

10

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 many organizations 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 chapter 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 the United Kingdom (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 believed 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 nearly 60 times more common than the former.¹

Using a professional language corpus (Cobuild) I discovered that ‘information’ is in fact the most frequent collocater 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, namely, ‘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, for example, ‘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 the 16 October is written by the chief executive, scarcely more than a week later.

Mr Anthony also requested a copy of the video tape of the ‘Local Resolution Meeting’ because he felt this tape contained important information about his treatment. Referring to the video tape 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’. As with ‘absent on leave’, which we discussed earlier, there is some redundancy in the phrase ‘wiped and reused’. Why would the tapes be ‘wiped’ if they were going to be ‘reused’?

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 of an interesting area of linguistics referred to as *Grice’s Conversational Maxims*. This curious sounding theory considers the issue of ‘how much’ information speakers/writers give and what the quality of that information is. At an informal level it seems to me, regarding the tapes, that the Chief Executive is protesting a little too much (quantity) while the ‘quality’ aspect also seems poor: ‘Once this material has been completed then practice is for the tapes to be wiped and reused. Regrettably, we are unable to provide you with copies of the tapes’. Here I would have expected the Chief Executive to begin with his regret, followed by the explanation, for example ‘Unfortunately we cannot provide you with copies of the tapes because they have been re-used, as is our practice’. Informally, it seems to me that the Chief Executive is weighing his words with a little too much 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 some 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. These strategies are tabulated in Table 10.1.

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

Table 10.1. 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 <i>himself</i> .
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.

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 as institutional abuse.

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.

Note

1. Internet search on Google on 27 May 2008 showed: 'disclose a record' 30,100; 'disclose information' 1,720,000.

Reference

- Grice, P. (1975). 'Logic and Conversation'. In Cole and Morgan (1975), pp. 41–58.

14

FITTED UP BY A 'PROFESSIONAL': FALSELY ACCUSED

Bill Johnson was a Midwest businessman who once enjoyed nothing more than getting into his privately owned aeroplane and flying off to Mexico to visit his holiday home. He used to be a member of a certain flying club in the Midwest, often assisting other members by flying them on pleasure trips, advising them about aircraft purchases and even helping one or two of them to build their own hangars on the grounds of the airport. Flying clubs, just like fishing clubs, chess clubs or car clubs can, however, be hotbeds of jealousy and rivalry, as groups of members vie with each other for positions on committees, better parking spots for their beloved machines, or the favour of the chairman. Bill Johnson had long since decided to keep clear of the internal politics of the club – he was just there to fly. As a highly successful contractor, with clients all over the United States he was in the fortunate position of not needing to jostle for position at the club. Mostly he flew during the weekends.

At one point in time he had been offered a plot of land on the grounds of the airport with a 25-year lease. He bought the lease and built a state of the art hangar on it. Imagine his horror when he arrived at the airport one Sunday morning to discover that his aircraft and hangar had been impounded by the owners of the flying club. He was not even allowed into the hangar to collect his personal belongings. The club's reason for this extreme step was that Johnson had been secretly authoring an internet journal which the management viewed as detrimental to their organization. They showed Johnson the internet journal they claimed he had been editing. I will refer to it as the X document. It contained articles mildly critical of the management of the flying club, but could hardly be said to be detrimental. In fact it was no more than a light-hearted, if on occasion semi-literate, look at the foibles of certain members, the vanities and ambitions of others, and the general muddle that occurs whenever you put more than ten people on a patch of earth

and expect them to get along with each other. Johnson was offended because of its somewhat poor grammar, spelling and punctuation.

Protesting his innocence, Bill Johnson was told he had 'no chance'. Several club heavies escorted him to the perimeter fence and told him that not only were they going to keep his aircraft and hangar, but they were also going to sue him for every penny. They smiled gleefully as they began to describe different properties he owned, which they seemed to know quite a lot about. Growing more suspicious by the minute as to their real motives, Johnson immediately contacted his attorney to issue counter-proceedings. Realizing that they would have a fight on their hands, the club then invited a certain professor to assess whether Mr Johnson was the likely author of X. The report was duly completed and I was contacted by Mr Johnson's attorneys to see whether 'anything could be done'.

First, let me say that the professor in question is quite well known within his field, and I believe quite popular on his own campus. However, he did not appear to have a very detailed understanding of either forensic linguistics or authorship attribution. I could find no references to any publications of his on forensic linguistic matters or on authorship attribution, and to my knowledge he is not, nor ever has been, a member of the International Association of Forensic Linguists, which represents the academic interests of the field, publishes its own peer-reviewed journal and regularly holds international conferences.

This lack of experience began to show at an early stage. For example, the professor's method of text selection seemed flawed, and I thought he was making some quite wild assumptions about the way words distribute in the language. The way he used statistics was also questionable. However, what would the courts think? That was the issue.

I also noticed that Professor Willerby (not his real name), seemed to lack objectivity in writing his report. For instance, he stated that his task was to 'provide a report on the linguistic similarities between a set of texts (letters and emails) written by Mr. Bill Johnson and a different set of texts taken from X . . .' It seemed that the report writer had already made up his mind that there would be significant similarities and was just looking for evidence to support this view. Investigators into such issues should be completely impartial, and should state their terms in much more neutral language, for example, ' . . . to provide a report on any possible linguistic similarity between x and y'. I also noted Willerby did not indicate that his report was being written for the benefit of any possible court or other trier of fact. Forensic linguists do not work for a client, even though they have to be commissioned by somebody – they work to assist the court.

In his preamble Willerby stated that

[the] field of forensic linguistics that focuses on such issues is known as Authorship Attribution. This field has a long history, known to the general public mainly through cases of contested authorship involving famous writers such as William Shakespeare. However, these methods have also been used in forensic studies such as the identification of the Unabomber based on a comparison between his personal notebook and his widely distributed manifesto.

There are several disturbing points about the above quotation if it is the view of an expert. The first is that forensic linguistics as a field is relatively new, and does not have a 'long history'. The term 'forensic linguistics' was first coined by Jan Svartvik in 1968. I once asked him how he had come up with the phrase and he said he used to watch a television series in the 1960s called *Quincy*. The word 'forensics' was always being mentioned and he thought it could be applied quite successfully to 'linguistics' as an adjective if the final 's' in forensics were dropped. It was in fact the title of the first ever paper on the subject.¹ However, the term was not in general use until about 1994 when the first academic journal with these words in its title was published. Moreover, authorship attribution is not exclusive to forensic linguistics. It is also common to literary poetics, where it has a much longer tradition. The authorship of Shakespeare is not a forensic matter: it is a populist dispute, with very little academic interest in that dispute (although there is some interest in the authorship issue), and, lastly, canonical literary authorship is an area of academic study which has no legal dimension whatsoever, and hence cannot be described as 'forensic linguistics'. Even Willerby's comments about the Unabomber are incorrect: he was initially identified on the basis of stylistic factors found in his letters to his brother and sister-in-law and not, as Willerby claimed, on the basis of his notebooks, which were only found after his arrest. I happen to have this information first hand from the distinguished FBI special agent who worked on the case, Jim Fitzgerald.

Talking about authorship attribution Willerby claims there are three procedures, which he cites as (i) informal analysis of linguistic idiosyncrasies, (ii) statistical analysis of the distribution of grammatical function words and (iii) the training of neural networks using artificial intelligence techniques.

Actually, Willerby appears to be confusing his terminology here. He begins by saying there are three main types of procedure. He then refers

to 'general method', which he subsequently calls a 'procedure', then 'method' again, finally settling on 'technique'. This confusion of terminology leads me to believe that Willerby does not understand the difference between methods, procedures and techniques. He has not even discussed *approaches*.

It seems proper to begin with the idea of an *approach* to a problem. The most basic approach is a subjective assessment of idiosyncratic features, that is to say features which are or appear to be peculiar to a given author. This approach has its uses, but some consider it to be 'unscientific' because it cannot be easily quantified. The alternative approach is a quantitative one: any one of a number of quantitative techniques can be applied to any one of a number of linguistic features. The three basic areas of quantitative approach concern the measurement and statistical analysis of (i) the lexis (or vocabulary) of a text, (ii) grammatical (or function) words² and (iii) language modelling, for example by the use of compression algorithms, neural networks and so on. For each of these three approaches there are many possible methods of measurement and analysis.

Referring to his second 'method' Willerby says: 'The second general method involves statistical analysis of the distribution of "context free" words'. He then cites two authors, but although bibliographic references are given at the end, no actual page numbers are recorded. In a report, as with most academic work, it is essential to provide readers with the exact location of reference material. It is not sufficient to provide the name of a reference work, for the simple reason that others will not have all the information necessary to evaluate the report author's claims.

Willerby then claims, with reference to his second 'method', that 'this procedure has gained wide acceptance in the field', but I would dispute this. No single procedure has gained wide acceptance in the field. And, in any case, which field? If he means the field of forensic linguistics, this is simply not true: no linguist that I am aware of has presented authorship attribution to a court based on function word counts. If he means the field of computational linguistics, which has a strong interest in authorship questions, but is not in a position to make any input into forensic linguistics, that may be so. However, it should be noted that computational linguists usually work with very long texts, such as novels, long essays and other lengthy works. In forensic linguistics the reality is that most texts which are analysed are very short. It is not uncommon to undertake an inquiry with only three texts, each no longer than a few hundred words, and some even less than 50 words. For this reason forensic linguistic techniques tend to be somewhat different from those found in literary attribution work. In any case I was not aware

of any computational linguist having given expert evidence in relation to a forensic authorship attribution claim.

In the context of expert reports it sometimes happens that authors will make a bold claim which cannot stand up to close cross-examination. Thus Willerby writes: 'Recently it has been argued that a complete syntactic analysis of both texts is a more direct method of achieving the same goal, although it is far more labour intensive and was not possible in the time allocated to the present analyses'. This seems to imply that Willerby has the necessary knowledge and software to undertake such a task and, most crucially, has experience at doing so. As far as I know, few analysts would ever attempt a 'complete' analysis on a 20,000 word document or set of documents, and in any case, carrying out syntactic analyses is a highly specialized task of which few linguists are capable.

Willerby also says: '[I]t is generally accepted that the frequencies of a large number of these types of words form a kind of fingerprint for each writer'. However, although this may be 'generally accepted' it is not accepted by linguists who have had anything to do with authorship attribution. In fact it is contrary to much of what we understand about how language is acquired and used, and how it develops and atrophies throughout the human life cycle. As with his other claims Willerby gave no authorities for his claim that this notion is 'generally accepted'.

At one point in his analysis Willerby refers to a notorious authorship method which was rejected by the courts more than 10 years ago. It was known as the Cusum method and was developed by a vicar and a computer scientist. It was roundly attacked by linguists and psychologists alike. Under one of his headings: 'Function word analysis' Willerby claimed the Cusum technique had been rejected because 'the set of 2- and 3-letter words, plus words beginning with a vowel is not a natural linguistic category'. In fact, this is far from being the main reason the Cusum test was rejected. It was primarily rejected because it violates several basic scientific principles and is utterly ignorant of a number of elemental tenets of linguistics. I would have expected Willerby, given his position at a major educational institution, to know this. It also surprised me that an 'expert' would write a report on authorship attribution and then talk about a method that has long since been rejected by the entire forensic linguistic community.

Later, while still on the topic of function word analysis he says: 'In the early stages of authorship attribution research, it was thought that mean sentence length would be of value'. The irony here is that he is talking about work carried out by mathematicians a hundred years ago, beginning with Augustus de Morgan, TC Mendenhall and later Udney Yule. None of these mathematicians – distinguished though they were – had

any connection with linguistics or even a serious interest in studying language. Therefore, to describe their work as 'the early phases of authorship attribution research' borders on the inventive. It was also puzzling to me as to why, once again, Willerby introduced a method, and then dismissed it by saying that it was not appropriate to the task of authorship attribution.

Still under the heading of 'function word analysis', Willerby then mentioned a measure known as type-token ratio. The type-token ratio is one of a number of ways linguists measure the richness of vocabulary in a specified length of text. The total length is the number of tokens, and the types are the total number of different words in the text being measured. However, type-token ratio is rarely a test undertaken of function words, mainly because it is used as a measure of a text's lexical richness. For this reason, I was surprised he mentioned it in the context of function (i.e. grammatical) words. More generally, it seems curious – once again – that he is describing a measure that has, as he admits, very little utility in authorship analyses.

Willerby says the Johnson texts and the X Material were supplied in hard copy format. Considering that the X Material was readily available on the internet, it seemed surprising that he then took printed text and scanned it using OCR techniques and then used various proofreading processes to ensure its accuracy. It seems somewhat strange for any analyst to imagine that he/she should transcribe an electronic document by first printing it, then scanning it and then proofreading it and in this way produce a version whose accuracy would be as good or better than the original electronic text. Willerby claims that 'each page was proofread by the OCR assistant'. He does not state whether this OCR assistant is a person or part of the software. OCR assistants exist in a number of software packages and generally 'help' with such matters as layout, text organization and so on. He says that a 'spell checker was used to correct errors if they did not appear in the original text'. I am not sure what this means. Does he mean that the spell checker used a standard spelling dictionary to remove errors, or does he mean that the spell checker was used to make sure that original errors were retained? I am not sure how a spell checker would do this: rather, this should have been carried out by Willerby himself.

The above points also apply to the known Bill Johnson texts. These were mostly email texts written by Mr Johnson at an earlier stage. If these were already electronic documents, why print them out and then scan and proofread them? Some earlier OCR (which stands for Optical Character Recognition) scanning packages were notorious for being

processor intensive and for producing multiple errors, especially from documents such as emails. Incidentally, the word 'error' has to be used advisedly here. An error in this context does not mean a grammatical, orthographic, spelling or punctuation error. Rather, it refers to whether the scanning process produces text which is *different* from the source text. If the scanner faithfully re-produces an item of text which is itself an error, then – in the forensic context this is not an error. It is correct. On the other hand, if the scanner 'corrects' an error, through – for example – its inherent spell-checking device – then this is an error, even though what is produced is grammatically, orthographically, and in other ways correct. The usual purpose of a spell checker in OCR scanning software, is to correct the spelling mistakes which the software finds: however, in the forensic context it is important *not* to 'correct' anything, since a forensic text must always be presented as is.

I very much doubt that anyone is able to proofread 20,000 words to such a high standard that it would be, in the context of scanning, error-free, especially if working under time constraints (which Willerby hints at elsewhere in his report). Even the most careful author will produce errors of various kinds, and if the OCR scanning process did not identify these, or misrepresented them in some way, then Willerby's version of the X Material was quite possibly not an accurate representation of the original source material. The same may apply to Mr Johnson's own texts. Willerby did not provide copies of the texts he used to compile his report. Johnson requested sight of the emails he wrote which were used as the texts to be tested against – which are termed 'exemplar texts', but Willerby refused this request. This is a most unusual procedure, given that it means that at this time the test documents in this case cannot be verified either as to provenance or accuracy.

Willerby's next heading was 'Selection of samples'. As far as I could tell the only methods he had so far described also happened to be ones he rejected. He had still not stated what methods he intended to use to carry out his 'attribution'. Moreover, in forensic work, the questioned texts are never described as *samples*. They are the obligatory set of texts, since the whole object of the exercise is to discover the identity of their author or, more scientifically, to discover what basis one would have for rejecting or failing to reject one or other candidate author.

But the most serious linguistic weakness in Willerby's report was yet to come. It was to do with the text types of the exemplar texts. Willerby said that these were all examples of 'expository prose'. Expository prose is simply a technical term for writing which is intended to explain something – usually technical. Some of Willerby's non-suspect exemplars do

indeed fall into the category of 'expository prose': for example he includes in his corpus of non-suspect exemplar texts an undergraduate essay and a portion of a doctoral thesis in linguistics (his own perhaps?). However, we may question whether Bill Johnson's emails – even if on fairly formal topics – would entirely qualify as 'expository prose', since texts written in asynchronous electronic media³ are usually much less structured than more formal prose, and are likely to be informal in addressivity and tone. Moreover, we may even question whether one of the texts included in Willerby's corpus, a lecture given to undergraduates which he himself wrote, qualifies as prose in the traditional sense of the word, since it was written to be spoken or read aloud to a group, rather than written to be read by individual readers. I wondered why he had chosen 'expository prose' as his criterion of classification. The X Material was definitely not expository prose. If anything, it was mild political satire.

I was also somewhat surprised he had used his own work in an authorship investigation. In nearly 15 years of forensic linguistic work I had never heard of anyone doing this. In the context of forensic work, this is completely unprofessional, and seems to demonstrate inexperience in the forensic field. A comparable example would be the election researcher who, in order to increment the apparent support among a population of sample voters for a particular political party, records his/her own voting intentions as part of the election poll he/she is researching. We would certainly consider that such a procedure on the part of that researcher would have a prejudicial effect on the poll results. I doubt very much whether any academic researcher would ever include his/her own work in a corpus to be studied, unless very particular circumstances applied, which I do not believe to be the case here.

So, what we are seeing here is that there is a complete mix of text and genre types, leading to a somewhat mixed register, which I will describe in more detail below. Before I go on to talk about register, however, it is worth noting the diversity of text types, as well as the genre mix, in Willerby's corpus of sample texts. As regards text type, we have lectures, emails, an essay and a thesis. In most authorship attribution analyses I have carried out or studied it is usual to ensure homogeneity of text type among exemplar texts, unless there are practical reasons for not doing so, for example if the questioned text is of a particularly rare type. This does certainly not apply in the present instance. However, I can understand that Willerby used Mr Johnson's emails simply because he had no other exemplar texts of Mr Johnson's writings available – but in that case he should have confined himself to emails or, at worst, just used emails and web journal texts for comparison purposes. The genre mix is also worth noting: we have

academic texts combined with texts relating to committee administration, and these in turn are being compared with anonymous items critical to the management of a flying club. This can scarcely be described as an homogenous corpus with regard to genre. Finally, Willerby did not appear to take on board the notion that he was dealing, not just with texts whose authorship was unknown, but with anonymous or concealed authorship. Willerby does not appear to consider the possibility that anonymous authors may attempt to disguise their work, just as anonymous phone callers will try to disguise their voice. There is no doubt that authorial attempts at disguise can affect the outcome of any testing procedure.

To come back to *register*, then, it is a further term that we need to discuss to understand where Willerby was going wrong. The concept of register concerns such questions as how a text is produced, to whom it is addressed, whether it is formal in its structure and content and so on. Register is considered to consist of three sub-areas:

1. *Mode* refers to language channel and concerns such issues as whether a text was produced by means of writing, speaking or dictating; whether it was produced in one form but delivered in another (e.g. news bulletins written to be spoken, lectures etc); whether a visual or interactive module was delivered concurrently with the language (e.g. photographs, cartoons in a newspaper, or charts and tables in an academic journal). Hence, when comparing texts for authorship, or any other linguistic analysis, it is important to have uniformity of *language mode*. There has been a considerable amount of research on asynchronous electronic media, such as emails and cell phone texts, in recent years, and these are considered to have many features in common with spoken language. As a result, I do not believe we can safely say that a useful comparison can be made between emails and dissertations, or emails and web journal text, for example. The fact that one of the exemplar texts is a lecture which was written to be spoken aloud to an audience, while another is an email addressed to a colleague, while yet another is an undergraduate essay, shows that Willerby does not appear to have taken register factors seriously which, given his claimed linguistic background, is somewhat surprising. He appears to be asserting that all of his sample texts are 'expository prose' and that this is sufficiently narrow a classification to ensure some kind of corpus homogeneity. I think any linguist would contest this point.
2. *Tenor* is a sub-area of register which is concerned with the relationship between the text producer and the recipient, that is, writer-reader,

speaker-hearer, lecturer-audience and so on. In the previous section I described how the mix of language modalities would mitigate against useful comparisons being possible between the exemplar texts and the X Material, and I maintain that the same point applies to issues of linguistic *tenor*. Thus, a student writing an essay addresses that essay to his/her professor for grading. The writer of an email may or may not know his/her addressee, but the writer-reader relationship in this instance is not necessarily going to be directly comparable to the student-professor relationship which is itself an apparent inverse of the professor-student audience relationship applicable to the lecture environment (given that Willerby includes a lecture in his exemplar corpus). For this reason, in authorship attribution analysis it is usual to choose exemplars which match the questioned texts with regard to *tenor* issues.

3. *Field* is a critical register issue. It concerns, broadly, the topic area of a text or set of texts and, what the language of the text is designed to do (in the performative sense). The field properties of the exemplar texts chosen by Willerby are as varied as are their mode and tenor properties. For example, we have advanced academic prose about linguistics (in the graduate thesis), emails about committees and administration, a lecture on linguistics and an undergraduate essay on an unspecified topic. The suspect candidate's exemplar texts consist of emails to his flying club about aviation and club matters. This is a fairly close match with the questioned texts, the X Material, but no match at all with the non-suspect exemplar texts mentioned above: however, the match only extends to *field* not to text type, and as previously commented on, the variety of text type is likely to mitigate against a useful authorship attribution comparison being possible. By constructing his corpus in this way, Willerby has biased his test procedure in favour of choosing Johnson as the most likely author.
4. *Other sociolinguistic issues*: We do not know the respective ages and gender/s of most of the candidate authors, nor their level of education. More critically, no mention is made in Willerby's report of any attempt to match sociolinguistic criteria. It is the usual practice, in authorship analysis, where a likely pool of candidates is not available alongside a suspect candidate (e.g. other personnel at a company or organization where the chief suspect candidate is also employed), to then attempt to match the candidate pool to the suspect, just as one would do in an identification parade. For example, we could not easily imagine an identification parade with

a tall, bald man of middle age, a short, fat hairy, young one, a young woman, an elderly person of either gender and so on. Further, in an ideal investigation we would probably want to ensure that our candidates came from the same geographical area, that they followed fairly similar occupations at similar levels, and that all of the texts were written within a fairly narrow timespan. Willerby makes no mention of any of these points.

In his report Willerby appears to accept implicitly that the authorship of the X texts is single in character. He does exclude 'letters to the editor' and other clearly non-candidate aspects of the texts, but he fails to mention the possibility that there might be more than one author of the X texts. However, we should bear in mind that it is not uncommon for publications of all types to attribute authorship to one individual, yet have several editors or contributors who are not specifically named. Moreover, it is also not unusual, in journals of all kinds, for two or more authors to collaborate in the writing of one text and for texts to be edited for a particular house style. Therefore, a lay reader of Willerby's report might think that it is safe to assume that the authorship of the text is single, and this could act on reader sensibilities in a prejudicial way since the possibility of any multiple-author configuration has not been considered.

In summary, Willerby's work did not constitute a professional report of a standard expected of courts as an aid to triers of fact. There were several reasons for this:

- His text selection is flawed because he mixes his own work with that of others. He pays no regard to questions of genre, text type, mode, field or tenor. He disregards sociological issues such as age, gender and level of education. He altogether avoids questions of dual authorship and author disguise.
- He did not seem to be aware of the fact that function words are not wholly context free, and he seems to consider both mean sentence length and type-token ratio are related to questions of function word analysis. In addition, his method of text preparation was suspect. Why did he decide to scan in electronic texts? How accurate was the product?
- No theoretical input is given to support his claims: we do not know why he believes authors can have a linguistic fingerprint when, in fact, there are strong reasons to believe the opposite: for example, the realization that language is a socially acquired property not an

inherited one, that it is subject to influence and change, and that it is susceptible to macro-socially homogenizing influences such as education and the media.

In my professional view, Willerby's understanding of the topic and selection of his corpus rendered the entire report worthless. I will, nevertheless demonstrate, in the following sections two further areas of gross error, namely that his method of linguistic analysis is flawed, as is his method of statistical reporting. In order to explain my position I will now talk about the two main types of word we find in language. These are *lexical words* and *function words*. Lexical words are assumed to hold meaning, and this class includes nouns, adjectives, verbs and certain adverbs. Function words, on the other hand, do not contain independent meaning, but, effectively, carry the grammar of the language, and include such categories as prepositions, determiners and function adverbs. This is the traditional division of words into types or levels of semantic structure, as originally advanced by Henry Sweet as 'sense-units' (i.e. lexical words) and 'form-words' (i.e. function words), (Sweet, 1891: 22).

In his section 'Selection of samples' it is somewhat difficult to follow Willerby's procedure. Apparently he took the 50 most frequent words for each text set, and from these retained only those which were 'context-free'. He says he discarded 'context-dependent' words such as 'taxi-way'. I am not clear here whether he means that he discarded all lexical words, or only those lexical words which he deemed to be context-dependent. What was his basis for, or definition of context dependence? In any case, the whole issue of context-dependence in terms of function words needs to be challenged. We may think that a function word would not depend on context for its distribution, but this is not true of all of them. Some function words show the same distribution whatever the genre or text type, and hence are context free in terms of distribution. On the other hand, some have a different distribution, dependent on text type or genre, and so are not context-free.

If, then, the test corpus consists of different text types, it follows that the test procedure will skew the results of those frequency words and not paint an accurate picture when comparing the questioned with the known text. Below, I illustrate this point by giving some examples of function word frequencies for different text types based on some simple word-count software. In the first instance I looked for word frequencies in a medium similar to the questioned texts, namely newspaper articles. I then compared word frequencies in this medium with that of emails – a text type

Table 14.1. Distribution of 'the' and 'a/an' in a small corpus of news articles and email texts

Word	News	Email
the	0.074	0.044
a/an	0.027	0.026

Table 14.2. Distribution of selected singular/dual pronouns in a small corpus of news articles and email texts

Word	News	Email
I	0.00	0.04
you	0.00	0.02
he/she	0.01	0.01

used in Willerby's corpus. I decided to use just the most common function words, namely the definite and indefinite determiners. The results were as shown in Table 14.1. As can be seen from Table 14.1, although the 'a/an' distribution is similar across the two text types, distribution of 'the' across the two text types is significantly different. In fact there are sound linguistic reasons for this, since newspaper articles are most often about something other than first and second grammatical persons (i.e. 'I', 'you' etc), whereas emails are most often either about 'you' and 'I' or necessarily include 'you' and 'I' in order to carry out a conversation about something else. This is clearly demonstrated when we come to measure the same test corpus for selected personal pronouns (singular/dual) as seen in Table 14.2.

From Table 14.2 we see that the selected first and second personal pronouns are virtually non-existent in news articles, but of relatively high frequency in email texts. Hence, results from Tables 14.1 and 14.2 appear to indicate that if comparing texts of different types then we should *not* rely on frequency counts of some common function words for authorship purposes: the text type and genre influences are likely to skew the result. For example, if we took emails and news articles for the same writer, there is no reason to believe that we would not find that exemplars of each text type would follow the above pattern.

This is why we need to treat terms such as 'context-free' and 'context-dependent' with some caution. While function words may be context-free in theory, or less context-dependent than lexical words, it does not seem that the concept of 'context-dependence' is a useful one when we are undertaking such a precise task as authorship attribution.

Authorship attribution is a serious task, especially where people's reputations, liberty and even, on occasion, lives, are at stake and definitions which may be useful in academic discussions, are not necessarily useful when applied to precise and important tasks such as authorship attribution. Following presentation of my analysis to the court, the flying club withdrew their allegations against Mr Johnson and paid him substantial damages for the distress they had caused him. Nobody knows quite what the management of the club were thinking when they launched this action against one of their most loyal members. At one point Mr Johnson observed officials from the club photographing his young grandchildren playing in the garden from a parked car, which is equally inexplicable behaviour.⁴

As to Professor Willerby I have not heard of any further ventures of his into forensic linguistics, and while I wish him a long and successful career in his chosen field, I trust that that will remain the case until such time as he acquires the appropriate knowledge and experience to deal with issues which are, at the present time, outside of his customary domain.

Notes

1. Svartvik, Jan (1968). *The Evans Statements: A Case for Forensic Linguistics*. Gothenburg: Acta Universitatis Gothoburgensis.
2. Linguists consider that the lexicon (vocabulary) is divided into two main categories: lexical words and function words. A lexical word is a content word, such as 'happy', 'table', 'love' and so on. A function word helps to convey the grammar and has no intrinsic meaning. Thus, 'the' is a function word, as is 'of', 'any', 'into' and so on.
3. Asynchronous electronic media are messages to which the reply is delayed (hence 'asynchronous'). This would refer to emails, phone text messages, ICQ and chatroom messages and so on.
4. I am not suggesting for a moment that Prof. Willerby had anything to do with this.

Reference

- Sweet, H. (1891). *A New English Grammar*. Part II, Syntax. Oxford: Clarendon Press (Reference here is to the 1968 impression).

other side of the road was a driveway with large white rocks, and nearby was a dark wooden fence. A little further down was a white wooden fence. I hitched a ride back to the Circle K on Highway 17 near the Mazda dealership. I got my bag and my coat from behind the Circle K. I hitched another ride to the Montague Exit off I-26. I walked across to the on-ramp and caught another ride with someone in a short van and he took me 5 miles down the interstate to another exit, which had a Taco Bell, a Home Depot and a K-Mart. Then I caught another ride up I-26 for about 20 or 30 miles at least. I ended up at a BP Mart and went to the on-ramp and stayed in the woods overnight. Then on Thursday morning at about 8:30 a.m. I got arrested by a Dorchester County Sheriff's Deputy. I cooperated and offered no resistance. End of Statement.

Notes

1. It is standard practice throughout most jurisdictions to caution suspects before questioning them. The caution usually takes the form 'You are being arrested on suspicion of ---. You do not have to say anything, but anything you say may be taken down and used against you in a court of law. Do you understand?'
2. Olsson J. (2004). *Forensic Linguistics: An Introduction to Language, Crime and the Law*. Continuum.
3. Mr Reed elected to be executed by electric chair on 20 June 2008. He was therefore a 'volunteer'. In my view he was not competent to make this decision, given the linguistic evidence of his lack of mental powers, and a history of mental illness added to the inevitable deprivations of 12 years on death row. On the day before he was executed, a judge in Illinois ruled, in another case, that just because a defendant was competent to stand trial did not mean that defendant was competent to represent themselves, which is what Mr Reed had done after firing his lawyers at his trial in 1996. This ruling should have given Supreme Court judges pause for thought in Mr Reed's case, but it did not. An Australian friend, Charles Willock, and myself made representations to the governor right up to the last moment of Mr Reed's life, but we were unsuccessful. Our attempts at mobilizing local media in South Carolina also failed. None of the lawyers I spoke to believed that James Earl Reed was mentally competent to be executed or that he deserved to die. The private investigator who tirelessly fought for James' rights in the last years of his life, Phillip Upton, told me that the prison guards at the death row facility where he was incarcerated before being moved to the death house treated him with unfailing kindness and compassion, mindful of his mental state and his efforts to establish his innocence.

Part 3 16

BETRAYED BY A FULL STOP

Sandra Weddell was a popular mother of three young children. She was well liked in her community, where she belonged to a number of voluntary organizations. A highly qualified nurse, and a person with strong religious beliefs who was widely known for her kindness to others, it came as a shock to her local community to learn that she had, apparently, killed herself. But had she?

Consider the short letter below, the 'suicide' note said to have been left by Mrs Weddell. What is unusual about it? Normally I would not think of asking a reader to do this, but you might like to try transcribing this letter yourself before reading further. You may be surprised at the result.

Garry.

I am typing this note, because I know that if I were to hand write it and leave it for you, then I know that you wouldn't read it.

I am so sorry for all the hurt I have caused you Garry. I never meant to hurt you or to cause you so much pain.

I made a stupid mistake and I betrayed your trust, and I betrayed my family at the same time. I don't know what made me do what I did. I wish the whole thing had never happened. It all got out of hand. I have ended up with nothing.

You are kind to want to forgive me. I don't deserve your forgiveness.

When you think of me, just try and think of the happier times.

Sandra Jane Weddell

On the last day in January 2007 in suburban Bedfordshire, not far from London, Garry Weddell, a police inspector, knocked on his neighbour's door and asked him to help find his wife, Sandra. He told the neighbour his wife had been missing since the previous day. After a short time

Sandra Weddell was found dead in the garage of the family home. She had apparently died of asphyxia. A cable tie was found around her neck. Near to the victim's body was a single A4 printed sheet with the above note. On examining the national records on types of murders, police found that all previous deaths involving a cable tie had been murder – there were no suicides. This in itself was not necessarily conclusive, but when the circumstances of the discovery of the body were taken into account, police started to become suspicious.

The candidates for authorship of the 'suicide note' were Sandra Weddell herself and her husband, Garry Weddell. I had worked on a suspicious death case for Bedfordshire Police several years earlier and after the discovery of the body I was asked by the same force to look at the alleged suicide note.

Initially, I made some subjective observations on the text. As the reader will have seen, there is nothing very unusual about the language in the above letter, but observant readers who transcribed the text, as invited, will probably have noticed that after the opening salutation, which is simply 'Garry', there is a full stop. It is such a 'small' detail that several people did in fact fail to notice it. In any case, this full stop turned out to be highly significant, for a number of reasons, which I will go into later.

Another thing you may have noticed is that the writer's name is centred on the page, written as 'Sandra Jane Weddell'. Again, how many of us will have transcribed it as such? Most writers typically print or write their name in the left hand margin, as did almost everyone invited to transcribe the above letter.

At this point we need to take a foray into a branch of linguistics known as *pragmatics*. Linguists use this word to describe how speakers make meaning, sometimes to say more than they mean – but in general *what* they say in order to mean something. In the eighteenth century it might have been common for a wife to sign her name in full when writing to her husband, but in modern times it's very unusual. Why didn't she just write her name 'Sandra'. It's not as if her husband would have said to himself 'I wonder which Sandra could have written this to me'.

Here is another word you will find useful: *prescriptive*. In the context of language it means 'proper' language or 'correct' language, language which follows all the rules or *prescriptions* of grammar. If you look at the letter above you will see that aside from one or two very minor glitches it follows the rules of traditional grammar. There are no grammatical errors in it, unless one were being particularly picky. For example, *garry*

is written with a lower case 'g', and there is no comma before *garry*. Other than this, the letter is entirely grammatical. Bear in mind that these minor omissions may be no more than the product of haste, rather than lack of knowledge about the correct forms. So, these points aside, I think we can agree that the letter follows the rules of prescribed grammar.

Now this may not seem very important, but actually it is much rarer than we think. I get letters from people all the time, and in my experience many letters are written with less than perfect grammar. With today's emphasis on communicative competence rather than grammar, many people nowadays have difficulty with issues such as spelling and punctuation.

Other people commented on the oddity of the letter being typed out rather than handwritten. You will have seen the apparent explanation for this: 'I am typing this note, because I know that if I were to hand write it and leave it for you, then I know that you wouldn't read it'. To the extent that this point – typing *vs* writing by hand – concerns motive for an action, it is a psychological question, and hence outside of the linguist's domain. On the other hand, insofar as it concerns the form in which the letter is produced it is a linguistic question, because of something we refer to as *mode*. Strictly speaking *mode* refers to whether language is in the form of speech, writing, dictation and so on. There are several speech modes: speech in casual conversation, language spoken by a teacher to a class, a college or university lecture, the speech of a news reader – this latter being speech which is written in order to be read aloud. Similarly, there are a number of ways of producing written language. We can write out something manually, with a pen or pencil, we can type it on a typewriter, we can use a word processor, and so on. In the course of working as a forensic linguist I have noticed that when the medium changes – for example, handwritten text *vs* word-processed text – there are always a few minor changes in the writer's style. In one case I noticed some changes in a series of letters which I could not explain. The forensic computer expert who also worked on the case later told me that some of the letters had been typed on a laptop while others had come from an office desktop machine. As anybody who has used both will know, laptops are not as convenient to use as conventional desktop machines – the keyboard is smaller, as is the screen and there is usually no mouse. These differences of medium or mode can cause minor changes in style. Given that this was a highly personal letter, and that it could have been written out in less than two minutes, and given the explanation as to why it had not been handwritten, I felt that *mode* was a factor in its authorship.

After considering the content of the letter it seemed that the likeliest candidates for authorship were Sandra Weddell herself and her husband Garry Weddell, the police inspector. Mrs Weddell was a senior nurse in a nearby hospital who also worked part time as an examination invigilator for the local examination authority. She had attended a morning examination session at a local school and had returned home for lunch on 30 January 2007, and was then due to return to the school at approximately 2 p.m. on the same day. From the school she was then due to pick up her own children from another school before returning home at approximately 4 p.m. Mrs Weddell did not return to the examination for the afternoon invigilation session and nor were her children picked up from school. The school called her husband at work and he picked up the children. Mrs Weddell was not found until the next day.

In many forensic cases we have very few examples of the language of candidate authors. In a kidnap case, for example, you might have only one or two letters. People who intend to commit crime are usually careful to commit as little to paper as possible. However, in the present case we had plenty of examples of both writers. In one letter, shortly after his wife's death Garry Weddell wrote:

Please don't send any more letters to any of my family members. They are all just as grief stricken as i am over this matter. We are meeting up regularly to allow me to get what i need to get off my chest. Family support is the best therapy at this time. I have that support in place.

What struck me immediately about this letter was the brevity of the sentences. The average sentence length here is just under 9 words. Looking at the 'suicide' letter we see that the sentence length is not much more than this – just under 12 words. This was a pattern that was to be repeated right across Mr Weddell's letters. Sandra, on the other hand, tended to write quite long sentences. At one point she had to write to her child's school in connection with a lapse in their security arrangements regarding the collection of children after school. She wrote: 'However, on Wednesday 18th January, which is the date that I brought to your attention, Fred¹ was collected by Mr Arbuthnot, who has never collected Fred before and I don't know who handed Fred over on that occasion or indeed why.' This sentence (with names changed to protect the anonymity of others) is over 40 words long. It is not at all unusual for Mrs Weddell to write such long sentences. In fact one of

her sentences was over 130 words long. Her average sentence length was nearly double that of her husband's. Moreover, she appeared to have a fondness for commas, dashes and semi-colons, sprinkled somewhat liberally across her letters, and not always with sound grammatical reasons. It was her habit to string whole sentences together, separated by no more than a series of commas.

One question that comes up in the popular media from time to time is the idea of the 'linguistic fingerprint'. According to this idea, each of us has a unique, identifiable way of using language. However, this idea needs to be thought about very carefully. Balanced against the 'fingerprint' idea is the concept of 'individual variation'. What kinds of factors might contribute to variation in our writing style? Above, I referred to *mode*, the form in which language is produced, and I said that depending on whether we hand write a text, type it on a desktop or a laptop, or write it on a blackboard using chalk, that these different methods and environments of production, would probably produce differences in our style of language use. But, in addition to *mode*, there are a number of other factors which could increase how much variation an author shows. I will first describe these, and then I will relate this question of variation to the texts in this case – both those of the victim Sandra Weddell and her husband, Garry Weddell.

- *Vocabulary*: We use different vocabularies (also called 'lexicons') depending on who we are writing to, what we are saying, and what the circumstances are of the communication. Formal letters will contain very precise words which are usually absent from informal communications. When writing to someone we know, for example, we would probably be informal, but when writing to a stranger the opposite would apply. A report would probably use technical vocabulary, while – by contrast – a birthday card would use very general, common vocabulary.
- *Time*: Another factor likely to produce variation in someone's style of language is the time lapse between two communications. Over a period our vocabulary changes, and so do other factors, like sentence length, phraseology and so on. The longer the period between any two texts, the more likely we are to find differences.
- *Personal circumstances*: A number of changes in our circumstances can produce changes in our use of language, such as bereavement, changes in employment, marriage, having children and so on. Some of these changes can have catastrophic effects on the way we write – and I am not referring to handwriting, though that too can be affected.

- *Cultural changes*: Our culture is changing all the time, though we do not usually notice the differences on a day-to-day basis. As a result of these changes words come and go, old phrases sometimes fall out of fashion, and new ones come along. One area in which you can particularly notice changes is in the way people text on mobile phones – the famous abbreviations, for example, '4u' for 'for you' are now no longer as startling as they first were. Mobile phone texting has continued to evolve, so that we can now easily pick out differences from 4 or 5 years ago. Thus, young texters used to be content with writing 'know' as 'no', but they now frequently write 'na', and for many texters 'dinner' became 'dinna' and has now become 'dina'. Frequently, letters which can be taken for granted are omitted, for example, 'remember' becomes 'remeba'. Even the word 'texting' has modified from 'txtn' to 'txn', for example, 'i ws rxn u ls nyt' (I was texting you last night).

Hence, as we can see from the above, there are many reasons why an individual writer's use of language can vary. When we add these possible causes of variation up, they are referred to as **within author variation**. But what about how authors vary from each other?

If two authors have similar backgrounds, similar levels of education, come from the same geographical area, and have similar occupations, then the possibility is that they will not vary much from each other. Conversely, if two authors have widely different social backgrounds, with one being educated to a high level while the other's education was somewhat limited, if they come from different geographical areas, and have completely different types and levels of occupation or profession – then the chances are they will probably have different styles of language use. Collectively, these different social factors are referred to as **inter-author variation**.

So, on the one hand we have within author variation and on the other hand we have inter-author variation.

What makes the linguistic fingerprint idea difficult is this: suppose for a moment we have an inquiry where one of the authors is showing a lot of within author variation, and suppose also that there is very little inter-author variation across the different authors? Then it will be clear that finding significant differences between these different authors would be very difficult.

So, how do these observations relate to the alleged suicide letter, and the letters written by Sandra Weddell and Garry Weddell?

In the present case, some factors do contribute to variation. For example, not all of the texts for each author are of the same type, or

written to the same addressee. Some of Mr Weddell's texts are emails, some are letters written in a reporting style, and some of them are business letters. Among his letters there was also one personal email, to a relative.

With regard to Mrs Weddell's texts there was a variety of addressees and text types. However, there are no personal communications among Mrs Weddell's known texts, and since the alleged suicide note can be considered to be a personal communication this could be a complicating factor in assessing her style.

An interesting question that comes up from time to time is that of 'convergence of style'. Do married couples tend to adopt each other's language habits over the years? Again, this will depend on early influences on their language to a large degree: if they are of similar ages and backgrounds then there might already be strong similarities. However, I have seen no evidence that married couples start to write like each other, in terms of the language they use. They might, over the years, adopt some of each other's phrases in speech and other speech habits, but merging written language styles is another question altogether. From what I could see of Sandra's language style there was no evidence that she wrote like Garry, or that Garry wrote like her, even though the couple had been married for a number of years.

In fact, despite the possible sources of variation referred to above, it is noticeable that some basic features of the texts of each author are quite consistent. Whoever Sandra Weddell is writing to, she tends to use the same kind of concatenated, rambling sentence structure as we saw above, and – similarly – whoever Garry Weddell is writing to, we see that he always seems to use short, sharp sentences. In this respect he is like a typical company executive or other senior official – say what you want and get it over with, no messing about, no wandering off the point.

The important point is that these characteristics are found across all of the letters: short sharp sentences for Garry, long, rambling sentences for Sandra. No matter the type of communication each is writing, this is what we find.

After that I reported my findings to the police. They had other evidence as well and arrested Garry Weddell in June 2007.

Tragically, Mr Weddell was given bail and told to keep clear of the area where his relatives lived. However, he appears to have felt vengeful to some of his relatives and, apparently, shot his mother-in-law, Traute Maxfield, before shooting himself. These were the coroners' findings in March, 2008. There seems little doubt that he had killed his wife and then tried to fabricate her suicide.

There is one further interesting detail about the full stop after the word 'Garry' in the opening line of the letter. Forensic scientists had to test the ink on the letter, to see if it was compatible with that found in the printer attached to the family computer. In order to do this they had to remove a sample of the ink from the letter. The sample they chose to remove was – the full stop! Fortunately, they had taken very good photographs of the letter before doing so, but even so, I sometimes wonder whether the defence team would have made anything of this point if the case had come to trial.

Note

1. Names changed.

17

A BLAND PAEDOPHILE

It sometimes happens that I am asked to make a comparison between two sets of letters or emails and on looking at the documents I find myself completely stumped for something constructive to say – for the simple reason that the suspect's language style appears, on first sight, to be completely featureless. This happened some years ago in the case of a businessman who was accused of downloading child pornography images from the internet. A highly educated, intelligent man in his fifties, one would not have expected him to be a person who made basic mistakes, but people do unexpected things. His mistake was that, on finding the photos and images he had downloaded from a certain website, which were not to his liking, he wrote a letter of complaint to the company providing the pornography service.

The police do not normally give information out about their sources, but I am led to believe that it was the company that, rather than refunding him his money as he requested, informed the police – anonymously one imagines – that Mr Sowerby had been downloading illegal material. Mr Sowerby was duly visited at rather an early hour in the morning several weeks later and his computer was seized. The offending images were found, as was a copy of the letter of complaint. Mr Sowerby's defence was that the images had been downloaded by a friend of his, who had – unknown to Mr Sowerby – also used his credit card to pay for the transaction. Asked about the letter of complaint Mr Sowerby said that his friend must have panicked when he realized that he, Mr Sowerby, would eventually notice the transaction. Thinking to get the money returned to his account, the friend had obviously concocted the letter of complaint. Asked to name the 'friend' Mr Sowerby named a local plumber who, however, had recently died. What could not be argued was that the plumber and Mr Sowerby had been acquainted. However, inquiries revealed that none of the plumber's friends or family believed that he knew how to use a computer, especially with regard to the sending and receiving of emails, making payments over the internet, or being able to download pictures or videos. What was more, nobody in the plumber's