'criminal' activity, and concludes his entry with reference to a telephone call to 'Mary'. I suggest it would have been difficult to collate existing diary entries with allegedly 'new' entries in this way.

The second possibility was that Mr Smith had in fact just made normal diary entries, of both incriminating and non-incriminating events, and treated both types in the same way – using similar style feature sets more or less in the same way whether the entry was of an incriminating event or not. So far, I could not see any reason to believe that the incriminating and non-incriminating entries had been made separately.

I next considered the question of *cohesion* and *coherence* in the diary entries. Cohesion is a textual observation of grammatical data. In using language we frequently use words of cohesion in order, as the term suggests, to bind together different parts of the text. Thus our first reference to an individual in a text may be through the use of that individual's name. Subsequent references would probably be by use of personal pronouns, such as *he, she,* etc.: the use of pronouns is one way of realizing cohesion in text. Other cohesion devices include adverbs of time, e.g. *then, now, before,* adverbs of place, e.g. *there, somewhere,* etc. Related to *cohesion* is the phenomenon of *coherence.* Coherence is less powerful than cohesion because it does no more than appear to indicate such logical phenomena as cause and effect, sequence(s) of related events, etc. Nevertheless, coherence does have an important role to play in language since we often link events which we consider related to each other, but without necessarily signalling a direct cause and effect relationship at the grammatical level. I will first consider some examples of coherence before going on to discuss instances of cohesion in the diary entries.

Coherence

A possible instance of textual coherence is the relationship between the diary entries for 1 May 2003 and 2 May 2003. In the entry for 1 May, Smith writes that he gave his neighbour 'George' a lift home and is told by George that 'Charleston' has been claiming Mrs Smith was working. On 2 May 2003 he tells how he cut off Mrs Charleston's communal lighting and threw her bin into the river. In the first entry he refers to Mrs Charleston as an 'old swine', and that he was 'furious'. In the second entry he refers to Mrs Charleston as an 'old bat' and says that he *hates* her.

It is possible to infer from these entries that the events of May 1 and his reaction to them prompted the acts of vandalism he writes about. More importantly, however, if the diary had only been re-written to detail criminal acts (culpability for which Mr Smith later denied) then it is worth asking how descriptions of those criminal acts happen to link – on a different date – to mundane, non-incriminating events, such as giving a neighbour a lift and his

reactions to news about another neighbour. Rather, it seems more probable, from the degree of coherence between the two entries, that the events of May 1 relate to the events of May 2, and thus that the events of May 2 were not added, as Mr Smith claims, at a later date.

There are other instances of coherence between incriminating and non-incriminating entries. These entries relate to what I would describe as a sub-type of coherence which is sometimes referred to as *texture*, which is the way in which writers link related events across clause and sentence boundaries in order to add to the overall coherence effect of the text. An example from the diary of this phenomenon of *texture* is the entry for 2 June 2003:

Text Sample 7.2
Deb's 13th driving lesson ☺
"Stalked the night" No tyres tonight, but set fire to a litter bin on Dew Street and (from a discreet distance) watched the fire brigade put it out. ☺
Decided not to "stalk the night" any more. I've had enough fun.
Posted all the documents Olivia will need to apply for her visa.

In this entry Mr Smith details the incriminating activities between two items of non-incriminating text, 'Deb's 13th driving lesson' and 'Posted all the documents'. It would seem that, since the incriminating text is located in this way, then his contention that he added incriminating text to existing entries cannot easily be supported: here he inserts the incriminating text between two innocuous items. This would imply that the 'original' diary – if it existed – had been edited rather than just added to. Moreover, it is curious that he begins his descriptions of the night's activities with a negative 'No tyres tonight'. Since he claimed that his alleged purpose in writing the incriminating entries was to incriminate himself, it does not make sense to begin with a negative – which is, effectively, a denial. Rather, I suggest that he begins with a negative because the statement in the negative is actually true – in other words that he is simply stating a truth here.

Moreover, the phrase 'decided not to stalk the night anymore' also has implications for the contemporaneity issue. First, it shows a possible association in the diary of the phrase 'stalked the night' with criminal acts or, at the very least, an intent to commit criminal acts (although not every criminal act described in the diary is accompanied by the phrase 'stalked the night'). Second, his claim that he decided not to 'stalk the night' anymore is, like the claim relating to 'no tyres tonight', a negative: it describes an intention not to do something. Again, if the incriminating diary entries were about incriminating himself,

there would be no reason for Mr Smith to deny an intention to commit a crime. Since we often find the phrase 'stalked the night' occurring in otherwise non-incriminating entries, I suggest that there is no reason to believe that entries containing 'stalked the night' were not contemporaneous with the non-incriminating entries. This phrase 'stalked the night' occurs for the first time on 29 May 2003, after several incriminating entries had already been made (including those of 19 May and 22 May. It does not occur at all in the 2002 diary, though incriminating entries are made in that diary (*18 February 2002*: 'Had a late night walk down Shakespeare Street & did Green's tyres ...'; *8 November 2002*: 'Did 2 of Green's tyres tonight. This is the last time I'll "visit" him ...'). Innocuous night-time activities are never described by Mr Smith using any reference to 'stalking the night', e.g. 'Visited Mary this evening', 'Went to see Dad ...', etc.

The entry for 20 June 2003 also shows an example of texture:

> "Stalked the night" before bed and torched 3 bins; Dew St, Coulson Avenue & Albright Terrace. "Fireman Sam" was soon in attendence ☺

Here, the phrase 'before bed' is out of temporal sequence to the two other events. In effect the sequence is 'stalked the night' 'and torched 3 bins', and then 'bed'. By placing 'before bed' in between the other two events or actions, the writer is integrating a non-incriminating activity, i.e. going to bed, with the incriminating events. Since the purpose of the diary being re-written was to record incriminating events I suggest that this kind of textural integration shows that the recording of the incriminating entries was actually a part of the writer's daily existence which he 'weaves' in with other, innocuous, events, and that this is a further indication of the contemporaneity of the incriminating with the non-incriminating entries.

Cohesion

Examples of cohesion include this entry from 3 June 2003:

> Ha! ☺ I'm "famous". Stupid bastards. "Vandals". Meanwhile; my weight is up to 15 st. I've got to do something about it.

In this entry the use of the adverb of time 'meanwhile' links the two comments or observations as overlapping events, and I suggest that this indicates contemporaneity of recording.

There is also this diary entry from 6 May 2003:

> Bat-face Charleston has returned & is immediately stirring about the communal lights not working. Ha ☺ I hope she breaks her neck. I'm sure "the Council" will be round soooon, but they can't prove a thing. I was (a little) worried that Charleston walked past Flora's (a shop in the town) just as Deb was inside, but, Deb said she didn't see her. I hope so. My knee is hurting. I'm worried about that too.

This entry opens with a mention of 'Charleston', refers back to the communal lights incident previously described, and then continues by mentioning the possibility of Charleston having seen Deb at work (as he mentions elsewhere in the diary, 'Deb' is not supposed to be at work as she is claiming benefits). In my view this kind of cohesion shows that the non-incriminating incidents in this entry cohere with the incriminating incidents. Charleston in 'worried that Charleston' coheres with previous references to Charleston. The use of 'too' in 'I'm worried about that too', refers back to the previous instance of 'worried'. The cohesion between the incriminating and non-incriminating text appears to be very strong and does not suggest that the incriminating events were recorded later.

As a result I was not able to find anything in the diaries to suggest that Mr Smith had made the two types of entry separately from each other. Thus, I could not see anything to support the contention that the incriminating material was added later. For this reason I believed little credence could be given to the claim that there was ever an earlier diary which had been destroyed by Mr Smith. Later, Mr Smith admitted that he had in fact carried out the acts of criminal damage referred to in his diaries. For this reason there was no trial, but police officers said that without the linguistic report Smith would not have admitted to anything. He was found guilty of 35 acts of criminal damage.

How not to inform: a case of medical disinformation

It seems that large corporations and organizations have, in recent years, taken advantage of modern technology to keep members of the public – especially those who are making a complaint – at arm's length. It is almost impossible to reach decision makers on the telephone, and it can take months, or even years,

to receive a reply to written complaints. Many requests are for information about matters that concern the claimant or complainant in important ways. Yet corporations operate in clandestine ways, often quite outside the spirit of the law, and have, in many cases, devised complex linguistic strategies to avoid providing information that consumers need. This section will illustrate one such case and conclude by discussing some of the strategies used.

Case history

A patient at a large NHS (National Health Service) hospital somewhere in England, we will call him Mr Anthony, was attempting to get copies of his medical records because he believed that a major operation he had undergone in 2003 was unnecessary. He believe that if the surgeon who operated on him in 2003 had properly familiarized himself with his medical history, which related to an earlier operation in 1995, he would not have operated and thus caused him, Mr Anthony, further medical problems.

After a number of phone calls and letters Mr Anthony had still not received all of the information he wanted from the hospital, and – in fact – the hospital had notified him that they were not able to do any more. He then wrote to the complaints unit of the hospital to report the situation and eventually received a letter from the hospital chief executive which exonerated the hospital of any blame.

The paper trail certainly seemed to indicate several inconsistencies. In a letter of 22 April 2003 the hospital chief executive wrote advising Mr Anthony that he had already received his complete medical records. However, according to Mr Anthony the records he had received contained no mention of the 1995 operation. He wrote to tell them this, but did not receive a reply until 16 October 2003, in which the hospital said they were writing to 'provide copies of the records you have requested'. Clearly, then, Mr Anthony could not have received his complete medical records in April. Closer examination of the April letter inadvertently appears to admit as much, because it claims 'we advised that you had received copies of your medical notes ... *in accordance with your request*'. This suggests that there may have been certain records that Mr Anthony was not aware of, and that the hospital was adhering strictly to his non-technical request, in which he would quite probably have failed to request certain records, perhaps because he did not know all of the records which were held – hence 'in accordance with your request'.

Quite often when large corporate organizations deal with private individuals, they are able to use apparently innocuous phrases like 'in accordance with your request' to deny access to information which may not have been in the letter of Mr Anthony's request, but which a reasonable person would have interpreted as having been intended by Mr Anthony. This happens because most complainants are ordinary citizens with no technical knowledge of how large organizations work, of how records are kept, and indeed have little concept of the bureaucratic machinery with which they are dealing. Furthermore, it seems from the April letter that the chief executive is distinguishing between *records* and *notes*, a distinction Mr Anthony may not have entertained as being significant. Although in some cases the distinction between 'medical records' and 'medical notes' may be significant, it could be that – in order to avoid releasing certain types of documentation – the chief executive was exploiting this possible lack of conceptual clarity on Mr Anthony's part to claim that all requests had been acceded to. However, although the phrase 'medical notes' appears innocuous and non-technical, if used in this way – as a type of documentation to be distinguished from 'medical records' – it is clear the chief executive is using the phrase technically, and possibly as a way of denying Mr Anthony information contained in the *records* rather than the *notes*.

Professionals working within such organizations know that they can protect the activities of their organization with such linguistic strategies. When the chief executive says, also in the April letter, that 'We also advised that there were other records relating to treatment in 1995 . . . not disclosed . . . because you did not request them . . . ' this sounds disingenuous since, as it turns out, the treatment Mr Anthony had in 2000 was linked, medically and historically, to the treatment he had had in 1995, and therefore the earlier records were germane to his quest for information. The use of the verb to *disclose* is interesting in this context – '. . . these [records] were not disclosed . . . ' – because to *disclose a record* is a much less frequent use of the word to *disclose* than to *disclose information*. An internet search reveals that the latter usage is approximately 37 times more common than the former.

Using a professional language corpus I discovered that 'information' is in fact the most frequent collocate of 'disclose'. There were no instances of 'records' with 'disclose'. It is also noticeable that the sentence which includes the use of the word 'disclose' is in the passive, viz. 'these were not disclosed'. The sentence does not say by whom the records were not disclosed – this is known as an *agentless passive*. Using an agentless passive enables the writer, the chief executive, to distance himself from that which was not 'disclosed'. Moreover,

since 'disclose' is much more common when used with 'information' it suggests that the chief executive knows that there has been a failure to 'disclose' information. By using the passive, especially an agentless passive, he is able to distance himself from the process. This is also evident in other phrases used by the chief executive in the April letter, for example: 'I understand that you have not made reference to any other records'. This implies that the writer does not know for sure, only that he 'understands', probably through a third party. A similar note is struck with 'I was sorry to learn that you are unable to accept my findings'. This suggests that the chief executive did not 'learn' this direct from Mr Anthony, but from someone else. All of these instances point to the chief executive attempting to distance himself from Mr Anthony's complaint. In his closing sentence the chief executive says 'Regrettably, in the circumstances I must conclude that there is nothing further we can add which might help resolve this matter to your satisfaction'. This suggests further distancing strategies, e.g. 'I must conclude' implies that the writer has done everything in his power, but this is partly contradicted by 'nothing . . . we can add which might help resolve this matter', which implies that the chief executive does not know exactly what will resolve the matter. However, Mr Anthony has been very specific about what would resolve the matter: full sight of his records.

In the letter of 8 October, the chief executive writes: '. . . Mr Smith, Chief Consultant, whom as you know chaired the Local Resolution Meeting and has been absent on leave . . . on his return we shall look further into your comments'.

I wondered why the writer says he has been '*absent* on leave'. 'On leave' implies that the person being referred to *is absent*. It could of course be that he 'has been' on leave, has finished his leave, but is still officially 'absent'. This suggests that the writer might not wish to state that the Chief Consultant is actually available for discussion, and could therefore be a device for delaying the access to Mr Anthony's records which, in any case, become available by the time the letter of 16 October is written by the chief executive, scarcely more than a week later.

Mr Anthony also requested a copy of the videotape of the 'Local Resolution Meeting' because he felt this tape contained important information about his treatment. Referring to the videotape the chief executive states that the purpose of the tape was to facilitate the preparation of notes of the meeting, and that thereafter the tapes would be wiped. He actually says: 'Once this material has been completed then practice is for the tapes to be wiped and reused'. The chief

executive does not state that the tapes were actually wiped, just that it is the practice to do so. In fact he does not say it is 'the practice', but that 'practice is'. In my experience, when words such as articles (*the, a, an*) are omitted it is sometimes a sign of evasion. I would suggest it is highly possible that, given the degree to which Mr Anthony had been complaining to the Trust, that in fact the tapes may not have been 'wiped', but could still exist. Moreover, as with '*absent* on leave' there is some redundancy in the phrase 'wiped and reused'. Why would the tapes be 'wiped' if they were going to be 'reused'? It is not usually necessary. If it is not usually necessary, why is it 'practice'? Finally on this point, why does the writer then add 'Regrettably, we are unable to provide you with copies of the tapes'? He has already implied this. This could indicate that the tapes, or copies of them, do exist. I offer this suggestion because a frequent issue in veracity testing is what is known as 'quantity of information'. Quantity of information becomes an issue in veracity testing because people who either deliberately dissemble, or who mask the truth in indirect ways, frequently provide more information than is strictly required. This would be particularly apt when considering the text of a highly educated individual such as the chief executive. Given the responsibilities and duties of such an official, it is likely that he would know, better than most other people, how to weigh his words with a high degree of care.

However, we can see that Mr Anthony has still not received all of his records, because in the 16 October letter the chief executive says: ' ... I am advised (by whom? In what context?) that you have now had full access to all documents that are applicable under the given circumstances'. This indicates that there are still records which Mr Anthony has not received, only those that are 'applicable'. Doubtless there is a hospital or National Health Service guideline about what is applicable under what circumstances – but how would Mr Anthony know this? The chief executive's words indicate that there may be records which are not being given to Mr Anthony, because they are not documents that conform to the 'given circumstances'. However, he does not state what these documents are, nor what the circumstances are, nor, crucially, how Mr Anthony could find out how to obtain any documents he does not already have.

Moreover, the excerpt 'Your request for other records in any physical, electronic or other forms as permitted by the Data Protection Act has been addressed ... ' could be taken to mean that the Data Protection Act restricts, in Mr Anthony's case under 'given circumstances', access to certain records, when in fact the sentence – as I read it – seems to mean nothing more than that

the Data Protection Act restricts the forms which records may take (physical, electronic, etc.). By then coupling this with ' ... I am now advised that you have had full access to all documents that are applicable under the given circumstances' the reader could be forgiven for thinking that there are restrictions under the Data Protection Act and that the writer is claiming he is following the Data Protection Act, and is in fact mandated by it to restrict access to further records.

There was little doubt that Mr Anthony had suffered delay and prevarication at the hands of the chief executive of the hospital concerned, who through his use of language appears to have employed stratagems designed to obscure the truth as to what Mr Anthony was entitled to know, and which records he was entitled to view, and thereby avoiding a timely resolution to vital matters relating to Mr Anthony's health and well-being. At the very least the organization lacked transparency. Large organizations have the ability to hide behind rules, regulations and procedures which are opaque to the lay person. Even the process of communicating with them is made difficult by modern technology (e.g. telephone systems with a confusing array of input options 'If you want X please press 1, 2, 3', etc.). Organizations can obfuscate, delay, confuse, resort to technical language which sounds like everyday language, and generally play the corporate game with little fear of being brought to account. Fortunately, close linguistic analysis can reveal the kinds of linguistic strategy in use by corporations. In the present instance the strategy appears to be fourfold: to (i) use common language which is also technical language in the context of the organization's activities; (ii) imply, by ambiguity, that there is legislation which controls or restricts the kinds of information to which the consumer is entitled; (iii) employ definitional categories using semantic terms with which the consumer is unfamiliar and (iv) avoid an implied meaning by resorting to literal terminology.

Common linguistic techniques to achieve the above strategic aims include the use of the following: lexical and clausal ambiguity partly by exploiting polysemy, agentless passive constructions, long sentences consisting of multiple clauses often with deferred verbs or objects and clauses with 'fuzzy' scope. Clearly the hospital has violated all the basic precepts of plain English usage and has used linguistic means to achieve what can only be termed some form of institutional abuse.

Postscript: Mr Anthony pursued the hospital trust to court and, at the last minute, the hospital agreed a settlement with him. Hopefully, forensic linguistics played a small part in this procedure.

Table 7.1 List of strategies to achieve non-informational ends

Linguistic device	Example	Comment
Uses common language which is also technical	'notes', 'records': is there a difference?	The hospital is able to exploit a semantic distinction which the consumer may not know about.
Implication that legislation restricts the permitted response	'Your request for other records in . . . other forms as permitted by the Data Protection Act has been addressed . . .'	The Data Protection Act is designed to protect the consumer – yet here we have an ambiguity that suggests the 'request' is permitted by the Act, rather than the form in which the record is being held. Legislation is being invoked to restrict what the consumer is entitled to know about *him*self.
Employ definitional categories using semantic terms with which the consumer is unfamiliar	'I am now advised that you have had full access to all documents that are applicable under the given circumstances . . .'	What kind of documents are applicable under what circumstances? How would Mr Anthony necessarily know what those circumstances are? Hence 'applicable' and 'given circumstances' are being used to illustrate categories of which Mr Anthony is not aware.
Avoid an implied meaning by resorting to literal terminology	'. . . in accordance with your request'	Here the hospital is taking Mr Anthony's request literally, yet knowing that there are technical senses in which he has used words which he takes to have only an ordinary meaning in the context.

Investigating claims of memory for 'verbatimness'

A landlord in Manchester, England, rented out a property to a care home company owned by, let us call them, Mr and Mrs Artemis. After a while, the care home company stopped paying rent to the landlord, let us call him Mr Wheeler. On ceasing to receive rental for his property Mr Wheeler sued his tenants. The tenants counterclaimed that Mr Wheeler had made false representations to them, in effect inducing them to enter into the leasing of his property. Mr Wheeler, a retired barrister (attorney) with an impeccable record, was

stunned at this allegation. Among their witnesses the tenants produced a local shopkeeper and his assistant. These witnesses claimed that Mr Wheeler had asked them to persuade the tenants to enter into the lease with him and to purchase the care home from its previous owners. Mr Wheeler submitted copies of the witness and defendant statements which contained examples of the language used by those individuals in their counterclaim.

Among the claims by the defence were these alleged quotations of Mr Wheeler's in his efforts to ensure that his property was leased by Mr and Mrs Artemis:

> (a) '. . . it's a very *good* thriving nursing home and an opportunity not to be missed . . . '
>
> (b) 'You will be missing out on a *good* opportunity . . . '
>
> (c) That the care home 'was a *good* little business . . . '
>
> (d) '. . . was a *good* business opportunity not to be missed . . . '
>
> (e) '. . . that the property was in a *good* state of repair . . . '

What interested me from the outset about these quotations was that they were quotations at all. How was it that the tenants used direct quotations and so did their witnesses? How was it that the wording was so alike? It seemed to be clearly the case that the defence were intending to signify that (a) to (e) were direct representations of Mr Wheeler's words, that is to say the exact words he supposedly used. This raised several important issues:

(1)   The nature of 'lawyer' language
(2)   Limitations of memory for exact wordings
(3)   Generic quality of quoted wordings
(4)   Persuasive content of words.

Lawyer language

In general, I have found lawyers to be somewhat cautious in their use of language, especially lawyers of Mr Wheeler's generation. Not only are they usually cautious about the content, but they are also somewhat conservative about the kinds of structures they use. Thus, in the present instance I would find it somewhat surprising for a barrister to say something like 'a very good thriving nursing home and an opportunity not to be missed', since (a) the first clause renders the second virtually redundant and (b) the qualifications to the noun

phrase 'nursing home' are somewhat awkward in structure, viz. 'very good thriving'. In addition, I also find the phrase 'very good thriving nursing home' something of a tongue-twister and not easy to say aloud, especially when 'and an opportunity not to be missed' is added to it. In fact, it has more of a written than a spoken feel about it. Moreover, most lawyers have a wide vocabulary, by reason of education, training and experience. I would be surprised if a senior lawyer of more than 30 years' experience was not able to offer anything more apt than 'good'. Some of the expressions even border on the trite, e.g. 'good little business'. All in all it did not seem credible that a lawyer would use this kind of language – not once, but many times. It certainly bore little resemblance to the telephone conversations I had had with Mr Wheeler, and even less with his written correspondence.

Memory for language

It is not controversial among linguists that memory for exact wording of what has been said is poor. Most people are unable to remember their exact words even a few minutes after having said them. Many psychologists and linguists have reported on this phenomenon, including MacWhinney *et al.* (1982), Hjelmquist (1984) and others. By contrast, the defendants' memory for times when important matters relating to the negotiations took place appears poor, with phrases like 'on several occasions', 'a later discussion', 'later proved to be untrue', etc. I asked Mr Wheeler to try to obtain accounts in the defendants' own words – unaided – either written or spoken, but preferably spoken. However, the defendants and their witnesses, on the advice of their solicitor, declined to provide this information.

Generic quality of quoted wordings

The alleged exact memory for the wordings listed above is all the more remarkable when one considers that the wordings themselves are entirely unexceptional. They are in fact somewhat generic with 'opportunity', 'business' and 'good' appearing several times. In support of this most corpus studies will reveal that 'good' and 'opportunity' are high-frequency collocates, as are 'missed' and 'opportunity'. Typical examples of generic language used in cases where veracity is an issue include precise times (e.g. 9am, 5.30, etc.), exact periods of time (e.g. 18 months, a fortnight, etc.), generic personal aliases (e.g. John Smith, Fred Brown, etc.), generic meals (e.g. 'pizza and chips') and generic excuses (e.g. 'my car won't start', 'I've got a headache', etc.).

Persuasive language

The defendants claimed that Mr Wheeler's words swayed them into buying the business and renting his property. However, there is little in the words themselves to persuade, especially given their content. There is no claim that Mr Wheeler showed the defendants documents to illustrate the success of the business or sought to persuade them in other ways of the necessity to make the investment or otherwise attempted to induce them to enter into the contract with him.

Section conclusion

Many witnesses, including on occasion police officers, will claim to be able to memorize precisely what another person said. In one case, for example, a number of police officers claim to have remembered verbatim, independently of one another, a complex utterance by a suspect several hours – and in one case three months – after the utterance was made. However, studies show that once you reach approximately seven words, it is very difficult for humans to recall exact wordings. It is also advisable to take into account any other linguistic circumstances, for example the generic nature of the language in the above example, the context, the educational and professional level of the speaker, etc.

In the next section I will discuss the last case in this chapter. This concerns some of the issues raised so far, including memory of language and similarities between apparently independently produced documents.

An affront to justice: Prosecutor memo leads to abuse of process ruling

Introduction

This was an important case because it was the first time, as far as I know, that forensic linguistics was used in an abuse of process case in a criminal trial in the UK. Abuse of process cases are themselves relatively rare, since what is at stake is the reputation of the judicial system itself. In an abuse of process case the judicial system is essentially examining itself, its own values and ethics. It is not something a court undertakes lightly. For this reason, I will be going into the case in some detail.

Mary Smith[15] was driving her car down a two-way street through the town of Middle England when, owing to a temporary distraction, she swerved and went over the central dividing line. A police officer was in a patrol car behind her and responded by signalling her to stop.

It appeared to the officer that Ms Smith's breath smelled of alcohol, so he requested Ms Smith to accompany him to the police vehicle where he could breathalyse her. She got into the back of the vehicle and was handed the mouth-piece. However, despite several attempts she was not successful in giving a satisfactory sample. As a result, the officer arrested Ms Smith and took her to Middle England police station where the breathalysing process was again attempted, but this time using the station equipment, known as a Lion Intoxilyzer 6000. Again, Ms Smith failed to provide a sample. According to the officers several attempts were made to breathalyse Ms Smith, but she was not co-operative. According to Ms Smith she was unable to provide a sample because she was hyperventilating and in a state of panic.

The officers charged Ms Smith with the offence of failing to provide a sample.

The officers' statements

The case came to the Middle England magistrates' court and was heard in front of three lay magistrates. The alternative in the UK for some hearings of this kind is that the case is heard before a district judge (this functionary used to be known as a 'stipendiary magistrate'). In this case, however, the case came before three lay magistrates.

Among the evidence tended by the prosecution were two statements from the two officers who had attempted to breathalyse Ms Smith. Upon disclosure (prior to the trial) what struck the defending barrister (attorney) was the similarity between the two police statements.

He questioned the first police officer as to the similarities. The officer maintained that the similarity was purely co-incidental. Asked if he could account for this 'co-incidence' the officer replied that he could not. Asked if he had shown his own statement to the second officer the first officer said he had not and that there was no possibility that the second officer could have seen his, the first officer's statement. The barrister was nothing if not determined. He pressed the officer for a satisfactory answer. Eventually, the officer said the only thing he could think of which might account for the similarity was a memo he had received from the prosecutor. The barrister asked him what memo he was referring to, pointing out that no memo had been disclosed to the defence.

The memo

The officer reiterated that he had received a memo from the prosecutor. The defence barrister requested the court to ask the prosecutor to disclose the memo. The prosecutor said he was unaware of any memo. He appeared to look through the bundle of documents in front of him but could apparently find no memo. The defence barrister offered to help him. This offer was declined, but the chair of the bench then asked if perhaps the witness could look through the bundle since he would surely recognize the memo more easily. The bundle of prosecutor documents was handed to the officer as he stood at the witness box and he immediately found the memo. The memo was copied and all parties were given a copy.

What the memo said

The memo purported to be a request from the prosecutor to the disclosure officer to make disclosure of several items which the defence required in order to run the case. This list included the mouthpiece, any notes the officers may have made in the course of Ms Smith's detention at the police station and, crucially, a request to the officers to provide Section 9 statements (i.e. witness statements) on the breathalysing process as it had applied to Ms Smith. These were the statements upon which the defence barrister was examining the officer when the fact of the memo's existence was first mentioned.

However, the memo was extremely derogatory to the defendant and her legal representatives, and in fact represented the task of collecting this evidence as a 'pain' – effectively a nuisance to the officers and the prosecution.

The defence immediately took issue with the memo and requested permission from the court for an adjournment so that an expert could be called to look at the memo and the two statements. This adjournment and the request for an expert was immediately granted by the bench.

The analysis of the memo

The memo was given to a forensic linguist, along with the two statements and an analysis was made. The key points of the analysis were these:

(1) The memo does not appear to be a request for a disclosure, but a demand for the disclosure to follow the prosecutor's prescription. For example, the memo requests the officers to provide statements '*that* Ms Smith was not suffering from any form of hyperventilation or panic attack'. The linguist in the case considered that this was not an objective, impartial request, but a

direction to the officers to state '*that* Ms Smith was not suffering from any form of hyperventilation', etc. Similarly, the memo asks the officer to provide details of his experience in operating the breathalyser equipment, but this is also phrased as a demand. The prosecutor says 'I want to show you are an experienced operator [of this equipment]'. However, at the time of writing there was no way that the prosecutor could have known if this were true or not. The memo also contains a presumption of the doctor's findings: 'Can I have an explanation as to why the doctor was called to rebut the defence assertion that it was due to the "shock/hyperventilation".' The linguist considered that this is really a question followed by a statement of intent: (i) Can I have an explanation as to why the doctor was called? And (ii) there is a purposive infinitive 'to', which can be rendered as '[in order to] rebut the defence assertion that it was due to . . . hyperventilation'. In other words, properly parsed, the memo appeared to say: 'Can I have an explanation as to why the doctor was called [because I wish] to rebut the defence assertion that . . . ' The legal view was that the prosecutor should not have disclosed the defence's arguments to the officer because the officer was also a witness. The linguistic view was that the prosecutor was informing the officer of how he intended to counter the defence case, but also that he had apparently done so in a confusing and ambiguous way which could have been designed to obscure this fact from the officer's notice.

(2) In this memo the prosecutor also tells the officer that he should expect questioning regarding the issue of whether he, the officer, acted unreasonably in refusing to allow the defendant the opportunity to give another sample.

(3) Additionally, the prosecutor also gives the officer unnecessary information regarding defence disclosure.

(4) Finally, the memo talks about the defendant and her defence team in the most disparaging and insulting terms, and, effectively, invites the police officer – who, it must be recalled – is a witness to the alleged offence – to share this view. For example, he writes: 'Enclosed is a copy of a defence case statement received from the defendant's solicitor (a lot of nonsense in a vain attempt to avoid the inevitable) which sadly we have to respond to'. There are several points at issue: (i) the prosecutor should not, unless under exceptional circumstances, be disclosing a defence case statement to a witness. Regrettably, the officer is not only a witness he is also the disclosure officer. This, it was claimed, is not best practice, (ii) at the linguistic level, the prosecutor uses inclusive 'we', thus inviting the officer to share his view that the defence case is 'nonsense' and hence that the task of collecting the evidence is a nuisance. Added to the unnecessary information the prosecutor gives the officer in connection with the defence arguments and evidence, this invites the officer to take a hostile attitude to the defence and forewarns him how the defence will be run, possibly creating an opportunity for tainting his evidence.

The trial re-convened

The trial was reconvened at the Middle England magistrates' court in front of the same bench as before. The linguist was called and asked to comment on the memo. The linguist outlined the arguments in his report and was examined and cross-examined on them. There was no prior examination on the linguist's admissibility as a witness because the prosecution had not objected to the linguist appearing.

The linguist explained the linguistic issues, which were as follows:

(1) A linguistic analysis of the memo was valid because (i) it is not only what we say that counts, but the context in which we say it. Thus, it might not be wrong for the prosecutor, if having a conversation with a colleague (who was not also a witness) to exclaim privately out of earshot from any of the parties that 'the defence case is a lot of nonsense'. Also, it would not be wrong for the prosecutor, if cross-examining the defendant, to say, 'Madam, that is a lot of nonsense' or to say to the bench, 'Your worship, in my view the defence case is a lot of nonsense'. However, the prosecutor was writing to a witness in the case, and was inviting that witness to share his view, with the use of inclusive 'we'. This is the second linguistic point: speakers can include or exclude others by the use of 'we'. In the present case both the prosecutor and the witness have certain duties which they have to carry out under the law. This is certainly a valid use of 'we'. However, the prosecutor went beyond the strict allocation of functions to the witness and invited him to share a derogatory, partial, view which was – crucially (in terms of prosecutorial codes of conduct) – anything but dispassionate. A prosecutor may have contact with a witness prior to a trial, but in his/her dealings with a witness must be dispassionate. This applies particularly to comments about a defendant. The prosecutor must not contaminate a case by commenting on such matters to a witness.

(2) The linguist claimed that the way in which the prosecutor requested the evidence hardly constituted a request, but was in effect a demand for the presentation of evidence in a particular way. This applied to the 'that' clause previously referred to, and to the officer's status as an 'experienced operator'. Nobody disputed that the officer was probably experienced, but the prosecutor should have been more circumspect in his phrasing, saying, for example, 'Could you please provide me with any information as to your experience as an operator of this equipment' or, even more plainly, 'What is your experience as an operator?' In court the linguist claimed that the prosecutor's sentence, 'I want to show you are an experienced operator', was the equivalent of a leading question. Few judges or magistrates would allow counsel to say something like, 'Constable X, tell us about your qualifications with this breathalysing equipment because I would like to show the court you are an experienced operator.'

(3) It was also noted that the prosecutor, by using deprecating language to an official who was, in the sphere of such issues as 'chain of command' an inferior functionary, and therefore not at liberty to contradict the prosecutor, effectively put the officer under pressure. This was, in a manner of speaking, the equivalent of a junior employee having to put up with a boss's racist or sexist jokes. All of these points together, it was claimed, meant that the officer was under pressure to conform to a set of instructions in a particular way, and was not free to gather his evidence in an impartial, objective manner.

(4) The defence asked whether the close lexical and phrasing similarities across the two statements showed causation, in the sense that the statements – which used some of the same phrases as the memo and a number of phrases which were identical to each other (one or two of which were very rare phrases) must have been influenced by the memo. The linguist replied that there were strong similarities, but that causation here was a legal issue, rather than a linguistic one, but that insofar as *both* witness statements followed the structure and wording of the memo, there were strong grounds for believing that the language of the memo influenced the language and structure of the statements.

The cross-examination

It should be noted that the prosecuting attorney in this 'failure to provide a specimen' charge was not the same prosecutor who had written the memo.

The prosecutor raised several very interesting points:

(1) He pointed out (in the form of 'Wouldn't you agree with me . . .') that the fact that the statements contained many similar phrases could be attributed to the fact that police officers speak and write in a 'jargon-like' manner, i.e. that they use police register.

(2) He stated that, since police officers are a hardworking community, it was quite reasonable that an officer witness could also be a disclosure officer in a case. He suggested that police officers can make up their own minds about such matters and would not necessarily be liable to influence.

The linguist replied to these two points as follows:

(A) It is true that practitioners of different professions develop a kind of jargon or register, and this is particularly true of members of highly institutionalized professions such as police officers. Nevertheless, this could not explain the joint occurrence of several phrases of more than six words in length, as well as one or two very unusual phrases, e.g. 'non-verbal signs'. The officer who had appeared on the witness stand had not used the phrase 'non-verbal signs', but had spoken of

'body language'. Moreover, there was a very high common lexical content between the two statements.

(B)   The linguist had no comment on the propriety or otherwise of witnesses acting as disclosure officers, since this was an operational or even a legal point rather than a linguistic one. Nor would he comment on whether police officers are any more able or not able to make up their own minds regarding their views of a case, a defendant in a case, or that defendant's legal representatives. He would simply point out that there is a memo, the memo is written in certain terms, and that the statements follow the structure of the memo almost precisely, that they contain many similar and identical phrases, and that the two statements are so like each other as to make the probability of their having been independently produced exceedingly low.

About 'abuse of process'

Following the linguist's evidence the defence requested that they be able to file a motion for abuse of process. Defence described how the question is laid out in the Police and Criminal Evidence Act, 1984 (Section 78). This section allows the exclusion of evidence if the result would mean unfairness to a defendant. The defence barrister argued that admission of the police statements would contradict the provisions of Section 78, and would result in an unfair trial. This provides that officials cannot abuse their role in a prosecution and, crucially, that they cannot act in 'bad faith'. The Code for Crown Prosecutors states that prosecutors 'must be fair, independent and objective', that they should provide 'guidance and advice to investigators' and that they have a duty to put 'all relevant evidence ... before the court'. However, prosecutors 'should ... tell ... police if they believe that some additional evidence may strengthen the case'. An important phase in the decision process of whether to go forward with a case is the evidential stage. A crucial aspect of this is the way in which evidence is gathered. This is because the court can apply a number of rules to test whether evidence should be excluded or not. When considering an application under Section 78, the court will first consider how the evidence was obtained. A breach does not automatically entail exclusion of evidence. Rather, the breach has to be 'significant and substantial'.

In R vs Derby Crown Court, *ex parte* Brooks, Lord Ormond defined abuse of process as a case where 'the prosecution have manipulated or misused the process of the court so as to deprive the defendant of a protection provided by law ... '

In reality, the granting of an abuse of process is a rare legal event. According to Dilhorne, in DPP vs Humphrys (1977), proceedings are stayed only 'in exceptional circumstances'. This comment was echoed by Lord Lane when he was Attorney General in 1990, while Lord Justice Stuart-Smith noted that 'it is a power that should only be exercised sparingly'. One Canadian decision noted that in order for a stay to be granted 'the fundamental principles of justice which underlie the community's sense of fair play and decency' would have to be violated.

In the present case the issue was whether misconduct which undermines the rule of law had taken place, and whether this in turn would constitute an affront to justice.

The prosecution countered by claiming that any such issue could be dealt with in the course of the trial, and that the memo, though clumsily worded, was not an abuse of process because there was no way of knowing what effect it had on the officers' conduct. He also claimed to quote case law stating that an abuse of process ruling to exclude evidence could only be granted if the evidence had not yet been adduced. Since, however, the officer had already begun his evidence, it could not be applied in the present instance.

Defence was able to respond to these points and to put the memo in a much stronger light, especially given the linguistic analysis. He reiterated the previously mentioned linguistic points about context, inclusive 'we', declarative demands, and embedded multiple implicatures. He pointed out, as did the justices' clerk (herself a solicitor), that prosecution was not correct to say that evidence could only be excluded before it was given, since this would in itself violate natural justice.

The ruling

The magistrates considered the application over a two- to three-hour period, and were able to avail themselves of the services of the justices' clerk in the matter. On returning to the court they ruled that the memo had abused the trial process and that an affront to justice had occurred. The case was dismissed and costs were awarded to the defendant, Ms Smith.

The significance for forensic linguistics

This is a significant ruling for forensic linguistics for a number of reasons. First, it is believed to be the first time that linguistic evidence was given in an abuse

of process case. What was demonstrated with this case is that when we speak or write we also *act*. If we are in a position of power we can get others to act. Therefore, to consider the memo as harmless would be to miss the fact that language is action, and – in this case – that it is a type of action which causes other actions. If those actions are illegal or unethical, then it is the causative action to which we should first turn our attention.

8 | Authorship profiling
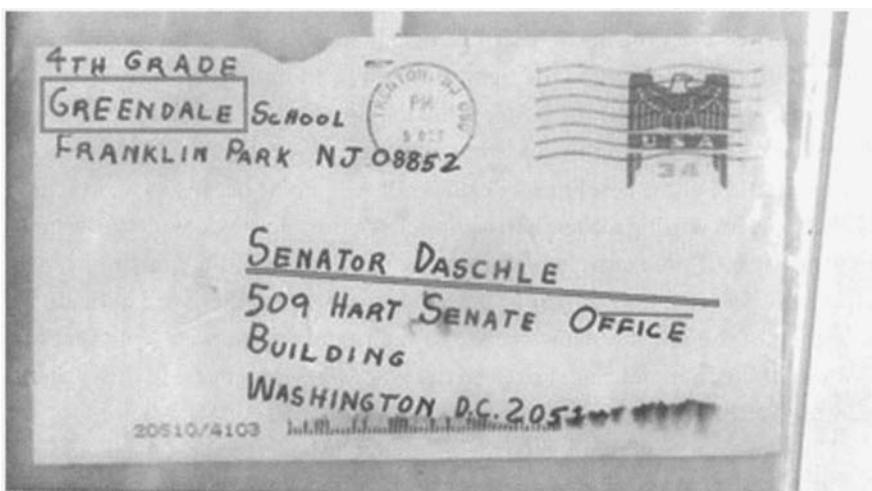
A further tool in the forensic linguist's authorship armoury is now considered: authorship profiling. The reader will be aware that this is not the same as psychological or personality profiling. In other words the focus is not on exploring the writer's 'personality' or 'character'. These constructs have no meaning for the forensic linguist, whose sole interest here is in obtaining linguistic information about the writer, some of which may be helpful in the course of an investigation. In the following section this will be illustrated with an example.

The anthrax 'scare'

The reader will be aware of the anthrax scare which followed the 11 September 2001 attack on the Trade Center and the Pentagon. At that time envelopes – disguised to look like harmless communications from school children – were received by several prominent public individuals: these envelopes contained the poison anthrax.

At the time it was thought by a number of commentators that the anthrax attacks formed part of the alleged Al-Qaida assault on America, and among the first questions asked – regarding the envelopes and their written content – was whether the writer was (i) a native speaker of English or some language which used the Roman alphabet – possibly English – (abbreviated L1 = 'E'), or (ii) a speaker of one of the languages which uses Arabic writing for its script – a list which includes Arabic, Farsi, Urdu and Punjabi as spoken in Pakistan – (abbreviated L1 = 'A'). A key part of the question is, of course, the role of linguistic evidence in such an inquiry.

It was decided to attempt to answer this question with reference to the envelope and text sent to the leader of the US senate, Senator Daschle, a copy of which is shown below:

Note that some consideration of the handwriting here is necessary, but this should not be taken as a handwriting analysis, or as an exercise in what is popularly known as 'graphology'. Those subjects are not usually of any interest to forensic linguists. Rather, what is of interest is the connection between handwriting and language – and the use of both as *linguistic* evidence. The following paragraphs should clarify the distinction.

Style of printing

The first task was to examine the style of printing: we might be tempted to think that L1 = 'A' because of the apparently very simple style of printing,

i.e. large, rounded letters sloping downward. It might be imagined that this is an Arabic-script L1 speaker who does not have much experience of English and therefore needs to print, rather than write cursively. However, the printing is far from uniform in style: it shows a mixture of large and small upper case letters and, indeed, some of the letters are almost lower case in appearance, e.g. the 'o' letters in *school.* This mixture of large and small capital letters might indicate that the true writing style of this writer could be a 'proper' cursive style in which upper and lower (or 'title') case letters are used. This could signal that the writer is possibly a native speaker of English (i.e. L1 = 'E'), or has been well educated in English and is simply printing in order to disguise his/her handwriting or to create the impression of a child's handwriting (hence the text relating to '4th Grade' in the upper left-hand corner of the envelope).

However, there is another factor which might contribute to our thinking of this writer as a non-native user of English and that is the apparent overall inclination of the writing to the right. It might be thought that, since Arabic is written from right to left the writer is having problems adapting to a left-right orientation, thus causing the baseline to wander down to the right. However, the inclination of the baseline angle does not seem to be because of a difficulty with left–right writing rather than right–left writing, it is to do with the writer's natural style. If an example of this writer's genuine writing (in the Roman alphabet) is ever found, it seems almost inevitable that there will actually be a sloping of the script downward and to the right in some way. Note that the bottom of the right descender of the letter A – always upper case in this writer's printing of the envelope – is always lower than the left.

The same is true of other letters with what may be termed right descenders, e.g. R, H, N, K. These all demonstrate, not that they intrinsically slope downward or rightward in the case of this writer, but merely that their right descender is lower than their left. Note that some of the lines, when ruled horizontally, are not that steeply inclined: it is not as if there were a continuous sloping down to the right. It is rather that the writer appears to have the tendency to finish letters with the right descender lower than the base position of the left descender.

Lack of punctuation

The fact that a return address (albeit fictitious with an incorrect or false zip code) is provided neatly top left shows that the writer seems comfortable with writing letters of different sizes. Moreover, the shapes of the letters are consistent

across the different sizes. This writer does not have any difficulty with the Roman alphabet.

The false return address, incidentally, might be evidence of the writer's overall deceit, i.e. to use the evidence of the text of the envelope to deceive: this indicates that the style of handwriting is also being used to deceive, i.e. textually – that is to say it appears to be confirmatory of the style of writing/printing being disguised rather than natural. One other indication of the writer possibly being Arabic is an apparent underuse of punctuation. Other analysts, e.g. Jonathan Charteris-Black have observed that Arabic speakers (among others) of English tend to underuse punctuation in English. However, the Daschle envelope does not contain many opportunities for punctuation, other than, for instance, 'N.J.' and 'D.C.' both of which appear to be punctuated as abbreviations in any case. This means that in reality there is no basis for believing that this writer is either of Arabic origin or has an Arabic script L1, such as Farsi, Urdu, Pakistani Punjabi, etc. However, there is one clue which has so far not been discussed in any detail.

Mix of large and small upper case letters

Above, the appearance of the mix of large and small upper case letters in the address was discussed. This may have come about because of the kind of address found, for example, on a number of internet pages, i.e.: 'Senator Thomas A. Daschle. 509 Senate Hart Office Building … ' Here the text is a mix of upper and lower case (a style known as title case). This could be the kind of candidate material for the source text of the address on the Daschle envelope. The writer may have felt there was not enough room for the Senator's first name or initials, hence these were omitted. However, even despite such an address being a possible candidate for the source of the envelope, it still does not answer the question – why did the writer transpose large and small upper case letters for title case?

Familiarity with Roman alphabet?

The apparent transposition of upper and lower case printed letters, into large and small caps (i.e. upper case letters), a completely different kind of printing, seems very curious: it requires much more thought than straightforward

upper/lower case copying, which is in any case closer to 'normal' handwriting. This is hardly 'least effort'. However, the fact that the writer combines large and small caps so well is almost definite proof of the writer as an L1 English speaker. Here there is evidence of someone who is able to transpose upper and lower case into large and small caps, seemingly without thinking about it. This is clearly a person who knows how to manipulate the language, even down to its most basic elements, in this case something evidently quite banal and simple, i.e. the use of printed letters in a particular way.

The contents of the envelope are interesting, in particular the way the date is written: month + day + year. This is the standard American way of writing the date. The British, for example, use the day + month + year format, while the Chinese use the year + month + day format. Others often use the year + day + month format. However, most Arab countries use the same form as that used in the USA. The full text of the letter inside the envelope (according to www.dawn.com) reads: '09-11-01. You can not stop us. We have this anthrax. You die now. Are you afraid? Death to America. Death to Israel. Allah is Great.'

This is very similar to that received by Tom Brokaw, which reads: '09-11-01. This is next. Take penacilin[16] now. Death to America. Death to Israel. Allah is Great.' Note the terseness of the style. It is far from easy for a learner of English to use the language in this concise, precise way. Moreover, it is probably indicative of someone with a good education and – paradoxically – someone who is used to doing a lot of writing. The misspelling ('penacilin') and the pseudo-pidgin style 'You die now' are probably just red herrings and should be ignored.

Chapter conclusion

In this chapter a brief demonstration of author profiling was given. Given the necessary emphasis on some aspects of presentation, the analysis may have sometimes appeared not unlike an analysis of handwriting – which it is not.

It should also be stressed that this analysis did not relate to psychology or personality profiling, topics as to whose merits the author offers no opinion. Rather, author profiling relates to the author's use of language and what it tells the analyst about the writer linguistically. As such it is still an inexact science, and one which – like other branches of forensic linguistics – will require the benefit of case examples in order to be of value. However, it may have value

in certain kinds of investigations where risk needs to be assessed, for example, stalking, terrorism, hate mail, etc.

Discussion

The texts being discussed here bring up the issue of credibility in text in general. When we read something, what do we believe? What does the text attempt to make us believe? Does it succeed or fail? Why/why not? Attempt to answer these questions with the excerpt from the anonymous smear mail note below:

DEAR BILL,
I SUPPOSE YOU THOUGHT I WOULD FORGET BUT YOU ARE WRONG HOW COULD I FORGET A RAT LIKE YOU. I HAVE SENT A LETTER WITH ALL YOUR PAST DETAILS TO THE PRESIDENT. ALL YOUR DEBTS AND PAST MISSDEMEANOURS. IF YOU DON'T RESIGN FROM THE COUNCIL IMMEDIATELY THE PRESS WILL PRINT A LIST OF ALL YOUR DEBTS BOTH LOCALLY AND NATIONALLY .... YOU MIGHT BE ABLE TO FOOL SOME PEOPLE BUT NOT ME. YOU FORGET I HAVE KNOWN YOU FOR ALL OF YOUR LIFE.

9 Plagiarism

Chapter Outline

Popular definitions of plagiarism are irrelevant to linguistics although the subject is important to linguists – whether working in the forensic field or teaching. Hence in this chapter I am less concerned with moral questions of plagiarism and will instead focus on what happens in plagiarism linguistically: what do plagiarists *do*? – that is the question. We could put this question another way: what is plagiarism linguistically? One way of finding out is to consider not just whether someone has plagiarized from someone else, but who has plagiarized from *whom*?

Direction of plagiarism

Most plagiarism inquiries are concerned more with detecting plagiarism than with its direction, which is often taken as a given. In fact, observing differences between prior and subsequent texts can be instructive in looking for the kinds of change which occur in language as a result of a plagiarism process.

Because most plagiarism investigations are concerned with detecting whether plagiarism has taken place, and because it is usually known which of two texts preceded the other, plagiarism direction research is by and large an unexplored country. Deardorff attempted to apply a Bayesian model to the question of whether the Matthew gospel had been plagiarized from the Talmud of Jmmanuel or whether plagiarism had occurred the other way around (see Deardorff, 2003, OL). However, this method requires long texts, and in the forensic context I doubt that it has been used. In order to avoid boring the reader I will consider just three well-known plagiarism claims, with a view to understanding what plagiarism is *linguistically*.

Margaret Canby and Helen Keller

The first claim I want to consider is that famous blind author Helen Keller's *Frost King* was plagiarized from Margaret Canby's somewhat earlier *Frost Fairies.*

At the time this accusation surfaced Helen Keller was a child of 11 years of age. In sending the story to a friend as a thank you present, Helen had not asked for it to be published. It was the well-meaning action of the recipient that caused all the trouble. Helen claimed that she must have read the story years before and unwittingly used its contents.[17]

Excerpts from Margaret Canby's text are given on the left of the table below, with Helen Keller's on the right:

Text Comparison 9.1: Canby and Keller excerpts

Margaret Canby	Helen Keller
He has two neighbours, who live still farther north; one is King Winter, a cross and churlish old monarch, who is hard and cruel, and delights in making the poor suffer and weep; but the other neighbour is Santa Claus, a fine, good-natured, jolly old soul, who loves to do good, and who brings presents to the poor, and to nice little children at Christmas.	
Well, one day King Frost was trying to think of some good that he could do with his treasure; and suddenly he concluded to	Well, one day King Frost was surveying his vast wealth and thinking what good he could do with it, he suddenly bethought

<div align="right">(cont.)</div>

Text Comparison 9.1: Canby and Keller excerpts (*continued*)

Margaret Canby	Helen Keller
send some of it to his kind neighbour, Santa Claus, to buy presents of food and clothing for the poor, that they might not suffer so much when King Winter went near their homes. So he called together his merry little fairies, and showing them a number of jars and vases filled with gold and precious stones, told them to carry those carefully to the palace of Santa Claus, and give them to him with the compliments of King Frost. "He will know how to make good use of the treasure," added Jack Frost; then he told the fairies not to loiter by the way, but to do his bidding quickly.	him of his jolly old neighbour, Santa Claus. "I will send my treasures to Santa Claus," said the King to himself. "He is the very man to dispose of them satisfactorily, for he knows where the poor and the unhappy live, and his kind old heart is always full of benevolent plans for their relief." So he called together the merry little fairies of his household and, showing them the jars and vases containing his treasures, he bade them carry them to the palace of Santa Claus as quickly as they could.
From that time, I suppose, it has been part of Jack Frost's work to paint the trees with the glowing colours we see in the autumn; and if they are not covered with gold and precious stones, I do not know how he makes them so bright; do you?	Ever since that time it has been King Frost's great delight to paint the leaves with the glowing colors we see in the autumn, and if they are not covered with gold and precious stones I cannot imagine what makes them so bright, can you?

Keller's method of copying is the most common one used by plagiarists, and will be familiar to university instructors who have taught first-year courses. The sequence of the narrative is unaltered. All Keller does is to make slight alterations to common words and phrases. So, in Canby's text a character called King Frost tries to imagine what to do with his treasures. He decided to give them to Santa Claus for the benefit of the poor. He calls his fairies and asks them to convey the treasures to Santa's palace. We have the following phrases from (C)anby (with their (**K**)eller parallels given below):

C: One day King Frost was trying to think of some good that he could do with his treasure
K: one day King Frost was surveying his vast wealth and thinking what good he could do with it
C: suddenly he concluded to send some of it to his kind neighbour, Santa Claus
K: he suddenly bethought him of his jolly old neighbour, Santa Claus
C: So he called together his merry little fairies
K: So he called together the merry little fairies of his household

In each case we can see that Keller uses less common words to say the 'same' thing – 'vast wealth' instead of 'treasure' (approximately 230 times less common in the language according to Google[18]), 'bethought' instead of 'concluded' (approximately 450 times less common). We see in the last example that Keller has King Frost calling 'together the merry little fairies of his household' whereas Canby just has 'called together his merry little fairies', the first being a somewhat clumsy construction.

Later Canby's Frost 'told them' (i.e. the fairies) whereas Keller's Frost 'bade them' – (about 30 times less common); Frost believes – according to Canby – that Santa 'will know how to make good use of [the treasures]' whereas Keller's Frost is sure that Santa 'is the very man to dispose of them satisfactorily'. I found on Google that to 'make use of' is more than twice as common as 'dispose of' – this aside from the fact that Keller's construction is longer and that 'dispose of' is ambiguous, a semantic flaw that 'satisfactorily' does little to repair. 'Good', in turn is three times more common than 'satisfactory', which is about five times more common than its adverb.

In the last paragraph, each story concludes with the fable that each year King Frost brings the colours of autumn to the leaves and trees, and that this is his real treasure. Canby uses the phrase 'from that time' while Keller writes 'ever since that time': the former is approximately 50 times more common than the latter. Keller writes 'I cannot imagine …' whereas Canby has 'I do not know …'. *Do* constructions are almost twice as common as *can* constructions in the language – both in full and contracted forms. 'Know' is about ten times more common than 'imagine'.

In conclusion we note that Keller's method is to take original expressions and alter them – in some cases slightly, and in others in major ways. She is forced, for the most part, to rely on a lexis which is significantly less common than that of Canby. Her text is less accessible as a result. Later I will give Flesch and Flesch–Kincaid readability statistics for all of the text excerpts referred to here (see note at end of chapter on origins of readability in plagiarism detection). As may be expected from my previous remarks, Canby scores much higher on the reading ease scale than Keller and obtains a much lower grade level too.

On the basis of the above, therefore, I believe there is little difficulty with the proposition that the Canby text shows more signs of originality than the Keller text. The differences are mainly at the lexical, and occasionally at the phrasal level.

Richard Condon and Robert Graves

The literal plagiarism between these works is not extensive. Condon (*The Manchurian Candidate*) seems to have 'borrowed' from a wide range of works, of which Graves' *I, Claudius* was only one.

In the source text, Graves writes: 'He knew that the marriage was impious: this knowledge, it seems, affected him nervously, putting an inner restraint on his flesh.' Condon's contribution is: 'Johnny knew in his superstitious heart of hearts that his marriage to Raymond's mother was an impious thing and this knowledge, it seems, affected him nervously, putting an inner restraint upon his flesh.'

Those who have read *The Manchurian Candidate* will know that Condon's work is rich in clichés, e.g. 'in his superstitious heart of hearts', and in flabbing out words into phrases – hence 'impious' becomes 'an impious thing'. Other examples of Condon's Faulkneresque attempts include: 'he decided that his mind must be bent or that he was drunk with compassion, or something else improbable like that' and 'owing to his endemic mopery, this one had to work nights, because, by now, it must be dark in St. Louis'.

Where Keller and Condon diverge is that Keller shuns common source words and phrases in favour of rare ones, while Condon pads out existing words into phrases and existing phrases into longer ones, producing somewhat purple prose in the process. Looking at Graves and Condon, I believe the reader will have little difficulty in attributing a greater originality to Graves than Condon.

Archibald Carey and Martin Luther King Jr

King is well known for having plagiarized almost half his doctoral dissertation from another theology student, rejoicing in the name of Jack Boozer, at Boston University (*Wall Street Journal*, 9 November 1990). What few recognized until recently was that his famous 'I have a dream speech' was also partly borrowed – from a speaker at the Republican National Convention in 1952. Thus where Carey has:

> Let Freedom Ring. Not only from the Green Mountains and White Mountains of Vermont and New Hampshire; not only from the Catskills of New York; but from the Ozarks in Arkansas, from the Stone Mountain in Georgia, from the Blue Ridge Mountains of Virginia

On the other hand King has:

> And so let freedom ring from the prodigious hilltops of New Hampshire. Let freedom ring from the mighty mountains of New York. Let freedom ring from the heightening Alleghenies of Pennsylvania. Let freedom ring from the snow-capped Rockies of Colorado. Let freedom ring from the curvaceous slopes of California. But not only that; let freedom ring from Stone Mountain of Georgia. Let freedom ring from Lookout Mountain of Tennessee. Let freedom ring from every hill and molehill of Mississippi – from every mountainside.

King's style of plagiarism here seems to have been to change the actual names of the mountains – he doesn't refer to any of the mountains mentioned by Carey by name. He then uses alliteration ('mighty mountains') or a kind of assonance ('heightening Alleghenies'). He enters a purple vein when he talks about the 'curvaceous slopes of California' and the 'prodigious hilltops of New Hampshire'.

What is interesting when comparing King with Condon is that both writers rely on 'purple' devices and both pad out the existing text. However, King differs from Condon in that he departs further from his source than Condon does. I would label King's style of plagiarism as elaboration rather than padding. Given the highly specific kind of oratory King was creating, this kind of plagiarism is relatively rare – a kind of 'poetic elaboration'. Looking at both King and Condon we find little difficulty in believing that, respectively, Carey's and Graves' texts were the source texts – they are shorter, pithier and relatively simple in structure.

Readability scores

One readily available way to turn the above observations into reality is to grade the excerpts according to their readability scores. Most computer users have this facility within their word-processor. It is a powerful method and uses the Flesch Ease of Reading Formula and the Flesch–Kincaid grade formula, whereby the appropriate reading level of a text as applied to the US school grade system (Years 1–12) can be assessed. These formulae are used here in preference to other, probably equally excellent, formulae because of their ready availability and ease of use.

In all three cases referred to above the source authors had higher ease of reading scores than their imitators and/or lower grade level scores. This was entirely as predicted. We would expect, on the basis of the foregoing arguments, that those who imitate succeed in making their texts complicated and, as a result, less accessible to readers. The results of the readability tests are given in the tables below, first for the originators and then for their copyists:

Table 9.1 Readability scores for the originating authors

The Originators	Flesch Ease of Reading	Grade Level
Canby	98.2	1.0
Graves	52.5	11.1
Carey	76.5	5.3

Table 9.2 Readability scores for the imitating authors

The Imitators	Flesch Ease of Reading	Grade Level
Keller	79.1	5.0
Condon	40	12
ML King Jr	57.2	7.5

As can be seen from the last two column of the tables the originators' texts have a higher ease of reading score than the imitators' texts or a lower grade level score.[19] The difference between Condon and Graves is slight – probably because Condon's imitation is virtually a literal copy of what Graves wrote.

Types of text dating and sourcing – what plagiarists do with language

By text dating I mean the process of attempting to determine which of two texts preceded each other temporally. In the case of texts where there are clear signs of plagiarism, it may be necessary to 'date' them, i.e. to assess which of the two authors copied from the other.

I suggest, based on previous experience, that text dating or 'sourcing' can be considered to be of three types. The first is *diachronic dating*, which is where a significant time period has elapsed between the earlier and later text. A suggested minimum time lapse for the observation of significant cultural changes to feed through to the language in the form of discourse styles, lexicon, phrasing and politeness forms would probably be about 20 years for most genres. However, one cannot be inflexible: a single event can change perceptions across the world, and so a lesser time period could be considered. On the other hand some genres change much more conservatively and so in those cases 30 or 50 years may be more appropriate. In such a case one would look for 'earlier' language in a later text, for words which had become redundant.

The second type of text sourcing to consider is *cultural sourcing* (i.e. the process of text dating by reference to culture), where one author may originate from a different culture than the other author, for example a UK author plagiarizing from a US author or vice versa. In a case like this we might expect to find instances of different lexical tokens for the same entity, different modes of address, a greater commonality of a certain phrase in the one culture rather than the other, etc. The issue here would be an inappropriate or infrequent cultural token appearing in one or other of the texts relative to its appropriacy or frequency in the other. We should also note that culture, in this sense, need not be 'national' culture – it could be based on age, gender, political affiliation, ethnic or religious group, class, etc. Lastly, I would like to mention what I term *archaeological sourcing*, which is the process of excavating two synchronous or almost synchronous texts to find which of the two presents symptoms of originality and which presents symptoms of copying – relative to each other.

Relating the above observations on types of text dating to the three examples given here, which are not dissimilar to the kinds of plagiarism found in educational settings, all three appear to be examples of archaeological dating. Examples of diachronic sourcing, on the other hand are relatively rare: we do not often find plagiarists ransacking texts from decades before in order to produce fresh-looking texts – the Hitler Diary forgery of the 1980s would, however, be a good example of diachronic sourcing. As regards cultural dating of texts, the only case that comes to mind was the allegation that an American senator hoping for nomination to the presidential election campaign had plagiarized from a speech given by the British politician, Neil Kinnock. So, in general, most plagiarism claims centre around what I have termed archaeological plagiarism, that is to say, with the linguist acting as a 'dater' of textual material in much the same way that an archaeologist might. As suggested here, readability

indexes have proved to be a very useful way of formalizing observations found by archaeological dating.

So, what *do* plagiarists do then?

In answer to this question, posed earlier, it seems on the basis of the above discussion on types of plagiarism, that plagiarists do three things (allowing for some overlap between the three types):

(1) Archaeological plagiarsts – the most common type – take an artefact and try to disguise its surface by substituting some of its parts and by re-arranging others.
(2) Diachronic plagiarists take an artefact from an earlier period and try to disguise its chronicity, by translating it into an artefact of their own time.
(3) Cultural plagiarists transpose elements of their own culture onto a cultural artefact of another culture or, alternatively, try to take cultural artefacts from elsewhere and convert them into own culture substitutes.

Needless to say, each of these three activities of plagiarism has its pitfalls for the unwary. The archaeological plagiarist is in danger because it is relatively easy to exhume common lexes and warped structures from the dust of two superficially different objects. The diachronic plagiarist is in danger because it is relatively easy to overlook small but critical differences which occur in a given culture over a period of time, while the cultural plagiarist would need a close familiarity with both source and target cultures to make sure that values from the source culture were not falsely imposed on the parasitic text, or that elements from the target culture were not 'false friends' of the source culture.

A note on readability and plagiarism

Interestingly, the connection between readability and plagiarism is not entirely new. Glatt and Haertel (1982) developed a way of using the cloze test to check for plagiarism. They blanked out every fifth word of a suspect text and asked the suspect plagiarist to fill in the blanks. They found that those who had plagiarized consistently produced lower scores on the cloze test than those who had not plagiarized. Much earlier, W. L. Taylor (1953) had suggested the use of

the cloze test as a readability test: if a panel of ordinary readers were unable to fill in above 40 per cent of blanked out words correctly, it was a sign that a text was 'unreadable'. Anything above 60 per cent meant that the text was highly readable. More recently Knight *et al.* (2004) found that using sentences with lower readability scores from suspect texts was more likely to assist in internet searching for plagiarism. What Knight *et al.* did was to measure the readability score of each sentence in a text, then put to one side those sentences with grade level scores below 10 and submit those to internet searches. They found this technique highly successful, and the work I have reported here (during most of which time I was unaware of their work) bears out their findings, though I would suggest the grade level of 10 is somewhat arbitrary: in general it would be better to take each text on a case by case basis, determine the mean grade level and then find those sentences in that text which deviate most from that mean. Alternatively, the authors could take the mean for a class or group of students and find deviations in individual students' sentences from that mean.

The above observations about the different types of text dating in plagiarism are really cultural, and not strictly speaking linguistic observations. For the linguist what is much more interesting is the question of what happens to the text at the lexical level when it is plagiarized. I go into more detail about this elsewhere, but for now it is sufficient to observe that the plagiarist has to model the copy lexicon on the source lexicon in a way which is counter to how we usually make our lexical selections. Briefly, most lexical choices are made unconsciously when we compose a text. We are usually unaware of the many reasons why we make specific choices. However, the plagiarist has to take the source lexicon into account, in effect to avoid many of the source lexical choices. Unconscious selection of the lexicon usually results in prototypical items being chosen (Rosch *et al.*, 1976), but the plagiarist has a problem in this case: the available prototypical lexis has already been used – to make the plagiary opaque it is necessary to choose different items from those found in the source text. Selection is thus conscious, rather than unconscious and – I suggest – forced, and sometimes even strained. This could be the cause of the plagiarist's text often being less readable than the author's text.

10 Veracity in language

True or false?

How do we react when we hear that someone has been accused of child abuse? Can we detect whether the witness is telling the truth or lying? This was the dilemma that faced Swedish psychologist Arne Trankell in a well-known 1950s child abuse case. Trankell developed his method, using the statements of the 5-year-old alleged victim and his mother to analyse what had happened (Rogers, 1990, online). In his reconstruction of the incident behind the accusation, Trankell was able to expose many inconsistencies in the evidence, and as a result the defendant was acquitted. Trankell later published an account of what had led to the accusations and the trial (Trankell, 1958), and in this way the science of 'witness psychology' was born (Rogers, *ibid.*). Its aim was to evaluate witness statements – in whole or in part – as true or false.

In this chapter I hope to demonstrate several techniques for handling statements. These are primarily of investigative rather than forensic value.

The origins of statement analysis

Even before Trankell proposed analysis of witness statements as a way of evaluating witness veracity, Undeutsch had already begun to develop Statement Validity Analysis (SVA), the core proposal of which was that memory-based experiences and fantasy-based experiences differ from each other (Undeutsch, 1967). Statement Reality Analysis (SRA) and Criteria Based Statement Analysis (CBSA) are broadly similar to SVA. Although SVA began with child abuse cases, it has since been adapted for use in all kinds of criminal and civil investigations.

Memory

The tenets of SVA and the related disciplines mentioned above are psychology-based. The key element is memory. A genuine, memory-based experience will differ significantly from a manufactured, falsified or fantasized experience. The two important dimensions are content and quality (Undeutsch, *ibid.*) The false experience will lack both. Psychologists score witness statements on a range of criteria – from characters and themes to sequence of events. Veracity is determined not only by the presence of such criteria, but is also based on whether criteria are weakly present, strongly present, etc. Criteria include logical structure, unusual or extraneous details, etc. (Triandafilou *et al.*, online).

Statement analysis in linguistics

Statement analysis in linguistics is somewhat different from the SVA of psychology. For one thing linguistics cannot easily base its findings on concepts such as *memory*, *fantasy*, *truth* or *veracity*. If we are to analyse statements as linguists we need to look at purely linguistic issues: narrative structure, sequence of events, association of narrator with events, etc.

Some time ago, police forces in the United States found that statements written out by witnesses were easier to evaluate than verbal evidence obtained from interviews. As this practice spread across the country a number of basic 'rules of the game' were discovered, enabling police officers to distinguish between the factual and the invented. Not surprisingly, many of the observations made by law enforcement professionals about statement analysis turned out to have close parallels in narrative theory and other branches of linguistics.

Narrative analysis and report analysis

The two basic types of statement analysis relate to *narratives* and *reports*. Witnesses and defendants narrate their knowledge of an incident, while in other cases people report a missing person, a robbery, etc. Some incidents lend themselves to a combination of both genres, e.g. you come home and find your house has been burgled. You relate the incident as a *narrative*, but then you need to make a *report* of any stolen or damaged items, perhaps for insurance purposes. The witness narrative tends to be in the first person, while claim reports usually background the first person and concentrate on the missing person or the stolen property.

The structure of first person narratives

In attempting to find the structure of first person narratives there are two approaches we can take. We can collect a corpus of narratives, and then survey them in order to draw up a style profile, or we can consider each

narrative on its own merits and look for stylistic inconsistencies within each narrative.

The first, top-down approach assumes that there is such a thing as an ideal or model first person narrative, to which all narratives should conform, and that failure to do so on the part of a narrative makes it 'ill-formed' in some way. The second, bottom-up approach makes no assumption about the structure of narratives in general, but looks for a norm for each of a range of values within the individual text. Significant deviation from this individual text norm might lead us to believe that the narrative is ill-formed. In this chapter we will confine ourselves to the 'top-down' approach. It is easier and more practical, and there is good reason to believe that first person 'fact' narratives are fairly consistent as to their properties, thus enabling us to observe and record inconsistencies without too many problems.

First person narratives

So, what is a well-formed witness/defendant narrative like? Evidence suggests that usually it is spontaneous and concise, and appears to conform to a number of criteria, the most important of which are summarized in the following table[20] (to be explained in detail in the section 'Categories in narrative' towards the end of the chapter).

Table 10.1 Characteristics of a well-formed witness narrative

	Criterion/Category	Description
1	Time	Events take place in given time period without gaps
2	Place	Key places in narrative are introduced as they appear
3	Sequence	Events are described in sequence
4	Descriptions, superfluity	Extraneous and superfluous details are absent
5	Tense	Simple past tense

Principles of evaluating narrative

Above we listed a number of categories which we could use to judge the 've-racity' of a witness or defendant narrative.[21] However, veracity is a psycho-logical construct, and as such is not directly observable. Instead I prefer the notion of **speaker commitment** to appraise veracity in language. How is speaker

commitment assessed? Broadly speaking, if we are committed to what we are saying we tend to be brief and to the point, we tell it 'like it is', as the saying goes – we do not hedge. On the other hand if we do not believe what we are saying or if we do not believe *in* what we are saying, we will probably beat about the bush, we will be hesitant and non-committal, and even the sequence of events we are describing will most likely be somewhat disordered. We will certainly not be demonstrating speaker commitment. The table below summarizes the key differences between what we may term *associative language* strategies on the one hand and *distancing language* strategies on the other.

Table 10.2 Speaker/writer distance

Observation	Deduced language strategy	Probable speaker commitment	Probable relationship with veracity
Speaker/writer uses simple, congruent forms, logical, sequential structures	Associative	High	High
Speaker uses elaborate 'polite' or formal forms, convoluted, roundabout non-sequential, 'illogical' structures	Distancing	Low	Low

The idea behind this table is to show that when the speaker or writer uses simple, congruent forms of language, in logically ordered sequence, then this individual is 'associating' with the narrative and with the reader. Under these circumstances the speaker relationship with veracity is high because speaker commitment is high. On the other hand, when speakers wish to distance themselves from their language, this is best achieved by using formal language, by placing narratives out of sequence, and by avoiding logical, ordered structures. In this case the relationship between speaker and veracity is probably low.

Methods of statement taking

Great Britain

How are police statements taken in Britain? Has it always been the case that the defendant's words were faithfully recorded? Traditionally, statements were

dictated by witnesses and defendants to police scribes. If carried out legally (according to Judge's Rules, 1952), a dictated statement could be an excellent way of narrating an incident. At its best it was spontaneous, brief and concise. It gave no more details than were required to investigate the crime further, and because the defendant was at liberty to stop dictating at any time, the chances of self-incrimination were low. However, the system was open to abuse if those taking the dictation were unscrupulous, and several famous cases of altered statements surfaced from the 1960s onwards. Eventually, audio and later video-recording of interviews became mandatory, so that nowadays, when the interviewing process is completed, a written summary of the tapes is produced by the police. Prosecution and defence legal teams are given copies of these tapes and may dispute the written version if they wish prior to trial. The final version is agreed by all sides and serious disputes as to content are rare.

United States

In the US police forces use a number of different methods, from contemporaneous interview notes to audio and video-recording, to the current practice adopted by many states, of getting the witness to write down their own statement. The statement is then analysed category by category (see next section for explanation of categories). A photocopy of the statement is taken, and inconsistencies are marked in colour-codes, or some other method. In important cases two or more officers are each given their own copy of the statement and they mark it up independently of each other. In the days before statement analysis, however, other methods were used to take statements from defendants in the US, and some of these were clearly open to abuse, just like their British counterparts.

In the next section each statement category will be explained and illustrated with examples, and the reader will be shown how to mark up a statement.

Categories in narrative

At this stage I will give examples of the previously listed categories (see Table 10.1 above) from genuine witness statements – both American and British, and I will show how statement analysis can be used to identify inconsistencies.

Categories of witness narratives: *time*

Time is important both in terms of scale and in terms of continuity. In a well-formed narrative we expect the narrator to be fairly precise about when events occurred, but not over-precise. For example, in this statement allegedly dictated by Timothy John Evans, he refers to events which occurred over a two-day period several weeks earlier:

Text 10.1 (excerpt:)
I went to the pictures—A.B.C. Lancaster Road, known as Royalty, at 4.30 p.m. I came out when the film was finished, I think about 7.15 p.m. I went home sat down and switched the wireless on. I made a cup of tea. My wife was nagging till I went to bed at 10 p.m. I got up at 6 a.m. next day, made a cup of tea, My wife got up to make a feed for the baby at 6.15 a.m.

There are two comments we can make about this excerpt and the frequency with which times of day are given in it. First, most people can scarcely remember what time they did something on the previous day, let alone the time of day they did something several weeks earlier, and this applies particularly to events outside our normal daily routine. If I go to the movies, it is unlikely I will first look at my watch to see what time I enter the cinema, and similarly, I am equally unlikely to look at my watch again the moment I come out. Even if I do note the time, why should I remember it several weeks later?

In fact the whole passage is suspicious. How would Evans remember – so long after the event – that he came home, sat down, switched the wireless on and then made a cup of tea, in that order? Even if he could remember these petty details, why would he narrate them? What relevance do they have to the crime?

Now let us look at the gaps in the times given and the information he left out. If he went to the cinema that afternoon, and if it was important for him to remember the time he went in and the time he came out – how is it he does not mention the name of the film, or whether he bought popcorn, or who he sat next to or spoke to, or how much he paid for the ticket?

Furthermore, are we to believe that Evans sat in his chair drinking the one cup of tea from the time he got home until he went to bed 'at 10 p.m.'? Now look at the following excerpt from the same statement. Analyse it in terms of

its treatment of time, and see what else you can observe about it, with reference to whether it is well or ill formed:

Text 10.2: Excerpt from Timothy John Evans' statement
I got home about 10.30 p.m. I walked in she started to row again so I went straight to bed. I got up Tuesday morning and went straight to work. I come home at night about 6.30 p.m. my wife started to argue again, so I hit her across the face with my flat hand. She then hit me back with her hand. In a fit of temper I grabbed a piece of rope from a chair which I had brought home off my van and strangled her with it. I then took her into the bedroom and laid her on the bed with the rope still tied round her neck. Before 10 p.m. that night I carried my wife's body downstairs to the kitchen of Mr Kitchener's flat as I knew he was away in hospital.

Marking-up: before you begin

If your text is on a word processor or web page, print it out first and work with pen and ink, preferably in different colours for each of the analysis categories. Otherwise, if working from a hard copy, make several photocopies of the text, retaining the master copy safely to one side. Practise typing and writing the text out several times, just to familiarize yourself with it. Make careful notes of any unusual spellings, punctuation methods, layout styles, etc.

Marking-up time

In marking up a narrative statement for *time*, first note actual times given, e.g. '4.30 p.m.'. Look for gaps in 'real-world' time. Look for sections of the text where there are clusters of stated times as opposed to sections of the text where all mention of specific times is absent. Where a specific time is mentioned, ask yourself what its relevance is. At this stage do not mark the text up for *sequence*, though you can make notes about it for later mark-up.

Categories of witness narratives: *place*

Here is an excerpt from a statement by Texas prisoner Max Soffar, who has been on death row since 1981 in connection with the Houston Bowling Alley Murders, in which three people died and one was seriously injured.

Text 10.3 (excerpt:)
One thing that I didn't tell the truth on was that Lat Bloomfield and I did this thing when we first got to the bowling alley, not like I said about being there in the parking lot for awhile.

Lat drove in and we were in his brown thunderbird. Lat pulled right to the front door so that the passenger side was next to the bowling alley. I think that there was a couple of cars in the parking lot when Lat pulled to the door.

It may not be immediately apparent but the above excerpt contains several opportunities for self-incrimination on the part of the narrator. As will be seen, each of these opportunities relates to *place*.

First we note that the narrator begins with an admission of untruth, 'not like I said about being there in the parking lot . . . ' I have not listed *admission of untruth* as a separate category within statement analysis, because it is relatively rare and at this stage it is probably better for students to concentrate on the most common categories.

However, it is almost always suspicious, because it invariably throws the narrative out of sequence – either in terms of time or, as here, in terms of place. The next mention of place relates to 'the front door'. However, first we have mention that it was Lat who was driving, and that they were in Lat's 'brown thunderbird'. Why do we need to know the Thunderbird was brown? This is an unnecessary detail, and moreover, it creates a gap between Lat driving in and Lat pulling to the front door. But what is the relevance that the passenger side is next to the bowling alley? Why are we told this at this stage, unless it is to show that the narrator himself would logically have been the first person to enter the bowling alley, since he was the passenger? Also, what is the relevance of there being 'a couple of cars in the parking lot', unless it is to provide the opportunity for the production of a witness later on? So, when we analyse this section of the text closely, we can see at least three opportunities for self-incrimination:

- Soffar claims to have lied about being in the parking lot 'for awhile' [sic] before entering the bowling alley.
- Soffar positions the passenger side of the car (his side) as nearest the front door of the bowling alley (where the crime took place).
- Soffar allegedly gives details of the car's colour and type: since this is irrelevant to the actual crime (especially the colour of the car, in view of the fact that the crime took place at midnight) we can only conclude that it was part of the self-incrimination process, as was possibly the claim of 'other cars' in the car park.

Analyse the following text excerpt from the same Max Soffar statement. Ask yourself why some details are being included and what is being omitted. Confine your answer to matters relating to the theme of *place*.

Text 10.4: Excerpt from Max Soffar's statement
Lat drove and we had the windows down to his car. He made a right on the highway and drove down for a little bit and then turned around and came back past the bowling alley. I asked him why he shot the dudes and he said he shot the dude for raising up and playing hero. He said he made me shoot the other two so that I would be as guilty as him if we got caught. I put the gun under the front seat after I reloaded it and it only had one live bullet in it before reloading. I don't know where the gun is now.

Marking-up place

Whenever a location is mentioned, or an event that occurs at a location, note it carefully and ask yourself why that event or that place is being mentioned at just this point in the text. Is it really relevant to the narrative? Is the detail excessive?

Marking-up and text observation tips

As you will have printed out several copies of the target text, do not hesitate to make as many hand-written notes as you want on each copy. Be open to all possible ideas, but be prepared to argue for and defend each one of these ideas rigorously. Discuss text details with other students. Learn to present an oral argument and to defend it. Your arguments must be substantial and structured, not slick. Imagine you are already a forensic linguist having to defend your position to a cross-examining barrister or attorney. Look for similar examples from other texts.

Categories of witness narrative: *sequence*

In the following excerpt, Frank Kuecken is supposedly relaying the events of a night on which a pizza bar assistant was killed. Later, after Kuecken 'confessed' several other youths unconnected with Kuecken admitted their part in the crime. Kuecken was eventually released and all charges were dropped.

Text 10.5: Excerpt from statement attributed to Frank Kuecken
Q: Who were you with on October 21, 2000 around 10 p.m.?
A: F.J.
Q: Were you in a vehicle and if so, who's [sic] vehicle?
A: Mine, "68" Ford F-100 black.
Q: Did you and F.J. go to Mancino's Pizza in New Baltimore on October 21, 2000 and if so what time?
A Yes. 9:50 p.m.

Q: What did F.J. do at Mancino's Pizza?
A: Went in and ripped the kid off.
Q: How do you know F.J. ripped the kid off?
A: He told me when he came out.

In the above excerpt, the questioning supposedly begins with a question relating to a specified time, namely '10 p.m.' but then later refers to an earlier time, namely '9.50 p.m'. While it is not difficult to understand the importance of a question like 'where were you at 10 p.m.?' the fact that the text then refers to an earlier time is puzzling, and while it is relatively easy to think in terms of a time on the hour, or even on the half hour, times such as 50 minutes past the hour, or 10 minutes to the hour, are not necessarily that salient that we would remember them weeks or even months later. So, in this excerpt it is the fact that the times are given out of sequence that highlights the spurious nature of at least some of the information the excerpt supposedly imparts. Notice also the 'thinness' of the information in this sequence: there is no substance or depth to the answers which have a rehearsed feel to them.

Analyse the following excerpt from the Derek Bentley statement for sequence. In terms of other categories, what mental process does the phrase 'I should have mentioned' illustrate? Do you think this section has been inserted? Bearing in mind that this is where the text ends, what is the significance of 'killed'?

Text 10.6: Excerpt from Bentley text
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.

Marking-up sequence
We can tell out of sequence sections fairly easily: when there is loss of focus, such as in the appearance of private verbs, and – as we will see shortly – when there is an alteration of tense, or an inconsistency of tense. When sequence is lost, ask yourself what the inserted text is designed to achieve.

Categories of witness narrative: *descriptions, superfluity*

In fact-narratives detailed descriptions are rare, just as they usually are in casual conversation. In fact, detailed descriptions are something we most often find

in such environments as fiction, academic journals and documentaries. This is why, when we do come across them in fact-narratives they seem out of place, and so we mark them up as suspicious. One description that has always intrigued me is the following from the Bentley statement:

'I could see he was hurt as a lot of blood came from his forehead just above his nose.'

The reader will be aware that the incidents related in the Bentley statement took place on a dark winter's night on a rooftop in south London. The speaker was at least 30 feet from the police officer who had been shot. He could not possibly have seen blood, far less exactly where it came from. Moreover, we all know a person's forehead is 'just above [his] nose', so why tell us? It is superfluous because it is common knowledge.

Rather, it seems to be the case that this part of the statement is being designed to paint a picture of horror. Put simply, details such as these were inserted into so-called defendant statements to shock juries and eventually, newspaper audiences. They seem to have borne little relationship to what was actually said by the defendant.

It is not possible from 30 feet away to say that blood is coming from 'just' above the nose, simply because 'just' in this sense implies precision, the kind of precision that would not have been available to an observer from that distance under those conditions.

In this sentence, the excess of detail – inserted to add realism as well as horror – makes the narrator lose focus. Attention is distracted from the narrative to produce a separate cognitive effort.

Look at the following excerpt from Evans' NH1 statement, and see if you can find any examples of detailed descriptions or superfluous information.

Text 10.7: Excerpt from Timothy John Evans' NH1 statement
I strangled her with a piece of rope and took her down to the flat below the same night whilst the old man was in hospital. I waited till the Christies downstairs had gone to bed, then took her to the wash house after midnight. This was on the Tuesday 8th November. On Thursday evening after I came home from work I strangled my baby in our bedroom with my tie and later that night I took her down into the wash house after Christies had gone to bed.

Marking-up: descriptions, superfluity

Look for information which the listener or reader does not need to know, or for information which is intended for the *overhearer* (Goffman, 1981: 133–4),

especially a jury, the media, etc. Previously we referred to examples of time and place which were superfluous or 'excessive'. The above examples are similar in nature.

Categories of witness narrative: *tense/aspect*

First person fact-narratives are almost entirely in the simple past tense. Where sentences are not in this tense then the lack of sequence could indicate an insertion of some kind. Consider this from the Bentley statement:

> 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 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.

Note that all the sentences use the simple past, except the second last one: 'Up to then Chris had not said anything.' This is a spontaneous denial – what could its purpose be, except to set up the narrator for his later denial about knowing the gun was going to be used, which in turn was followed by a denial of knowledge of a gun?

Consider the following excerpt from Max Soffar's statement. Look at tense changes and ask yourself why those changes occur. What do they add to the narrative? The last sentence of this excerpt is in fact the end of the statement. What impression is the overhearer left with?

Text 10.8: Excerpt from Max Soffar's statement
We went to my house and did some preludin and Lat said he was afraid someone had seen his car so he went and took it home. He walked back over to my house that night and we did the rest of the pills. We stayed up all day and went out to the park the next day. I was scared and that is the reason that I did not tell the whole truth before and I feel like shit and feel bad about what happened and ought to take my punishment for it. I think Lat and me both ought to pay for what we did.

Marking-up tense/aspect
The most common non-simple past tense in narratives is past perfect. However, it is not always a cause for suspicion: only usually where clauses relate to something outside of the present focus of the narrative. It is really the combination of change of tense and loss of focus that is most likely to signal narrative inconsistency.

Other categories

As previously indicated there are many possible categories for the analysis of statements – the above five constitute what seem to be the most significant for evaluating veracity. In the following paragraphs I illustrate several others which, though significant, have not been encountered as frequently as those mentioned in the previous sections.

Character naming: Elsewhere I mention 'Craig' becoming 'Chris Craig' which then becomes 'Chris' – in the Bentley text. Manifestly this is highly suspicious. In a fact-narrative of less than 600 words, it seems very strange to have three different ways of referring to a close friend.

Register: Also in the Bentley text we see several curious instances of mixing register. The *police* are mostly *police* except for one part when Bentley supposedly refers to a *copper*.[22] Generally there are no contractions in the text – but where they do appear, there is a cluster of them. In other words we appear to have more than one register in the text – a formal one and an informal one. Mixing register in any genre, except perhaps fiction, is rare, and it is certainly unusual in fact-narratives.

Pronoun omission: Remember these sentences from Frank Kuecken's statement?

Q: What did F.J. do at Mancino's Pizza?
A: Went in and ripped the kid off.

Why does the speaker drop the first person pronoun in 'Went in and ripped the kid off...'? This appears to be a distancing feature. It seems that either the speaker himself has omitted the pronoun because he does not really endorse his own words, or the person transcribing the text has made this alteration. If you look closely at the Kuecken interview you will see that the answers are very terse and seem prepared, almost mechanical. The statement seems to read almost like a Bogart script.

Chapter conclusion

In this chapter the reader has been shown several techniques in the handling of statements, whether oral or written, uninterrupted or in interview form. A number of categories under which statements could be analysed have been given. As pointed out, there are many further possible categories,

and some of these were illustrated. It is hoped that this chapter will equip the student of forensic linguistics with the raw material of statement analysis and handling.

Practice text

Below is a text on which to practise your statement analysis techniques. The first half of the text has been marked up for you. You might wish to photocopy the text and mark up the second half. While marking up the text you might like to think about answers to the four questions below. Note that question 4 is not a psychology/mental health question – you are being asked to evaluate the veracity of the speaker's words, nothing more than that.

(1) Does the statement narrate the crime?
(2) Does the author admit the crime unequivocally?
(3) Does the author face up to the crime?
(4) Do you believe the author when she says she was 'an absolute mental case' at the time and what is the significance of the adjective 'absolute'? What is the evidence?

Text 10.9: The Susan Smith written statement
When I left my home on Tuesday, October 25, I was **very emotionally** distraught. I didn't want to live anymore**!** I felt like things could never get any worse. When I left home, I *was going to* ride around a little while and then go to my mom's. As I rode **and rode and rode**, I felt **even more anxiety** coming upon me about not wanting to live. I felt I couldn't be a good mom anymore, but I didn't want my children to grow up without a mom. I felt I had to end our lives to protect us from any grief **or harm**. I had never felt so lonely **and so sad** in my entire life. I was in love with someone. **very much**, but he didn't love me and never would. I had a **very** difficult time accepting that. But I had hurt him **very much**, and I could see why he could never love me. When I was @ John D. Long Lake, I had never felt **so** scared and unsure **as I did then**. I wanted to end my life so bad and was in my car ready to go down that ramp into the water, and I did go part way, but I stopped. I went again and stopped. I then got out of the car and stood by the car a **nervous** wreck. Why was I feeling this way? Why was everything **so** bad in my life? I had no answers to these questions. I dropped to the lowest when I allowed my children to go down that ramp into the water without me. I took off running and screaming 'Oh God! Oh God, no!' What have I done? Why did you let this happen? I wanted to turn around so bad and go back, but I knew it was too late. I was an **absolute mental** case! I couldn't believe what I had done. I love my children with all my (a picture of a heart). That will never change. I have prayed to them for forgiveness and hope that they will find it in their (a picture of a heart)

to forgive me. I never meant to hurt them!! I am sorry for what has happened and I know that I need some help. I don't think I will ever be able to forgive myself for what I have done. My children, Michael and Alex, are with our Heavenly Father now, and I know that they will never be hurt again. As a mom, that means more than words could ever say.

I knew from day one, the truth would prevail, but I was so scared I didn't know what to do. It was very tough emotionally to sit and watch my family hurt like they did. It was time to bring a piece of mind to everyone, including myself. My children deserve to have the best, and now they will. I broke down on Thursday, Nov. 3, and told Sheriff Howard Wells the truth. It wasn't easy, but after the truth was out, I felt like the world was lifted off my shoulders. I know now that it is going to be a tough and long road ahead of me. At this very moment, I don't feel I will be able to handle what's coming, but I have prayed to God that he gave me the strength to survive each day and to face those times and situations in my life that will be extremely painful. I have put my total faith in God, and he will take care of me.

[Signed]Susan V. Smith
[Dated]11/3/94 5:05 p.m.

Endnote on statement analysis of *reports*

In many respects **reports** are very similar to **narratives** except that narratives are almost always undertaken using the simple past tense. Reports usually deal with missing persons, burglaries, stolen goods etc. In the previous section you looked at Susan Smith's statement. Prior to making her confession Susan Smith claimed that her children had been kidnapped by a carjacker. This is what she said on videotape:

Text 10.10: Susan Smith's claim regarding her children's abduction
all he ever told me was shutup or i'll kill you (break) and i just screamed i said what are you doing he said shutup and drive and had a gun and he was (...) poking it in my side you know and told me to drive (.) and I (.) so I drove

Later Susan Smith and her estranged husband appeared on television at a news conference. This is what they had to say:

Text 10.11: Smith news conference appealing for the return of their children
SUSAN: I have been to the Lord in prayers every day with my family and by myself with my husband it just seems so unfair that somebody could take such two beautiful children

HUSBAND: please do not give up on these two little boys and the search for their return safe
home to us

As an exercise, read the statement below, in which the author reports an alleged case of sexual assault. Consider whether the statement is genuine or false. Give reasons for either view. Debate your views with a classmate.

Text 10.12: Statement alleging sexual assault[23]

Upon arrival at Joe's Club, (date April X[th], 200- <......day>) I had entered the lounge area to have a drink and dinner. I was approached verbally by a man (mid 30's) as walking to a seat just 2 away from him. The words I remember most from any conversation that may have taken place that evening were "where you been? What took you so long?" I do no recall if/what I responded with in turn.

I then ordered a bourbon and coke while waiting for my order of "crab rolls." I wish I had more recolection of people or words exchanged, as this is where my mind is having trouble remembering. My food had arrived, enjoyed it with a glass of water topped with a lemon, finished my dinner and ordered 1 last drink, with all intentions of leaving at that point. if my memory is treating me well I had been offered another drink – hesitated but accepted offer. At this time I remember clearly going to the rest room. Upon my return, I casually had my 3rd Bourbon and Coke, while conversation was minimal and do remember paying most of my attention to the television directly in front of me.

Some time in the evening I do receall speaking to, one which was a bartender, James Smith & younger brother, Ben ?? These 2 are childhood – neiborhood acquaintances – I wish I could recall any of which conversations in the evening but unfortunatelly cannot as of today.

From this point, honestly things are neither here nor there in my head. Can not seem to decifer any actual instances.

I can say that approx. times were last recolection of 8 pm or 9 pm roughly. I woke with my alarm clock staring the time of 10:03 am. I frantically went to check on my son, he was not there. After becoming somewhat aware of my surroundings I then realized the only clothing I had on was my sleevless dress, sweater, which was worn the day of April X[th] under my blue dress jacket, which all clothing now in pocession/evidence. No under clothing as well. My sheets were soiled in what appeared, to me, as to be a mixture of slight blood and my own urine. In discust and pure confusion I stripped the fitted sheet from my bed and replaced it with a fresh one . . . first phone call was to my mother, with my little knowledge at that time our conversation was brief, just enough to know my son, Hugh, was safe and at daycare . . . (this info as of 4-12, mom stated it was 10 or 15 min after 10 am when she received my phone call.)

She was angry and we ended the phone call. Various phone calls including Mary Williams and Sam Walker, 2 close personal friends . . . after a few conversations and tears shed with them, I then decided to call the local police to report this incedent.

(rewind: first call after mom was to close friend Darleen Brown –> she gave me a number to the crisis line. moments later I discussed the situation "BRIEFLY" with a lady there . . . She

(not purposely) gave me the intention if I were to report, nothing could be done on the little info I had.) So back to the police call I got disconnected, strangly, as another call was coming in. I had called my neighbor in #94 a sally Edwards to see if she had heard or seen any thing -> she came home and stated times home and awake 8:30 pm to 11 pm -> nothing. Sally stayed w/me and that's when (all within moments) the 2 police officers arrived. Spoke to Ofcr. Macnamara.

All of stated information IS true & correct to the best of my knowledge. 4/12/02
Wendy M. Mullins

11 Forensic text types

Chapter Outline

This chapter is primarily concerned with investigative linguistics, that is to say the aim is to provide the investigator with methods of text analysis on a text type basis rather than to provide the court with scientifically evaluated evidence.

In referring to text types found in forensic linguistics we are really referring to a number of different contexts of situation, each of which generates one or more text types. For instance, a kidnap situation is typically seen to generate at least one ransom demand. It may also generate threats not directly related to the hostage. In that case the context of the situation is not just the kidnapping itself, but might include several other factors, for example the kidnapper's previous criminal activities, relationship with law enforcement, etc. The text types which will be discussed in this chapter are as follows:

- emergency calls
- ransom demands and other threat communications
- suicide letters

- final death row statements
- confessions and denials by public figures

Not every type of forensic text is discussed here, because many are dealt with extensively in other chapters or sections of the book. For example, the chapter on statement analysis included extensive commentary on witness and defendant statements, confessions and claim reports.

In this chapter we will present some of the main features of each of the types listed above, some aspects of a 'typical'[24] context of situation applicable to that text type, and a comparison between genuine and simulated texts and/or situations pertaining to that text type, where applicable. The amount of space devoted to each section will depend in part at least on my own experiences with that text type.

The first situation we will look at is that of the emergency call. The ability to distinguish genuine from simulated emergency situations is a necessary skill for any emergency operator, because of the need to deploy resources effectively. Learning about emergency calls is a good introduction to this whole idea of genuineness vs simulation in text types.

Emergency calls

Before we begin our study of this section, you might like to consider the following: Imagine your car has caught fire. Luckily you managed to get everybody out of the vehicle and you are all now standing at a safe distance from it. You get out your cell phone and start to call the emergency services. Write down the first thing you would say to the emergency operator. Head the example 'Forensic Linguistic Exercise'.

Until now we have mostly been discussing written text but looking at emergency calls will give us an opportunity to analyse spoken forensic language. Additionally, several of the text types to be discussed in this chapter sometimes take a spoken rather than a written form. The reader will therefore see that this book presents plenty of opportunities to analyse spoken, as well as written data.

The most important feature of any emergency call is probably *urgency*. The situation the caller is in seems urgent to that caller, even if the caller may have misjudged the situation. However, the linguist will generally find that it is not particularly helpful to attempt definitions of urgency, or 'sense of

urgency' or related concepts. This is because notions such as *urgency* are essentially psychological constructs which can manifest themselves in a variety of ways.

Rather, it seems useful to focus on something more familiar and perhaps slightly less subjective to linguists, namely the notion of information.

But first let us consider the role of the emergency operator, an individual we can conceive of as, typically, very experienced, and with good visualization of a number of scenarios, and one who – in most circumstances – can speedily evaluate the quality of information being received. Note that the concept of 'quality of information' in the abstract cannot compare with the day-to-day experience of the operator who is directly involved in the actual operation of collecting intelligence moment by moment in an emergency situation: it means, quite simply, the ability to sift primarily linguistic information in potentially deadly circumstances and to come up with the required response in a timely manner. The emergency services operator needs to evaluate the degree of threat, within seconds if possible. These situations are made more fraught by what can only be described as a great social evil – the making of hoax or malicious calls. People who make what are known as malicious calls can sound just as distressed as genuine callers, or they can portray an apparent calm in the face of adversity or fear: this is why it is not useful for the linguist to think in terms of notions such as *urgency* at the emotive level.

Below is a conversation[25] between a caller, a male security guard, and an emergency services operator. Inaudible speech is indicated by a dash.

Emergency Call 1

1. Hello can I help you?
2. Yes security guard here XYZ Electronics at the ABC can*al*.
3. *Yes*?
4. *I've* just been walking on the back well there's **smoke** coming out of one of our **roofs**. So, could you tell the fire service our ***curt***ain shop there's a lot of **chem**icals and what **have** you in the*re*.
5. *Uh*. Right. Is it *XYZ* Electro*nics*.
6. *Yeah*.
7. *ABC* Indu*strial Estate*.
8. *That's right*.
9. *By* the *ABC* Can*al*?
10. Yeah. What I'll do I'll make my way round the **back** and I'll leave the gates **open** for *them*.
11. *Right* and it's **smoke** is *it*?
12. Yeah. Smoke coming out of the ***roof***.
13. *From* the ***roof***?

14.   *I haven't* been there to investigate but I'll –
15.   Right – are you. . . .any of the chemicals that are *kept there*?
16.   *I **could**n't* really tell *you*.
17.   Are they just –
18.   chemicals from the –
19.   So –
20.   phuric **ac**id is in *there*.
21.   *It's what sorry*?
22.   *Sulphuric **ac**id is in there.
23.   Okay, don't *worry* now, we're on our way there.
24.   Thank you.
25.   Thank you. Bye.

One way to evaluate a conversation like this is to take a piece of paper and, laying it across the page of the book, cover the entire conversation except the first line. Study that line, and then expose the next line. Do this line by line. In this way you will be able, in part at least, to see the text as it unfolds. As you do so, ask yourself what kind of information is being passed between the caller and the operator, and what the quality of that information is. What does the operator know at any one moment in time? Is the operator's information being increased significantly with every turn? This latter point is very important: there must be a sense in which information continues to be added to throughout the exchange. You will see that at a particular point in time the operator has enough information and makes moves to terminate the call.

In notating these calls I have concentrated on just three elements of each conversation. Firstly, I am interested to know the extent to which there is co-operation between caller and operator at any one time. I take it as axiomatic that co-operation means the full, frank and timely answering of any questions in addition to the spontaneous offering of any information regarded as essential. Hesitations, incomplete or overly short answers, and any signs of evasiveness on the part of the caller must indicate that there is something wrong: i.e. that the caller might be making a false or hoax call (medical or other reasons apart).

One of the most striking symptoms of a genuine call is the interlocking of turns: what we seem to be looking for is a slight overlap between turns, probably measured as no more than a few hundredths of a second, between speakers.

In my analysis of these calls I indicate this kind of overlap by italicizing the last few letters or syllables of the outgoing turn and the first few letters or syllables of the incoming turn (depending on the amount of overlap):

> *From* the *roof?*
> *I haven't* been there to investigate but I'll –

In this sequence the operator is trying to ascertain the visible source of the smoke. She says: 'From the *roof* ?' Note that '*roof* ' is italicized here. This means that the next turn overlaps with this word, or part of it, because we see that what the caller says is also italicized: '*I haven't* been there . . . '

So, what this notation means is that the caller does not wait for the operator to finish the information request. Rather, we can imagine that the caller is listening very intently, and is ready with an answer even before the operator has finished asking the question. This would appear to show a willingness on the part of the caller to co-operate, particularly as the content of the turn appears relevant to what the operator would seem to need to know at that point. We can surmise that if the caller were not willing to co-operate then in all probability attendance to the operator's needs would be much less. What does co-operation mean, in the narrow sense in which we are using the word here? Strictly speaking it is a term, not a word, in this context. It means the caller is continuously adding or seeking to add to the information being given to the operator. The caller knows that for the call to be successful, information must be incremental.

An emergency call seems, therefore, to have at its core a relationship of trust between caller and operator. The operator trusts the caller to give accurate and timely information, and the caller in turn trusts the operator to ask only the most pertinent questions. The entire call moves on a trajectory from nil knowledge on the part of the operator to the maximum amount of knowledge that the caller can impart in the minimum possible period of time. This makes an emergency call unlike any other type of service encounter. Although it is true that both operator and caller have needs and can supply needs, over and above this the caller and the operator have a common goal: to activate the appropriate assistance as speedily as possible. There is a clear understanding that the key to this process is information. Note that I do not qualify information here, by calling it 'good', or 'accurate', for example. If an exchange in which information is to be imparted lacks these characteristics then what is being exchanged is not information. Information is per se 'accurate', 'good', etc.

Regarding the mark-up process – my own somewhat minimal version of the complex art of CA (conversation analysis) – the second element of the conversation I concentrate on is that of intonational emphasis (i.e. verbally stressing a word or syllable):

I've just been walking on the back well there's **smoke** coming out of one of our **roofs**.

As can be seen from the above excerpt from the call, the speaker emphasizes two words *smoke* and *roofs*. I would contend that this is part of the speaker's commitment to information. He is emphasizing the two most informative (pertinent, relevant etc.) words in the sentence. We can imagine that even if the operator had somehow missed every other word in the sentence, the two words *smoke* and *roofs* would have been heard. Speech emphasis is very important in emergency calls, and is quite unlike 'normal' conversational emphasis.

Finally, there is rising voice pitch, which I highlight because it tells us two things. If the caller repeatedly uses rising pitch at the end of each turn, then it may indicate lack of commitment. If the operator uses rising pitch it may mean either that the operator doubts the call, or is seeking clarification. When someone is doubtful of what we say, or appears to need clarification, then in 'normal' conversation we have the opportunity to commit firmly to what we want to say, or we have the option of modifying what we have said, or even backing away from it. Therefore, it would seem when the operator uses rising pitch, the caller has the opportunity to commit or to withdraw.

I have confined my analysis to these three elements of phonetic output: overlap, emphasis and rising pitch. There are many other possible aspects of emergency call phonetic output which could be highlighted. However, I have avoided mention of these with the specific aim of keeping the analysis as uncomplicated as possible, while still retaining essential information. I believe these three elements alone can give us most of the information we need when analysing emergency calls. This does not exempt the student forensic linguist from pursuing the topic at a deeper level and finding out more about how to analyse emergency calls.

In the following paragraphs there will be several opportunities to see each of these, and other non-phonetic features of emergency calls in action. Let us consider the first two lines of the above emergency call:

1. Hello can I help you?
2. Yes security guard here XYZ Electronics at the ABC cana*l*.

Turn 1 is the operator speaking. Note that the greeting 'can I help you' is something of a politeness ritual. It could be argued that the caller is 'helping' the fire service by providing it with information, rather than that the operator

is actually helping the caller. Similarly it could be argued that 'can I help you' is an invitation by the operator to the caller to begin.

In Turn 2 the caller identifies himself. He does so in a way which seems brief and concise. He first states his function, then the name of his company, and finally where the company is located. Even if, at that moment, the caller were cut off, the operator would know where to send assistance – assuming the call to be genuine. The fact that the assistance is apparently required at an electronics company might give the operator further valuable information as to what kind of help might be required.

The operator next says:

> **3.** Yes?

This 'yes' seems to mean several things: 'Yes, I've got that' and 'Yes, carry on'. It could also mean, 'Yes, fine, now give me some details.'

Note that the brevity of the answer could mean the operator is looking favourably on the call, i.e. disposed to treat it as genuine. We could also say that now that the caller has apparently established his credentials, the operator wants him to continue. Hence the perfunctory 'Yes?' Using a speech act model we could say that a number of moves have been made. The operator first invites the caller to give his information, the caller takes up the invitation, gives some basic information and then waits for the operator to issue another request. The operator does so, with this perfunctory 'Yes?'. In some ways this is a very unusual service encounter. The operator is actually behaving like a customer and the caller is behaving like someone providing a service. Curiously, both speakers seem to take this situation entirely for granted. They both appear to understand that this is what an emergency call consists of, or rather that this is what *this* emergency call consists of at just *this* moment.

The caller continues:

> **4.** *I've* just been walking on the back well there's **smoke** coming out of one of our **roofs**. So, could you tell the fire service our *curt*ain shop there's a lot of **chem**icals and what **have** you in the*re*.

Turn 4 seems slightly garbled. Although the caller knows what he is saying, he seems to want to get the information out fairly quickly. Do you think he is genuine, or does this indicate that he is anxious to get off the line? Again, I would suggest questions asked in this way do not interest the linguist. What we are really interested in is the information being given. We need to look at

the current information in the context of the information previously given. Previously, in Turn 2 the caller gave his job title or function, his company and location. Now he is giving his reasons for calling. He is describing the real world conditions which are of relevance to the service he is implicitly requesting. Note how, as he comes to the end of what he is saying, he is no longer providing a service, i.e. information. Rather, he is requesting help. What has been achieved so far? The caller has provided the background information to the call: his function – which impacts on his reason for calling – and the location of the incident. Only then does he give the actual reason for his call: '. . . there's smoke coming out of one of our roofs.' Compare this with:

Emergency Call 2

1. Fire service.
2. Ah mm our shed is on **fire** at the *bott*om of the *garden*.
3. Right, what's the address *please*.
4. *172* Smith Street Mt Pleasant.
5. 172 Smith Street Mt *Pleas*ant?
6. *Yeah*.
7. *Right* what's your *tele*phone *number*?
8. *555 – .*
9. *Yeah*.
10. *5555*.
11. *Right* okay we'll be along right *away*.

What are the main differences between the previous call and this one? In the previous call the security guard first gives his own credentials by stating who he is, perhaps because he may consider that he is dealing with fellow professionals, and they would give him immediate credibility if they know who they are dealing with. In the second call the speaker, a female householder, appears to hesitate momentarily, but then launches right into her reason for calling. She does not give her name, but she implies that her concern in the matter is one of ownership, '. . . *our* shed . . . ' this is tantamount to saying: 'I am the owner of a property where the shed is on fire.' In other words she states her relationship to the event just as the security guard did in the previous call.

The second caller therefore appears to have a vested interest in calling, a valid reason, which can be compared with the security guard's reason: he is an employee of the factory, one who is charged with special responsibilities, i.e. *security*. That is his reason for calling. However, the operator asks the second caller for her telephone number. She did not ask the first caller for his telephone number. Is she being sexist, or is it possible that she has some doubts

about the second caller? However, these doubts seem to be dispelled by the fact that the second caller gives her telephone number without any hesitation. Note that the operator actually acknowledges the first part of the caller's number before the caller is able to give the whole number:

7.	Right what's your telephone number?
8.	555 –.
9.	Yeah.
10.	5555.

It is possible that the operator is acknowledging not just the number, but also the caller's right to make the call: in other words the operator is accepting the call as genuine. Why do we say this? Because the caller co-operates immediately – she gives her telephone number unhesitatingly. The operator interjects with 'yeah' after the code portion has been given, indicating her awareness of this co-operation.

Just as with other types of conversation, the caller and the operator co-operate with each other. They work together for a common goal. It is understood by both parties that the real power must reside with the operator. This is after all an emergency service. In practice this means that the caller must co-operate with the operator: the operator is the authority.

However, the operator makes this as easy as possible for the caller. She does not put obstacles in the caller's way. The caller may choose to give information in any order, as long as it is given promptly. Sometimes, though, depending on the level of threat to the caller, the order will be important. All of this the operator makes allowances for.

Any request for information must be responded to immediately, so that if the operator asks for the telephone number and the telephone number is not forthcoming, or there is undue hesitation, the operator will become suspicious, unless there is good reason for the delay.

Consider the following call. At what point does it become apparent that there is something unusual about this call?

Emergency Call 3

1.	Hello fire service can I help you?
2.	[ ]
3.	Hello?
4.	Hello there's a fire in Bennets Bowling *Green*.
5.	In *where*?
6.	*Bennets Bowling Green*.

7.   *Bennets Bowling Green?*

8.   *Yes.*

9.   Uh. Right and where in Bennets is it?

10. In the *site*.

11. Is it in the *street*?

12. No the **Bowl**ing Green.

13. Yeah, uh, is the Be- is the Bowling *Green* in the street?

14. No.

15. No. It's not in the street at *all*?

16. No.

17. Is it in the *park*?

18. No, it's just in the Bo- in the actual Bowling Green.

19. Right and it's just in Bennets is *it*?

20. *Yes.*

21. And what what's on fire there?

22. Whaa?

23. What's on **fire** in the Bowling Green?

24. Yes. Um. The shelter *thing*.

25. *The* shelter?

26. Yes.

27. Uh. Right and you're definitely telling me the **truth** now?

28. Yeah, yeah, I'm **tell**ing you the tru*th*.

29. *Ok*ay then we'll send them a*long* straight away.

In this call (the caller is a young child) we start to see signs of something wrong almost immediately. Turn 2 is what we call an empty turn. An empty turn means nothing is said in answer to a greeting or request. There could be valid reasons for this, for example deafness, or some other disability on the part of the caller, so at this stage we keep an open mind. Instinctively, when an operator gets an empty turn, the operator will start to listen for other signs – for example unexplained hesitations and pauses. In the next turn the caller says: 'Hello there's a fire in Bennets Bowling Green.' Not only is a third 'hello' now being exchanged, but the actual topic of the sentence, the reason for the call is further delayed, postponed by this apparently superfluous greeting. Moreover, it is not the case that something is *on* fire, rather that there is '*a* fire', i.e. a fire exists. The phrase 'a fire' – as mentioned above – is much less urgent than 'on fire'. 'A fire' is impersonal, even non-threatening, but if something is 'on fire' then there would appear to be a strong suggestion of actual danger.

Hence in this example, the perceived level of threat on the part of the caller seems very low, while the level of caller co-operation is also low: it is not until Turn 5 to Turn 6 that we see any overlap between turns.

Later we see that the caller is giving very little information about the 'fire'. It takes to Turn 24 to ascertain that it is the 'shelter' which is on fire, and then it turns out not to be a shelter as such, just a 'shelter thing'. Note also the following sequence, as an example of the function of rising voice pitch:

    4.   Hello there's a fire in Bennets Bowling *Green*.
    5.   In *where*?
    6.   *Bennets Bowling Green*.
    7.   *Bennets Bowling Green*?
    8.   *Yes*.
    9.   Uh. Right and where in Bennets is it?
    10. In the *site*.

In this sequence of turns we note, firstly, that there are few overlaps between turns. This seems to indicate an undue slowness, possibly due to lack of co-operation, on the caller's part. However, the operator is aware that the caller is a very young child, possibly less than 10 years of age. She must, therefore, tread carefully. However, as the turns progress, it seems that not only is there a lack of overlap, but that the caller keeps raising the pitch of her voice at the end of each turn. In other words it seems that the operator keeps trying to get commitment out of the caller, but fails to do so.

This call is characteristic of hoax calls: there are long pauses between turns, some turns are not at all informative, and the accumulation of information is painfully slow. Finally, the operator – who appears to have been making allowances for the youth of the caller up until now – asks: 'Right, you're definitely telling me the truth now?' to which the caller hastily asserts 'Yeah, yeah I'm telling you the truth.'

It is evident, however, that the operator gives no credence to the call: she has not bothered to take the caller's name or address. Note that she does not ask for the caller's telephone number because the call is being made from a public telephone, so we cannot use the absence of this question as an indicator that the operator accepts the call as genuine (as in the case of the security guard).

Emergency call model: a sketch

Previously we suggested that the best way to evaluate emergency calls might be to construct some kind of information model, and then to test this model against what we actually encounter. Having now looked at a few emergency calls, we can perhaps sketch out a possible sequence of events in the model emergency call:

Emergency centre response
Reason for calling/role of caller
Location of emergency
Nature of emergency

Before the caller can begin to relate the event the emergency centre must respond. Hence 'Emergency centre response'. This can take many forms, but is often phrased as a greeting, e.g. 'Hello' followed by an apparent offer to help such as, 'Can I help you?'

Next the caller proceeds to give details of the emergency. This information can be in any order, and may possibly best be summed up in the six wh-words *who, what, why, where, when, how.*

Who: The person who is calling, either by name, function of person calling, or
    relationship to the incident, e.g. security guard, owner, neighbour, passer-by.
What: The (detailed) nature of the emergency, e.g. shed fire, factory fire,
    etc.
Why: Why the person is calling (if not previously covered).
Where: The location of the emergency.
When: When the emergency began or when the caller became aware of it.
How: How the emergency started or was caused.

Sequence of details, implied information, conflict avoidance

The sequence of details given to the operator may not always be important to our model, for instance it probably does not matter if we know where the fire is before we know what kind of fire it is. However, the sequence is important from the point of view of what it tells us about the incident and the caller's relation to the incident. For example, in Emergency Call 1 the caller begins with his relationship to the incident. In other words he foregrounds his own relationship to the emergency. In Emergency Call 2 the caller does the same but does it in a different way, she implies her relationship to the incident with the use of the word 'our', as in 'our shed is on fire at the bottom of the garden'. In numerous other calls I have analysed we hear callers beginning with their own involvement in the situation. This can take the form of stating a job function, claiming ownership, being the neighbour of a victim, being a victim or potential victim, or even just being a passer-by.

Of course, the caller could begin with stating the location of the incident. There is no reason for this not to happen. However, the operator would very rapidly expect the caller to state or imply their own relationship to the incident.

In the following call excerpt we see that, as in the second call above, the caller does not state her precise role in the incident – in this instance because the call has come through on a secure line.

Emergency Call 4
1. Um we've got a - - - running into into a house.
2. Right anyone *trapped*?
3. *Uh no* not that we're aware of. um one woman has actually gone through the window and has landed in the lounge of the house so obviously there could be problems with um petrol coming from the vehicle and I'm sure there's – damage to the house.
4. Right but. But you're not aware there's anybody trapped there?

Note that the gravity of the emergency does require that the caller gives considerable detail to the operator. Curiously, whereas the police officer is focused on the possibility of a fuel spillage into the house, the operator is more concerned with whether any of the crash victims are trapped in the vehicle. Each side has their priorities, but each side accommodates the other and lives with the other's priorities, without conflict. Contrast this with line 28 of Emergency Call 3, where the caller impatiently asserts that she is telling the truth 'Yeah, yeah.' Clearly, in that situation there is the potential for conflict, whereas in a genuine call the potential for conflict is generally much less, or is avoided. The police officer takes account of what the operator needs to know, but is anxious to put over her point at the same time.

As far as the model is concerned, therefore, co-operation is critical.[26] Part of this co-operation is, as we have just stated, lack of confrontation. We could surmise that people who are genuinely co-operating with each other in an emergency will tend to avoid confrontation wherever possible. Bear in mind the emergency call is part of the emergency event.

Emergency calls: conclusion

In this section we have looked at some aspects of emergency calls, specifically fire emergency calls, though police, ambulance and other emergency service calls could also be considered. We have seen that there is great variety in the presentation of information in emergency calls. We cannot propose a fixed order of information, and it is not always easy to tell the difference between

genuine and malicious calls. For the operator at the end of a telephone line, having to make instant decisions on where to deploy resources, it is even more difficult. However, we can propose that a genuine emergency call has several constituents, which we can consider under three headings:

(i)   components of the call
(ii) attitude of the caller
(iii) aspects of phonetic output

Call components
- information about the incident, especially *where* and *what*;
- the relationship of the caller to the incident.

Attitude of the caller
- commitment by the caller to relay the details of the emergency;
- co-operation of the caller with the operator – a willingness to answer questions fully, promptly and accurately;
- absence of conflict or confrontation from the caller, even if confronted as to the genuineness of the call.

We can also, to some extent, predict some of the physical or phonetic features of the emergency call:

Aspects of phonetic output
- overlap between turns, especially from the caller, indicating co-operation;
- emphasis of important, salient items of information;
- avoidance of sentence-final rising voice pitch except where repetition or clarification is required or where there is immediate cause for alarm.

Ransom demands and other threat texts

Threat texts (which can be spoken or written, or even videoed) have something in common with malicious emergency calls: the sender of the text or the person making the call is usually anonymous. The recipient of the malicious text and

the person receiving the malicious call are both at a disadvantage: violence of one form or another is being perpetrated against them by someone who is unseen.

It seems axiomatic that the essence of a ransom demand is the demand for a ransom, but the truth is – in my view at least – that it is the *threat* component which is really the most important part of any ransom note, and is therefore that part of the ransom note to which we should pay the closest attention.

Threats are the counterparts of promises. If one person threatens to cause death or injury to another unless *x* is done, then the person making the threat is implying that they will not carry out the threat if the required condition is fulfilled. This is why ransom demands are so complex: they appear to contain a conditional promise: 'If you do X, or pay Y, we will return Z to you.' However, it is very important that the wording of the condition be considered carefully, because very often the kidnapper has no intention of returning the hostage, alive or dead. In the Lindbergh case the family is asked: 'It is realy [sic] necessary to make a world affair out of this, or to get your baby back as soon as possible.' Making allowances for the writer's difficulties as a non-native speaker possibly with some kind of aphasia, we note that the return of the child is backgrounded to the second clause of the sentence, rather than, for example, 'To get your baby back . . . ' This backgrounding of the 'claim', i.e. that the baby will be returned, makes this a somewhat vague commitment, to say the least.

However, even if the threat were more direct, e.g. 'if you want your baby back . . . ' we note that this is still not a commitment to return the kidnap victim. It is an equivocal conditional clause: the result is not stated in terms of certainty. It is very different from 'You will get your baby back if you do *x, y* or *z*' or, for example, 'Do *x, y* or *z* and you will get your baby back . . . ' These two latter examples can be referred to as *cause and effect* clauses, *Do A and B will follow*. I would suggest that usually a ransom demand which does not contain an unequivocal *cause and effect* clause could indicate a detrimental result for the hostage.

Therefore, we need to bear in mind how far removed '. . . to get your baby back . . . ' is from 'You will get your baby back if you do this' or 'Do this and you will get your baby back'.

Clearly in the Lindbergh case, one cannot gather any commitment from the ransom demand as to the child's return. It has long been suspected that the child was dead before he was removed from the premises by Hauptmann – having been dropped from the ladder. If so, Hauptmann would have had to conceal the child's death, and so continued to press for a ransom payment

(which he received). So, it seems that the real, underlying move in the Lindbergh kidnap notes is one of making a threat while appearing to make a promise. There is a real threat to the child in the language. This is bound up with the conditional 'promise' of the child's return and the false claim as to the child's state of health. As an exercise you might like to write several threats disguised as conditional promises and award them varying degrees of merit as regards their actual vs apparent commitment. Label them 'Class Exercise: forensic linguistics' at the top, just to make sure that they cannot be inadvertently misinterpreted!

In the Lindbergh ransom note the author stated: 'We warn you for making anyding public or for notify the polise the child is in gute care.' Here the kidnapper claims the child is in good hands. However, the note would have to have been written before the perpetrator entered the premises, and so we see immediately that the claim is false, since the kidnapper had not even encountered the child when he wrote the note. How did he know that the child was not ill on that day, for example?

When kidnappers (or perpetrators of other, similar violent crimes) make claims they are not in a position to verify at the time of making them, we need to respond quickly and decisively: we are dealing with a perpetrator who is lying to us and will therefore not hesitate to break any so-called 'promise'.

Hate mail

Ransom demands are not the only kinds of threat text. In fact threat texts probably require a book by themselves since they are so diverse in character and, unfortunately, all too common. They include threats to expose public officials or company executives unless such individuals resign, for example:

Forensic Text 11.1: An anonymous hate-threat letter

DEAR BILL,
    I SUPPOSE YOU THOUGHT I WOULD FORGET BUT YOU ARE WRONG HOW COULD I FORGET A RAT LIKE YOU. I HAVE SENT A LETTER WITH ALL YOUR PAST DETAILS TO THE CHIEF EXECUTIVE. ALL YOUR DEBTS AND PAST MISSDE-MEANOURS [sic]. IF YOU DONT RESIGN FROM THE SENATE IMMEDIATELY THE PRESS WILL PRINT A LIST OF ALL YOUR DEBTS BOTH LOCALLY AND NATIONALLY. I ALSO KNOW OF THE BRIBES PAID TO THE UNDESIRABLES TO VOTE FOR YOU. YOUR BEHAVIOUR IN MEXICO ALSO LEAVES ME WONDERING WHY YOU BOTHERED TO GO. IT CERTAINLY WAS,NT TO PAY HOMAGE TO THE DEAD AS
                                          ⟶

> YOU WERE NEVER SOBER. WHAT WOULD THE LOCALS THINK OF THAT. I ALSO
> KNOW OF THE MONEY YOU HAVE POCKETED FROM THE MONEY RAISED FOR
> THE CELEBRATIONS YOU MIGHT BE ABLE TO FOOL SOME PEOPLE BUT NOT ME.
> YOU FORGET I HAVE KNOWN YOU FOR ALL OF YOUR LIFE. GO BACK INTO YOUR
> RAT HOLE WHERE YOU BELONG AND STAY THERE.
> 　　?????

The above letter was received by a client of mine some years ago, and contained a number of apparent accusations. However, if we really want to accuse someone of something – instead of just smearing them – then usually we would be more direct about it: 'You stole such-and-such an amount of money from so-and-so' or, 'You owe money to so-and-so and such-and-such'. If someone claims to know of 'money you have pocketed' then they should know something about the amount, when it was taken, etc., otherwise we are entitled to think they are just gossiping or propagating scandal. In fact, if someone wants to accuse another person of theft, all they have to do is to report the crime to the law enforcement authorities. In other words, the writer's intention in sending this letter does not seem to be the substantiating of 'claims' as to genuine wrongdoing, but rather to distress the addressee with threats.

So, it seems that the threat in the above letter is not like the threat in the ransom note: the writer is not in possession of anything that the addressee will want returned. There is a conditional, which can be summed up as: 'Resign or I expose you.' The real threat here is to the recipient's peace of mind and health. The most the recipient can do is to resign, i.e. give in to the writer's demands, but even that is no guarantee that the recipient will be left alone by the writer. On the other hand in a hostage situation, if the ransom is paid and the kidnap victim is returned, that is generally the end of the matter as far as the person paying the ransom is concerned, at least in respect of any contact between kidnapper and ransom payer. Suppose 'Bill' in the above letter did resign? Would he be left alone? The chances are the writer would subsequently demand he left the area, or sold his business etc.[27] We should be aware that there is no limit to what blackmailers can and will demand of their victims.

That is the key difference between the letter to 'Bill' demanding his resignation and, say, a hostage note. A hostage note is not a threat accompanied by blackmail, it is a threat accompanied by a demand. It appears to hold out

a promise. When, however, we evaluate the 'promise' (i.e. to return the victim) we very often find that it is not a real promise, just the appearance of a promise. The real object of hate mail is just that, to express hatred by appearing to threaten someone. However, there is no actual threat because in reality the person making the threat is not in a position to carry it out. Just as the writer of the hostage note often fails to deliver on the promise to return the victim, so too the hate mail writer seldom delivers on the threat to 'expose' the recipient.

Earlier we looked at the anthrax letters. At first glance these seemed to be almost pure examples of trick mail, i.e. the recipient was tricked into believing the letter was harmless, and therefore opened it. No demand was made to the addressee, no promises were made to the addressee. The threat component in those letters was not within the letters themselves, but was really directed at members of the public, any one of whom, it was implied, could receive such a letter – hence the mass mobilization of postal workers to prevent the successful transmission of such letters through the US postal system. People were genuinely afraid they might be sent such a letter, and in fact several were. The campaign was very successful in its aim of causing widespread fear, partly helped by the media. The anthrax letters contained elements not only of trick mail and threat mail, but also terror mail, that is to say they used a relatively few letters to conduct a nationwide campaign of terror. The person who sent them can truly be described as a terrorist, since he/she used all of the methods and ingredients employed by terrorists in carrying out acts of terrorism: surprise, fear and publicity, in all of which the terrorist (as we have noted) was very successful. When you have analysed many anonymous letters of all types, you begin to see a common pattern emerging. After a while the terrorist, the kidnapper and the sender of hate mail begin to seem remarkably similar to each other: their communicative aim is identical, namely to victimize others through terror and pain or the threat of terror and pain. The only difference is one of degree. The common component is *threat*, as summed up in the table below.

Table 11.1 Threat type in relation to context

Context	Primary threat	Secondary threat
Kidnap	Death or injury to victim	Negative effect on the recipient of the ransom demand
Anonymous hate mail	Exposure or embarrassment	Possibility of repeat letter
Trick mail	Death or injury to recipient	Perceived mass danger

Conclusion

I believe the foregoing indicates that different types of threat mail – of which we have considered only a few – have far more in common with each other than otherwise. Any form of threat mail constitutes a danger to those to whom it is sent, to those to whom it refers, and – in some cases – to other individuals as well.

Suicide Letters

In this section we will evaluate several known genuine and simulated suicide texts.[28] A suicide note, to have credibility, must – in my view – be seen to be making a definite proposition although the situational context of the proposition may not always be clear to the casual reader, as for instance in the text below:

Forensic Text 11.2
I hope this is what you wanted.
(Schneidman and Farberow, 1957)

This is the entire text of a genuine suicide note. We know nothing about the circumstances of the victim's death, other than that the Los Angeles County Coroner found the death to be one of suicide, and that Schneidman and Farberow (1957) appear to have evaluated the text in the same way.[29]

For the addressee of the text, the context behind the sentence 'I hope this is what you wanted' may or may not have been apparent, but the casual reader – i.e. one who has no knowledge of the participants, the events or the circumstances – would probably have to take such a note almost on trust.

However, even though the context may not be apparent to us we can, I believe, recognize the elements of a proposition. Although this may entail using a degree of imagination, this is a licence we can to some extent allow ourselves in investigative linguistics, though we cannot do so in evidential forensic linguistics.

First of all, though, we need to distinguish between the apparent locution and illocution of the text. We can imagine that the writer does not literally intend to provide the addressee with what the addressee wants, since the phrase 'I hope this is what you wanted', said in a certain way, and under certain circumstances, can mean 'I know this is *not* what you want, and that is why I am giving it to

you'. In other words, it may be the intention of the writer in this text to thwart or spite the addressee rather than to provide the addressee with something the addressee has stated he/she wants. The proposition, therefore, behind this text could be an intention to make the addressee suffer or feel guilt as a result of the writer's death.

Further – as regards the proposition itself (if it is a proposition) – whereas normally we might say 'I hope this is what you *want*', in this text the writer says '. . . what you want*ed*', implying that there is nothing the addressee can do now because it is too late. This really does seem to show the writer's intention *not* to provide the addressee with what the addressee wants: the verb is in the past tense.

By contrast with the previous seven word text, the following text is much longer and, apparently, more descriptive:

Forensic Text 11.3
Dear Mom, In the last week a number of occurrances have forced me into a position where I feel my life is not worth continuing. Friday I lost the job I have held for the past seven years. When I told my wife she packed her bags and left me. For six years she has been living with me, not for me but for my money.

Mom please take care of Mary for me. I'm leaving and I don't want Betty to have her.

I have nothing left to live for so I'm just checking out – I love you Mom, Bill
(Schneidman and Farberow, 1957)

Consider the first sentence of the text: 'In the last week a number of occurrances [sic] have forced me into a position where I feel my life is not worth continuing.' This sentence does appear to contain a proposition, namely: '. . . my life is not worth continuing.' But is this really a proposition, or is it just informational? One thing that may tell us that this is informational rather than propositional is the fact that the clause apparently containing the proposition is not thematic to the sentence. The sentence consists of two clauses 'a number of occurrances [sic] have forced me into a position' and its dependent clause 'where I feel my life is not worth continuing'. The first clause is preceded by an adverbial phrase of time 'in the last week' and the subject of the clause is an indefinite subject 'a number of occurrances'. In fact the second clause is not 'my life is not worth continuing', nor even 'I feel my life is not worth continuing', but is a dependent clause 'where I feel my life is not worth continuing'. Thus, if the proposition is part of a non-thematic, backgrounded subordinate clause, I am sure we are legitimately able to view the proposition with some scepticism, since it is axiomatic – in the model we have been proposing throughout this

chapter – that propositional content is frequently in proportion to thematicity. Moreover, consider the phrase 'my life is not worth *continuing*'. This is very different from 'my life is not worth living'. In the latter case, the verb is much more powerful and personal: 'to live'. But in the former case 'continuing' is an impersonal, middle verb, i.e. 'to continue one's life'. Even the choice of tense weakens the propositional value of the sentence: '. . . my life is not worth continuing' as opposed to 'I don't want to continue my life' – itself not as strong as 'I don't want to carry on living' or the even more direct 'I don't want to live' or – to take matters a step further – 'I want to die'.

The previous two texts, the one genuine and the other simulated, were perhaps not that easy to 'read' as to their genuineness. However, consider the following note:

Forensic Text 11.4
Dearest Mary. This is to say goodbye. I have not told you because I did not want you to worry, but I have been feeling bad for 2 years with my heart. I knew that if I went to a doctor I would lose my job. I think this is best for all concerned. I am in the car in the garage. Call the police but please don't come out there. I love you very much, darling.
Goodbye.
Bill
(Schneidman and Farberow, 1957)

Here we have I believe, a classic suicide note: it is brief, concise and highly propositional and yet there seems to be a degree of evasiveness. It begins: 'This is to say goodbye.' The context of 'this' is 'this note', 'this communication', but also, I believe, 'this act', i.e. of suicide. This is a very direct opening, one which is not easy to misinterpret. The letter states what the problem is, the fact that the writer felt he could not discuss the problem with 'Mary', and – as sometimes happens in suicide contexts – the desire not to be seen dead by the loved one. As with many suicide notes we sense – too late of course – that there was a perceived communication impasse between writer and addressee. Indeed, it is almost as if the suicide note is the only way in which this impasse could be overcome, or even just addressed. As with many other suicide notes, it seems that the text and the suicide act are one, with the suicide note having something of the status (in the writer's mind at least) of the act itself. In examples like this where we find formulations like 'I did not want you to worry. . .' it sometimes turns out that the writer – here claiming to have had heart trouble – is actually hiding something from the addressee, for example a debt, an undisclosed crime, an affair, etc. Whether or not the writer is telling the truth about the reason for

the suicide, the text is still highly propositional in regard to the act of suicide, and is very clear about the purpose of the note, about what he could or could not tell the addressee, and his resolve: 'I think this is best for all concerned.' In summary, I believe the following points can be made about suicide notes.

- A suicide note should contain an unequivocal proposition, even if it is apparently bizarre or unusual.
- The proposition is normally to do with the act of suicide itself and/or with the communication (i.e. the suicide note).
- The proposition needs to be simply phrased: it should be thematic (i.e. the first clause of the sentence/clause complex); it should be direct; it should be directed at the addressee, and relevant to the writer's relationship with the addressee.
- Where mention is made as to necessity for (or even 'desirability' of) the act of suicide, the suicide note should be unequivocal about the writer's view that this is not just the best course of action, but the only course of action.
- Genuine suicide notes tend to be short, mostly less than 300 words in length. There is usually little or no extraneous material.
- The situational context of the suicide note is not always obvious – as in Forensic Text 11.2, for example – and usually the reader has to deduce it. It is rarely stated directly.

Final death row statements

Final death row statements (also just called 'final' or 'last' statements) are a relatively recent addition to text types in forensic linguistics, and stem from the long-standing American tradition of allowing the condemned person to say a few words immediately prior to execution. This practice was not, to my knowledge, carried out in the UK, Australia or elsewhere in the English-speaking world.

Even though final death row statements represent a fairly small corpus over-all, yet there is a surprising degree of variety among them. The category is included here because it seems to constitute a discrete text type: despite the differences you will find in the following texts, there are certain crucial simi-larities which mean, in my view at least, that they constitute a single, and quite important class of text. The following is a fairly typical *tacit* admission of a crime:

Text 11.1: Dennis Dowthitt (Texas death row prisoner)

I am so sorry for what y'all had to go through. I am so sorry for what all of you had to go through. I can't imagine losing two children. If I was y'all, I would have killed me. You know? I am really so sorry about it, I really am.

I got to go sister, I love you. Y'all take care and God bless you.

Gracie was beautiful and Tiffany was beautiful. You had some lovely girls and I am sorry. I don't know what to say.

All right, Warden, let's do it.

In this statement Dowthitt does not admit the crime directly, but by saying that had he been the family of the victims then he would have had himself killed he effectively appears to be making an admission. Even though his admission is indirect, the statement itself is very direct, brutally so: 'I would have killed me.'

By contrast Mitchell, also a Texas death row prisoner, makes a more direct admission in his final statement. He also asks for forgiveness from the victim's family:

Text 11.2: Texas death row prisoner Gerald Mitchell's last statement (killed 20-year-old Charles Marino)

Yes, sir. Where's Mr Marino's mother? Did you get my letter? Just wanted to let you know, I sincerely meant everything I wrote. I am sorry for the pain. I am sorry for the life I took from you. I ask God for forgiveness and I ask you for the same. I know it maybe hard, but I'm sorry for what I did. To my family I love each and every one of you. Be strong. Know my love is always with you . . . always. I know I am going home to be with the Lord. Shed tears of happiness for me. I love each and everyone of you. Keep on living.

Betty, you have been wonderful. You guided me to the Lord. You have been like a mother to me. Sean, Rusty, Jenny, Marsha, God Bless each and every one of y'all.

Jesus, I confess you as my Lord and Savior. I know when I die, I'll have life in heaven and life eternal everlasting. I am ready for that mansion that you promised me.

Take care.

It's alright Sean, it's alright. I'm going to a better place.

Mitchell's 'I am sorry for the life I took from you' is also very direct. Unlike the previous statement, however, in Mitchell's text the mention of religion is significant. In the next statement Doughtie barely makes even an indirect admission:

Text 11.3: Jeffrey Carlton Doughtie (murdered elderly couple aged 80 and 76)

For almost nine years I have thought about the death penalty, whether it is right or wrong and I don't have any answers. But I don't think the world will be a better or safer place

without me. If you had wanted to punish me you would have killed me the day after, instead of killing me now. You are not hurting me now. I have had time to get ready, to tell my family goodbye, to get my life where it needed to be.

It started with a needle and it is ending with a needle.

Carl, you have been a good friend, man. I am going to look for you. You go back and tell your daughter I love her. Tell her I came in here like a man and I will leave like a man. It's been good, dude. Thank you, Shorty. I appreciate you. I came in like a man and I will leave like a man. I will be with you. I will be with you every time you take a shower. If you leave crying you don't do me justice. If you don't see peace in my eyes you don't see me. I will be the first one you see when you cross over.

They got these numbers that I called today. Calling my family.

That is it. Ready, Warden.

Rather, Doughtie makes the death penalty the issue, about how – in his view – it is wrong and ineffective as a punishment. Others, by contrast, refer to the conditions of death row inmates, and specifically condemn either the death penalty process itself or the way in which prisoners are kept. For example:

Text 11.4: Last statement of Jose Jesus Ceja
Arizona's death row has become a swamp of inhumane treatment with men driven to various degrees of madness and suicide. Isolation, noise, mistreatment by guards and public indifference take a terrible toll on the human psyche.

Still others deny to the end the crimes of which they were accused:

Text 11.5: Last statement of Basil McFarland
I owe no apologies for a crime I did not commit. Those who lied and fabricated evidence against me will have to answer for what they have done. I know in my heart what I did and I call upon the spirit of my ancestors and all of my people and I swear to them and now I am coming home.

What do these texts have in common? Can we sensibly group them into one category, not just because of the circumstances in which they were produced, but also because of their linguistic structure?

At first sight it would seem not: death row final statements do not initially seem to represent a single 'genre' – despite the circumstances of their production – but rather appear to form several types of text, including very particular types of admission, denial, public condemnation, appeals for forgiveness, etc.

However, all of the statements have two particular features in common, which override any of their differences. The first is that these speakers share

something crucial, regardless of innocence or guilt, admission or denial, condemnation or contrition: they are all about to die, they know it, and they are not (apparently) resisting it. They also have something else in common: they all seem to want to die with dignity or virtue of some kind. It is perhaps all they have left. This is why, I believe, they generally either

- admit the crime – and so perhaps attempt to leave the witnesses with an impression of 'honesty', 'forthrightness', etc.;
- deny the crime – and thus attempt to leave at least some witnesses with an impression of innocence;
- condemn the death row process as inhumane or ineffective – and in this way take attention away from the crime and/or any part they may have played in it, or distract from the painfulness of the moment;
- or denounce witnesses as dishonest, law enforcement as corrupt, etc., and so portray themselves as innocent victims, or extract some small measure of revenge in their last moments.

Note that the above points are purely issues of linguistic content and structure: from the linguistic point of view the actual innocence or guilt of the speakers is immaterial, as is the question of whether the system which operates the death penalty is corrupt or not. Rather, what is of importance is how these various speakers respond linguistically to the moment of execution and whether, in doing so, they exhibit some kind of unified linguistic behaviour which may have resulted not just from their common fate as people about to be executed, but also from the entire death row process through which they have lived, some for many years.

American executions have to be viewed as public acts. Even though the general public is not invited to them and they take place behind closed doors, there are always witnesses, perhaps as many as 20 or 30 in some cases. This means that anything the condemned person says at such a time is also public. Condemned speakers need to create or make an impression, to be seen to be standing up for their corner. The context is not only public, but is – specifically – set within the judicial framework: the language used – e.g. 'a better or safer place', 'inhumane treatment', 'public indifference', 'crime I did not commit', 'fabricated evidence' – is articulate and appropriate, with just a hint of legalese. It too reflects the setting. These are not simple farewell texts, but complex commentaries on society as seen from death row. They constitute a type because only someone who has been in that situation – i.e. awaiting execution for some

years and then knowingly facing the moment of death at the hands of the state – can produce such a text. Suicide notes and ransom notes also constitute scientific types, but they are not usually within the heavily institutionalized setting of death row prison units, whereas final statements are. It is a matter of debate as to whether final statements are of any use in determining guilt or innocence. However, of one thing we can be certain. Providing the provenance of the final statement text checks out, we are never in doubt as to the genuineness of the text.

Confessions and denials of public persons

Previously we looked at Derek Bentley's alleged statement from an authorship point of view, while elsewhere we evaluated some aspects of veracity in Timothy John Evans' statements as part of a general investigation into statement analysis. These texts are different from those of the final statement variety, and those we have considered under the headings of authorship and statement analysis: these texts were written or spoken by public figures. Francis Bacon, for example, wrote:

> I am ready to make an OBLATION of myself to the King, in whose hands I am as Clay, to be made into a vessel of Honour or Dishonour. . . . Yet with respect to this Charge of Bribery I AM INNOCENT. I never had Bribe Or Reward.
> (Letter to the Duke of Buckingham, 1617)

In this text Bacon offers a sacrifice of himself to save the king's 'honour', while maintaining his own innocence. In his statement from the gallows Henry Garnet, head of the Jesuits in England, confessed to his part in the Gunpowder Plot as follows:

> Good countrymen, I am come hither this blessed day of The Invention of the Holy Cross to end all my crosses in this life. The cause of my suffering is not unknown to you. I confess I have offended the King, and am sorry for it, so far as I was guilty; which was in concealing it, and for that I ask pardon of his Majesty.
> (Harrison, 1941)

Finally, let us consider the closing words of Nelson Mandela's statement at his treason trial, at the end of which Mandela was sentenced to life imprisonment on Robben Island.

> During my lifetime I have dedicated my life to this struggle of the African people. I have fought against white domination, and I have fought against black domination. I have cherished the ideal of a democratic and free society in which all persons will live together in harmony and with equal opportunities. It is an ideal for which I hope to live for and to see realised. But, my lord, if it needs be, it is an ideal for which I am prepared to die.
> (The British Library National Sound Archive)

Why are these public denials and confessions important linguistically? These are public figures responding to their accusers, sometimes in public and sometimes in private. In these texts – diverse as they are – we have prominent people, some of whom are members of the establishment, facing and responding to the establishment. Bacon is prepared to save the king's honour, because he believes that this is what a loyal subject should do, but he insists on his innocence as a matter of principle. Garnet admits to having known that the Gunpowder Plot was to take place. He is confessing for one reason only: to save people of his own religion from persecution. Finally, we have the statement of Nelson Mandela, who states that he is prepared to die for his beliefs. All of these individuals believed they were acting from principle. There are surprising linguistic similarities between these three excerpts. Both Garnet and Mandela specify aspects of a struggle, Garnet referring to 'crosses' and 'suffering', Mandela to 'this struggle of the African people'. Bacon states his preparedness to making a 'vessel' of himself. The idioms are different, and there are different shades of meaning, but the communicative purpose is very similar across all three texts, making them worth further study as part of a possible genre.

Forensic phonetics **12**

Forensic phonetics is the use of phonetic techniques primarily in the analysis of the voice applied to criminal investigations. It includes technical voice comparisons, *naif* voice recognition, transcription of spoken language including transcription of disputed utterances, speech signal enhancement and the authentication of recordings. In this chapter I am going to depart from the general structure of previous chapters. I am aware that some readers will be more familiar with basic phonetics concepts than others. Therefore, in order to assist those who are not experienced and to avoid plaguing those who are with unnecessary material, this chapter will contain 'general information' panels. The more experienced reader can simply disregard those and move on. Having said all of that let us begin with *naif* voice identification. Just remember to ignore the 'grey' panels if you already have at least a basic knowledge of phonetics.

Information Panel 12.1 Basic information about phonetics
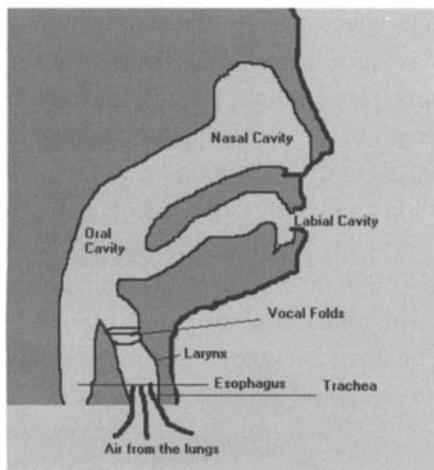
- Speech consists of sounds made with our speech apparatus, i.e. lungs, vocal folds, oral and nasal cavities, lips, etc. The smallest sound unit of speech is known as the phoneme, for example the /f/ in *forensic*. When we write phonemes using phonetic symbols they look not unlike the letters we use every day to write words in the language, but there are some important differences. A phonetic symbol represents one and only one phoneme in the language. Any single letter of the 'normal' alphabet, however, can represent several different sounds in the language. For example, the letter 'a' is pronounced differently in *father, ape, any, fall, etc.* However, in phonetic script the 'a' sound in *father*, for example, is written /ɑ/, but the 'a' sound in *ape* is transcribed /eɪ/.
- The same applies to a number of other vowel sounds in English. In fact English has a very complicated vowel structure, so that although we have only five letters to represent our vowel sounds, there are in fact about 20 actual vowel sounds and diphthongs in the language (depending on which dialect you speak). Compare this with the five vowel sounds in Spanish and Italian. Consonants, too, are not that straightforward to represent. Consider, for example, the different sounds for the letter 's' in i**s**, ki**ss**, plea**s**ure, **s**how. As you can see, what we think of as 's' is actually a number of different sounds.
- However, phonetics is about much more than phonetic transcription. It concerns, for example, how and where speech sounds are made (articulatory phonetics); how speech sounds are perceived (auditory phonetics); and speech as wave structure (acoustic/experimental phonetics).

How do we recognize voices?

We are all familiar with the idea of 'line-ups', the process whereby a group of otherwise unconnected people are lined up in a room to be looked over by a witness for the purposes of identification. As a procedure the suspect line-up has its detractors, with a number of researchers demonstrating its deficiencies, e.g. Malpass and Devine (1981), Loftus (1975). Nevertheless, the judicial fondness for eye-witness identification is well known and for many courts, voice line-ups have also proved popular. Superficially, the voice line-up resembles the more familiar identification parade. The investigator obtains recordings of the suspect voices and a number of 'foil' voices which vary in the degree to which they resemble the suspect voice. The voices need to be recorded under similar conditions to each other. The witness is then given the opportunity to listen to the voices and identify, if appropriate, one of them.

Unfortunately, as Hollien points out, there are some substantial differences between visual and auditory memory, including: '... differences between how the auditory/visual memories are processed ... how a voice is structured ... as opposed to how visual features are assessed and stored ... the ways in which fear [etc.] can affect the two ... processes ... the ways in which poor eyesight and hearing disabilities relate, and ... the abilities ... which ... make some ... [people] better at identification than others' (Hollien, 2002: 95).

Information Panel 12.2: The speech process

Nasal Cavity

Labial Cavity

Oral Cavity

Vocal Folds

Larynx

Esophagus   Trachea
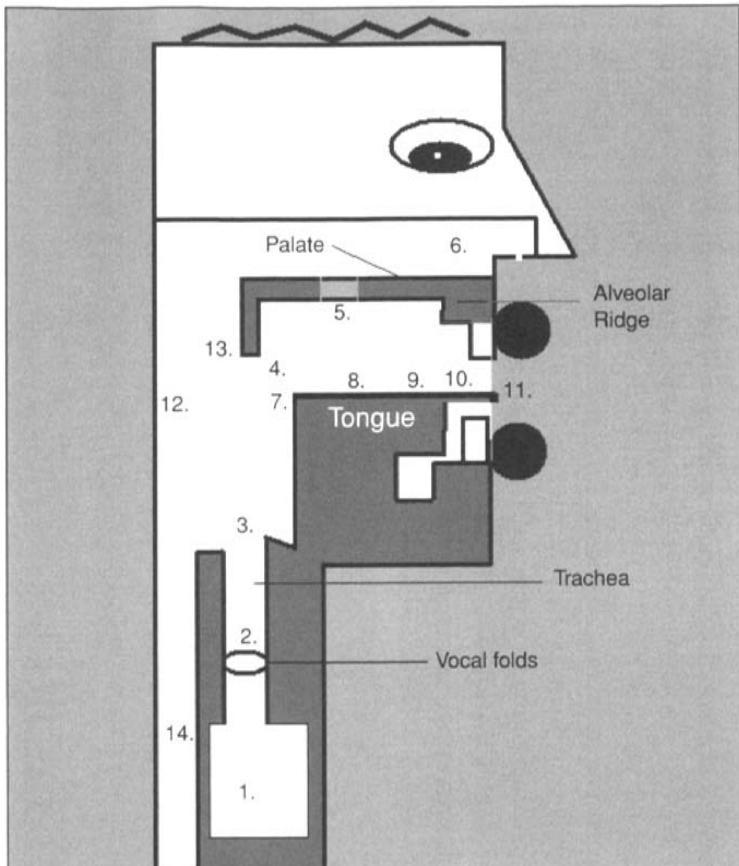
Air from the lungs

The speech process begins with air pulsing upwards from the lungs, after which it enters the larynx (wind pipe) prior to reaching the vocal folds (vocal cords). These are a pair of elastic-like muscles which are capable of being vibrated at an astonishing range of frequencies, from very low to very high. In women and children they are usually shorter than in men, and so women and children mostly have higher pitched voices than men. When the vocal folds vibrate to produce speech the sounds so produced are called *voiced*, otherwise they are *voiceless*.

According to van Lancker and Kreiman (1987) voice recognition is a pattern-matching exercise, and begins in the womb when prenatal foetuses can distinguish the mother's voice from that of other speakers (Kisilevsky *et al.*, 2003). On the other hand, perceiving an unfamiliar voice is more a case of analysing and categorizing features (Hollien, 2002: 97). Thus, when we receive a phone call from a voice that sounds familiar, and before the other speaker says 'It's Fred', we will already have begun our attempts to recognize the voice based on a number of stored patterns. When the voice is unfamiliar, we will probably begin by discarding the features we hear unless they are significant for some reason. However, if the call is important psychologically – as in the case of a threat call – we will probably store some of the features without realizing that we are doing so. It is also important to realize, as McGehee (1937) discovered, that memory for voices can fade rapidly, as opposed to memory for faces which is much more durable.

Memory for voices and voice line-ups

The classic study on memory for voices was carried out following the kidnapping and death of Charles Lindbergh Jnr. The child's father, an American hero, claimed in court to have recognized Bruno Hauptmann's voice after a lapse of several years. According to McGehee (1937) memory for voice was accurate in 83 per cent of listeners after one day, but reduced over a period of two weeks to 68 per cent, 35 per cent after three months and down to 13 per cent after five months. In a more recent study Hollien cites a number of factors including whether the witness believes the suspect voice is present in the line-up (cf. the instruction given to eyewitnesses at identity parades in the UK, viz. 'the person you saw may or may not be present'), length of spoken sample, and quality of spoken sample (Hollien, 2002: 97–100). Clifford (1980) found that the time lapse between hearing a voice and holding an identification session was also one of the factors that could affect accuracy of recall. Researchers in this field concur that voice line-ups can be successful, so long as they are held without delay, and are managed by investigators who understand the phonetic and linguistic issues, including the nature and quality of samples, the number of foil voices (Hollien suggests between five and eight) and the emotional state of the witness. It is important to realize that lay voice identification can be influenced by many factors and is not infallible.

1. Air flows up from the lungs.
2. It passes through the glottis more easily if the sound is voiceless – the glottis is the open space between the vocal folds.
3. The air continues up and reaches the epiglottis which only folds down over the windpipe when food is going to enter the esophagus (14).
4. The air flows into the vocal tract and along the top of the tongue if the sound is oral, but if it is nasal, then
5. The velum (also called the 'soft' palate) drops and lets some air into the
6. Nasal cavity. Otherwise the velum (5) stays raised and blocks off the nasal cavity (6).
7. The tongue starts at the root, we then have the back or dorsum (8).
9. This is the front of the tongue.
10. Then we have the blade of the tongue.
11. Finally, we have the tip of the tongue.
The big dark circles are the lips. The white rectangles behind them are the teeth.
12. This is the pharyngeal wall.
13. This is the uvula.
14. This is the esophagus where food goes down into the stomach.
Remember that (3) the epiglottis (a flap of sorts) will close the windpipe (trachea) off when food is in the offing – to prevent food getting down the windpipe (trachea).

Figure 12.1 Speech articulators

Information Panel 12.3 Types of English speech sound by place of articulation

Labial: Sounds produced by or at the lips (/p/, /b/). Labio-dental: Sounds produced by the upper teeth coming in contact with the lower lip (/f/, /v/). **Dental:** The tongue touches the teeth or protrudes through them (as in 'th' sounds) (/θ/, /ð/). **Alveolar:** The tongue comes in contact with or is near to the aleveolar ridge just behind the teeth (/s/, /z/). **Palatal:** The front of the tongue is in contact with or in proximity to the front of the palate (the 'y' sound in 'yes', written /j/). **Velar:** The tongue is in contact with the velum (soft palate) (/k/, /g/). **Glottal:** The glottis is closed (glottal stop, /ʔ/. **Nasal:** The velum opens to admit air into the nasal cavity (/m/, /n/). This is not really a place of articulation, but is included here for completeness (see Figure 12.1). See Appendix 2 for all phonetic symbols (IPA Chart).

Notes on transcription

There are two kinds of phonetic transcription. When we wish to render a phoneme or a word into phonetic script we enclose it within forward slashes, for example /fə'netɪks/ is the word *phonetics*. However, when we wish to show the detail of how a particular individual says the word *phonetics* on a particular occasion, we enclose the transcription in square brackets. For example, an American friend of mine says what sounds like 'phonedics' rather than what sounds like 'phonetics'. To show an example of him saying *phonetics* I need to enclose my transcription in square brackets. It is an actual instance of someone saying something, rather than an abstract idealization of how a word is pronounced. Hence I would write:

[fə̃'neɾɪks]

In this transcription you see the character /ɾ/. It is known as the alveolar tap. Note also that above the first vowel, the schwa, written /ə/, is a tilde sign. This means that the vowel is nasalized – it is anticipating the manner of articulation of the following consonant /n/. We could add more detail. Supposing, the speaker was slightly unusual in that he voiced the /k/ sound, we would then have:

[fə̃'neɾɪg̬s]

Here we see the 'v' symbol under the /k/ to indicate the phoneme is being voiced. We could continue for quite some time modifying the transcription to reflect, in as much detail as we want, the speaker's utterance. No transcription is ever really complete. The two types of transcription are known as phonemic and phonetic. A phonemic transcription is enclosed in forward slashes. We call it a phonemic transcription because it is a transcription of the phonemes, whereas in a phonetic transcription we are transcribing what we believe a person actually said on a specific occasion – the next time they say the word in question, the transcription will be different, even if only in apparently minor ways.

Summary of transcription types

Phonemic transcription is the transcription of the language's phonemes. The phonetician assumes a 'standardized' value for the speech sounds of the language and transcribes them according to those values. What you get in a dictionary next to a head word is a phonemic transcription of that word. It is an idealization of the pronunciation according to a 'standard' dialect, such as RP (received pronunciation). A phonemic transcription is usually shown inside forward slashes, e.g. /həˈləʊ/

Phonetic transcription is a phonetician's attempt to capture a particular individual's pronunciation of a particular word or utterance at a particular moment. It can be a **broad phonetic transcription** or a **narrow** one. If it is **broad**, then the phonetician is attempting to convey only the most essential features (just a bit beyond the phonemic transcription), but if it is **narrow**, then it means the phonetician is trying to convey as much detail as possible. We can also talk of a **focused phonetic transcription**, which is where the phonetician is attempting to highlight a feature of perceived importance without, however, giving too much detail of other features. A phonetic transcription is usually shown in square brackets as in the examples given above.

Transcription of disputed utterances

Sometimes an entire case can hang on one word. It is critical, therefore, that the correct interpretation of the disputed word is made – with the proviso that this is possible, since the quality of many recordings, especially those made covertly, is often poor. In one case two suspects were discussing a crime. The one was heard to say '. . . weapons that we used' or 'weapons that were used'. Simply listening

to the tape over and over yielded little, although it was suspected that there was a little colouring of the vowel with an /r/ sound. The accent in question was a rhotic one, meaning that the speaker would produce /r/ between vowels unlike, say, a typical south of England accent which is non-rhotic. In this particular case the use of the spectrogram proved invaluable. A brief description is in order. I will then show what the spectrogram revealed about this example.

How sound travels

Sound travels in waves. Waves have amplitude and frequency. Amplitude translates roughly to 'loudness' and frequency translates to perceived pitch.

Sound illustration 12.1: Sound wave of a voice rising in pitch and volume
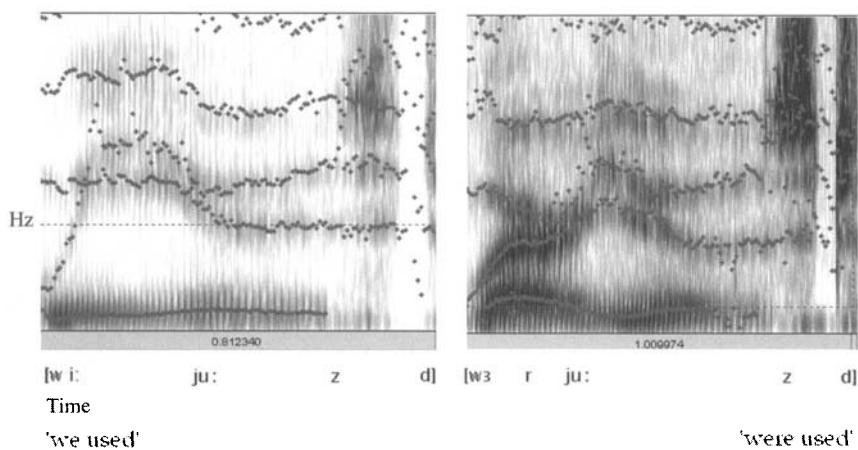
The above illustration is of four sung notes. You can see the four different shapes from left to right. The first note is low and soft. The notes get progressively louder as you move along the horizontal axis (time). The louder the sounds get the longer the vertical lines also get. This is *amplitude*. I hope that you can see that the solid looking 'blob' shape on the right has much darker lines than the wavery slivery shape on the left. The lines on the right are much darker simply because they are closer together, and they are closer together because they are being sung at a much higher pitch – a much higher note, if you wish. This is *frequency*.

When we hear someone sing, we may imagine that we only hear one note, namely the note which that person is singing, for example a note beloved of opera fans is 'High C'. However, we are not hearing just one note, but actually several notes. First, we are hearing the base note of the voice – which phoneticians call the 'fundamental frequency' or the first harmonic. If you stripped

away all the talking apparatus and could just hear the vocal cords vibrating, that would be your singer's 'fundamental frequency', the first harmonic of the voice. The other notes you hear cannot really be called notes as such. They are the frequencies around which the voice is centring most of its energy. They are based on the shape of the oral tract – which is to say the vocal tract and the nasal tract. They are called formants. The exact value of each formant will depend on the shape of the oral tract at the time, and this depends on what speech sound we are making or attempting to make. You can imagine therefore that different sounds we produce will necessitate the articulators to be in different positions, and that this will change the shape of the oral tract, and that this in turn will change the position of the formants.

So, to return to our example, here is what 'were used' and 'we used' look like on the spectrogram:

Sound illustration 12.2: Spectrogram of 'we used' and 'were used'

| [w i: | ju: | z | d] | [w3 | r | ju: | z | d] |

Time

'we used'   'were used'
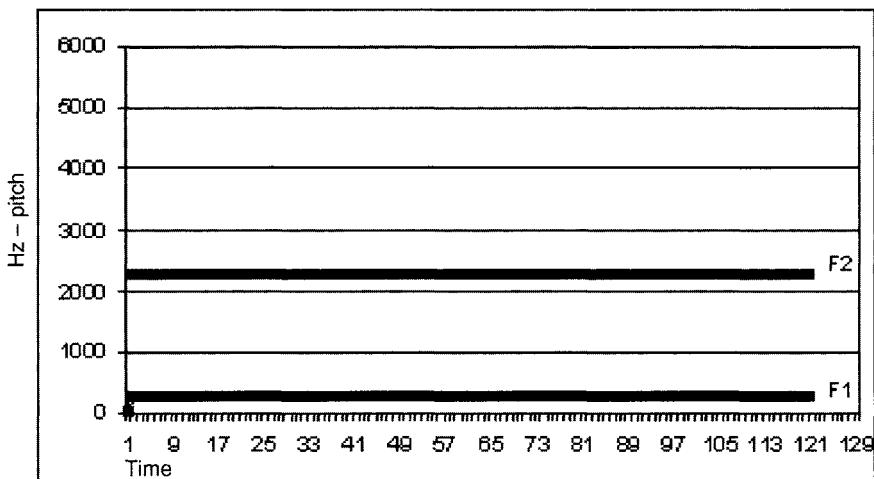
For the moment, ignore the details – just observe that there is a difference. Now let's talk about formants, and then we will return to this illustration.

The most easily visible formants are those which correspond to vowels. There are two simple reasons for this. First, to make a vowel sound you always have to use the voice – the vocal folds (forget 'vocal cords', they're called 'folds'). Second, there are no obstructions when we make a vowel – the tongue stays out of the way, and nothing is being pressed against the teeth, the palate or the lips. This is why vowel sounds are so clear (they belong to a group of sounds termed 'sonorants') and why their formants are so clear. If we had to
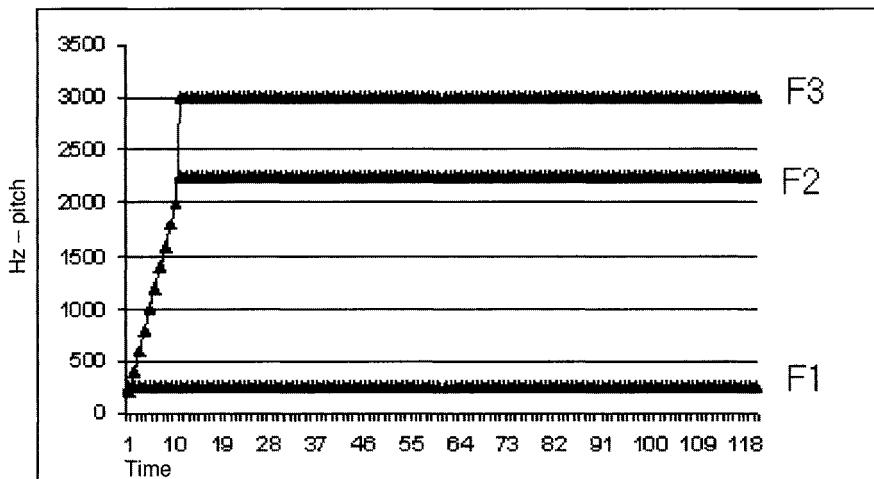
produce a graph of the first two formants of the sound /i:/ (like 'ee' in 'bee')
the spectrogram would look something like this:

Sound illustration 12.3: Graph version of spectrogram values of /i:/
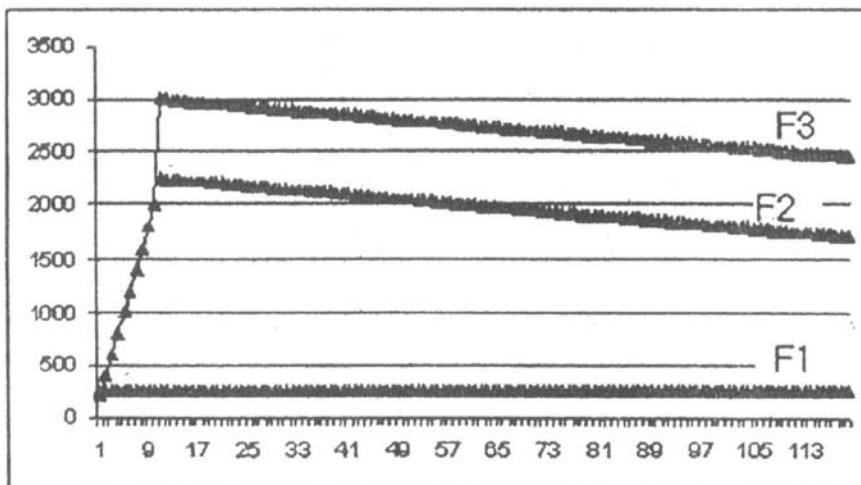
In this graph F1 (the first formant) is at 250 Hertz (the measurement for
frequency) and F2, the second formant is at 2250 Hertz (abbreviated Hz).
When we see a spectrogram and it looks like this, we are probably looking at
/i:/. If we introduce a /w/ before the /i:/ sound it looks like this:

Sound illustration 12.4: Graph of spectrogram values of 'we'

What happens is that as the mouth is forming /w/ it changes shape, altering the position of the formants. As we go on to make the /i:/ sound the formants assume the position you saw in the previous graph, except this time I've added the third formant, F3 (F1 is at the bottom, then F2, then F3 at the top). For the next illustration I'm going to add the /r/ sound, as it occurs in a sequence like 'we're' or 'were', depending on the dialect of the speaker. I'm going to exaggerate the effects slightly for illustrative purposes.

Sound illustration 12.5: Graph of spectrogram values of 'were'

What we see happening in this graph illustration of the spectrogram values is that after forming /w/ the formants 'dip' in order to get ready to form the /r/. After a short period of time (about 150 milliseconds or so) they will re-assume (or approach) the values that we saw in the previous graph. Now go back to Sound illustration 2 and look at the difference between the left and right spectrograms. You will see that the one on the right dips very much as the graph above does. This was how we knew, or rather my colleague Dr Peter French knew (since he did this particular analysis), that the words being spoken were 'weapons that were used' and not 'weapons that we used'. Had it been the latter, it would have meant that the speaker was quite possibly admitting having committed the crime (in this case one of murder).

Information Panel 12.4: Manner of articulation

A sound can be like an explosion or 'popping' noise: we call these **plosives** - /p/, /b/, /t/, /d/, /k/ and /g/ are the English plosives, as well as the glottal stop (/ʔ/). Above I referred to **nasal** sounds in the 'place of articulation' information panel. However, a nasal speech sound refers to the manner in which the sound is articulated, i.e. 'nasally'. **Fricatives** are also an important type of speech sound: think of the friction produced when you say 'sssss'. The space between the tongue and the alveolar ridge becomes constricted, increasing the air pressure – hence the hissing sound of 'sssss'. Other fricatives are: /f/, /v/, /θ/, /ð/, /z/, /ʃ/, /ʒ/. We also have **affricates** – where there is friction after release, but where the tongue barely comes in contact with the inside of the mouth, for example the 'tch' and 'dj' sound in 'change', written /tʃ/ and /dʒ/. An example of a **lateral approximant** sound is /l/ - the air is forced out of the sides of the tongue. The /r/ sound in English, more correctly transcribed /ɹ/ (i.e. 'upside down r'), is an **approximant** (actually an alveolar approximant if you think about where the tongue is when /r/ is produced).

What is a spectrogram?

A spectrogram (also called a spectral view) is a way of representing certain aspects of sound – any sound, including speech. It is particularly useful when used for representing speech because it tells us several things at once: the horizontal axis gives us time. This way we can tell what is happening in the speech signal at any point in time. The vertical axis represents the frequency of the sound (is it high-pitched or low-pitched or something in between?). The dark, roughly horizontal, bands represent intensity, the amplitude (roughly translated as 'loudness') of a particular sound. The dark bands are, in fact, the formants. For most purposes the first four formants are sufficient to analyse speech, referred to as F1, F2, F3 and F4. You will also see F0 mentioned. F0 is in fact the rate at which the vocal folds are vibrating (the *fundamental frequency* of the voice referred to earlier). It is not usually given on the spectrogram unless you have a software program which you can also set for 'pitch'. For vowels, F1 and F2 are the most relevant. F1 is associated with the degree to which the vowel is open or closed. The closed vowels have the lowest F1 frequencies. Most of the illustrations here are in chart form with the formant values entered as

graphs. This is to make them easier to read. The only actual spectrogram so far is Sound illustration 12.2 above.

Here is a vowel chart (courtesy and copyright of the IPA).
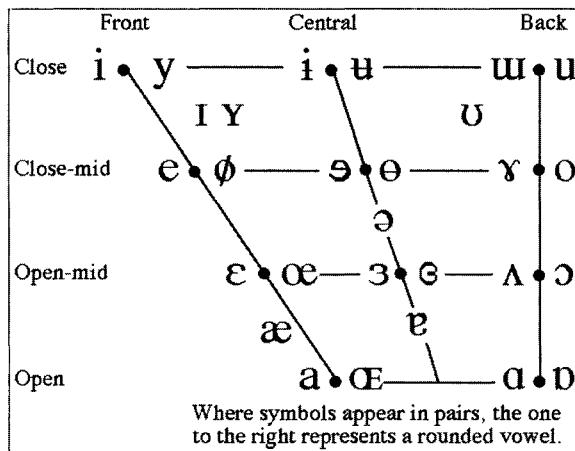
Figure 12.2 IPA vowel chart
© IPA
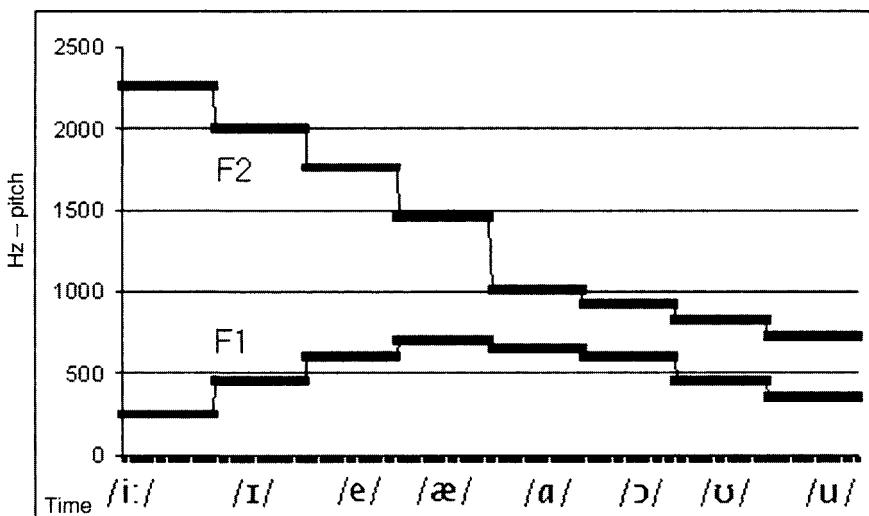
In terms of most English dialects the two close vowels are /iː/ and /u/ (*peat* and *tooth*). F1 for /iː/ lies at about 250–280 Hz for most speakers. F1 for /u/ lies at about 300–350 Hz for most speakers. These are the close vowels found in English.

Next are two vowels which are about halfway between the close and close-mid vowels. These are /ɪ/ and /ʊ/ (*pin* and *put*). F1 lies at about 450 Hz for these two vowels. The more open the vowel, the higher F1. Take a look at the vowel second from the bottom on the left in the above vowel chart. This is /æ/ as in *pat*. For most English speakers this is the most open vowel. F1 for this vowel is at about 700 Hz.

As we go from front to back, F2 *decreases*. F2 apparently has a different function from F1. Whereas F1 relates to open vs closed vowels, F2 relates to front vs back vowels. Again, /iː/ is the most front vowel, with /u/ right at the back. Thus, /iː/ has the highest F2, while /u/ has the lowest F2. Here are the graph values for a spectrogram of F1 and F2 for the vowels of English:

Sound illustration 12.6: Graph of spectrogram of F1 and F2 for some English vowels

Note: spectrogram values are approximate

The first four vowels shown are front vowels (/iː/, /ɪ/, /e/, /æ/); the last four are back vowels (ɑ, /ɔ/, /ʊ/, and /u/). Note that for the front vowels F1 increases as F2 decreases, while for the back vowels F1 and F2 decrease together. Put another way we see that F2 decreases steadily from front to back, while F1 initially increases then decreases.

Another disputed utterance

A recent case involved covert recordings of a number of conversations between a police officer and a suspect in a moving vehicle. One of the topics of conversation between the officer and the suspect was a third man, known as 'Ernie'.

However, in one section of the tape a new name appeared to surface: 'Ronnie'. Police officers were not aware of anyone involved in the case by the name of *Ronnie*, and suspected that the name was actually *Ernie*, but that the poor signal of the recording made it sound like *Ronnie*.

The tape was listened to and as many as possible of the instances of undisputed 'Ernie' were excerpted. A sample of these was selected and spectrogram values were obtained for the first phoneme, pronounced 'er', written

phonetically as /ɜː/. The same procedure was followed for disputed instances of the name.

The surveillance tapes in this case present a number of acoustic problems including, in some instances, an intrusive electronic-sounding crackle, the sound of the car engine, the playing of the car radio, the movement of the target vehicle, etc. The intrusive noise sometimes coincided with the first syllable of the disputed name in this case, which complicates acoustic analysis.

Sources of auditory confusion

It may be asked why as hearers we might confuse 'Ronnie' with 'Ernie'. Surely, it will be asserted, there are clear differences between these two names? However, there are a number of similarities.

First, both names are bi-syllabic, they have a similar stress profile, with the accent falling on the first syllable and they have identical second syllables.

The first vowel in 'Ronnie', written phonemically as /ɒ/ is an open back vowel. It is not very far from the /ɜː/ of 'Ernie' which is an open-mid central vowel. We can see this clearly from the vowel chart below:
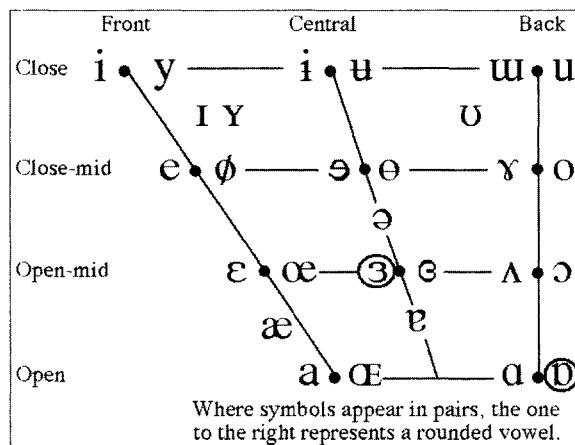
Figure 12.3 IPA vowel chart, with target vowels circled

In the figure above, the target vowels, the 'o' in 'Ronnie' and the 'er' in 'Ernie' have been circled. As can be seen the two vowels are not that far apart on the

vowel chart. Additional auditory confusion can arise because as listeners we are susceptible to influences from a number of sources, including the difficulties of separating *signal* (what we want to listen to) from *noise* (everything we do not want to hear), the presence of co-articulation effects (i.e. the sounds surrounding the target sound), and subjective factors such as fatigue. Moreover, the speaker who appears to say 'Ronnie' when – as I will argue shortly – he is, I believe, saying 'Ernie' speaks a dialect of English in which the /r/ is not emphasized. When some speakers of English appear to pronounce the /r/ this is referred to as *rhoticization*. In the male speaker's case in the present instance he speaks a dialect of English where this is not usual. Thus, there would sometimes not be a great deal of difference to the hearer, in rapid conversation, between 'Ernie', 'Ronnie' and even 'Rinnie' or 'Arnie'. In both 'Ernie' and 'Ronnie' the first vowel is in transition to /n/, which is a further similarity between the two names. In the present instance I would argue that the initial phoneme of what sounds like 'Ronnie' is in fact a vowel in transition rather than being /r/ itself. Since both /r/ and /n/ are alveolar sounds this may be adding to any possible auditory confusion.

However, I suggest that any auditory confusion is only applicable to 'over-hearers' to the conversation, not the participants themselves. As overhearers we are not aware of all the contextual factors relevant to the conversation, and, so – effectively – it is of no contextual importance to us whether the name is 'Ronnie' or 'Ernie'. However, the participants in the conversation were probably aware of the contextual factors, and so would be able to make the necessary semantic distinctions at the time of the conversation. It should be noted that if the name were 'Ronnie', what is surprising is that the police officer did not comment on it or seek any information on it.

Could a speaker say 'Ronnie' when he/she meant 'Ernie'?

As speakers of a language, in this case English, we 'know' the rules of our language, including how to pronounce particular words. However, as humans we are subject to the limitations of performance, and so can easily misarticulate words. All kinds of factors, including lack of concentration, age, fatigue, emotion, etc., might conspire to make this happen. It takes only fractionally different positions of the articulators and very slight mistimings to make a word like 'Ernie' appear to sound more like 'Ronnie'. When two speakers know each
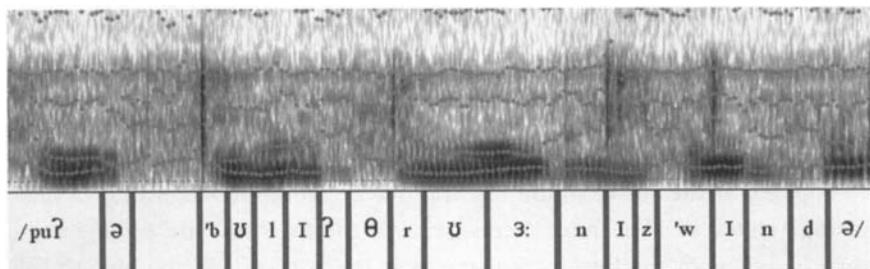
other well, and converse frequently – as was the case between the police officer and the male in question – it is not unusual for them to speak rapidly and pronounce their words with a certain degree of carelessness. It is symptomatic of their familiarity with each other.

At one point the male target says (according to my transcription):

'...put a bullet through Ernie's window', transcribed as [puʔ ə ˈbʊlɪ θˈu ʷɜːnɪz wɪndəʰ].

The defence claimed that this was actually '...through Ronnie's window'. However, I was able to show that this confusion was brought about largely through the transition from the unstressed vowel /u/ in 'through' to the /ɜː/ in 'Ernie'.

Sound illustration 12.7: Spectrogram illustration of 'put a bullet through —'s window' (broad transcription)

What sounds like 'through Ron-' is in fact 'through Ern-', but, unusually, there is no stressed syllable in 'Ernie' and there being no intervening consonant between the vowels the formants do not show much movement from the first to the second vowel. In fact, the vowels /u/ and /ɜː/ are linked, i.e. there is no break between them, with /ʷ/ serving as a linking transition.
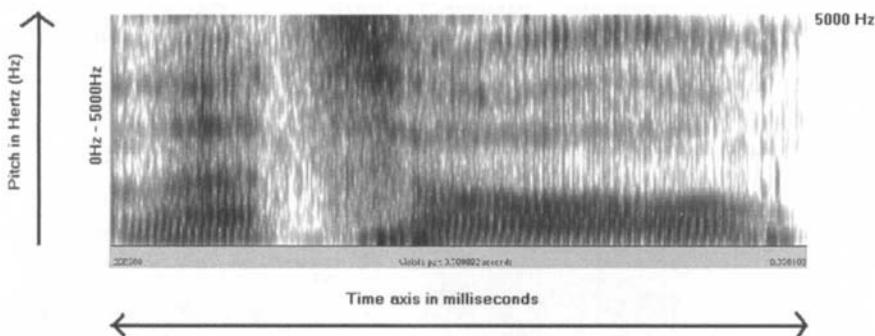
In addition, I took a number of spectrogram measurements of the disputed and undisputed sections of the tapes. These showed that the formant measurements corresponded to an open-mid vowel rather than an open-back vowel. I also looked at how the speaker produced /r/ and found that this was quite definite on the spectrogram. I was therefore able to conclude that the disputed word was more likely to be 'Ernie' than 'Ronnie'. The defence did not oppose this opinion.

In the next section we will revisit spectrograms. If you already know about spectrograms feel free to skip this section.

Anatomy of spectrograms

Consider the spectrogram below. In it a male speaker is saying: 'That's all'. This is a reasonable quality recording into a computer microphone. It is not the best quality, but it is better than the mobile phone one we will also be seeing in this section.

Sound illustration 12.8: Main axes of a spectrogram
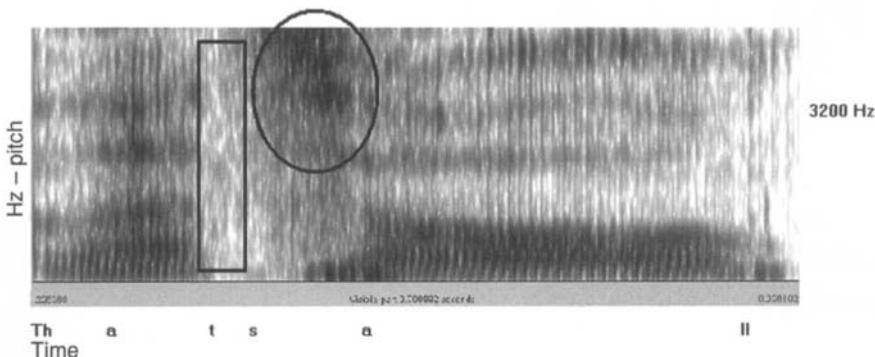
Looking at the above sound illustration, we see a spectrogram of a male saying 'That's all'. The pitch frequencies of the formants are given on the vertical axis with the lowest frequencies at the bottom and the highest frequencies at the top. These correspond to the harmonies of the voice. The voice, as we hear it, may sound as if it is on one note or pitch only, but in fact many 'notes' are being sounded at any one millisecond. The illustration shows pitch values between about 20 Hz and 5,000 Hz. The horizontal axis shows the length of time it took to say 'That's all'.

What else can we tell about this spectrogram? The fact that we can see a good contrast between the darkest and lightest areas tells us that the sound is of reasonable quality. The dark areas at the bottom indicate the voiced parts of the utterance. The 't' and 's' in 'that's', for example, are not voiced (say them aloud to yourself and you will see what I mean). Also, the very earliest part of the spectrogram, on the left, shows an unvoiced part – this is the initial part of the 'th' in 'that's', before voice onset begins.

The contrast detail is helpful in breaking down the constituents of the spectrogram and shows that the sound is certainly good enough to enable us to do an analysis. The vertical striations are also a good indication. They give us some further information about pitch. The closer they are together, the higher

pitched the voice will be at the fundamental frequency. In the present instance we see a medium- to low-pitched male voice. Now let us introduce some more detail:
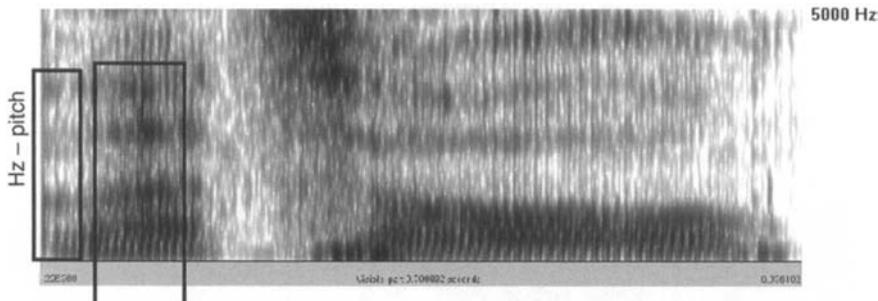
Sound illustration 12.9: Showing a voiceless stop and a fricative in the spectrogram

For ease of reference I've enclosed two areas. Look at the first one, a rectangle. You can see I've transcribed the utterance (using conventional orthography – except I've omitted the apostrophe). Underneath the rectangle you will see the second 't' in 'that'. Note that there are no dark areas in the rectangle. This is not just because 't' is not voiced. It is also because at this point nothing is happening: 't' has already been 'released' – the final stage of issuing a plosive sound. Time is waiting for something else to happen in these few milliseconds. Now look at the second enclosed area, an ellipse. Within the ellipse I've drawn you can see (starting at about 3,200 Hz – look over to the right of the illustration and you will see 3,200 Hz) a dark area. In case you thought all voiceless sounds on a spectrogram showed virtually nothing, as was the case with the 't' we have just been talking about, this shows that it is not so: the dark area in fact is nothing to do with voice, but with the hissing found in 's' – the final phoneme of 'that's', transcribed (/ðæts/). This kind of energy banding – note its irregular shape – is the aperiodic (i.e. irregular) noise you get from fricatives like /f/, /s/ and so on. The /s/ sound is particularly noisy in this respect. Notice that immediately after the /s/ sound the dark, irregular banding at the top of the spectrogram disappears – the /s/ has come to an end and we are now into the vowel, the 'a' in 'all', actually transcribed as /ɔ:/.

But this isn't all! There is a lot more to see on this spectrogram. Let us look at the left-hand side of the spectrogram, at the consonant and vowel in 'tha', written /ðæ/.

Sound illustration 12.10: Looking at the consonant and vowel in /ðæ/.

First, the consonant /ð/. This is shown on the extreme left, with the rectangle overlapping the edge to make it more visible. Note that there are four dark grey horizontal bands in that first rectangle.

Those are the formants which are being made for /ð/. In fact, as I've drawn the first rectangle I've overlapped slightly onto the vowel. This is because it is quite difficult to see a voiced fricative on a spectrogram and so I am including some of the vowel to make it easier. Notice that the horizontal bands in that first rectangle are not as dark as the horizontal bands in the second rectangle.
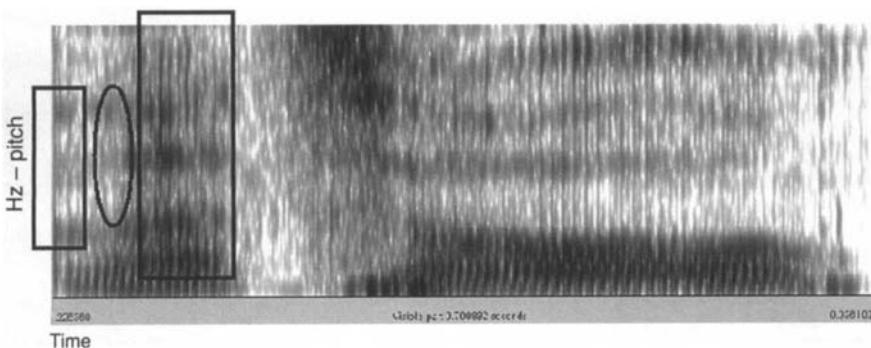
Information Panel 12.5: Information about the tongue

The tongue is the most flexible organ in the human body. It can assume thousands of shapes, it can retract, advance, it can become hollowed in the middle, it can be a convex shape or concave, it can reach almost any point in the mouth, it can elongate or become, in effect, shorter. For speech purposes, it consists of several parts: the root, which is embedded in the front wall of the pharynx, the back or dorsal part of the tongue which lies below the velum, the front which lies below the palate when at rest, the blade which lies below the alveolar ridge, and the tip of the tongue which gets the most action at the front of the mouth – with dental and alveolar sounds, for example. Poets have written odes to the tongue, but phoneticians rarely say much about it. It is the most exciting, mysterious organ imaginable: it can obstruct parts of the mouth to produce fricatives, it can vibrate to produce trills, it can tap like a hammer to produce flaps and taps, it can hover just below the alveolar ridge to produce /s/ and /z/, it can move like lightning or effortlessly in slow motion, untense itself as it forms the release part of an affricate. It can fill the mouth or take up very little space according to need.

This is because in this second rectangle we have only a vowel /æ/. Vowels should always show good contrasts on spectrograms depending on sound

quality. However, what about transitions between consonants and vowels? Again, it helps to have a good quality illustration, which in turn depends on the actual sound quality of the recording in the first instance. In this next spectrogram we will look at the first consonant–vowel transition in the utterance. Look for the different degrees of darkness in the horizontal bands across the spectrogram – the formants.

Sound illustration 12.11: Transition from /ð/ to /æ/.

In the first rectangle we see the consonant. In the second shape, the elongated ellipse, we see the transition from the consonant to the vowel. Finally, in the third shape, the bigger rectangle, we see the vowel itself. Note that the horizontal bands get darker as we go from left to right. Why is this? Vowels show much higher energy at each formant than other types of sound. Recall /s/ from an earlier illustration. It shows a lot of energy, but it's not regular – it's not shown as separate horizontal bands, i.e. formants, as is the case with the vowel /æ/ shown here. Thus, when we say /ð/ we are phonating, and hence producing some formant energy, but not as much as for a vowel. As we transition into the vowel the fricative loses its power and the vowel begins to take over – hence we have more vocal energy. Finally, as we get into the vowel proper we see even darker horizontal bands. Another important transition is between the vowel and the consonant in the word 'all', written /ɔːl/. What happens here is interesting. The vowel /ɔ/ is rounded. We cannot produce a typical lateral /l/ here unless we first spread the lips. Speech is nothing if not economical, so this does not happen. Rather, what we do is modify the /l/ sound by simply lifting the tip of the tongue upwards. The body of the tongue is below the velum, and this gives this particular version of /l/ its usual name, velar /l/, also written /ɫ/.

This is all very well about vowel to consonant transitions, I hear you saying, a back vowel to a velar /l/, and so on, but how on earth do we see this on the

spectrogram? Well, if you look at the right hand side of the spectrogram and the two very solid bands at the bottom, moving across you will see that they are not perfectly horizontal. In fact, they seem to wave a bit and then dip down – that is where the transition to /l/ occurs. Note also, that as we go into /l/ the bands are lighter than during the vowel sound. This is because vowel formants have more sonority than the consonant formants.
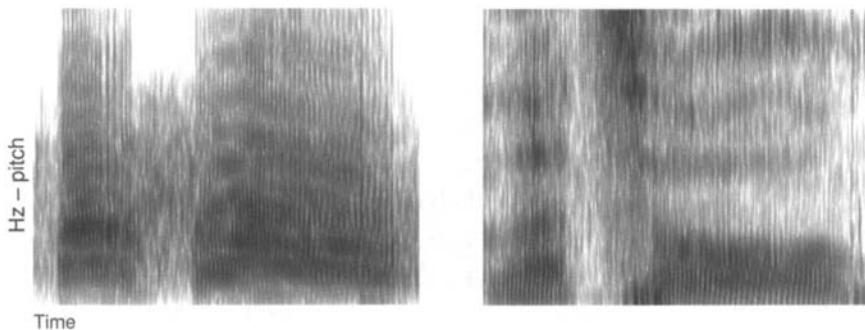
Information Panel 12.6: Formants

A formant is simply a concentration of intensity and amplitude at a particular frequency. The frequency of the particular formant depends on the shape of the oral tract at the time the sound is being made. In speech analysis we generally find four formants sufficient for study, though there are more. The most important formants for the study of vowels are the first and second formants.

How we write formants: F1, F2, F3 and F4 are formants 1 to 4, the most usual formants for study purposes. F0 is not a formant – it is the fundamental frequency (the 'basic pitch' of the voice minus all the speech apparatus: the rate at which the vocal folds vibrate).

We could say a lot more about the spectrogram of 'That's all'. However, I now want to turn to the main difference between academic phonetics and forensic phonetics, the difference between what happens in a phonetics laboratory at a university and speech in real life. I can perhaps best illustrate this with a spectrogram of someone else saying 'That's all', this time in an actual criminal context. This is a real live forensic utterance, being uttered on a covertly surveyed telephone. I will place the two spectrograms – the one you have been looking at and the real world one side-by-side.
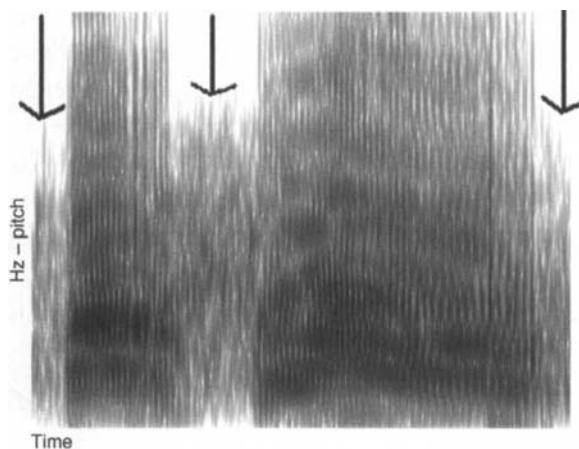
Sound illustration 12.12: Two spectrograms of 'that's all'

Now, you have been reading about spectrograms for the past two hours, and you have been learning all about how the quality of the recorded sound affects the appearance of the spectrogram. So, two questions: (i) which of the above represents a better quality recording and (ii) which of these spectrograms is the one you have been looking at for the past several hours?
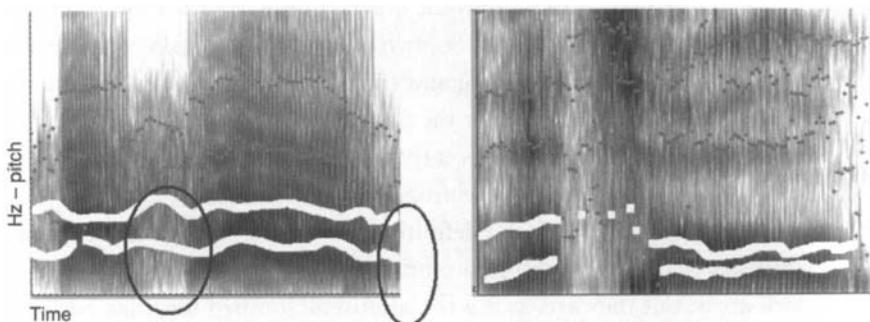
Well, maybe as this is all new to you these are not very fair questions. I will rephrase them. Which of the two spectrograms you see here contains the most contrast? Which of them has clearer definition between lighter and darker areas? I hope you agree it is the spectrogram on the right. The one on the left does have dark areas, but they are, as it were, almost all lumped together. Second, notice that the one on the left is missing detail in the upper frequencies, as per the arrows I have marked in below:

Sound illustration 12.13: Showing missing upper frequencies in a spectrogram

What do these missing frequencies tell us? I suggest they tell us that what we are looking at is not a very high quality sound. The possibility also exists that at least some of the rest of what we are looking at in the upper frequencies might just be noise. All of this tells us that we are looking at a spectrogram which represents a sound recorded on a device with a somewhat narrow frequency band and/or with a poor microphone. In fact the spectrogram is typical of a mobile phone representation: large areas of white space, compressed banding of formants and poor greyscale contrast. When we superimpose the formants on the illustrations we notice something further:

Sound illustration 12.14: Further differences between the two spectrograms

In the left-hand spectrogram (from the mobile phone recording) I've placed two large ellipses. These reflect parts of the spectrogram where in the first case there is no phonation, and in the second there is less phonation than the preceding vowel. In the spectrogram on the right you can see that the voiceless part has no phonation, and the less energetic area which occurs as the /l/ is produced also shows no phonation. The white bands are the formants I have drawn in. The one on the left shows formants all the way, the one on the right shows that there are no formants in the voiceless and partially devoiced areas. The mobile phone signal is relatively poor. There are all kinds of GSM effects (interference on the line). There is a residual formant trace in voiceless areas anyway. What seems to happen with the mobile phone signal is that the signal picks up noise at the formant trace levels. Hence, we end up with formants showing where there is no voicing going on. This is typical of phone signals and mobile phone signals in particular.

Speech signal enhancement

Audio enhancement involves the removal of unwanted noise from a recording. When police and security agencies install surveillance equipment the operator has to work with existing placements, and 'studio' quality sound will not be possible. The quality of surveillance audio is rarely as clear as the movies would have us believe. There are many potential sources of noise because microphones have to be placed discreetly, and because there is often no prior method of surveying a target location to minimize the possibility of extraneous noise interfering with the desired signal. Here are some examples of the kinds of noise one is likely to encounter in surveillance material:

Other speakers: If the target location is a public area speech from other speakers can interfere with the desired signal. Undesirable speech is difficult to remove because it occurs at similar frequencies to the target signal. The worst types of target area are airport and railway station waiting rooms, railway platforms, public parks, bars, restaurants and night clubs.

Music: Music can be a severe hindrance to recovery of signal, especially music with a strong beat or percussion.

Generators: Hotel rooms often have generators located near them, and this may not be obvious to the person installing the surveillence equipment.

Trains, planes, cars and boats: The inside of any moving vehicle presents its own problems, with a variety of noises, often of the impulse type.

Prevention and techniques

Selection of the right kind of microphone is essential. Omnidirectional microphones are often used because the operative can never be certain exactly where a noise or speech signal will originate. However, if it is possible to install several microphones, then a cardioid microphone might be better. In any event, installing two or more microphones can assist with recovery. It is important to be able to expect the unexpected: noises from children, people unwrapping sweets or candy or opening crisp packets, cats scraping at doors or on litter trays, washing machines and dishwashers, refrigerator motors starting and stopping, traffic noise from outside, radios and televisions, speakers with weak voices, etc.

To remove unwanted noises you need to begin by isolating the noise type: is it regular or irregular? Does it have a pulse? What is its frequency or frequency range? Next you need to understand which types of speech sound it is most likely to interfere with. If in the 2,000 to 6,000 Hz range it will probably impact on fricatives, especially /s/ and /z/ sounds. Background dishwashers and washing machines make it quite hard to distinguish some plosives and nasals, for example. A child's crying could affect the signal throughout the speech range, as could a dog's barking, or a cat's meowing. It is important not to try and remove all the unwanted noise. For the newly graduated phonetician used to quality recordings, discretely comprehensible phonemes, etc., the effect of real-world noise on speech can come as something of a shock. University courses rarely prepare you for forensic work as it actually happens. If you try to restore an audio to studio quality you will usually be disappointed. The best you can hope for – and *all* that you should aim at – is comprehensibility of the speech signal. There is another reason, in addition to frustration avoidance, for this

policy: it is very easy to introduce unwanted artefacts into the recording. These can actually cause severe distortions of the subtle kind: that is to say, you get something that sounds recoverable and you transcribe it. It makes complete sense, but in fact bears little resemblance to what is on the tape. Why? Because you have unwittingly introduced an artefact which masks the actual speech in the recording. It might be something as simple as causing a voiceless sound to appear voiced, or it might make a dental phoneme sound like an alveolar or palatal sound. This is why audio restoration requires a very light touch and considerable experience. It is also why, in my view, it requires linguistic training in addition to phonetic and acoustic training: the audio restorer who has a background in linguistics will be sceptical of what s/he hears, will want to test it against contextual cues. If this individual is phonetically trained as well, so much the better. The phonetician understands about issues such as linked vowels, the presence of a bilabial nasal when an alveolar nasal was expected, etc.

In this section I have deliberately avoided being too specific. I do not anticipate that this book will get into criminal hands, but nevertheless it pays to be careful. There are many techniques for recovery of almost any type of speech signal under a wide range of signal-to-noise circumstances. Naturally, discretion prevents me from discussing them here in any further detail.

Information Panel 12.7: Basic points about voice identification processes

- Although the spectrogram and other visualizations of sound are important in the voice identification process, the task of listening is just as crucial. Note that it is not sufficient simply to possess professional software. Unless the user is familiar with phonetics issues and has been trained in linguistics, it is likely that important details and questions will be overlooked.
- There is no such thing as a completely reliable method of voice identification. Voice identification is an exacting, though not yet exact science. It requires considerable technical knowledge and skill, not only of linguistics in general and phonetics in particular, but also some knowledge of statistics and mathematics.

Voice comparison

The crux of speaker identification is voice analysis, and the heart of voice analysis is learning to recognize voice types under a variety of recording conditions.

Some phoneticians favour an automatic procedure for doing these tasks, largely by the use of spectrograms, cepstral analysis, linear predictive coding and the like. Analytical methods include algorithms based on Bayesian statistics and a host of others. However, the International Association for Forensic Phonetics and Acoustics favours a combined approach: use not just software and instrumentation, but your ear as well. If you cannot hear the difference between two voices, and describe that difference technically, how would you describe those voices to a court?

Types of voice

Broadly speaking, there are only a handful of voice types, but a single voice can be a combination of types, and this can manifest itself in many different ways (cf. Laver's 1991 compound phonation types). However, it is probably more accurate to speak of types of phonation rather than types of voice. Broadly, phonation is just making a speech sound, but more particularly phonation is the way in which an individual's vocal folds work, i.e. phonate during speech, specifically during the voicing of speech sounds. The way in which a voice phonates gives it its characteristic timbre.

How the voice works

The voice works because the vocal folds vibrate. In speech we produce a mix of buzzing, popping and hissing. The buzzing is the voicing of speech sounds, the popping is what happens when we produce plosives and the hissing comes from sibilants. We get voiceless plosives and sibilants too. Voice quality varies depending on how open or closed the glottis is during phonation. Voicelessness is just that – air is passing freely through the glottis from the lungs: the glottis is in its maximally open position. A breathy voice is one in which the glottis remains open and in which the vocal folds flap loosely. A modal voice is one in which the glottis is slightly open, while a creaky voice means the glottis is almost completely closed.

To imagine a breathy voice think of Marilyn Monroe, or Margaret Thatcher at her most hypnotically persuasive. With a breathy voice you will get a hushing sound not unlike wind rushing through an open window throughout the speech sample. On the other hand, to imagine a creaky voice think of the typical expert sitting back in a chair, pontificating to the audience in a low voice, somewhat like Dr Henry Kissinger's voice. A creaky voice is a creaky type sound at a much lower pitch than normal. Imagine a door that could speak and creak

simultaneously. These descriptions are very subjective, but they may help you visualize different voice types.

Think of a musical instrument: a cello sounds deeper than a violin because of size, proportions, structure and materials used. You can tell them apart even if they are playing the same note. It is the same with the voice: two speakers might be speaking at the same pitch, but they will sound different because of different shape, proportions, size, etc. – one speaker will have longer vocal folds than another, a bigger oral cavity, etc. Voice quality is also changed by atmospheric conditions – a violin in damp conditions will sound different in dry conditions, and the human voice will vary according to the conditions in which speech is being made, according to the health of the speaker, tiredness, and so on.

Phoneticians have viewed the question of voice quality differently over the years. Ladefoged (1971) developed a simplified continuum of glottal constrictions, while Laver (1980) saw possibilities in combining voice types, for example a whispered modal voice. This seems intuitively right, because people do combine different voice styles. In one case a female suspect had been recorded talking on the telephone to her husband. Her voice was mainly modal but when she came to stressing syllables, these were often in a creaky voice. On the other hand she would often begin an utterance in a breathy voice. On the police interview tape, however, matters were somewhat different. For a start, she spoke in an entirely different register: the informal, casual, word-initial h-dropping, and word-final '-ing' pronounced as '-in' was gone. The voice sounded huskier, more persuasive. The responses were delayed, perhaps carefully thought through. The emotional range was different, too. When talking to her husband she was alternately happy, sad, annoyed, enthusiastic, dispirited, bored, etc. In the interview she was much more controlled, formal, mostly polite, although still displaying some emotion at times. The police interview tape, too, was recorded under completely different acoustic conditions: it was a face-to-face interview in a room with fairly flat acoustics. When we speak to people we know well and are close to, we tend to be much more informal – we adapt to their register and they adapt to ours. Moreover, talking on a telephone usually involves raising the voice. Mobile (cell) phones are particularly tricky because the distance between the speaker's mouth and the microphone is not constant. The person is often in a public place, may be walking around, moving their head, etc. In an interview the speaker is likely to be tense, controlled, sitting fairly stationary, keeping a more constant distance from the microphone.

Human hearing

In the case I am discussing, the spectrogram of the interview tape showed virtually the full range of human voice pitch. On the other hand, the phone conversations were recorded at the landline end of a conversation with the female voice speaking through a mobile phone. The frequencies thus cut off at about 3,200 Hz, as opposed to having up to 5,000 Hz available in the interview recording. As we saw from an earlier spectrogram, there were sizeable areas of white space in the spectrogram, which told us that the captured signal had lost some of its upper frequencies. In circumstances like this it can be very difficult to make a valid comparison: essentially, you are not comparing like with like. One of the first things to do is to try to emulate the acoustics of the phone voice: this is done by filtering sections of the interview tape through what is known as a bandpass filter. To envisage all of this, let us begin by reviewing what humans can hear as opposed to what microphones usually filter to the ear:
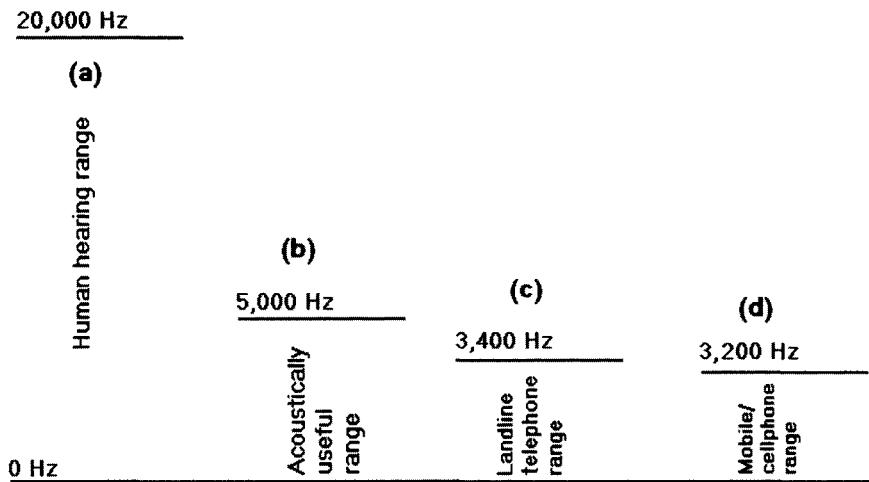
Figure 12.4 Human hearing range vs filtering ranges of recording devices

What I have tried to show in this diagram is – from left to right – (a) what most of us can hear, (b) what we mostly need to hear, (c) what landline telephones tell us we are hearing and (d) what mobile phones tell us we are hearing. This is not strictly accurate – I am using lay terms to describe these phenomena, designed to give you a feel for what is happening. The essential point is that there is a continuous filtering process – even for humans. Just because we cannot hear beyond 20,000 Hz does not mean nothing is happening above that

pitch level. What our brain does is to filter out what we don't need or can't cope with. Dolphins, for example, can hear a much wider range of frequencies than we can: approximately 75 Hz to 150,000 Hz! However, such sensitivity has its downside. Dolphins, and whales too, cannot tolerate prolonged exposure to some types of human noise, and in fact often have to surface too rapidly in order to escape certain human-made marine sounds. This can result in them suffering something like divers' bends and dying. So, perhaps nature has done us a favour by giving us less sensitive hearing.

Electronic recording devices and human pitch ranges

In fact, as it turns out, we don't need anything like 0 Hz–20,000 Hz to hear conversations. Most of the information we need is in the 20 Hz–5,000 Hz range, as the illustration shows. Police interview recording equipment takes this into account, and so, in an interview recording we are likely to be able to capture most of what we need to carry out an analysis. However, in a phone conversation, a great deal of the frequencies at the top end are cut out, and this is even more acute with a mobile phone and even some portable landline telephones. What we have to do when comparing interview acoustics with mobile phone acoustics, therefore, is to re-filter the better electronic signal, and reduce it to the level of the inferior mobile phone signal. Instantly, the comparison becomes more viable. We are now comparing two sets of sounds which have more in common with each other than before. There are a number of reasons for this. By reducing the overall frequency range, we are squeezing the available spectrum of sound. This tends to have an effect on the voice pitch (which in any case may be altered because of the tendency of humans to raise their voices when speaking on phones, especially mobile phones).

Effects on formant values in (mobile) phones

Byrne and Foulkes (2004) found that mobile phone signals affect formant values. Recall that earlier we referred to F1 and F2, formants 1 and 2. These are the two formants which are mostly concerned with vowels. You will remember from our illustrations that F1 was very low for closed vowels, rising significantly for open vowels. What Byrne and Foulkes found was that in the mobile phone environment F1 can be raised by as much as 30 per cent. F2 is not significantly affected. Byrne and Foulkes built on work done by Künzel (2001) who found that F1 is raised by an average of 14 per cent in telephone transmissions, for both male and female speakers for some vowels. Hence the effect on mobile phones

is even more dramatic than ordinary landline telephones. You will also recall we spoke about F0, the fundamental frequency of the voice – its basic pitch. This is also affected by mobile phone transmission. In the case I am referring to the female's voice in the interview context had a low, urbane quality to it. On the phone she sounded much higher-pitched, which was assisted by the fact that in conversing with her husband she was a lot livelier and more ebullient than when talking to the police officer. Byrne and Foulkes found that in mobile phone conversations the voice's fundamental frequency can be as much as 30 Hz higher than in face-to-face recordings. There are other complications, such as the quality of the signal, variation among different phone types and networks, electronic artefacts introduced by the phone signal, etc. All of these factors can make a comparison between a face-to-face recording and a telephone recording (especially a mobile phone recording) somewhat technically difficult. However, now that phoneticians and acoustics engineers are aware of these problems, we can make allowances for the differences and do our calculations accordingly. Nevertheless, great caution needs to be exercised: it is important that we learn to hear differences and similarities and not just rely on measurement devices, though it is important to have some idea of the physics of the acoustic signal in order to be able to make those measurements accurately.

There is much more to say about forensic phonetics than I have so far been able to do in this chapter, so I am going to end by talking about a few issues we have not touched on so far: first, the kind of sound samples forensic phoneticians have to deal with and second, some ethical considerations in voice comparison and voice identification.

Sample length and phoneme stock

It is often the case that suspect samples are very short and this means that of necessity the amount of different speech sounds we are going to get to test a known voice against a questioned one will be relatively low. Moreover, in many cases, we may have only one example of a phoneme in a given recording. The known sample is likely to be much richer in its range of phonemes, its phoneme stock. For argument's sake let us say there are 40 phonemes in English. The known sample is likely to have dozens of examples of most of those. The questioned sample may only have 20 or so of the 40, and perhaps only one example of some of the less frequent phonemes, such as affricates, for example. This is a definite problem and anyone writing a report for a court on this issue

would usually point this out as being a limiting factor in being able to make a comparison, although sometimes the phoneme stock one has available may be enough to reach an opinion. In some countries phoneticians will not report on a questioned sample of less than 30 seconds. This is understandable, but I feel that a better yardstick of quality is phoneme stock, which is fairly easy to compute.

However, we have to bear in mind something else that we referred to earlier. When the female in the interview was talking, she was somewhat formal, speaking in a relatively soft voice. This was different from the female on the telephone: this female was speaking fairly loudly, as we expect when people talk on telephones (mobile or otherwise). Effectively the two recordings contain two different registers of the language: a more formal one and a less formal, much more colloquial one. This is an important consideration to bear in mind because articulation of consonants, for example, is affected by register, and when we add this to the differences between telephone and face-to-face signal quality, we certainly have our hands full as to whether the two voices we are listening to actually belong to one person or two.

Ethical considerations in forensic phonetic analysis

Voice analysis has come a long way since the days when Charles Lindbergh famously declared that he recognised Bruno Hauptmann's voice in a New Jersey court. However, there are still some 'practitioners' who insist that the spectrogram is nothing but a 'voiceprint' which can infallibly determine speaker identity. This is plainly not so. Over a period of many years a number of those working in forensic phonetics have developed ways of doing voice analysis which take into consideration the crucial question of ethics, asking such questions as: when can we undertake an analysis, who should undertake such analyses, what limits are there to an individual phonetician's abilities, when should we not undertake an analysis, what kind of levels of probability can we attach to our findings, etc. In Appendix 3 to this book you will find a document entitled IAFPA Code of Practice. This is the set of guidelines developed by the International Association for Forensic Phonetics and Acoustics. It deals with all of these questions in a simple, straightforward manner and is currently the definitive word on the subject. It is worth studying in some detail.

Chapter conclusion

In this chapter I have done no more than introduce the topic of forensic phonetics. Reading and understanding this chapter will hopefully give some readers an appetite for learning more, but what we have here is no more than an outline, a general guide to some aspects of the field.

13 Notes on forensic transcription

The text is the evidence

The task of transcribing forensic texts is frequently underestimated by professionals in the legal profession, law enforcement authorities and even among linguists themselves. It is often assumed that linguists, by virtue of their extensive training in language, are automatically effective at the task of transcription. The complacence that exists about transcription often results in poorly transcribed text, text which is summarized rather than transcribed, and text which is not faithful to the original. There can be no more important task within forensic linguistics than the rendering of a text to a high standard of accuracy. All further analysis of forensic text depends on the care and precision with which that text was initially transcribed. This is beginning to be recognized by some professionals and specialist services are starting to emerge. It is suggested that in the future the forensic scribe will be a specialist in high demand by law enforcement and the legal profession, as well as by some forensic linguists. All of these professions will require the services of highly trained and experienced scribes. Employment opportunities are limited at the moment, but as the importance of this service is realized it is possible that new opportunities will occur. Ultimately, it is believed, the contribution of the forensic scribe to forensic linguistics will be substantial. Just ten years ago, few could

have predicted the degree to which forensic linguistics itself would be in demand within the justice system today, nor the interest that it would generate as an area of academic study. As an integral part of forensic linguistics, forensic transcription is set to fulfil a significant role within the justice system.

The purpose of this chapter is to increase your awareness of the need for accurate and ethical transcription of all types of text from written, printed and audio/video sources. Linguists, especially those interested in transcription, should aim:

- To be able to transcribe audio files of varying degrees of difficulty, including broadcasts, telephone calls, conversations, covert or surveillance recordings, and recordings of non-native or distinctive, disabled, inebriated or other speakers of varying dialect and regional groups.
- To be able to render transcriptions of a variety of written documents to a high standard of accuracy, whether produced in handwriting, on a typewriter, printed, word-processed, or originating as image files.
- To be able to recognize particular difficulties attached to specific text and audio types, including the difficulties of transcribing conversations containing overlapping speech, distraught or indistinct speakers or speakers in difficulty, the transcription of barely legible handwriting, the problems associated with various types of electronic text.
- To strive to produce ethical (i.e. fair, impartial) transcriptions and to state in any report or other evidence the limits of their transcriptions.

Some general notes on transcription

Accurate, reliable text transcription is important because in any inquiry, whether for academic research purposes, teaching purposes, criminal investigation or judicial purposes, the text *is* the data, it *is* the available evidence. If we get a transcription wrong, we are unwittingly altering the evidence. If we fail to transcribe the entire text or required excerpt, regardless of whether we think what we are transcribing is relevant or not, we are, again, unwittingly altering the evidence. So, it is important to emphasize from the beginning that the text is the evidence. If you are transcribing, for example, an audio file from the internet or elsewhere, and you are also given a transcription of that audio file *never assume that the transcription you have been given is accurate.* In all probability it is not. This applies to files which originate with major news

organizations, government departments, commercial corporations or private individuals.

Care and handling of texts: It is important that paper and audio texts (in whatever format) are stored in a safe, secure, clean, dry environment and that backup copies (where authorized) are made. Electronic texts should be separately copied to a 'read-only' folder and each file in that folder should be marked 'read-only', thus preventing the file from being changed. Moreover, further backup copies of the electronic data should be made and stored in a non-magnetic, fireproof, theft-proof environment. They should not be stored on websites unless the security of those sites has been assured. Moreover, texts of a confidential nature should not be sent as email attachments except to a secure server, and if faxing such texts, ensure that non-security-cleared personnel will not have access to the receiving fax equipment at the time of transmission. Always obtain a fax receipt. Confidential texts should not be stored on the root drive of a computer, the machine's Desktop or the 'My Documents' folder. On computers avoid generic filenames like 'Directory' or 'Texts', as files with these names are routinely attacked and corrupted by viruses of all types. Computers where confidential texts are stored need to be password protected. The room in which copies of texts are kept should not be accessible to non-security-cleared personnel unless such texts are locked away securely. When using texts for training purposes they should be anonymized and not used until the relevant case has been disposed of by the courts.

Texts sent by mail should be sent by recorded or registered mail, should be insured and a receipt should be requested. If texts feature in criminal or civil cases particular care must be taken of them. In the case of old texts, or texts produced on paper which is not in good condition, advice must be sought from paper and text conservationists regarding care and maintenance of texts. It is also important to emphasize that the work of the forensic scribe requires absolute confidentiality, secrecy and trust. Never discuss work with those outside your immediate laboratory team, and even when discussing work with colleagues be careful to ensure that discretion is maintained at all times.

What kinds of transcription are there?

There are two main types of transcription. The transcription of a written document and the transcription of a video or audio record. Each type of transcription presents its own special problems. For example, a handwritten document may contain illegible handwriting, deletions which render it hard to comprehend,

illustrations and drawings which could prove difficult to describe, unusual spellings which result in ambiguous meanings. Scanning typed or printed documents is not always easy either: you will have to ensure that the scanned version does not alter anything from the original. When asked, as a forensic scribe to scan a document, you are always doing more than that: you are really undertaking a transcription task, to ensure that the final document is a correct rendition of the original. It is often no easier scanning a short document than it is typing it in yourself: in other words, scanning is not necessarily an easy option.

Audio or video recordings also present problems: very often speakers mumble or are indistinct, repeat themselves, hesitate, speak nonsense, etc. They may also produce non-linguistic sounds, such as laughter, crying or other noises which cannot easily be transcribed. Audio recordings vary enormously with regard to quality: either quality of equipment or quality of the recording process (distance of speakers from microphone, ambient room noises, echoes, etc.).

Sometimes the linguist is called on to transcribe using phonetic characters from the IPA (the International Phonetics Association). This is a specialist task you will have learned about on a linguistics course. This is a different task from most forensic transcription work, but even so the forensic scribe or linguist with expertise in phonetics may be called on to do this as well, if it is a matter of some importance how a word is pronounced, or where a word or utterance is disputed in a criminal or civil case.

Conversational analysts (CAs) also have a highly specific way of transcribing conversations. This can be very detailed and demanding work, but in general the forensic scribe's task is somewhat easier than that of the CA, although there may be occasions when a forensic scribe will need to call on the services of a CA in order to clarify a transcription task, or to improve a transcription.

Finally, we should consider the task of the Sign Language Speaker or Interpreter who may be called on to write out what a deaf person is saying through BSL (British Sign Language), ASL (American Sign Language) or any of the other sign languages.

Phonetic, conversational and sign transcription work is, for the most part, quite separate from forensic transcription work.

Most forensic transcription work deals with the transcription of spoken recorded language or of texts such as ransom notes, suicide notes and so on. The main requirement is accuracy, and implicit in the task is the necessity to make clear the limitations of the transcription: in other words to note where the task

was difficult or impossible, due to such matters as illegibility or inaudibility, difficult handwriting or unclear diction, and so on. The scribe should give an estimate of those words which were not ambiguous, those where some ambiguity existed, and those where transcription was not possible.

Note that it is not the task of the scribe to give a view on the meaning of a text or recording unless he/she is also the commissioned linguist or phonetician. The scribe's task is to render the text into a written or typed (word-processed) record.

Whatever the difficulties, the forensic transcriptionist (or scribe) must do his/her best to ensure accuracy, but – as previously noted – must state the limitations of accuracy which were possible in the given instance.

Transcribing audio recordings

What does transcribing audio (including video) entail?

Identifying simultaneous or clashing speakers: If there is more than one speaker you will need to ascribe each utterance to its correct speaker. This is not always easy, for example if the speakers are both of the same gender, have similar accents, etc. It is not always the case that two speakers have quite distinct voices.

Identifying the number of speakers on an audiotape: Sometimes you may need to take a head count of how many speakers take part in an exchange on an audio recording (or even in a video recording).

Transcribing the speech of an emotionally distraught speaker: Transcribing someone's speech when they are distressed or upset presents special problems. Consider using Conversation Analysis (CA) symbols for this purpose. See especially Local and Walker, 2005.

Transcribing the speech of other dialects: Even native speakers have problems with accents and dialects with which they are not familiar. You may find it beneficial to commission a transcription from a person with the same geographical profile as the speaker on the tape.

Transcribing the speech of a non-native speaker: The degree of difficulty will depend on your familiarity with the accent of the speaker concerned, the degree to which you have been exposed to English language speech of that community, the degree of proficiency and fluency of the speaker and the degree to which the speaker's English speech corresponds to that of a native speaker.

Transcribing telephone calls: In phone conversations because speakers cannot usually see each other there tends to be a great deal of speech overlap. To

compensate for this, they often produce hesitations and disfluencies. In addition, sometimes the signal is poor, or fluctuates in quality. Telephone sound frequencies are narrower than those found in other electronic equipment and modern telephone systems which use low-bandwidth software present particular problems. Hence, telephone transcription can give scribes some difficulty. This is particularly true of mobile (cell) phones where the frequency range is even narrower than normal telephones and there are additional GSM effects (signal noises produced by the cell phone infrastructure), discontinuous transmission effects, etc.

Transcribing speech in regard to veracity issues: In his period in office President Clinton had occasion to address the press with regard to his relationship with Monica Lewinsky. Regardless of whether you think the president was telling the truth or not (which is not the transcriptionist's concern) it is interesting to note that in certain parts of the tape, the president becomes very conscious of his pronunciation and does pronounce certain words very clearly. This often happens when speakers feel they are being challenged, or when they feel they have to 'spell out' a situation to a listener. Some linguists refer to this phenomenon as 'hypercorrection', the tendency to over-correct the way one speaks.

For all of the above categories you need to be aware of neutralization, assimilation and elision processes which occur in natural speech (see Akamatsu, 1995: 3–5).

Transcription of written language

A text may need to be transcribed for a number of reasons including (i) measurement of the text (word-length average, sentence-length average, punctuation density, etc.); (ii) to enable a linguist to undertake an analysis of its content; (iii) to enable a corpus search of words and/or phrases within the text; (iv) to add the text to an inquiry corpus for comparative and authorship attribution purposes.

Transcribing texts produced on old typewriters: Some transcription tasks involve interpreting the text produced on an old typewriter, sometimes with a poor ink-ribbon. This presents its own problems. An example of this is the will of Mrs Grundy, which you can find on the internet.

Transcribing electronically produced text: It is all too easy to imagine that all electronic text will be easy to transcribe. However, this is far from being the case. Also, it is sometimes assumed that there is no need to transcribe electronic

text. Nevertheless, it is a useful process because transcribing an electronically produced text gives the analyst insight into the text, its conventions, possible interpretation as to its meaning, and highlights particular problems which the writer may have encountered and attempted to solve.

Mobile phone text messages: There are several difficulties with mobile phone messages. First, the transcription task depends on the source of the message. Messages are sourced from (i) service providers, (ii) forensic communication or telecommunication services (such as FTS in the UK), or (iii) from transcriptions provided by non-linguists. The dangers with the latter are obvious: how can you be sure the message was correctly transcribed? In a recent case, a linguist had to examine messages sent by a client which failed to note the casing (upper vs lower) in the text, failed to punctuate as per the text, and had standardized the spelling. This meant that the linguist could not determine the different texting styles of the various candidates in an inquiry. Even when service providers provide text messages they may break the message across several lines. This means the spacing may be lost: in one inquiry an important feature of the texts was the user's tendency to put two or more spaces between words. This appeared to be the sign of an inexperienced user. However, because of the line breaks in the transcription provided it was not always possible to tell exactly where these double spacings occurred. Be sure to get the best physical copy of the messages before attempting transcription. Since messages often arrive by facsimile transmission legibility can be a problem.

Texts for scanning: As previously noted a hard copy of a text which requires scanning does not necessarily mean that you have an easy option. The scan has to be proofread very carefully. You will need to ensure that all aspects of the scan are as per the original, including casing, spacing, punctuation, orthography and spelling. This is very detailed and painstaking work.

Word-processed documents and .txt files: You may not always be able to transcribe a text into either of these types of software because of font problems or character recognition. Not all word processors contain every single possible ASCII or Unicode character, and if using a .txt program, such as Notepad, you will always be prompted to save the file as a Unicode file in case special characters have been employed. If in doubt, render a few lines of the text, save, close the program and reopen, just to ensure that your transcription is working.

Appendix 1: Texts

Alphabetical list of referents in the forensic texts

Use the texts in this appendix to learn more about forensic linguistics. Ask yourself such questions as: what gives a forged text away as forged; how would you gauge a sexual assault report; above all look closely at questions of witness veracity. Also, you should practise transcribing some of these texts. This will give you a very close insight into what the texts are about, and how they are structured. Often you will find your opinion of a text changing over a period of time if you analyse and examine it often enough.

Text A.1: Forgery by Dr Harold Shipman of Mrs Kathleen Grundy's will

I GIVE ALL MY ESTATE, MONEY AND HOUSE TO MY DOCTOR. MY FAMILY ARE NOT IN NEED AND I WANT TO REWARD HIM FOR ALL THE CARE HE HAS GIVEN TO ME AND THE PEOPLE OF HYDE. HE IS SENSIBLE ENOUGH TO HANDLE ANY PROBLEMS THIS MAY GIVE HIM.

MY DOCTOR IS Dr H. F. SHIPMAN 21 MARKET ST HYDE CHESHIRE SL14 2AF

9th JUNE 1998
SIGNED K. Grundy
SIGNED P Spencer
SIGNED Claire Hutchinson

(Author thanks the family of Mrs K Grundy for permission to print this text.)

Text A.2: Statement alleging sexual assault

Upon arrival at Joe's Club, (date April X^th, 200- <......day>) I had entered the lounge area to have a drink and dinner. I was approached verbally by a man (mid 30's) as walking to a seat just 2 away from him. The words I remember most from any conversation that may have taken place that evening were "where you been? What took you so long?" I do no recall if/what I responded with in turn.

I then ordered a bourbon and coke while waiting for my order of "crab rolls." I wish I had more recolection of people or words exchanged, as this is where my mind is having trouble remembering. My food had arrived, enjoyed it with a glass of water topped with a lemon, finished my dinner and ordered 1 last drink, with all intentions of leaving at that point. if my memory is treating me well I had been offered another drink – hesitated but accepted offer. At this time I remember clearly going to the rest room. Upon my return, I casually had my 3rd bourbon and Coke, while conversation was minimal and do remember paying most of my attention to the television directly in front of me.

Some time in the evening I do receall speaking to, one which was a bartender, James Smith & younger brother, Ben ?? These 2 are childhood – neiborhood acquaintances – I wish I could recall any of which conversations in the evening but unfortunatelly cannot as of today.

From this point, honestly things are neither here nor there in my head. Can not seem to decifer any actual instances.

I can say that approx. times were last recolection of 8 pm or 9 pm roughly. I woke with my alarm clock staring the time of 10:03 am. I frantically went to check on my son, he was not there. After becoming somewhat aware of my surroundings I then realized the only clothing I had on was my sleevless dress, sweater, which was worn the day of April X^th under my blue dress jacket, which all clothing now in pocession/evidence. No under clothing as well. My sheets were soiled in what appeared, to me, as to be a mixture of slight blood and my own urine. In discust and pure confusion I stripped the fitted sheet from my bed and replaced it with a fresh one ... first phone call was to my mother, with my little knowledge at that time our conversation was brief, just enough to know my son, Hugh, was safe and at daycare ... (this info as of 4-12, mom stated it was 10 or 15 min after 10 am when she received my phone call.)

She was angry and we ended the phone call. Various phone calls including Mary Williams and Sam Walker, 2 close personal friends ... after a few

conversations and tears shed with them, I then decided to call the local police to report this incedent.

(rewind: first call after mom was to close friend Darleen Brown –> she gave me a number to the crisis line. moments later I discussed the situation "BRIEFLY" with a lady there . . . She (not purposely) gave me the intention if I were to report, nothing could be done on the little info I had.) So back to the police call I got disconnected, strangly, as another call was coming in. I had called my neighbor in # 94 a sally Edwards to see if she had heard or seen any thing -> she came home and stated times home and awake 8:30 pm to 11pm -> nothing. Sally stayed w/ me and that's when (all within moments) the 2 police officers arrived. Spoke to Ofcr. Macnamara.

All of stated information IS true & correct to the best of my knowledge.
4/12/02

Wendy M. Mullins

(Author acknowledges the co-operation of a mid-western police department in the publication of this text.)

Text A.3: Pipe bombers text (Luke John Helder, 21-year-old college student)

(Text of note left in mailboxes in the Midwest of America where pipe bombs were found)

Mailboxes are exploding! Why, you ask?

Attention people.

You do things because you can and want (desire) to

If the government controls what you want to do, they control what you can do.

If you are under the impression that death exists, and you fear it, you do anything to avoid it. (This is the same way pain operates. Naturally we strive to avoid negative emotion/pain.)

You allow yourself to fear death!

World authorities allowed, and still allow you to fear death!

In avoiding death you are forced to conform, if you fail to conform, you suffer mentally and physically. (Are world powers utilizing the natural survival instinct in a way that allows them to capitalize on the people?)

To "live" (avoid death) in this society you are forced to conform/slave away.

I'm here to help you realize/understand that you will live no matter what!

It is up to you people to open your hearts and minds. There is no such

thing as death. The people I've dismissed from this reality are not at all dead.

Conforming to the boundaries, and restrictions imposed by the government only reduces the substance in your lives. When 1% of the nation controls 99% of the nations total wealth, is it a wonder why there are control problems?

The United States strives to provide freedom for their people. Do we really have personal freedom? I've lived here for many years, and I see much limitation. Does the definition of freedom include limitation? I've learned about the history of various civilizations in history, and I see more and more limitation. Do you people enjoy this trend of limitation? If not, change it!

As long as you are uninformed about death you will continue to say "how high", when the government tells you to "jump". As long as the government is uninformed about death they will continue tell you to "jump" Is the government uninformed about death, or are they pretending?

You have been missing how things are, for very long. I'm obtaining your attention in the only way I can. More info is on its way. More "attention getters' are on the way. If I could, I would change only one person, unfortunately the resources are not accessible. It seems killing a single famous person would get the same media attention as killing numerous un-famous humans. There is less risk of being detained, associated with dismissing certain people.

Sincerely,

Someone Who Cares

PS. More info. will be delivered to various locations around the country.

(Public domain by virtue of having been publicized by the perpetrator of a crime)

Text A.4: Complaint by an Australian aboriginal person about verballing

Yeh – well what happens is that the police – they make the statement out and they make you sign it. They write the statement out themselves Then all you do is sign it. Oh yeh they ask you questions but they just do the writing or the typing themselves. I mean in actual fact the true statement should be in writing in the hand of the person himself but of course it's not. It's accepted as long as he's got his signature there. And maybe nine times out of ten he couldn't read that bloke's writing anyway. He's probably never said half those things. Or the police says Sign this here now and you can go'. Of course he wants to go so he signs it. Many's the time I've come across statements that have been different from anything that could have been said. Many's the time, yes. Then

the prosecuting sergeant gets up there in court and says this is the statement he's given'. He'll read it out to the judge and that's it and that's the weakest thing and yet it's the vital thing. And it's very rarely thrown out – very rarely. Video or tape-recording should happen. It should be a right of every individual. But they don't want Aboriginal people to have rights.

(Found at: wwwmcc.murdoch.edu.au/ReadingRoom/Ashforth/Terthes6. html on Saturday, 30 March 2002. Author is grateful to Dr Teresa Ashforth for permission to publish this text.)

Text A.5: Susan Smith's admission of having killed her children

When I left my home on Tuesday, October 25, I was very emotionally distraught. I didn't want to live anymore! I felt like things could never get any worse. When I left home, I was going to ride around a little while and then go to my mom's. As I rode and rode and rode, I felt even more anxiety coming upon me about not wanting to live. I felt I couldn't be a good mom anymore, but I didn't want my children to grow up without a mom. I felt I had to end our lives to protect us from any grief or harm. I had never felt so lonely and so sad in my entire life. I was in love with someone. very much, but he didn't love me and never would. I had a very difficult time accepting that. But I had hurt him very much, and I could see why he could never love me. When I was @ John D. Long Lake, I had never felt so scared and unsure as I did then. I wanted to end my life so bad and was in my car ready to go down that ramp into the water, and I did go part way, but I stopped. I went again and stopped. I then got out of the car and stood by the car a nervous wreck. Why was I feeling this way? Why was everything so bad in my life? I had no answers to these questions. I dropped to the lowest when I allowed my children to go down that ramp into the water without me. I took off running and screaming "Oh God! Oh God, no! "What have I done? Why did you let this happen? I wanted to turn around so bad and go back, but I knew it was too late. I was an absolute mental case! I couldn't believe what I had done. I love my children with all my (a picture of a heart). That will never change. I have prayed to them for forgiveness and hope that they will find it in their (a picture of a heart) to forgive me. I never meant to hurt them!! I am sorry for what has happened and I know that I need some help. I don't think I will ever be able to forgive myself for what I have done. My children, Michael and Alex, are with our Heavenly Father now, and I know that they will never be hurt again. As a mom, that means more than words could ever say. I knew from day one, the truth would prevail, but I was so scared I didn't know what to do. It was very tough emotionally to sit and watch my family hurt like they did. It

was time to bring a piece of mind to everyone, including myself. My children deserve to have the best, and now they will. I broke down on Thursday, Nov. 3, and told Sheriff Howard Wells the truth. It wasn't easy, but after the truth was out, I felt like the world was lifted off my shoulders. I know now that it is going to be a tough and long road ahead of me. At this very moment, I don't feel I will be able to handle what's coming, but I have prayed to God that he gave me the strength to survive each day and to face those times and situations in my life that will be extremely painful. I have put my total faith in God, and he will take care of me.

[Signed]Susan V. Smith [Dated]11/3/94 5:05 p.m.

(Public domain by virtue of having been used in court proceedings)

Text A.6: Before Susan Smith confessed, she had claimed a black man had carjacked her and driven away with her children

FIRST VIDEO:
all he ever told me was shutup or i'll kill you .... (break) .... and i just screamed i said what are you doing he said shutup and drive and had a gun and he was (...) poking it in my side you know and told me to drive (.) and I (.) so I drove

(Public domain by virtue of having been used in court proceedings)

Text A.7: Appeal by Susan Smith at a press conference:

SECOND VIDEO:
SUSAN: I have been to the Lord in prayers every day with my family and by myself with my husband it just seems so unfair that somebody could take such two beautiful children
HUSBAND: please do not give up on these two little boys and the search for their return safe home to us

(Public domain by virtue of having been used in court proceedings)

Text A.8: One of John Hinckley's many letters to Jodie Foster

12:45 P.M.
Dear Jodie,
    There is a definite possibility that I will be killed in my attempt to get Reagan. It is for this very reason that I am writing you this letter now.

As you well know by now I love you very much. Over the past seven months I've left you dozens of poems, letters and love messages in the faint hope that you could develop an interest in me. Although we talked on the phone a couple of times I never had the nerve to simply approach you and introduce myself. Besides my shyness, I honestly did not wish to bother you with my constant presence. I know the many messages left at your door and in your mailbox were a nuisance, but I felt that it was the most painless way for me to express my love for you.

I feel very good about the fact that you at least know my name and know how I feel about you. And by hanging around your dormitory, I've come to realize that I'm the topic of more than a little conversation, however full of ridicule it may be. At least you know that I'll always love you.

Jodie, I would abandon this idea of getting Reagan in a second if I could only win your heart and live out the rest of my life with you, whether it be in total obscurity or whatever.

I will admit to you that the reason I'm going ahead with this attempt now is because I just cannot wait any longer to impress you. I've got to do something now to make you understand, in no uncertain terms, that I am doing all of this for your sake! By sacrificing my freedom and possibly my life, I hope to change your mind about me. This letter is being written only an hour before I leave for the Hilton Hotel. Jodie, I'm asking you to please look into your heart and at least give me the chance, with this historical deed, to gain your respect and love.

I love you forever,
John Hinckley

(Public domain by virtue of having been used in court proceedings)

Text A.9: Statement allegedly made by Derek Bentley

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 8 pm and 9 pm 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 7 pm. A little later Norman Parsley 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 after them. I walked up the road

with them to the paper shop where I saw Craig standing. We all talked together and then Norman Parsley and Frank Fazey left. Chris Craig and I then caught a bus 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 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. A plain clothes man climbed up the drainpipe and on 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 then 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 policemen 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 pushed me down the stairs and I did not see any more. I knew we were 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 one 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.

(Public domain by virtue of having been used in court proceedings)

Text A.10: Carlos the Jackal ransom note

http://www.crimelibrary.com/terrorists/carlos/text/12.htm 23/12/01
   (Reportedly dictated to British secretary Griselda Carey by Carlos the Jackal, born Ilich Ramirez Sanchez)

To the Austrian Authorities

We are holding hostage the delegations to the OPEC conference.

We demand the lecture of our communiqué on the Austrian radio and television network every two hours, starting two hours from now.

A large bus with windows covered by curtains must be prepared to carry us to the airport Of Vienna tomorrow at 7.00, where a full- tanked DC9 with a crew of three must be ready to take us and our hostages to our destination.

Any delay, provocation or unauthorized approach under any guise will endanger the life of our hostages.

The Arm of the Arab Revolution

Vienna 21/XII/75

(Public domain by virtue of having been publicized across media worldwide.)

Text A.11: Anonymous letter: smear mail

To Whom it May Concern: (and I hope that's ALL of you!)

I am a true lover of Green Yellow Falls and the Green Yellow Falls area and that is why I am writing to you. I have been very, very disturbed about something for a long time but waited to see if perhaps you would become aware of the problem and take care of it without my needing to write to you. This has not happened so I feel it my civil duty to write this letter. Please understand that I would not normally do this but I am extremely fond of Green Yellow Falls and hate to see things happen that reflect so badly on the area.

I am referring to the Green Yellow Falls Inn which is promoted by your Chamber. What IS the criteria you follow for recommending lodging in the Green Yellow Falls area? Do you screen these places in any way? Are you aware of the things that go on at this Inn? The woman who runs it – Louise – has been extremely rude to customers. The conversations that take place are very inappropriate for a public place or, as a matter of fact, for just about any place! There is talk of racism, violence, radical political topics, Nazis, etc. there is talk of many weapons owned by the proprietors. There is negative talk about the Green Yellow Falls area. The children are often treated badly by their parents (The Browns). These are not isolated incidences either but have occurred many times for many years and talked about by several patrons (now former patrons) of the Inn. I just hope the people who go there do not become former patrons of the Green Yellow Falls area!

I thought that you might have become awareof these things but apparently not since this Inn is not only still on your web site but even promoted this year in

the Fall Art Tour brochure!!!! This does not reflect in any way positively on the Green Yellow Falls area. Why is this allowed to go on? Does anyone care? Haven't you gotten other complaints about this? I hope others have written so that this will not seem like one person complaining about some single unimportant incident. Believe me, that is not what this is. This is VERY important and should be taken seriously.

I hope that my letter will fall into the hands of someone who will take this seriously and realize what a mistake it is to promote this Inn. Green Yellow Falls is much too nice a place to be dragged down like this.

(Author's collection)

Text A.12: root@primenet.com (found on internet – this was written by a 19-year-old student who was later arrested for having incited terrorism)

Instead of huntng Lions in California, let us declare open season
on State SEN TIM LESLIE, his family, everyone he holds near and
dear, the Cattlemen's association and anyone else who feels that LIONS
in California should be killed.

I think it would be great to see ths slimeball, asshole, conservative
moron hunted down and skinned and mounted for our viewing pleasure.

I would rather see every right-wing nut like scumface Leslie destroyed
in the name of politicl sport, then lose one mountain lion whose only
fault is havng to live in a state with a fuck-ed up jerk like this
shit-faced republican and his supporters.

Pray for his death. Pray for all their deaths.

(Public domain by virtue of having been the subject of criminal investigation)

Text A.13: Nikola's confession (A Serbian driver confesses to The Hague – translated from Serbian)

Because I had been a driver in the army, when I arrived I immediately got a freezer truck, empty, with the order to drive it to Kosovo, to a concentration camp east of Pristina. As soon as I arrived, a general began to interrogate me. For an hour, he asked me about my past, my political orientations, if I had ever traveled abroad and so on – was I a patriot, did I want to defend my country and things like that. Since I had already served my army term, I knew the answers to all the questions.

As we were talking, the freezer truck was loaded and sealed somewhere – I don't know where, they never allowed me to come close – and then it was brought back. My job was only to drive, without asking any questions.

I made ten of those trips. It was strange for a big truck to be traveling back and forth while fighting was going on in Kosovo – at a time when there were no people, no vehicles, no food reserves, I was driving an empty freezer truck back. It did not take me long to realize that something was not right.

I would drive the empty truck from Bor to the military camp [in Kosovo], which was filled with army, police and various paramilitary units I did not recognize. A policeman would take over from me and drive the truck away. He would return the truck fully loaded and sealed, and on the travel documents – breaking all the rules – he would write only '*Confidential!*' I always drove from Kazoo to Bor at night; I would give the truck to a policeman at the entrance of the copper works in the Bor Mining and Smelting complex, then I would wait at the entrance for the truck to be brought back.

It didn't take long for me to realize that I had been transporting corpses; you don't need much wisdom for that . . .

It was clear to me where the corpses came from, but I did not understand where they ended up once I delivered the freezer truck in Bor. I assumed that they were burned in the copper melting furnaces. Perhaps not . . . Perhaps they were buried somewhere near the copper mine. There were many mining surface fields around, much accumulation of mining waste, many places one can dig a mass grave.

I don't know what happened to the corpses. The only thing I know is that I could not stand it anymore. I began to have nightmares of driving the truck and someone inside who is not dead chasing me with a gun and trying to kill me.

I could not take it anymore. And I was also afraid that I would be killed once the job was completed – as a witness, sooner or later. It is nothing for them to kill a human being.

When I had made my decision, I asked two of my hometown friends whom I trusted to help me. We agreed that they would wait for me at a secret place close to Bor. Because I had to report to the police at the entrance of the smelting works by a specific time, I drove faster than I usually did so I could save half an hour for us to carry out our plan and still avoid suspicion. Like we had agreed, my friends were waiting for me. As I was changing into civilian clothes, they opened the freezer truck. It was full of corpses, almost touching the roof. My friends photographed the inside of the freezer truck, and then I ran away.

They drove the truck to a hidden place where they counted the corpses. There were 78: mostly civilians, among them one woman and three soldiers from the Yugoslavian Army. They recognized one of them – it was a kid from our town.

(Public domain by virtue of having been used in court proceedings)

Text A.14: Plagiarized text (with next text)

D.S.: It is important to note that there are many other things that can be done in addition to these options. For instance, T.S. may be enrolled in a sexual education class to enhance both options, or if the mother is to be told she may be given help to understand her daughter's condition. For the sake of argument on an ethical basis, however, the two courses of action listed above are what concern the problem of confidentiality. In this case there are many different principles that need to be discussed. Both T.S. and her mother are affected by these options therefore distinguishing them as two different parties with their own interests. This makes beneficence/malificence particularly important because one option may benefit one individual while maligning the other. It is necessary to also evaluate the importance of autonomy and justice in this case. As T.S. is only 13 years old, it is very important to evaluate if she is capable of making sound decisions. Though it may violate her autonomy to disclose the infection to her mother, it needs to be determined whether autonomy at her age trumps the other principles involved. As far as T.S.'s mother is concerned, one must also question if she is being discriminated against based on her social status. What if she was a doctor who was consistently on-call? If we are to avoid violating the principle of justice, then it must be determined that the decision to withhold information from the mother has nothing to do with the fact she is a waitress and not something else. Another thing that makes this case unique is that it affects more than just T.S. and her mother. In evaluating these principles it is important to look not only at what is best for those directly involved, but also those indirectly involved. If doctors do violate confidentiality in minors, then many adolescents that need help may be swayed to not seek it for fear of their parents finding out. Even if T.S. were ultimately benefited from her mother knowing, it must be determined if that is worth the consequences of numerous other girls her age that would not seek treatment for their conditions. When establishing policies based on these matters it is important to have the foresight to analyze the actions people may take. If patients know their doctor

may violate confidentiality then it is very likely that they will seek medical help less often.

(Author's collection)

Text A.15 (plagiarized text – see previous text)

V.L.: In this case there are many different principles that need to be discussed. Both T.S. and her mother are affected by these options distinguishing them as two different parties with their own interests. This makes beneficence/non-malificence particularly important because one option may benefit one individual while harming the other. However, since TS is our patient we should apply the ethical principles to TS only. It is necessary to also evaluate the importance of autonomy and justice in this case. As TS is only 13 years old, it is very important to evaluate if she is mature enough to make sound decisions. Though it may violate her autonomy to disclose her sexual activity to her mother, it needs to be determined whether she is an autonomous individual at the age of 13. As far as TS's mother is concerned, one must also question if she is being discriminated against based on her social status. What if she was a doctor who was consistently on-call? If we are to avoid violating the principle of justice, then it must be determined that the decision to withhold information from the mother has nothing to do with the fact she is a waitress and not something else. Another thing that makes this case unique is that it affects more than just T.S. and her mother. In evaluating these principles it is important to look not only at what is best for those directly involved, but also those indirectly involved. If doctor's do violate confidentiality in minors than many adolescents requiring help may be swayed to not seek it for the fear of their parents finding out. Even if T.S. were ultimately benefitted from her mother knowing it must be determined if that is worth the numerous other girls her age that would not seek treatment for their conditions.

(Author's collection)

Text A.16: Engima Machine, ransom demand

I have been asked by the current owner the above Enigma machine, who purchased it in good faith to say and tell you now today, the unwitting person has no ultimate desire of depraving [sic] your august self or anyone the pleasure to see it again. It is though also not his position to freely give the possession for nothing either as the large sum is not to be lost that has been paid (here the

police have blanked out the sum) but only on your full acceptance which are to be published nationally, with no conditions of escape on your part or any other person or official body involved in this matter.

[blanked out paragraph]

A guarantee as said before in this letter that no pursuit of the unwitting now owner shall be made, this to get had by published notice in television and newspaper. This condition is of utmost desire to him and must be done for any further word can be exchanged to you on this matter. It is also of utmost importance also in this matter the person who will be as the negotiating medium will be afforded the same freedom of entanglement in this matter as they are not involved only on my insistence that they contact you to make this offer and for no other reason.

If no notice is to be seen by the day of Monday as the 18th day in September then nothing else is to be said again.

(Public domain by virtue of having been publicized in the commission of a crime)

Text A.17: Enigma Machine, second note

I have been instructed to inform you that negotiations for the return of the G312 are hereby terminated, and further, that the machine will now be destroyed.

(Public domain: see previous text)

Text A.18: Timothy Evans, first statement made at Notting Hill police station

She was incurring one debt after another and I could not stand it any longer so I strangled her with a piece of rope and took her down to the flat below the same night whilst the old man was in hospital. I waited till the Christies downstairs had gone to bed, then took her to the wash house after midnight. This was on the Tuesday 8th November. On Thursday evening after I came home from work I strangled my baby in our bedroom with my tie and later that night I took her down into the wash house after Christies had gone to bed.

Signed T.J. Evans 9.55 pm 2/12/49

(Public domain by virtue of having been used in court proceedings)

**Text A.19: Timothy Evans, second statement made at Notting Hill
police station**

I was working for the Lancaster Food Products of Lancaster Road, W.11. My
wife was always moaning about me working long hours so I left there and went
to work for the Continental Wine Stores of Edgware Road. I started at 8 a.m.
and finished at 2 p.m. and the job was very nice there. In the meanwhile my
wife got herself into £20 debt so I borrowed £20 off the Guvnor under false
pretences, so he give me the £20 which I took home and gave it to my wife. I
asked her who she owed the money to but she would not tell me, so a week later
I got sacked. I was out of work then for two or three weeks. In the meanwhile
I had been driving for two or three days a week. I was earning 25/- to 30/- a
day. This was for the Lancaster Food Products I used to give her this money
and she was moaning she wasn't getting enough wages, so one of the regular
drivers at the Lancaster Food Products left so the Guvnor asked me if I would
like my regular job back at a wage of £5 15s. 0d. a week. I was doing quite a lot
of overtime for the firm working late, which I used to earn altogether £6 to £7
a week. Out of that my wife used to go to the firm on a Friday and my Guvnor
used to pay her £5 what she used to sign for. Perhaps through the week I would
have to give her more money off different people from which I used to borrow
it. I used to pay them back on a Friday out of my own pocket. I had to rely on
my overtime to pay my debts and then I had a letter from J. Brodericks telling
me I was behind in my payments for my furniture on the hire purchase. I asked
her if she had been paying for the furniture and she said she had, then I showed
her the letter I had received from Brodericks then she admitted she hadn't been
paying it. I went down to see Brodericks myself to pay them my £1 a week
and ten shillings off the arrears so then I left the furniture business to my wife.
I then found she was in debt with the rent. I accused her of squandering the
money so that started a terrific argument in my house. I told her if she didn't
pull herself together I would have to leave her, so she said "You can leave any
time you like," so I told her she would be surprised one day if I walked out on
her. One Sunday, early in November, I had a terrific row with her at home so
I washed and changed and went to the pub dinner time. I stopped there till
two o'clock. I came home, had my lunch, left again to go out, leaving my wife
and baby at home, because I didn't want any more arguments. I went to the
pictures – A.B.C. Lancaster Road, known as Royalty, at 4.30 p.m. I came out
when the film was finished, I think about 7.15 p.m. I went home sat down and
switched the wireless on. I made a cup of tea. My wife was nagging till I went

to bed at 10 p.m. I got up at 6 a.m. next day, made a cup of tea, My wife got up to make a feed for the baby at 6.15 a.m. She gets up and starts an argument straight away. I took no notice of her and went into the bedroom to see my baby before going to work. My wife told me she was going to pack up and go down to her father in Brighton. I asked her what she was going to do with the baby, so she said she was going to take the baby down to Brighton with her so I said it would be a good job and a load of worry off my mind, so I went to work as usual so when I came home at night I just put the kettle on, I sat down, my wife walked in so I said, "I thought you was going to Brighton?" She said, "What for you to have a good time?" I took no notice of her. I went downstairs and fetched the pushchair up. I come upstairs she started an argument again. I told her if she didn't pack it up I'd slap her face. With that she picked up a milk bottle to throw at me. I grabbed the bottle out of her hand, I pushed her, she fell in a chair in the kitchen, so I washed and changed and went out. I went to the pub and had a few drinks. I got home about 10.30 p.m. I walked in she started to row again so I went straight to bed. I got up Tuesday morning and went straight to work. I come home at night about 6.30 p.m. my wife started to argue again, so I hit her across the face with my flat hand. She then hit me back with her hand. In a fit of temper I grabbed a piece of rope from a chair which I had brought home off my van and strangled her with it. I then took her into the bedroom and laid her on the bed with the rope still tied round her neck. Before 10 p.m. that night I carried my wife's body downstairs to the kitchen of Mr Kitchener's flat as I knew he was away in hospital. I then came back upstairs. I then made my baby some food and fed it, then I sat with the baby by the fire for a while in the kitchen. I put the baby to bed later on. I then went back to the kitchen and smoked a cigarette. I then went downstairs when I knew everything was quiet, to Mr Kitchener's kitchen. I wrapped my wife's body up in a blanket and a green table cloth from off my kitchen table. I then tied it up with a piece of cord from out of my kitchen cupboard. I then slipped downstairs and opened the back door, then went up and carried my wife's body down to the wash house and placed it under the sink. I then blocked the front of the sink up with pieces of wood so that the body wouldn't be seen. I locked the wash house door, I come in and shut the back door behind me. I then slipped back upstairs. The Christies who live on the ground floor were in bed. I went into the bedroom to see if my daughter was asleep. When I looked in the cot she was fast asleep so I then shut the bedroom door and laid on the bed all night fully dressed until it was time to get up and go to work. I then got up, lit the gas and put the kettle on. I made my baby a feed and fed it.

I then changed her and put her back into the cot wrapping her up well so that she would not get cold, then went to the kitchen and poured myself out a cup of tea. I then finished my tea and slipped back into the bedroom to see if the baby had dropped off to sleep. It was asleep so I went off to work. I done my days work and got home about 5.30 p.m. that Wednesday evening. I come in, lit the gas, put the kettle on and lit the fire. I fed the baby, had a cup of tea myself, sat in front of the fire with my baby. I made the baby a feed about 9.30 p.m. I fed her then I changed her, then I put her to bed. I come back into the kitchen sat by the fire until about twelve o'clock, then went to bed. I got up at 6 a.m. next day lit the gas put the kettle on, made the baby a feed and fed it. I then changed her and dressed her. I then poured myself out a cup of tea I had already made. I drank half and the baby drank the other half. I then put the baby back in the cot, wrapped her up well and went to work. I done my day's work and then had an argument with the Guvnor then I left the job. He give me my wages before I went home. He asked me what I wanted my wages for. I told him I wanted to post some money off to my wife first thing in the morning. He asked me where my wife was and I told him she had gone to Bristol on a holiday. He said "How do you intend to send the money to her" and I said, "In a registered envelope." He paid me the money so he said "You can call over tomorrow morning for your cards." I then went home picked up my baby from her cot in the bedroom, picked up my tie and strangled her with it. I then put the baby back in the cot and sat down in the kitchen and waited for Christies downstairs to go to bed. At about twelve o'clock that night I took the baby downstairs to the wash house and hid her body behind some wood. I then locked the wash house door behind me and came in closing the back door behind me. I then slipped back upstairs and laid on the bed all night, fully clothed. I got up the following morning, washed, shaved and changed, and went up to see a man in Portobello Road about selling my furniture. I don't know his name. During the same afternoon he came to my flat, looked at it and offered me forty quid for it. I told him I would take £40 for it and then he asked me why I wanted to sell it. I told him I was going to Bristol to live as I has a job there waiting for me. He asked me why I wasn't taking the furniture with me. I told him my wife had already gone there and had a flat with furniture in it. He then asked me if it was paid for. I said it was. He said he would call Sunday afternoon to let me know what time the driver would call on Monday for it. I said I would wait in for him. Between 3 and 4 p.m. on Monday this man took all the furniture all the lino, and he paid me £40 which I signed for in a receipt book. He handed me the money which I counted in

his presence. I waited till he went then picked up my suitcase which I took to Paddington. The same night I caught the 12.55 a.m. train from Paddington to Cardiff and made my way to 93 Mount Pleasant, Merthyr Vale, where I stayed with my Uncle, Mr. Lynch. The rest I think you know. I have been asked to read this statement myself, but I cannot read. It has been read over to me and it is all the truth.

(Public domain by virtue of having been used in court proceedings)

Text A.20: New Baltimore Police Department. Written voluntary statement of Jonathan Kaled, 10/27/00, 1:40 a.m.

Me, Matt Dannells, and Frank drove up to Mancinos. I went in and lock the doors and told Justin to give me the money. Then I took him in to the walk-in to lock him in there and as I was walking out the gun went off so I turned off the lights and ran out the back. Then we left and went to Matt's brothers house.

I also I'm the one who make the fake call and order a pizza and a grinder and told them it was down County Line. He told me that they can't and I said that I was a friend of Ken.

Justin's wallet was through in a dumster at Van Pamle car lot. I didn't mean for this to happen. It was an accident.

(Public domain by virtue of having been used in criminal investigation)

Text A.21: Frank Kuecken's police statement, New Baltimore Police Department. Written voluntary statement of Frank Kuecken, 10/27/00, 1:30 a.m.

1. Who were you with on October 21, 2000 around 10 p.m.?
   F.J.
2. Were you in a vehicle and if so, who's vehicle?
   Mine, "68" Ford F-100 black.
3. Did you and F.J. go to Mancino's Pizza in New Baltimore on October 21, 2000 and if so what time?
   Yes. 9:50 p.m.
4. What did F.J. do at Mancino's Pizza?
   Went in and ripped the kid off.
5. How do you know F.J. ripped the kid off?
   He told me when he came out.
6. What was F.J. wearing when he ripped the kid off?
   Olive color cargo pants, burgandy hooded sweatshirt.

7. Did you and F.J. talk about what he was going to do at Mancino's Pizza that night?
    F.J. told me he was going to rip the kid off.
8. Where were you and F.J. coming from when he told you his plan?
    From Matt Daniels.
9. When F.J. got out of your truck and entered Mancino's Pizza, what did you see him do?
    He turned around and locked the door then turned off the lights and both of them went toward the back, then F.J. came from the back and entered my truck and said I ripped the kid off.
10. After F.J. got in your truck and told you what happened where did you go?
    Went to Matt's and picked him up and went to his brother's.
11. After you left Matt's brother's house, who was with you?
    Matt and F.J.
12. Where did you go?
    Back to Joe Goolins.
13. Did you drop F.J. off at his house?
    Yes.
14. Did F.J. say anything else to you besides what you have already said about the robbery?
    No.
15. When F.J. and Justin went to the back of the pizza store did you hear a gunshot?
    Yes.
16. Did F.J. tell you he shot Justin when he got into your truck after you heard the gunshot?
    Yes.
17. Is the gun F.J. used to shoot Justin in Matt Daniels' closet at far as you know?
    Yes.
18. After F.J. told you he shot Justin, you swore at him and did he then tell you he had to get rid of the gun?
    Yes.
19. When you and F.J. got to Matt's house did F.J. tell Matt he shot Justin?
    Yes.

(Public domain by virtue of having been used in court proceedings)

Text A.22: Excerpt from Unabomber text (The chapter entitled 'Revolution is easier than Reform')

Therefore two tasks confront those who hate the servitude to which the industrial system is reducing the human race. First, we must work to heighten the social stresses within the system so as to increase the likelihood that it will break down or be weakened sufficiently so that a revolution against it becomes possible. Second, it is necessary to develop and propagate an ideology that opposes technology and the industrial society if and when the system becomes sufficiently weakened. And such an ideology will help to assure that, if and when industrial society breaks down, its remnants will be smashed beyond repair, so that the system cannot be reconstituted. The factories should be destroyed, technical books burned, etc.

(Public domain by virtue of author's publication)

Text A.23: Mitchell's confession (Saudi Arabia)

My name is Alexander Hutton Johnston Mitchell. I am also known as Sandy.
I am of British nationality and I work at the Security Forces Hospital here in Riyadh.
I am the chief anaesthetic technician. I confirm and confess that I was ordered to carry out an explosion here in Riyadh, which took place on Friday, 17 November 2000.
The explosion was directed against Mr Christopher Rodway, who is of British nationality.
During this explosion I was assisted by Dr William Sampson, of Canadian nationality.
I placed the explosive device under the driver's seat of Christopher's car. That afternoon, Christopher came out with his wife, got into the car and drove off.
William detonated the remote control, which caused the explosion in Christopher's car.
We then turned south away from the scene of the incident. We found out later that Christopher had been killed in the explosion and his wife had been injured.
A second explosion was authorised and we carried out the order in a car which was similar to that of Christopher Rodway.

The second explosion was planned for 22 November. As Bill and I were discussing the results of the first explosion, a friend of ours, Mr Ralph Carl, a Belgian citizen, overheard our conversation.

It became necessary for us to involve Ralph in carrying out the second explosion to ensure his silence.

(Public domain by virtue of having been televised as a confession)

Text A.24: Sampson's confession (Saudi Arabia)

My name is William James Sampson, also known as Bill. I am a Canadian national working as a marketing consultant for the Saudi Industrial Development Fund.

I admit and acknowledge that I participated with Mr Alexander Mitchell in setting up an explosive device on the vehicle belonging to Mr Christopher Rodway, a British national.

Two days later, Mr Mitchell ordered me to set up a second explosion with the participation of Mr Ralph Skevins, a Belgian national.

The date for this explosion was set at Wednesday 22 November 2000. The explosive device and the remote switch were given to me by Mr Mitchell personally.

Dummy device

Under an agreement with Mr Mitchell, I constructed a dummy device which was given to Mr Ralph Skevins, as we both felt that his correct participation in the following events was uncertain.

I phoned Mr Skevins and told him to plant the device, his device, on the GMC Blazer, parked next to his . . . I waited in my vehicle and then proceeded to the target vehicle, where I removed the dummy device and planted the real explosive device.

(Public domain: see previous text)

Text A.25: Skevins' confession (Saudi Arabia)

My name is Ralph Carl Marya Skevins. I am of Belgian nationality. I work in King Fahd National Heart Hospital as a trauma co-ordinator.

I admit that I was in the house with Mr Alexander Mitchell, and also that I have heard a conversation between Alexander Mitchell and Dr William Sampson regarding the first car explosion where Mr Christopher Rodway died.

Later, Mr Alexander Mitchell informed me of his real involvement in the ex-
plosion. He also told me that he required me for the second explosion. My
role would be to actually place the explosion under the car, which would be
later appointed to me.
I received the explosive device from Dr William Sampson and I went to Al-Falah
compound ... Later on, I received a phone call from Dr William Sampson
to say that I have to place the device under the car parked on the right side
of me ...
[A short time later] I saw the car explode in front of me. I parked my car
immediately behind the car and I helped the people out to extricate them.
There was one severe injury and three are light injuries. Everything I write
down is in the books and is all in detail. Everything I say is a truth.

(Public domain: see previous text)

Text A.26: JonBénet Ramsey ransom note

Mr. Ramsey,

Listen carefully! We are a group of individuals that represent a small foreign
faction. We do respect your bussiness but not the country that it serves. At this
time we have your daughter in our posession. She is safe and unharmed and if
you want her to see 1997, you must follow our instructions to the letter.

You will withdraw $118,000.00 from your account. $100,000 will be in
$100 bills and the remaining $18,000 in $20 bills. Make sure that you bring an
adequate size attache to the bank When you get home you will put the money
in a brown paper bag. I will call you between 8 and 10 am tomorrow to instruct
you on delivery. The delivery will be exhausting so I advise you to be rested. If
we monitor you getting the money early, we might call you early to arrange an
earlier delivery of the money and hence a earlier pick-up of your daughter.

Any deviation of my instructions will result in the immediate execution of
your daughter. You will also be denied her remains for proper burial. The two
gentlemen watching over your daughter do not particularly like you so I advise
you not to provoke them.

Speaking to anyone about your situation, such as Police, F.B.I., etc., will
result in your daughter being beheaded. If we catch you talking to a stray dog,
she dies. If you alert bank authorities, she dies.

If the money is in any way marked or tampered with, she dies. You will
be scanned for electronic devices and if any are found, she dies. You can
try to deceive us, but be warned that we are familiar with Law enforcement

countermeasures and tactics. You stand a 99% chance of killing your daughter if you try to out smart us. Follow our instructions and you stand 100% chance of getting her back. You and your family are under constant scrutiny as well as the authorities. Don't try to grow a brain John. You are not the only fat cat around so don't think that killing will be difficult.

Don't underestimate us John. Use that good southern common sense of yours. It is up to you now John!

Victory!

S.B.T.C.

(Public domain by virtue of worldwide publicity)

Text A.27: E-Mail confession (excerpt) in the JonBénet Ramsey case

I feel so guilty for what I've done. I was there when the whole thing occurred. I never wanted any part in it, but they said if I didn't help I would be killed as well. I was only 14 when this took place, so I went along with the whole plan."

(Public domain: see previous text)

Text A.28: Lindbergh ransom note (Note 1)

Dear Sir!

Have 50000$ redy with 2500$ in 20$ bills 1500$ in 10$ bills and 1000$ in 5$ bills. After 2–4 days we will inform you were to deliver the Mony.

We warn you for making anyding public or for notify the polise the child is in gute care.

Indication for all letters are singnature and 3 holes.

(Public domain by virtue of having been used in court proceedings)

Text A.29: Lindbergh ransom note (Note 2)

Dear Sir. We have warned you note to make anything public also notify the police now you have to take consequences- means we will have to hold the baby until everything is quite. We can note make any appointments just now. We know very well what it means to us. It is realy necessary to make a world affair out of this, or to get your baby back as soon as possible to settle those affair in a quick way will be better for both- don't be afraid about the baby- keeping care of us day and night. We also will feed him acording to the diet.

We are interested to send him back in gut health. And ransom was made aus for 50000$ but now we have to take another person to it and probably have to

keep the baby for a longer time as we expected. So the amound will be 70000 20000 in 50$ bills 25000$ in 20$ bill 15000$ in 10$ bills and 10000 in 5$ bills Don't mark any bills or take them from one serial nomer. We will form you latter were to deliver the mony. But we will note do so until the Police is out of the cace and the pappers are quite. The kidnapping we prepared in years so we are prepared for everyding.

(Public domain by virtue of having been used in court proceedings)

Text A.30: Achim Joseph Marino confession

Dear Governor Bush Sir,

My name is Achim Josef Marino, #573514 and I am currently confined in the McConnell Unit of the T.D.C.J. – I.D., serving three life sentences plus three ten-year sentences for crimes committed at Austin, Texas in both 1988 and 1990. While in Austin in 1988, I also robbed, raped and shot a 20 year old woman at the Pizza Hut at Reinli Lane. This was in late October of 1988, after purchasing the murder weapon via the Austin American Statesmans classified section. The womens name was Nancy Lena Dupriest, and I have not been convicted for this crime. Approximately a month after this crime, I was arrested in El Paso, Texas, where the murder weapon was confiscated by the El Paso police department, however, the federal goverment ultimately convicted me for it. At the time of my arrest, I had the keys as well as two currency bags from the Pizza Hut with the name of Pizza Huts bank on the bag, in my possession and which remained in my personal property in the county jail for approximately 14 months. My friend, Janet Vaughn of P.O. Box 4973, El Paso, Tx., 79914, picked up my personal property after I was transferred to T.D.E.J.-I.D., for parole violation. She later took these items to my parents home where they remain to this day. Included with this confession to you is a B.O.T.F. report in conection with the confiscated murder weapon, and the purchase of it in Austin, Texas, shortly before the murder. IN 1990, after I was re-paroled by T.D.E.J.-I.D., I was once again arrested in Austin, Texas for robbery on approximately 5/30/90. While in the county jail, I was told by my cell mate, Raughleigh Lawson, that two men named Dansinger and Ochoa had been convicted for that crime. I told Raughleigh at that time that they had gotten the wrong people, that I knew the guy who had done it. He then told me that Dansinger and Ochoa had plead guilty to the murder. Governor Bush Sir, I do not know these men nor why they plead guilty to a crime they never committed. I can only assume that they must have been facing a capital murder trial with a poor chance of aguittal, but I tell you this sir, I did this awfull crime and I was alone.

Early last year, I wrote the Editor of the Austin American-Statesman, Chief Elizabeth Watson of the O.P.D., and Ms Susan Maldonado of the Austin office of the A.C.L.U. confessing to this crime because I believed that I was about to be killed here at the prison, and therefore I wanted to clear my conscience somewhat in regards to the lives of Dansinger, Ochoa and their loved ones. However, the confessions I'd made to these people was ultimately ignored. Now, I make this confession for a different reason. My life is no longer in danger, but my conscience still sickens me. I can not help Nancy Lena Dupriest or her family, but at least I can attempt to make ammends to Dansinger and Ochoa and their loved ones by doing my Christian duty and come clean about this terrible crime, a crime which has been enlarged and magnified by the arrest and conviction of two innocent men. Additionally, I have had a spiritual awakening and conversion, resulting in me becoming a Christian. This is a direct result of joining the Alcoholics Anonymous/Narcotics Anonymous Twelve Step Program, some 21 months ago and whose 12 steps and guiding principles caused me to have a spiritual awakening which ultimately lead me to the answer, Jesus Christ, His Father our Creator, the Holy Spirit, and of course, this confession. The Christian life-style and value system demands that I do this, even at the loss of my life, which I am fully pre-pared to lose and expect to loose. I 'am deeply sickened, disgusted and mortified for the crime I have committed, as well as my entire past life. I grieve for Nancy Lena Dupriest, her loved ones, as well as those of Dansinger and Ochoa, and also my familly. Prior to my Christian conversion and healing, I was insane. Never the less, there can be no excuses for my crime, because I knew exactly what I was doing. I am pre-pared to pay the prise for my actions. Governor Bush Sir, a copy of this letter/confession to you will also be sent to Ronny Earle of the Travis County District Attorney's Office. I wish to respectfully remind you, that in the event that you all decide to once again ignore this confession that you all are legally and morally obligated to contact Dansinger and Ochoa's attorneys and families concerning this confession. Thank you. God bless you and your family,

Yours in Jesus Christ,
Achim J. Marino
573514
McConnell Unit
3001 S. Emilly Dr.
Beeville, Texas 78102

(Public domain by virtue of having been used in court proceedings)

Text A.31: Confession of Anne Foster at Salem (1692)

The Devil appeared to her in the shape of a bird at several times, such a bird as she never saw the like before; and she had had this gift (viz., of striking the afflicted down with her eye) ever since. Being asked why she thought that bird was the Devil, she answered, because he came white and vanished away black; and that the Devil told her she should have this gift, and that she must believe him, and told her she should have prosperity: and she said that he had appeared to her three times, and always as a bird, and the last time about half a year since, and sat upon a table, had two legs and great eyes, and that it was the second time of his appearance that he promised her prosperity. She further stated, that it was Goody Carrier that made her a witch. She told her, that, if she would not be a witch, the Devil would tear her to pieces, and carry her away, at which time she promised to serve the Devil; that she was at the meeting of the witches at Salem Village; that Goody Carrier came, and told her of the meeting, and would have her go: so they got upon sticks, and went said journey, and, being there, did see Mr. Burroughs, the minis- ter, who spake to them all; that there were then twenty- five persons met together; that she tied a knot in a rag, and threw it into the fire to hurt Timothy Swan and that she did hurt the rest that complained of her by squeezing puppets like them, and so almost choked them; that she and Martha Carrier did both ride on a stick or pole when they went to the witch-meeting at Salem Village, and that the stick broke as they were carried in the air above the tops of the trees, and they fell: but she did hang fast about the neck of Goody Carrier, and they were presently at the village; that she had heard some of the witches say that there were three hundred and five in the whole country, and that they would ruin that place, the village; that there were also present at that meeting two men besides Mr. Burroughs, the minister, and one of them had gray hair; and that the discourse among the witches at the meeting in Salem Village was, that they would afflict there to set up the Devil's kingdom.

(Public domain by virtue of absence of copyright)

Text A.32: Ann Putnam, Salem (New England Historical and Genealogical Register, Volume XII, July 1858, page 245)

The deposition of Ann Putnam Junr who testifieth and saith some time in April 1692 there appeared to me the Apparition of an old short woman that told me her name was Martin and that she came from Amesbury who did immediately

afflict me, urging me to write in her book, but on the 2 May 1692 being the day of her examination Susanna Martin did most grievously afflict me during the time of her examination for when she did but look personally upon she would strike me down or almost choke and several times since the Apparition of Susanna Martin has most grievously afflicted me by pinching me and almost choking me urging me vehemently to write in her book. Also on the day of her Examination I saw the Apparition of Susanna Martin go and afflict the bodies of Mary Walcott, Mercy Lewis, Elizabeth Hubbard and Abigial Williams.

(Public domain by virtue of absence of copyright)

Text A.33: Fake suicide note from www.onion.com

I know you must be very angerey with me now but it had to be This way. Dont you see that now my pain is finnelley over forevr. My soul is set free. Tell Brittany I loved her truely. I just want something I can never have – Trent Reznor KIN 4Ever

    I apologilze to all whom I have brought sadness too.

    love, ethan

(Copyright onion.com 2001 – a text produced for its satirical value (accompanied by a humorous article criticizing educational standards in modern schools))

Text A.34: From David Ferrie (1) (suspect in JFK assassination)

Dear Al: When you read this I will be quite dead and no answer will be possible. I wonder how you are going to justify things.

    Tell me you treated me as you did because I was the one who always got you in trouble. The police arrest. The strip car charge. The deal at Kohn School. Flying Barragona in the Beech.

    Well, I guess that helps ease your conscience, even if it is not the truth. All I can say is that I offered you love, and the best I could. All I got in return in the end was a kick in the teeth. Thus I die alone and unloved.

    You would not even straighten out Carol about me, though this started when you were going steady.

    I wonder what your last days and hours are going to be like. As you sowed, so shall you reap.

(Public domain by virtue of absence of copyright)

Text A.35: Virginia Woolf

Dearest, I feel certain I am going mad again. I feel we can't go through another of those terrible times. And I shan't recover this time. I begin to hear voices, and I can't concentrate. So I am doing what seems the best thing to do. You have given me the greatest possible happiness. You have been in every way all that anyone could be. I don't think two people could have been happier till this terrible disease came. I can't fight any longer. I know that I am spoiling your life, that without me you could work. And you will I know. You see I can't even write this properly. I can't read. What I want to say is I owe all the happiness of my life to you. You have been entirely patient with me and incredibly good. I want to say that – everybody knows it. If anybody could have saved me it would have been you. Everything has gone from me but the certainty of your goodness. I can't go on spoiling your life any longer.

I don't think two people could have been happier than we have been.

V.

(Public domain)

Text A.36: Guy Fawkes (was he framed by Sir Robert Cecil or did he do it?) (Excerpt)

I confesse that a practise in general was first broken unto me against his Majesty for reliefe of the Catholique cause, and not invented or propounded by myself. And this was first propounded unto me about Easter last was twelve month, beyond the Seas, in the Low Countries of the Archduke's obeyance, by Thomas Wintour, who came thereupon with me into England

(Crown copyright)

Text A.37: Shoeless Joe Jackson (baseball player)

Q   You played in the World Series between the Chicago Americans Baseball Club and the Cincinatti Baseball club, did you?

A   I did.

Q   What position did you play in that series?

A   Left field.

Q   Were you present at a meeting at the Ansonia Hotel in New York about two or three weeks before – a conference there with a number of ball players?

A   I was not, no, sir.

Q Did anybody pay you any money to help throw that series in favor of Cincinatti?

A They did.

Q How much did they pay?

A They promised me $20,000 and paid me five.

Q Who promised you the twenty thousand?

A "Chick" Gandil.

Q Who is Chick Gandil?

A He was their first baseman on the White Sox Club.

Q Who paid you the $5,000?

A Lefty Williams brought it in my room and threw it down.

Q Who is Lefty Williams?

A The pitcher on the White Sox Club

Q Where did he bring it? Where is your room?

A At that time I was staying at the Lexington Hotel, I believe it is.

Q On 21st and Michigan?

A 22nd and Michigan, yes.

(Copyright, Chicago Historical Society. Permission gratefully acknowledged.)

Text A.38: Dennis Dowthitt (Texas death row prisoner)

I am so sorry for what y'all had to go through. I am so sorry for what all of you had to go through. I can't imagine losing two children. If I was y'all, I would have killed me. You know? I am really so sorry about it, I really am.

I got to go sister, I love you. Y'all take care and God bless you.

Gracie was beautiful and Tiffany was beautiful. You had some lovely girls and I am sorry. I don't know what to say.

All right, Warden, let's do it.

(Death row final statements do not carry copyright in the USA, Texts 38–46.)

Text A.39: Texas death row penpal request

My name is Dennis and I have been incarcerated since my birthday on June 20, 1990. I do not have any penpals and would like to start writing to a few people who are truly interested in a close relationship on paper.

Although I am where I am, I am an honest straightforward person with needs just like anyone else and the main need like most is companionship. I have a sister who has stuck by me through the entire ordeal of my trial and

incarceration and is now having many problems of her own, including not being able to walk, and can no longer come to visit. I really don't want to burden her any longer and try to make things easier on her by not telling her many of my problems. Truth is, I need a friend and someone to talk to and share with. However, I am one who likes to listen more than talk and need someone who would share their life and be honest with me.

I would really enjoy a family who would let me live the rest of my life with them.

I love kids and enjoy answering their questions and trying to help them. If there is anyone who would be interested in writing and sharing with me, please don't hesitate to write and we can take it slow and easy and get to know each other. Thank you.

(Public domain by absence of copyright)

Text A.40: Last statement of Basil McFarland

I owe no apologies for a crime I did not commit. Those who lied and fabricated evidence against me will have to answer for what they have done. I know in my heart what I did and I call upon the spirit of my ancestors and all of my people and I swear to them and now I am coming home.

Text A.41: Gerald Mitchell (Texas death row prisoner). His last statement (robbed and slayed 20-yr-old Charles Marino)

Yes, sir. Where's Mr. Marino's mother? Did you get my letter? Just wanted to let you know, I sincerely meant everything I wrote. I am sorry for the pain. I am sorry for the life I took from you. I ask God for forgiveness and I ask you for the same. I know it maybe hard, but I'm sorry for what I did. To my family I love each and every one of you. Be strong. Know my love is always with you … always. I know I am going home to be with the Lord. Shed tears of happiness for me. I love each and everyone of you. Keep on living.

Betty, you have been wonderful. You guided me to the Lord. You have been like a mother to me. Sean, Rusty, Jenny, Marsha, God Bless each and every one of y'all.

Jesus, I confess you as my Lord and Savior. I know when I die, I'll have life in heaven and life eternal everlasting. I am ready for that mansion that you promised me.

Take care.

It's alright Sean, it's alright. I'm going to a better place.

Text A.42: Jeffrey Carlton Doughtie (murdered elderly couple aged 80 and 76)

For almost nine years I have thought about the death penalty, whether it is right or wrong and I don't have any answers. But I don't think the world will be a better or safer place without me. If you had wanted to punish me you would have killed me the day after, instead of killing me now. You are not hurting me now. I have had time to get ready, to tell my family goodbye, to get my life where it needed to be.

It started with a needle and it is ending with a needle.

Carl, you have been a good friend, man. I am going to look for you. You go back and tell your daughter I love her. Tell her I came in here like a man and I will leave like a man. It's been good, dude. Thank you, Shorty. I appreciate you. I came in like a man and I will leave like a man. I will be with you. I will be with you every time you take a shower. If you leave crying you don't do me justice. If you don't see peace in my eyes you don't see me. I will be the first one you see when you cross over.

They got these numbers that I called today. Calling my family.

That is it. Ready, Warden.

Text A.43: Mack Hill (admitted to nothing but convicted of murdering a friend)

First, I would like to tell my family that I love them. I will be waiting on them. I am fine. I hope that everyone gets some closure from this. I am innocent. Lubbock County officials believe I am guilty. I am not. Travis Ware has the burden on him to prove that he did not commit felonies. He needs to be stopped or he is going to do it time and time again. The power is invested in you as a public official to do your job.

That's all Warden. I love y'all.

Text A.44: Miguel Flores (convicted of capital murder for kidnapping, sexually assaulting, and murdering a 20-year-old female)

I want to thank my attorneys, Father Walsh ... Sylvia, te quiero mucho y a Consulado, te quiero decir muchas gracias por todo. I want to say I am sorry and I say a prayer today for you so you can have peace and I hope that you can forgive me. God is waiting and God is waiting now.

Text A.45: Last statement: Jeffery Dillingham (contract-killed 40-year-old Caren Koslow)

I would just like to apologize to the victim's family for what I did. I take full responsibility for that poor woman's death, for the pain and suffering inflicted on Mr. Koslow.

Father, I want to thank you for all of the beautiful people you put in my life. I could not have asked for two greater parents than you gave me. I could just ask for two greater people in their life now. It is a blessing that there are people that they love so much but even more so, people that I love so much.

I thank you for all the things you have done in my life, for the ways that you have opened my eyes, softened my heart.

The ways that you have taught me. For teaching me how to love, for all of the bad things you have taken out of my life. For all the good things you have added to it.

I thank you for all of the beautiful promises that you make us in your word, and I graciously received every one of them. Thank you Heavenly Father for getting me off of death row and for bringing me home out of prison.

I love you Heavenly Father, I love you Jesus. Thank you both for loving me. Amen.

Text A.46: Last statement: Karla Faye Tucker (slayed 27-year-old Jerry Dean with a pickaxe)

Yes sir, I would like to say to all of you – the Thornton family and Jerry Dean's family that I am so sorry. I hope God will give you peace with this.

Baby, I love you. Ron, give Peggy a hug for me. Everybody has been so good to me.

I love all of you very much. I am going to be face to face with Jesus now. Warden Baggett, thank all of you so much. You have been so good to me. I love all of you very much. I will see you all when you get there. I will wait for you.

Text A.47: Ransom note to Sheriff Sheldon (The following note was received Nov. 28 via U.S. Mail at the Yoknapatawpha County Sheriff's Department. The postmark was dated Nov. 25 from Jackson, MS. The note arrived in a plain white envelope with no return address.)

"SHERIFF" SHELDON
Bust another card room and we will bust her head!

Leave it alone like all the others
Busting card rooms won't help make you a real sheriff.
Busting card rooms won't help her at all.
If you want her back, BACK OFF!

(Public domain by virtue of having been publicized in the commissioning of a crime)

Text A.48: Example of a fake suicide note

```
ITS THE COWARDS ROAD BUT IT SUITS ME. I'VE SACRIFICED
TOO MUCH TO LIVE AS AN HONORABLE MAN. I'D BE LYING TO
CONTINUE.
I'VE CHOSEN THIS ROAD MYSELF. NO ONE HAS FORCED ME TO
DO IT. ITS ALL MY OWN DOING AND I HAVE NO ELSE ONE TO
BLAME.
MY FAMILY, MY PRECIOUS MACY- I LOVE YOU, I KNOW I'VE
HURT YOU, BUT ITS THE ONLY WAY I COULD CHOOSE. ITS NOT
ABOUT YOU, ITS ME AND MY CONSCIENCE AND ITS A STRUGGLE I
HAVE TO END.
   Charles
```

(Public domain)

Text A.49: Excerpt – Elderly Male, 82, Allegheny County, Pennsylvania

I am at a situation where I cannot bear the future and subject you to the many months or years of trying to take care of an invalid who is suffering from emphysema and senility, and possibility of worse

(Public domain)

Text A.50: Excerpt – Elderly Male, 73, Allegheny County, Pennsylvania

It's time for my exit," his note said. "I can't breathe, I can't sleep, my legs are swollen and can't walk. My back has been killing me. . . . I'm trying to eliminate costly hospitals that will cut into finances

(Public domain)

Text A.51: Joel Rose (excerpt)

"I am so sad – I have ruined your life – but it will be proved, I didn't do these awful things, I hope the DNA test helps show that," Joel Rose wrote in a hastily composed suicide note to his wife Lois.

"Those men who invaded our house killed me, and those awful people bringing these bogus charges . . .

"When your integrity has been destroyed you have nothing . . .

(Public domain)

Text A.52: A.T. Renouf's suicide note

To Whom It May Concern

Last friday (13-October) my bank account was garnisheed. I was left with a total of $00.43 in the bank.

At this time I have rent and bill's to pay which would come to somewhere approaching $1500.00 to $1800.00.

Since my last pay was also direct deposited on friday I now have no way of supporting myself. I have no money for food or for gas for my car to enable me to work. My employer also tells me that they will only pay me by direct deposit. I therefore no longer have a job, since the money would not reach me.

I have tried talking to the Family Support people at 1916 Dundas St. E. their answer was:-" we have a court order." repeated several times.

I have tried talking to the welfare people in Markham. Since I earned over $520.00 last month I am not eligible for assistance.

I have had no contact with my daughter in approx. 4 year's. I do not even know if she is alive and well. I have tried to keep her informed of my current telephone number but she has never bothered to call.

I have no family and no friend's, very little food, no viable job and very poor future prospects. I have therefore decided that there is no further point in continuing my life. It is my intention to drive to a secluded area, near my home, feed the car exhaust into the car, take some sleeping pills and use the remaining gas in the car to end my life.

I would have prefered to die with more dignity.

It is my last will and testament that this letter be published for all to see and read.
Signed
A.T. Renouf

(Public domain)

Text A.53: Ocalan's Final Statement

(Text of Abdullah Ocalan's Statement just before the verdict was handed down at his trial.)

I reject the accusation of treason. I believe that I am struggling for the unity of the country and freedom. I believe that my struggle was for a democratic republic not against the republic.

I hope that the problem which has grown as a result of historic mistakes will reach a solution. I want this trial to contribute to that. I am repeating my call, the determined promise I made at the onset, for a fair and honorable peace and brotherhood in line with the democratic republic.

I call on humanity, the state and all societal forces to fulfill their duty. The future of the country lies with peace not war.

I greet you all.

(Copyright Associated Press)

Text A.54: Michael McBride's final statement (after receiving the lethal injection)

Thank you um I anticipated that I would try to memorize and recite beatitudes New Testament more or less Luke's beatitudes I should say and ah a chapter on love in first Corinthians chapter 13, ah I pretty much knew that I would not be able to memorize so much there was also a poem that went along with it and in anticipation of not being able to um fulfill that desire I provided a written statement that will be made available to anybody that will be made available to anybody that wants it I believe, isn't that correct? So uh I wanted you to hear me say that and I apologize and for any other grief I have caused you know, including that a what your about to witness now. It won't be very long as soon as you realize that appear I am falling asleep I would leave because I won't be here after that point I will be dead at that point. It's irreversible. God Bless all of you thank you.

(Public domain see 38–46, above)

Text A.55: Joel Rose – sent from David Morton, the news editor at the Cleveland Free Times, Colorado USA in Jan 2002. Joel Rose committed suicide after being accused of sending women innuendoes through the mail. Dennis Kacnler was his boss.

To Dennis Kacnler

Chief – check that garbage theft thing – I'm sure this was connected – I remember seeing one BUC pickup truck with cab picking it up one morning it was stolen – a CIZ employee was driving – I don't know who –

Chief, I did not do this but I cannot face anyone again –

There is a DNA test outstanding that was done Wed at Parma hosp, – I am sure the tesults will clear me but no one will believe it –

Bless you

JR

(Public domain)

Text A.56: Michael Sams – one of several ransom demands

Your employee has been kidnapped and will be released for a ransome of £175,000. With a little luck he should be still O.K. and unharmed, to prove this fact to you will in in the next day or so receive a recorded message from him. He will be released on Friday 31 January 1992, provided:

On Wednesday 29 January a ransome of £175,000 is paid, and no extension to this date will be granted.

The police are not informed in any way until he has been released.

On Wednesday 29th at 4 pm (on line 021 358 2281) you will receive a short recorded message from the hostage. To prove he is still alive and O.K. he will repeat the first news item that was on the 10am, Radio 2 news. He will then give further instructions. A second and more detailed message will be given at 5.05 pm the same day. Your watch must be synchronized with the 5 pm pips on Radio 2. The location of the second call will be given at 4 pm, so transport with a radio must be available.

The money must be carried in a holdall and made up as folows, precisely;

£75,000 in used £50. £75,000 in used £20. £25,000 in used £10 packed in 31 bundles, 250 notes in each.

Kevin Watts (if not the hostage) mst be the person to receive all messages and carry the money to the appointed place.

However, please note that all messages will be pre-recorded, so no communication or negotiations can be made.

YOU HAVE BEEN WARNED. HIS LIFE IS IN YOUR HANDS.

(Public domain by virtue of having been used in criminal and court proceedings)

Text A.57: Michael Sams – excerpt from another ransom demand

A young prostitute has been kidnapped from the Chapeltown area last night and will only be released unharmed if the conditions below are met if they are not met then the hostage will never be seen again also a major ciy centre store (not necessary in Leeds) will have a fire bomb explode at 5 am 17 July.

A payment of £140,000 is paid in cash (one hundred & fourty thousand)

(Public domain by virtue of having been used in criminal and court proceedings)

Text A.58: Dyfed-Powys Fire Service, emergency call

Hello can I help you?

Yes security guard here XYZ Electronics at the ABC canal.

Yes?

I've just been walking on the back well there's smoke coming out of one of our roofs. So, could you tell the fire service our curtain shop there's a lot of chemicals and what have you in there.

Right. Is it XYZ Electronics.

Yeah.

ABC Industrial Estate.

That's right.

By the ABC Canal?

Yeah. What I'll do I'll make my way round the back and I'll leave the gates open for them.

Right and it's smoke is it?

Yeah. Smoke coming out of the roof.

From the roof?

I haven't been there to investigate but I'll –

Right – are you . . . . any of the chemicals that are kept there?

I couldn't really tell you.

Are they just –
chemicals from the –
So –
phuric acid is in there.
It's what sorry?
Sulphuric acid is in there.
Okay, don't worry now, we're on our way there.
Okay.
Thank you. Bye.

(Grateful acknowledgements to the Dyfed-Powys Fire Service)

Appendix 2: IPA Chart

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)

CONSONANTS (PULMONIC) © 2005 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	N		
Trill	B			r					R		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	L			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks	Voiced implosives	Ejectives
ʘ Bilabial	ɓ Bilabial	ʼ Examples:
ǀ Dental	ɗ Dental/alveolar	pʼ Bilabial
ǃ (Post)alveolar	ʄ Palatal	tʼ Dental/alveolar
ǂ Palatoalveolar	ɠ Velar	kʼ Velar
ǁ Alveolar lateral	ʛ Uvular	sʼ Alveolar fricative
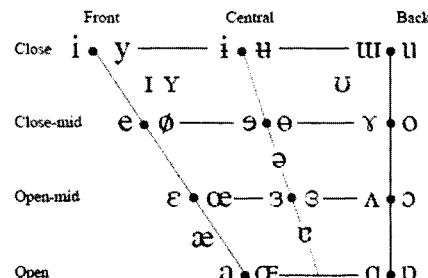
OTHER SYMBOLS

ʍ Voiceless labial-velar fricative

w Voiced labial-velar approximant

ɥ Voiced labial-palatal approximant

ʜ Voiceless epiglottal fricative

ʢ Voiced epiglottal fricative

ʡ Epiglottal plosive

ɕ ʑ Alveolo-palatal fricatives

ɺ Voiced alveolar lateral flap

ɧ Simultaneous ʃ and x

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.

k͡p t͡s

VOWELS

Where symbols appear in pairs, the one to the right represents a rounded vowel.

SUPRASEGMENTALS

ˈ	Primary stress
ˌ	Secondary stress
	ˌfoʊnəˈtɪʃən
ː	Long eː
ˑ	Half-long eˑ
˘	Extra-short ĕ
ǀ	Minor (foot) group
ǁ	Major (intonation) group
.	Syllable break ɹi.ækt
‿	Linking (absence of a break)

DIACRITICS Diacritics may be placed above a symbol with a descender. e.g. ŋ̊

̥	Voiceless	n̥ d̥	̤	Breathy voiced	b̤ a̤	̪	Dental t̪ d̪
̬	Voiced	s̬ t̬	̰	Creaky voiced	b̰ a̰	̺	Apical t̺ d̺
ʰ	Aspirated	tʰ dʰ	̼	Linguolabial	t̼ d̼	̻	Laminal t̻ d̻
̹	More rounded	ɔ̹	ʷ	Labialized	tʷ dʷ	̃	Nasalized ẽ
̜	Less rounded	ɔ̜	ʲ	Palatalized	tʲ dʲ	ⁿ	Nasal release dⁿ
̟	Advanced	u̟	ˠ	Velarized	tˠ dˠ	ˡ	Lateral release dˡ
̠	Retracted	e̠	ˤ	Pharyngealized	tˤ dˤ	̚	No audible release d̚
̈	Centralized	ë	̴	Velarized or pharyngealized ɫ			
̽	Mid-centralized	e̽	̝	Raised	e̝ (ɹ̝ = voiced alveolar fricative)		
̩	Syllabic	n̩	̞	Lowered	e̞ (β̞ = voiced bilabial approximant)		
̯	Non-syllabic	e̯	̘	Advanced Tongue Root e̘			
˞	Rhoticity	ɚ a˞	̙	Retracted Tongue Root e̙			

TONES AND WORD ACCENTS

LEVEL		CONTOUR	
e̋ or ˥	Extra high	ě or ˇ	Rising
é ˦	High	ê ˆ	Falling
ē ˧	Mid	e᷄ ˈ	High rising
è ˨	Low	e᷅ ˌ	Low rising
ȅ ˩	Extra low	e᷈	Rising-falling
↓	Downstep	↗	Global rise
↑	Upstep	↘	Global fall

Appendix 3: International Association for Forensic Phonetics and Acoustics (IAFPA) Code of Practice

This is the current IAFPA Code of Practice which was approved by the AGM in Helsinki, 2004. See www.iafpa.net/code.htm for any future updates.

1. Members should act in all circumstances with integrity, fairness and impartiality.
2. Recognising the varied array of casework subsumed under the interests of IAFPA (e.g. speaker identification/elimination, speaker profiling, voice line-ups, transcription, authentication, signal enhancement, sound propagation at crime scenes), Members should maintain awareness of the limits of their knowledge and competencies when agreeing to carry out work.
3. Members should not enter into any arrangements in which remuneration is dependent on the outcome of the case.
4. Members should make clear, both in their reports and in giving evidence in court, the limitations of forensic phonetic and acoustic analysis.
5. In reporting on cases where an opinion or conclusion is required, Members should make clear their level of certainty and give an indication of where their conclusion lies in relation to the range of judgements they are prepared to give.
6. (a) Members should exercise particular caution if carrying out forensic analysis of any kind on recordings containing speech in languages of which they are not native speakers.
   (b) In carrying out forensic speaker identification/elimination work, Members should exercise particular caution if the samples for comparison are in different languages.

(c) Members should exercise particular caution if carrying out authenticity or integrity examinations of recordings that are not claimed to be original.

7. Members undertaking forensic phonetic and acoustic analyses or operations of all kinds should state in their reports the methods they have followed and provide details of the equipment and computer programs used.

8. Members, in making their analysis, should take due account of the methods available at the time and of their appropriateness to the samples under examination.

9. Members should not attempt to do psychological profiles or assessments of the sincerity of speakers.

10. Members' reports should not include or exclude any material which has been suggested by others (in particular by those instructing them) unless that Member has formed an independent view.

Appendix 4: Notes on Daubert

From: KUMHO TIRE CO., LTD., et al. v. CARMICHAEL et al. In the SUPREME COURT OF THE UNITED STATES (1999)

'The court noted that *Daubert* discussed four factors – **testing, peer review, error rates, and "acceptability" in the relevant scientific community** – which might prove helpful in determining the reliability of a particular scientific theory or technique, *id.*, at 593–594, and found that those factors argued against the reliability of Carlson's methodology. On the plaintiffs' motion for reconsideration, the court agreed that *Daubert* should be applied flexibly, that its four factors were simply illustrative, and that other factors could argue in favor of admissibility.'

Also:

'The *Daubert* "gatekeeping" obligation applies not only to "scientific" testimony, but to all expert testimony. Rule 702 does not distinguish between "scientific" knowledge and "technical" or "other specialized" knowledge, but makes clear that any such knowledge might become the subject of expert testimony. It is the Rule's word "knowledge," not the words (like "scientific") that modify that word, that establishes a standard of evidentiary reliability. 509 U.S., at 589–590. *Daubert* referred only to "scientific" knowledge because that was the nature of the expertise there at issue. *Id.*, at 590, n. 8. Neither is the evidentiary rationale underlying *Daubert's* "gatekeeping" determination limited to "scientific" knowledge. Rules 702 and 703 grant all expert witnesses, not just "scientific" ones, testimonial latitude unavailable to other witnesses on the assumption that the expert's opinion will have a reliable basis in the knowledge and experience of his discipline. *Id.*, at 592. Finally, it would prove difficult, if not impossible, for judges to administer evidentiary rules under which a "gatekeeping" obligation depended upon a distinction between "scientific" knowledge and "technical" or "other specialized" knowledge, since there is no clear line dividing the one from the others and no convincing need to make such distinctions.'

Also:

'A trial judge determining the admissibility of an engineering expert's testimony *may* consider one or more of the specific *Daubert* factors. The emphasis on the word "may" reflects *Daubert*'s description of the Rule 702 inquiry as "a flexible one." 509 U.S., at 594. The *Daubert* factors do *not* constitute a definitive checklist or test, *id.,* at 593, and the gatekeeping inquiry must be tied to the particular facts, *id.,* at 591. Those factors may or may not be pertinent in assessing reliability, depending on the nature of the issue, the expert's particular expertise, and the subject of his testimony. Some of those factors may be helpful in evaluating the reliability even of experience-based expert testimony, and the Court of Appeals erred insofar as it ruled those factors out in such cases. In determining whether particular expert testimony is reliable, the trial court should consider the specific *Daubert* factors where they are reasonable measures of reliability.'

In the Kumho tire case, Carlson's testimony failed because of his 'failure to satisfy either *Daubert*'s factors *or any other* set of reasonable reliability criteria'.

Appendix 5: Selection of author's recent cases

Investigation	Linguistic issue	Charge/Matter/Action
Alteration of interview statements	Non-native Speaker ability	Terrorism
Interpretation of language in a song	Interpreting specific cultural expressions	Murder
Copyright infringement	Plagiarism	Copyright infringement
Authorship of anonymous mail	Authorship	Witness intimidation
Authorship of anonymous mail	Authorship	Attempt to pervert justice
Authorship of witness statements	Public/communal authorship of witness statements	Genocide
Authorship of anonymous mail	Authorship	Death threats
Copyright infringement	Authorship	Copyright infringement
Authorship of anonymous mail	Authorship	Malicious communication
Authorship of anonymous mail	Authorship styles of Management vs Trade union	Malicious communication
Authorship of anonymous mail	Authorship	Malicious communication
Authorship of alleged suicide note	Dictation vs written language	Murder
Interpretation of coded language	De-coding code and interpreting	Child sex abuse
Authorship of anonymous mail	Authorship of hate crime letters	Malicious communication
Copyright infringement	Plagiarism	Copyright infringement
Authorship of mobile phone texts	Mobile phone authorship style	Murder
Authorship of anonymous mail	Authorship	Malicious communication
Authorship of mobile phone texts	Mobile phone authorship style	Missing person
Text type determination	Interpretation of text type	Suicide/suspicious death
Copyright infringement	Plagiarism	Copyright infringement
Transcription and interpretation of recording	Transcription and interpretation	Sexual harassment
Transcription and interpretation of recording	Transcription and interpretation	Sexual assault

Investigation	Linguistic issue	Charge/Matter/Action
Provenance of diary entries	Written vs spoken language	Illegal tape-recording
Authorship of anonymous mail	Authorship	Malicious communication
Authorship of anonymous mail	Authorship	Malicious communication
Authorship of commercial letters	Authorship	Fraud
Authorship of examination paper	Plagiarism	Academic integrity
Authorship of anonymous mail	Authorship	Malicious communication
Authorship of child's letter	Child vs adult authorship	Child custody
Authorship of anonymous mail	Authorship	Malicious communication
Authorship of mobile phone texts	Mobile phone authorship style	Murder
Allegation of oppressive interviewing	Conversation analysis of interview tape	Reckless driving
Transcription of disputed utterance	Transcription and interpretation	Terrorism
Transcription of recording	Audio clean-up and transcription	Organized crime – narcotics
Threat assessment	Assessment of stalking threat	Harassment
Authorship of hate mail	Authorship	Malicious communication
Authorship of email	Authorship	Downloading child pornography
Authorship of hate mail	Authorship	Malicious communication
Authorship profile	Evaluation of language ability and cultural information	Product contamination
Denial of information to consumer	Language of power and corporate abuse	Breach of Freedom of Information Act
Authorship	Authorship and text type determination	Suicide/suspicious death
Interpretation of coded language	Interpreting specific cultural expressions	Illegal firearm possession
Interpretation of mobile phone texts	Authorship of mobile phone texts	Suicide/suspicious death
Transcription and statement analysis of interview tape	Interpretation of answers	Murder/suspicious death
Authorship	Authorship of hate emails	Malicious communication
Transcription and interpretation of coded language	Transcription and interpretation	Armed robbery, conspiracy
Authorship of hate mail	Authorship	Malicious communication
Authorship of hate mail	Authorship	Malicious communication

Investigation	Linguistic issue	Charge/Matter/Action
Transcription of disputed utterance	Acoustic phonetics	Conspiracy to pervert justice
Transcription of recording	Acoustic and auditory phonetics	Murder
Interpretation of mobile phone texts	Authorship	Blackmail
Assessment of expert report	Authorship evaluation	Malicious communication
Interpretation of coded language	De-coding code and interpreting	Child sex abuse
Comparison of police statements	Written vs spoken language	Perjury
Analysis of diary entries	Contemporaneity of written texts	Criminal damage, arson
Comparison of security officer statements	Written vs spoken language and collusion	Theft
Comparison of witness statements	Testing for language of power/abuse	People trafficking
Comparison of witness statements	Testing for language of power/abuse	People trafficking
Authorship of hate mail	Authorship	Malicious communication
Analysis of document found near crime scene	Interpretation of document relative to assault	Serious assault
Analysis of prosecutor and witness statements	Authorship influence	Drink driving charge
Analysis of police statements	Authorship influence	Reckless driving
Voice comparison	Acoustic and auditory phonetics	Malicious communication
Voice comparison and interpretation	Acoustic and auditory phonetics	Firearms, attempted murder
Voice comparison and interpretation	Acoustic and auditory phonetics	Arson
Statement comparison	Written vs spoken language	Murder
Statement analysis	Interpretation of answers	Sexual assault
Comparison of documents	Authorship, plagiarism	Fraud
Coded diary entries	Interpretation of code in context of diary entries	Fraud
Authorship of hate mail	Authorship	Malicious communication
Authorship of hate mail	Authorship	Malicious communication
Comparison of police statements	Collusion	Murder
Voice comparison and interpretation	Acoustic and auditory phonetics	Narcotics offences
Interpretation of coded language	Interpretation	Child sex offences

Investigation	Linguistic issue	Charge/Matter/Action
Authorship of hate mail	Authorship	Staff harassment
Authorship of hate mail	Authorship	Malicious communication
Authorship influence	Whether male partner influenced authorship	Assault
Authorship of hate mail	Authorship of political hate mail	Malicious communication
Interpretation of exam question	Semantics, ambiguity	Discourse analysis
Authorship of witness statements	Collusion	Child custody
Authorship of hate mail	Authorship	Staff harassment
Authorship of hate mail	Authorship	Malicious communication
Interpretation of coded language	Interpreting specific cultural expressions	Murder
Authorship of hate mail	Authorship	Malicious communication
Authorship of term paper	Plagiarism	Academic integrity
Authorship of hate mail	Authorship	Malicious communication
Authorship of hate mail	Authorship	Malicious communication
Transcription and interpretation of recording	Acoustic and auditory phonetics	Staff harassment
Spoken/written language	Comparison of statements	Commercial dispute
Authorship profile	Evaluation of language ability and cultural information	Product contamination
Witness statements	Witness collusion	Assault

Notes

1 Verballing is the process whereby police officers alter the statements of suspects and witnesses.
2 All online references are listed in the bibliography.
3 These remarks are made only with reference to the English language, see Barton 1994: 124–5.
4 At the beginning of 2001 there were an estimated one billion pages on the internet, six months later this figure had increased by 50%. In other words it took six months to achieve what had previously taken three years. In mid-2002 there were 2.01 billion pages on Google alone.
5 Others who have used multivariate analysis in authorship studies have included Holmes (1992) and Hoover (2001).
6 This is not to suggest that the linguistic motivation for spelling errors is itself superficial: we can study underlying spelling patterns.
7 Beginning with phonology, Jakobson's interest in universals moved on to theories of language acquisition and change, while Chomsky has sought to understand what kinds of structures underpin knowledge of language.
8 Assuming the individual had no physical or neurological speech difficulties.
9 It is important to emphasize that I am not applying any form of statement analysis here. There are no veracity issues in the inquiry I am discussing.
10 'Cause' is a frequent collocate of 'probable'.
11 *Kelliher (Village of) vs Smith*, ([1931] S.C.R. 672) (see: www.economica.ca/ew42p2.htm] (as at 23/02/02).
12 *R. vs Mohan*, ([1994] 2 S.C.R. 9, at 23)
13 This valuable *insight* into Canadian forensic expertise may be found at: www.economica.ca/ew42p2.htm (as at 23/02/02).
14 More information can be found at: www.umanitoba.ca/faculties/law/Courses/esau/legal_systems/outline20-99.html
15 At the request of the defence barrister in this case personal names and place names have been changed. If this case is featured in subsequent editions of law books, e.g. Blackstone's, then this document will be revised to enable readers to refer to the case more formally.
16 A WWW search using Google.com reveals 11 instances of 'penacilin' as opposed to 'penicillin' with AltaVista.com showing a similar quantity. The main instances are in pop, hip-hop music, rap and other emblems of 80s and 90s youth culture including a reference to LSD. This misspelling could simply be a red herring.
17 I could find no record of Margaret Canby's story having been transformed into a Braille text – the only method by which Helen Keller could have read it. This issue does not seem to have been raised before.

18 A note on the use of an internet search engine for the purpose of checking overall frequencies of words and phrases in the language is in order. The main issue is whether it is a reliable and valid method. One test is to apply the function word test. This involves checking the frequency of very common function words ('the', 'it', 'a/an', etc.) and seeing whether their occurrence on an engine like Google matches that of general language corpora. I found (on Google at least) that there was a close match. Extrapolating from this it seemed to me that using an internet search engine for this purpose is a reasonably safe method for making observations on function word frequencies in the language generally. There is, in the present case at least, an objection to this method, which is the argument that when analysing texts of a certain period – in this case the early twentieth century – it would probably be better to use a purpose-built corpus of the time, incorporating literary works from, say, 1880 to 1920.

19 Caution should be applied given the small size of the data set.

20 Many other categories are possible, including the following (some of which will be explained at the end of the chapter): character naming – should be consistent; register – should either be formal, informal or neutral; pronouns – subject pronouns should not be omitted before important clauses; participants – should be introduced as they appear; absence of private verbs – e.g. 'I think'; Gricean maxims, especially of quantity – as you read the text ask yourself whether excessive detail is being given; nicknames – if people are given nicknames after being introduced with their 'usual' name, this may be suspicious, could indicate a 'padding' of the text with 'realistic' details; negatives are much rarer than positives – also, look out for spontaneous, i.e. unforced denials; watch out for ascriptive adjectives, especially if they are 'stacked'.

21 See end of chapter for discussion on statement analysis of *reports*.

22 My suspicion about the word *copper* – unconfirmed at this time – is that it is a word police officers use to refer to themselves, not a word in common use by members of the public.

23 Names and other information are changed to protect the identities of parties involved.

24 The reader will appreciate that very few contexts in forensic linguistics are 'typical'.

25 Note this is not a full (CA) conversation analysis, but the author's own modified version. I am indebted to the Dyfed-Powys Fire Service, Wales, for their kind provision of this corpus of genuine and simulated taped calls.

26 Note that the operator, the security guard, the householder and the police officer are all native English speakers, and drawn from similar social backgrounds. Major social and linguistic differences, however, might place tensions on the model proposed here.

27 As a matter of fact this is exactly what happened. The recipient did resign, but several years later received further letters threatening 'exposure' of 'misdemeanours'.

28 As found in Schneidman and Farberow (1957) who founded the science of suicidology.

29 Note that under no circumstances should you ask people to create simulated suicide notes.

Bibliography and further reading

Akamatsu, T. 1995. 'On some neutralisations and archiphonemes in English allegro speech', in J. Windsor Lewis (ed.), *Studies in General and English Phonetics: Essays in honour of Professor JD O'Connor*. London: Routledge.

Ashforth, T. 'Silence in court'. Online at wwwmcc.murdoch.edu.au/ReadingRoom/Ashforth/Terthes6.html on 30/06/2002 at 16:18.

Baayen, R. H., H. Van Halteren, and F. Tweedie. 1996. 'Outside the cave of shadows: using syntactic annotation to enhance authorship attribution'. *Literary and Linguistic Computing*, XI, 121–31.

Baldwin, J. R. and P. French 1990. *Forensic Phonetics*. London: Pinter Publishers.

Barton, D. 1994. *Literacy: An Introduction to the Ecology of Written Language*. Oxford: Blackwell.

Battistella, E. L. 1996. *The Logic of Markedness*. Oxford: Oxford University Press.

Bax, R. Online at www.let.leidenuniv.nl/English/studiepad/SHT/corpus_linguistics.htm on 30/05/2002 at 15:39h UTC.

Biber, D. 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.

Bilgrami, A. 1993. *Belief and Meaning*. Oxford: Blackwell.

Binongo, J. N. and M. W. A. Smith 1999. 'The application of principal component analysis to stylometry'. *Literary and Linguistic Computing*, XIV, 445–66.

Bower, K. M. Using CUSUM Charts To Detect Small Process Shifts [online].

Burrows, J. 1987. *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. Oxford: Clarendon Press.

Byrne, C. and P. Foulkes 2004. 'The mobile phone effect on vowel formants'. *International Journal of Speech, Language and the Law* (formerly *Forensic Linguistics*), XI (1) 83–102.

Canter, D. and J. Chester 1997. 'Investigation into the claim of weighted Cusum in authorship attribution studies'. *International Journal of Speech, Language and the Law* (formerly *Forensic Linguistics*), IV (2), 252–61.

Charteris-Black, J. 1997. 'A Study of Arabic, Chinese and Malay First & Second Language Punctuation', in A. Ryan and A. Wray (eds) *Evolving Models of Language*. Clevedon, Philadelphia, Toronto, Sydney, Johannesburg: BAAL & Multilingual Matters, pp. 23–40.

Chaski, C. 1998. 'A Daubert-Inspired Assessment of Current Techniques for Language-Based Author Identification'. US National Institute of Justice.

—2001a. 'Language based author identification techniques'. *International Journal of Speech, Language and the Law* (formerly *Forensic Linguistics*), VIII (1), 1–65.

—2001b. 'Empirical evaluations of language-based author identification techniques'. *International Journal of Speech, Language and the Law* (formerly *Forensic Linguistics*), VIII (1), 1–65.

—2005. 'Who's At The Keyboard? Authorship Attribution in Digital Evidence Investigations'. *International Journal of Digital Evidence*, IV (1), 1–13.

Clifford, B. R. 1980. 'Voice identification by human listeners: on earwitness reliability'. *Law and Human Behavior*, On Eyewitness Behavior, IV (4), 373–94.

Coulthard, M. 1994. 'On the use of corpora in the analysis of forensic texts'. *International Journal of Speech, Language and the Law* (formerly *Forensic Linguistics*), I (1), 27–43.

—1994(b). 'Forensic Discourse Analysis', in M. Coulthard (ed.) *Advances in Written Text Analysis*. London: Routledge, pp. 242–58.

—1994 (c). 'Powerful evidence for the defence: an exercise in forensic discourse analysis', in J. Gibbons (ed.), *Language and the Law*. London: Longman, pp. 414–27.

Craig, H. 'Authorial Attribution and Computational Stylistics: if you can tell authors apart, have you learned anything about them?'. Centre for Literary and Linguistic Computing, University of Newcastle, NSW, Australia. Online at http://lingua.arts.klte.hu/allcach98/abst/abs7.htm on 02/04/2002.

Crystal, D. 1991. 'Stylistic Profiling', in K. Aijmer and B. Altenberg (eds), *English Corpus Linguistics*. London and New York: Longman, pp. 221–38.

—'Vanishing Languages'. Online at www.farsarotul.org/nl21_4.htm on 30/05/2002 at 17:42h UTC.

Davison, Alice 1984. 'Syntactic Markedness and the Definition of Sentence Topic'. *Language*, LX (4), 797–846.

De Morgan, S. 1882. *Memoir of Augustus De Morgan*. London.

Eichhorn, J. G. 1812. *Historische-kritische Einleitung in das Testament*. Jena.

Ephratt, Michael. 'Authorship attribution – the case of lexical innovations'. Online at www.cs.queensu.ca/achallc97/papers/p006.html on 02/04/2002.

Farringdon, J. M. 1996. with contributions by A. Q. Morton, M. G. Farringdon and M. D. Baker. *Analysing for Authorship. A Guide to the Cusum Technique.* Cardiff: University of Wales Press.

Forsyth, R. S. and D. I. Holmes. 1996. 'Feature-finding for test classification'. *Literary and Linguistic Computing*, XI, 163–74.

Gibbons, J. 2004. (ed). *Language and the Law*, London: Longman.

Glatt, B. and E. H., Haertel. 'The Use of the Cloze Testing Procedure for Detecting Plagiarism'. *Journal of Experimental Education*, L, 127–36.

Goddard, C. 1996. 'Can linguists help judges know what they mean? Linguistic semantics in the court-room'. *International Journal of Speech, Language and the Law* (formerly *Forensic Linguistics*), III (2), 250–72.

Goffman, E. 1981. *Forms of Talk*. Philadelphia: University of Pennsylvania Press.

Goutsos, D. 1995. 'Review article: Review of McMenamin, G. 1993 *Forensic Stylistics*, *International Journal of Speech, Language and the Law* (formerly *Forensic Linguistics*), II (1), 99–113.

Grant, T. and K. Baker 2001. 'Reliable, valid markers of authorship'. *International Journal of Speech, Language and the Law* (formerly *Forensic Linguistics*), VIII (1), 66–79.

Greenberg, J. 1963. (ed.) *Universals of Language*. Cambridge: MIT Press.

Hänlein, Heike 1998. *Studies in Authorship Recognition – A Corpus-Based Approach*. Frankfurt: Peter Lang.

Harrison, G. B. 1941. *A Jacobean Journal 1603–1606*. London: George Routledge and Sons.

Herbert, R. and W. Best 2005. 'Therapy for word-finding in aphasia: working at the syntactic level', in Clinical Aphasiology Conference: Clinical Aphasiology Conference (2005: 35th: Sanibel Island, FL : May 31–June 4).

Heselwood, B. 'Speaker Identification by Earwitness: A bigger picture'. Online at www.ling.ed.ac.uk/~anthonyb/undergrad_diss.pdf on 05/08/2002 at 02:56 UTC.

Hill, O. 'Who wrote Shakespeare? The authorship controversy'. Online at www.urbana.k12.oh.us/699/oh/authorship%20controversy.html on 06/04/2002.

Hjelmquist, E. 1984. 'Memory for conversations'. *Discourse Processes*, VII, 321–36.

Hollien, H. 2002. *Forensic Voice Identification*. New York: Harcourt.

Holmes, D. I. 1985. 'The analysis of literary style – a review'. *Journal of the Royal Statistical Society*. Series A (General), CXLVIII (4) (1985), 328–41.

Holmes, D. I. (1992). 'A stylometric analysis of Mormon scripture and related texts'. *Journal of the Royal Statistical Society*, Series A, CLV, 91–120.

Holmes, N. 1887. *The Authorship of Shakespeare*. Boston and New York: Houghton, Mifflin and Company.

Hoover, D. L. 2001. 'Statistical stylistics and authorship attribution: an empirical investigation', *Literary and Linguistic Computing*, XIV (4), 421–44.

Horton, Thomas Bolton 1987. 'The Effectiveness of the Stylometry of Function Words in Discriminating between Shakespeare and Fletcher' (Ph.D. dissertation; Department of Computer Science, University of Edinburgh, December).

Ingram, J. C. L., R. Prandolini and S. Ong. 1996. 'Formant trajectories as indices of phonetic variation for speaker identification'. *International Journal of Speech, Language and the Law* (formerly *Forensic Linguistics*), III (1), 129–45.

Jiang, M. 1996. 'Fundamental frequency vector for a speaker identification system'. *International Journal of Speech, Language and the Law* (formerly *Forensic Linguistics*), III (1), 95–106.

Johnson, D. E. 'Online phonetics course'. Online at http://www.unil.ch/ling/phonetique/api1-eng.html on 04/07/2002 at 03.45.

Joseph, A. 1995, *We Get Confessions*. New York: AJ Books.

Joshi, A. K., E. Prince and M. A. Walker 1998. *Centering Theory in Discourse.* Oxford: Clarendon Press.

Kenny, A. 1982. *The Computation of Style: An Introduction to Statistics for Students of Literature and Humanities*. Oxford: Pergamon Press.

Kerlinger, F. 1973. *Foundations of Behavioral Research*. New York: Holt, Reinhart & Winston.

King, G., R. Keohane and S. Verba 1994. *Designing Social Inquiry*. Princeton: Princeton University Press.

Kisilevsky, B. S., Sylvia M. J. Hains, Kang Lee, Xing Xie, Hefeng Huang, Hai Hui Ye, Ke Zhang, Zengping Wang 2003. 'Effects of experience on fetal voice recognition'. *Psychological Science*, XIV (3), 220–4.

Kniffka, H. 1981. 'Der Linguist als Gutachter Bei Gericht. Überlegungen und Materialien zu einer "Angewandte Soziolinguistik"', in G. Peuser and S. Winter (eds), *Angewandte Sprachwißenschaft: Grundfragen – Berieche – Methoden*. Bonn: Bouvier Verlag H. Grundmann, pp. 584–633.

Kniffka, H. 1990. *Texte zu Theorie und Praxis forensischer Linguistik*. Tubingen: Max Niemeyer Verlag.

Künzel, H. J. 2000. 'Effects of voice disguise on speaking fundamental frequency'. *International Journal of Speech, Language and the Law* (formerly *Forensic Linguistics*), VII (2), 149–79.

Ladefoged, P. 1971. *Preliminaries to Linguistic Phonetics*. Chicago: University of Chicago Press.

—-2001. *A Course in Phonetics*, 4th edn. New York: Harcourt Brace.

Lancashire, I. 1998. Paradigms of Authorship. *Shakespeare Studies*, XXVI, 298–301.

Laver, J. 1991. *The Gift of Speech*. Edinburgh: Edinburgh University Press.

Leech, Geoffrey and Michael H. Short 1981. *Style in Fiction. A Linguistic Introduction to English Fictional Prose*. London, New York: Longman [cited in Hänlein (30)].

Levi, J. 1994. 'Language as evidence: the linguist as expert witness in North American courts', *International Journal of Speech, Language and the Law* (formerly *Forensic Linguistics*), I (1), 1–26.

Lewin, Beverly A., Fine J. and L. Young 2001. *Expository Discourse*. London and New York: Continuum.

Lightfoot, D. 1979. *Principles of Diachronic Syntax*. Cambridge: Cambridge University Press.

Local, J. and G. Walker 2005. 'Methodological imperatives for investigating the phonetic organisation and phonological structures of spontaneous speech'. *Phonetica*, LXII, 120–30.

Loftus, E. F. 1975. 'Leading questions and the eyewitness report'. *Cognitive Psychology*, VII, 560–72.

MacWhinney, B., Keenan, J. M. and P. Reinke (1982). 'The role of arousal in memory for conversation'. *Memory & Cognition*, X (4), 308–17.

Malpass, R. S. and P. G. Devine 1981. 'Eyewitness identification: lineup instructions and the absence of the offender'. *Journal of Applied Psychology*, LXVI, 482–9.

McGehee, F. 1937. 'The reliability of the identification of the human voice'. *Journal of General Psychology*, XVII, 249–71.

McMenamin, G. 1993. *Forensic Stylistics*. Amsterdam: Elsevier.

—-2002. *Forensic Linguistics: Advances in Forensic Stylistics*. Boca Raton and New York: CRS Press.

Mendenhall, T. C. 1887. 'The Characteristic Curves of Composition', *Science*, IX, 237–49.

Moats, L. 2001. 'Overcoming the Language Gap'. *American Educator*, Summer.

Moravcsik, Edith, and Jessica Wirth 1986. 'Markedness – An Overview', in F. Eckman, E. Moravcsik, and J. Wirth (eds), *Markedness*. New York: Plenum, pp. 1–11.

Morton, A. Q. and S. Michaelson 1990. *The Qsum Plot*. Internal Report CSR-3-90, Department of Computer Science, University of Edinburgh.

Mosteller, F. and D. L. Wallace 1964. *Inference and Disputed Authorship: The Federalist*. Reading, MA: Addison-Wesley.

Nilo, J., G. Binongo and M. W. A. Smith 1999. 'The application of principal component analysis to stylometry'. *Literary and Linguistic Computing*, XIV (4), 445–65.

Olsson, J. G. 1997. 'The alteration and dictation of text'. *International Journal of Speech, Language and the Law* (formerly *Forensic Linguistics*), IV (2), 226–51.

—2000. 'Some aspects of dictation and text analysis: A study of language configuration and the structure and proportionality of text'. Unpublished M.Phil dissertation. University of Birmingham, England.

Owen, C. 1995. 'Review article: "Language and the Law"'. *International Journal of Speech, Language and the Law* (formerly *Forensic Linguistics*), II (2), 168–87.

Pappas, T. 1998. *Plagiarism and the Culture War: The Writings of Martin Luther King Jr. and Other Prominent Americans*. Tampa, FL: Hallberg.

Pennebaker, J. W. 1999. 'Psychological factors influencing the reporting of physical symptoms', in A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman and V. S. Cain (eds), *The Science of Self-report: Implications for Research and Practice*. Mahwah, NJ: Erlbaum Publishers, pp. 299–316.

— 'Natural language use and personality'. Online at www.psy.utexas.edu/psy/undergrad/courses/fall01/pennebaker2.html on 30/05/2002 at 15:39h UTC.

Radford, Andrew 1990. *Syntactic Theory and the Acquisition of English Syntax*. Oxford: Blackwell.

Rodman, R., D. McAllister, D. Bitzer, L. Cepeda P. Abbitt. 'Speaker Identification Based on Spectral Moments'. Online at: www.csc.ncsu.edu/faculty/rodman/Speaker%20identification%20using%20isophonemic%20sequences.pdf on 04/08/2002 at 18:06 UTC.

Rogers, M. L. 1990. 'Coping with alleged false sexual molestation: examination and statement procedures'. *IPT Online Journal*, II (2). Online at www.ipt-forensics.com/journal/volume2/j2_2_1.htm on 24/03/2002.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. and P. Boyes-Barem 1976. Basic objects in natural categories. *Cognitive Psychology*, VIII, 382–439.

Sandblom, G. 'Peter Ramus, I Serien Om Historiska Retoriker'. Online at www.utv.mh.se/~gorans/ramus.html on 04/01/2002.

Sanford, T., J. Aked, L. Moxey and J. Mullin 1994. 'A critical examination of assumptions underlying the cusum technique of forensic linguistics'. *International Journal of Speech, Language and the Law* (formerly *Forensic Linguistics*), I (2), 151–67.

Särndal, C. E. 1967. 'On Deciding Cases of Disputed Authorship'. *Applied Statistics*, XVI (3), 251–68.

Scherer, R. 'A basic overview of voice production'. Online at www.voicefoundation.org/VFScherervoiceprod.html on 25/08/2002 at 15:04 UTC.

Schils, E. and P. de Haan 1995. 'Mortons Qsum plot, exposed'. Online at http://lands.let.kun.nl/literature/dehaan.1995.2.html on 30/12/2001.

Schneidman, E. S. and N. L. Farberow 1957. *Clues to Suicide*. New York: McGraw-Hill.

Shuy, Roger W. 1997. 'Ten unanswered language questions about Miranda'. *International Journal of Speech, Language and the Law* (formerly *Forensic Linguistics*), IV (2), 175–96.

Somers, H. 'An attempt to use weighted cusums to identify sublanguages'. Online at http://ucrel.lancs.ac.uk/acl/w/w98/w98-1216.pdf

Stamatos E., N. Fakotakis and G. Kokkinakis 2000. 'Automatic text categorization in terms of genre and author'. *Computational Linguistics*, XXVI (4), 471–95.

Steller, M. and G. Koehnken. 1989. 'Criteria-based statement analysis', in D. C. Raskin (ed.), *Psychological Methods in Criminal Investigation and Evidence*. New York: Springer, pp. 217–46.

Taylor, W. L. (1953). 'Cloze procedure: A new tool for measuring readability'. *Journalism Quarterly*, XXX, 415–33.

Theobald, R. 1901. *Shakespeare Studies in Baconian Light*. London: Sampson Low & Marston.

Trankell, A. 1958. 'Was Lars sexually assaulted? A study in the reliability of witnesses and of experts'. *The Journal of Abnormal and Social Psychology*, LVI, 385–95.

Triandafilou, R., M. McCullough and M. Eslea 1998. 'Using criteria based content analysis to identify true and false allegations of bullying'. Symposium paper, British Psychological Society Developmental Section, Annual Conference 1998, University of Lancaster. Online at www.uclan.ac.uk/facs/science/psychol/bully/triand.htm on 25/03/2002.

—1944. *The Statistical Study of Literary Vocabulary*. Cambridge: Cambridge University Press.

Undeutsch, U. 1967. 'Beurteilung der Glaubhaftigkeit von Aussagen', in U. Undeutsch (Hrsg.), *Handbuch Forensische Psychologie*. Göttingen: Hogrefe, XIV (4), 445–65.

van Lancker, D. and J. Kreiman 1987. 'Voice discrimination and recognition are separate abilities'. *Neuropsychologia*, XXV, 829–34.

Williams, P. J. [online review] 'Riad Aziz Kassis: *The Book of Proverbs & Arabic Proverbial Works*'. Online at www.abdn.ac.uk/divinity/williams/ReviewofKassis.htm

Wright, Laura (ed.) 2000. *The Development of Standard English, 1300–1800: Theories, Descriptions, Conflicts*. Cambridge: Cambridge University Press.

Yegnanarayana, B., K. Sharat Reddy and S. P. Kishore. 'Source and system features for speaker recognition using AANN models'. Online at www.ee.columbia.edu/~dpwe/papers/YegRK01-aannspkrid.pdf on 08/07/2002 at 05.15.

Yule, G. 1944. *The Statistical Study of Literary Vocabulary*. Cambridge: Cambridge University Press.

Yule, G. 1938. 'On Sentence-Length as a Statistical Characteristic of Style in Prose, with Application to Two Cases of Disputed Authorship'. *Biometrika*, XXX, 363–90.

Zwicky, Arnold 1978. 'On Markedness in Morphology', *Die Sprache*, XXIV, 129–43.

Index

Page numbers in *italics* indicate sample forensic texts in Appendix 1