

# The Linguist as Expert Witness

Malcolm Coulthard

## Introduction

In the past fifteen years, there has been a rapid growth in the frequency with which Courts in a number of countries have called upon the expertise of linguists in cases where there is a dispute about aspects of a written text. The cases in which linguistic evidence has been used range from disputes about the meaning of individual morphemes in a trademark case and individual words in jury instructions, through the 'ownership' of particular words and phrases in a plagiarism case to accusations of fabrication of whole texts in certain murder cases. Usually the linguist uses standard analytic tools to reach an opinion, although few cases require exactly the same selection from the linguist's toolkit. Occasionally, though, cases raise new and exciting questions for descriptive linguistics, which require basic research, such as how can one measure the 'rarity' and therefore the evidential value of individual expressions, (see Coulthard in press,) or the reliability of verbal memory (see p XX below).

Once the analysis has been done and an opinion reached, the linguist is faced with two interactional problems: firstly, how can s/he best transmit the linguistic insights and findings in a written report to a lay audience and secondly, if called to give oral evidence in court, how to cope with the unusual interactional rules which involve lawyers asking questions, notionally on behalf of the court and the answers being addressed directly to the judge and/or jury. Even more difficult can be cross-examination which pits an expert, who has sworn to obey the Gricean maxim of *quality*, against a lawyer who is under no such constraints and can apparently say at will what he may 'believe to be false' and things for which he 'lacks proper evidence'.

All experts face these communication problems, but the linguist has the additional and unique problem that everyone is in some senses an expert on language. Indeed it is very difficult to call a linguist to give evidence on word meaning to a jury, because courts are mainly interested in two kinds of meaning: technical, legally defined meaning, as for instance the meaning of 'dusk' in a statute which says 'The park gates will be closed at dusk', and where 'dusk' will have been given a specific meaning elsewhere in the statute as something like '30 minutes after sunset'; and *commonsense* meaning, which is what a jury, being a collective representative of the common man, thinks a word means- so much so that juries are normally denied dictionaries within the jury room.

In cases where meaning is in dispute and a linguist is called, the role is often restricted to what Solan (1998) calls a 'semantic tour guide', that to providing the court with the linguistic insights about the nature of the problem, which will then put them in a position to arrive at an informed opinion themselves. However, linguistic evidence is by no means confined to meaning and in what follows I will try to give a sense of the range of the linguistic topics covered and the linguistic techniques used in reaching opinions.

### **1. Morphological meaning in a trademark case**

Shuy (2002: 95-109) reports his contribution to the case of McDonald's Corporation v Quality Inns International, Inc, which revolved around whether McDonald's could claim ownership of the morpheme 'Mc' and so prevent its use in other trademarks. The case began in 1987 when Quality Inns announced they were going to create a chain of basic hotels and call them McSleep, claiming, when challenged, that they hoped the 'Mc' prefix would evoke a Scottish link and with it the Scots' well-known reputation for frugality. McDonald's, who had previously successfully prevented the use of the name McBagel's, when a judge had decided that the prefix could not be used in conjunction with a generic food product, decided to challenge the McSleep mark, claiming it was a deliberate attempt to draw on the goodwill and reputation of the McDonald's brand.

In supporting their case McDonald's argued they had deliberately set out, in one advertising campaign, to create a 'McLanguage' with Ronald McDonald teaching children how to 'Mc-ise' the standard vocabulary of generic words to create 'McFries', 'McFish', 'McShakes' and even 'McBest'. Fanciful as this linguistic imperialism might seem to be to ordinary users of the language, particularly to those with Scottish or Irish descent, who would seem to be in danger of losing their right to use their own names as trademarks, the lawyers took the claim very seriously. Quality Inns lawyers asked Shuy to help with two linguistic arguments, firstly, that the morpheme 'Mc' was in common use productively, when it was not seen to be linked in any way to McDonald's and secondly, that such examples showed that the prefix, originally a patronymic, equivalent in meaning to the *son* in Johnson, had become generic and thus now had a meaning of its own, which was recognisably distinct from both of the other meanings 'son of' and 'associated with McDonald's'.

Shuy chose a corpus linguistic approach and searched to find real text instances of 'Mcmorphemes'; among the 56 examples he found were general terms like McArt, McCinema, McSurgery and McPrisons, as well as some items already being used commercially like the McThrift Motor Inn, a budget motel with a Scottish motif and McTek a computer discount store which specialised in Apple Mac computer products. On the basis of such examples, Shuy argued that the prefix had become in the language at large a lexical item with its own meaning of 'basic, convenient, inexpensive and standardized' (p 99). Rather than use corpus evidence themselves, McDonald's hired market researchers to access the public's perception of the prefix directly through interview and questionnaire and they reported that consumers did indeed associate the pre-fix with McDonald's, as well as with reliability, speed, convenience and cheapness. Faced with conflicting evidence, the judge ruled in favour of McDonalds, thereby giving them massive control over the use of the morpheme.

### **2. Syntactic complexity in a letter**

Levi (1993) reports a case in which she acted as expert witness, testifying on syntactic complexity. The plaintiffs' claim was that a letter was so badly written that it failed to inform them of their rights. Levi identified a series of syntactic features which, she argued, were likely to interfere with understanding, for example, 'multiple negatives, complex embeddings, nominalisations ..... passive verbs without subjects and difficult

combinations of logical operators like *and*, *or*, *if* and *unless*. (pp 7-8). She quotes the following extract from the letter as an example of the syntactic problems encountered:

If your AFDC financial assistance benefits are continued at the present level and the fair hearing decides your AFDC financial assistance reduction was correct, the amount of AFDC assistance received to which you were not entitled will be recouped from future AFDC payments or must be paid back if your AFDC is cancelled.

and ‘translates’ this as being equivalent to saying:

If X happens and then Y happens then either Z will happen [expressed in very complex terms including a negative with a relative clause] or – if R has also happened – then Q must happen (p 8).

and then calculates the syntactic complexity as consisting of ‘a complex internal structure built out of seven clauses, six passive verbs without subjects ... and several complex compound nouns, (e.g. financial assistance reduction), which themselves contain nominalised verbs without expressed subjects (pp 8-9).

Sadly she does not report the outcome of the case nor give any indication of what the judge thought of her evidence, but at least she was admitted as a witness on syntactic meaning, which is by no means always the case - indeed one judge in the United States refused to admit the linguist Ellen Prince as an expert explicitly on the grounds that it is the function of the Court to decide on meaning. Certainly, it is more difficult when the texts involved are legal texts, because lawyers and judges usually see themselves as the guardians of and adjudicators on such meaning. However, even here linguists are occasionally allowed to express a professional opinion, although it helps if they are lawyer-linguists and/or have a lawyer as co-author, as happened in the next case.

### **3. Lexico-grammatical ambiguity in a statute**

Kaplan et al (1995) report on an appeal case which went to the Supreme Court in 1994. The facts are as follows: a certain Mr. Granderson pleaded guilty to a charge of destroying mail, for which the maximum sentence was 6 months in prison. In fact the judge decided to fine him and put him on probation for 5 years, that is 60 months. Subsequently Mr. Granderson violated his probation by being caught in possession of cocaine. In such cases the law instructs the Court to ‘revoke the sentence of probation and sentence the defendant to not less than one third of the original sentence’. This presented the Court with a problem because, if it took ‘original sentence’ to refer to ‘probation’, imposing a sentence of ‘not less than one third’ could in fact reduce the penalty as he had not served 40 months of probation; in the end it was decided to sentence him to 20 *months in jail*, that is a sentence which rather than being ‘not less than one third’ was more than three times greater than the original maximum prison sentence.

Kaplan et al argued that this particular interpretation was inadmissible on linguistic grounds, because one cannot allow an admittedly ambiguous item to have both of its meanings simultaneously – the Court had interpreted the phrase *original sentence* as referring to ‘imprisonment’ for the purpose of determining the type of punishment, but to

‘the initial imposition of 5 years’ (of probation) for the purpose of determining the *length* of the sentence. One of the authors observed that what the court had done was the linguistic equivalent of a Frenchman taking the phrase *Pierre a fait tomber l’avocat* to mean, ‘Pierre did something to a lawyer [l’avocat1] and caused the avocado [l’avocat2] to fall’. The Supreme Court not only took note of the linguistic argument, but also cited it in their judgment and changed the interpretation to ‘a sentence of not less than 2 months in prison’ – as the accused had already been in prison for 11 months, almost double the original maximum sentence, he was released immediately.

#### **4. Obscure lexical meaning in jury instructions**

Levi (1993) reports a lexical analysis of a set of jury instructions concerned with imposing a penalty of death, which she undertook as part of an expert report in the case of *US ex rel. James P Free Jr v Kenneth McGinnis et al.* She was asked to express an opinion on the question ‘How well could [the language of the jury instructions] have served its purpose in communicating clearly to the jury the legal concepts they needed to understand for sentencing in a capital case?’ (p10). The instructions in question were (my highlighting):

If you unanimously find from your consideration of all the evidence that there are no mitigating factors **sufficient to preclude** the imposition of a sentence of death then you should return a verdict imposing a sentence of death.

If, on the other hand, you do not unanimously find that there are no mitigating factors **sufficient to preclude** the imposition of a sentence of death then you should return a verdict that the sentence of death should not be imposed

In considering *sufficient* Levi focused on the inherent vagueness of the word, which, she pointed out, has only a contextually derivable meaning and on the fact that the instructions themselves did not give the individual juror any help on how to decide what would count as a sufficient mitigating factor in the particular situation of sentencing someone to death, or whether a single factor, that was perceived to be sufficient, but only by one juror, would in itself be ‘sufficient to preclude’ – even though, according to the law, it would.

In considering *preclude* Levi’s approach was different; she pointed out that, while this word did have a context-independent meaning, most of the jurors were unlikely to have known it. She supported this assertion by testing some fifty undergraduate students attending one of her courses; only three students were able to provide the correct definition. She does not report how the court evaluated here evidence.

#### **5 Pragmatic meaning in an insurance proposal.**

Prince (1981) reports possibly the earliest forensic pragmatic analysis. It is a case where a 58 year old cement worker sued an insurance company, which was refusing to pay his disability pension, because they asserted that he had lied when he responded to four of the questions on the original proposal form. One of those questions read as follows

Have you any impairments?.... Loss of sight or hearing?.... Loss of arm or leg?...Are you crippled or deformed?... If so explain....

The insurance company argued that the man was lying when he replied in the negative, since 'he was overweight, had a high cholesterol level and occasional backaches', even though they did not dispute his assertion that none of these conditions had ever caused him to take time off work, (op cit:2). In her evidence Prince used the concept of a co-operative reader genuinely trying to make sense of the meaning of the document. For this particular question she focused on the vagueness of the word *impairment*, and argued that any 'co-operative reader' would reasonably infer, given the content of the phrases that followed the word *impairment*, which constituted the only textual clue to the meaning of 'impairment' in the insurance proposal, that it was being used in that particular context to mean a relatively severe and incapacitating physical condition. Given that *impairment* was not defined and that the examples given were at best unhelpful, if not downright misleading, given the meaning the insurance company insisted the word was supposed to have in the question, she argued that the man had indeed answered "no" 'appropriately and in good conscience' to the question he thought they were asking, (op cit:4). The judge ruled in favour of the plaintiff.

## **6 The recoding of interaction in written form – police records of interviews**

Many cases revolve around the accuracy of a written record of an interaction between the police and the accused. Converting the spoken to the written, as anyone who has attempted it is aware, is not an unproblematic task, but, even so, most police forces have no explicit guidelines about the procedures to use and what could or should legitimately be omitted, even when the aim is to produce a verbatim record in the interviewee's own words. In this context it is useful to consider Slembrouck's (1992) observations about the production of Hansard versions of proceedings in Parliament, where scribes similarly linguistically untrained are charged with the creation of highly important verbatim records of what was said. Slembrouck notes that

there is filtering out of 'disfluency' and other obvious properties of spokenness (e.g. intonation, stress). Repetitions, (even when strategically used...), half-pronounced words, incomplete utterances, (un)filled pauses, false starts, reformulations, grammatical slips, etc. are equally absent, (104).

In the typical police record the same rules seem to apply. For this reason the appeal of Robert Burton, *R v Robert Burton*, in the English Court of Appeal in 2002 was fascinating. Burton was captured red-handed with several companions trying to steal trailers loaded with £250,000 worth of whisky from a trailer park. He did not realise that his companions were in fact all undercover police officers. Burton's defence was that he had tried to call off the operation on several occasions, but the undercover police officers, who he thought were real criminals to whom he owed a lot of money for drugs, forced him to go through with the robbery – thus his defence was that the undercover police officers had been involved in an illegal action, incitement to commit a crime.

As part of their evidence the police submitted several records of telephone calls, which they claimed an undercover officer, using the codename Charlie, had written down from

memory immediately afterwards. Part of Burton's defence was that these records were too accurate to have been made from memory and thus must have been transcribed from tapes, which, of course, could not be submitted in evidence, because in the same conversations the police officer was also pressuring him to commit a crime.

The linguistic evidence confirmed that either the police officer had an amazingly ability to recall conversations verbatim or there was indeed a tape-recording. This opinion was based on two sets of features, firstly, the inclusion in the records of a set of spoken discourse items which are regularly produced by speakers, but which carry little or no significant content and which are therefore typically forgotten or at least not reported by those producing remembered accounts of what was said, These include

- a) discourse markers - items which typically occur at the beginnings of utterances - 'well', 'right', 'so';
- b) acknowledgements of replies to questions – what some call *third parts of exchanges* – realised by 'yeah', 'okay' 'alright' and repetitions of whole phrases from the preceding utterance;
- c) other kinds of cross-utterance repetition and reformulation;
- d) fillers such as 'like' and 'you know what I mean'.
- e) adverbial modifiers like 'just', 'really', 'actually' and 'fucking'
- f) slang items and non-standard grammatical forms like 'gonna'.

Secondly, and even more surprisingly, Burton had a marked stammer which he had learned to partially control by the use of a 'step word', which in his case was 'like' – this too had not only been reproduced but also in the kinds of linguistic contexts in which he himself used it. Many of these features are exemplified in the short extract below:

I said, "You **gonna** take something heavy, do **you know what I mean**, to make things easier in there."  
Bob said, "No, **fuck off, like** that's too much, I'll **just** have a blade, that'll do."  
I said, "**Yeah okay.**"  
Bob said, "I'm **just** a bit jittery **like** as its getting close **like.**"  
I said, "**Yeah, okay** but keep in touch."  
Bob said "**Yeah**, sorry about that Charlie, there's no problems honest, I'll chase that **bloke** up and find out what's happening whether there's 2, 3 or 4 there."  
I said, "**Okay** we may have to do it on two to get it done by Christmas."  
Bob said, "**Yeah okay.**"  
I said, "**Alright**, see you later Bob."  
Bob said, "**Yeah** later Charlie **mate.**"

My evidence was accepted unchallenged, but the appeal failed on other grounds.

## 7. Non-native English in a contested will

McMenamin (2002) reports the case of a contested will of a woman who died at the age of 85 in Alaska, having been born in Japan and grown up in Hawaii. The will apparently left everything to a couple of neighbours and was supported by photocopies of five letters on the topic of the will, which had supposedly been dictated by the deceased to a friend

called Kim and later discovered in the boot of a car. Kim was untraceable, as were the originals of the letters.

The 'Kim' letters had a series of typical creolized English features such as the deletion of articles, subjects, objects and some auxiliary and copula verbs, as well as the omission of plural and tense morphemes. However, by contrast, the known writings of the deceased, although they 'evidenced some features of Hawaiian Creole English' (op cit:132) were much closer to Standard English. In addition some of the creolisations found in the letters did not occur at all in the known writings. More worryingly, all the creole features in the suspect letters were deletions of grammatical elements in the standard language, whereas McMenamain notes that not only is there 'no variety of English known to be defined by a single process of variation like deletion' (ibid). To add further doubt, the known writings of the deceased did include other creole features that were not simply deletions, such as mismatch between verb and complementiser and mass nouns used as count nouns. The facts and derived opinions convinced the judge, who found that

'the .. "Kim" papers were prepared by the [neighbors] or at their direction... [and that] the language usage .... is concocted and a fraud' (op cit 135-6).

### **8 Uniqueness of expression and a disputed confession**

In this case, four men were accused, and subsequently convicted, of killing a 13-year old newspaper delivery boy, Carl Bridgewater, solely on the basis of the confession of one of them, Patrick Molloy – there was no corroborating forensic evidence and Molloy subsequently retracted his confession, but to no avail. He admitted that he did actually say the words recorded in the confession, but insisted that he was being told what to say, by a policeman, who was standing behind him. He also claimed that he had only made the confession after being physically and verbally abused for some considerable time, immediately beforehand.

The police, however, as support for the reliability of the confession, produced a contemporaneous handwritten record of an interview which, they claimed, had taken place immediately before the confession and which contained substantially the same information, expressed in the same language, as the confession statement. Molloy denied that this interview had ever taken place – in his version of events he was being subjected to abuse at that time. He counter-claimed that the interview record had been made up later on the basis of the by then pre-existing confession. As is evident from a cursory glance at the two extracts below taken, respectively, from the statement which Molloy admitted making and the interview record which he claimed was falsified, the similarities are enormous; I have highlighted identical items in **bold** and close paraphrases in *italic*.

#### **Extract from Molloy's Statement**

(17) **I had been drinking and cannot remember the exact time I was there but whilst I was upstairs I heard someone downstairs say be careful someone is coming.** (18) **I hid for a while and** *after a while I heard a bang come from downstairs.* (19) **I knew that it was a gun being fired.** (20) I went downstairs and **the three of them were still in the room.** (21) **They all looked shocked and were shouting at each other.** (22) **I heard**

**Jimmy say, "It went off by accident".** (23) I looked and **on the settee** I saw the *body of the boy*. (24) **He had been shot in the head.** (25) **I was appalled and felt sick.**

**Extract from Disputed Interview with Molloy**

- P. How long were you in there Pat?
- (18) **I had been drinking and cannot remember the exact time that I was there, but whilst I was upstairs I heard someone downstairs say 'be careful someone is coming'.**
- P. Did you hide?
- (19) Yes **I hid for a while** and then **I heard** the **bang** I have told you about.
- P. Carry on Pat?
- (19a) I ran out.
- P. What were the others doing?
- (20) **The three of them were still in the room.**
- P. What were they doing?
- (21) **They all looked shocked and were shouting at each other.**
- P. Who said what?
- (22) **I heard Jimmy say 'it went off by accident'.**
- P. Pat, I know this is upsetting but you appreciate that we must get to the bottom of this. Did you *see the boy's body*?  
(Molloy hesitated, looked at me intently, and after a pause said,)
- (23) Yes sir, he was **on the settee.**
- P. Did you see any injury to him?  
(Molloy stared at me again and said)
- (24) Yes sir, **he had been shot in the head.**
- P. What happened then?
- (25) **I was appalled and felt sick.**

Linguists from all persuasions subscribe to some version of the 'uniqueness of utterance' principle and so would expect that even the same person speaking/writing on the same topic on different occasions would make an overlapping but different set of lexicogrammatical choices. Most linguists would agree, on the basis of the number and length of the identical strings, that either one of the two documents was derived from the other or that both had been derived from a third. However, at the time of the original trial, no linguist was called to give evidence – in fact there were no forensic linguists practising in Britain at the time – so it was left to the lawyers to evaluate the linguistic significance of the similarities between the interview and the confession. As a result, the same phenomenon, massive identity in phrasing and lexical choice, was argued by the defence to be evidence of falsification, and by the prosecution to be evidence of the authenticity and reliability of both texts, on the grounds that here was an example of the accused recounting the same events, in essentially the same linguistic encoding, on two separate occasions.

Both the prosecution assertion that identity of formulation in two separate texts is to be expected and indicative of reliability and the apparent willingness of the lay jury to accept this assertion, depend on two commonly held mistaken beliefs: firstly, that people can and do say the same thing in the same words on different occasions and secondly, that people can remember and reproduce verbatim what they and others have said on

some earlier occasion. The former belief can be demonstrated to be false either by recording a person attempting to recount the same set of events on two separate occasions, or by simply asking a witness to repeat word for word what s/he has just said. The second belief used to have some empirical support, at least for short stretches of speech, (see Keenan et al 1977 and Bates et al 1980), but was seriously questioned by Hjelmquist (1984) and Hjelmquist and Gidlung (1985), who demonstrated that, even after only a short delay, people could remember at best 25 percent of the gist and 5 percent of the actual wording of what had been said in a five minute two-party conversation in which they had just participated.

Confirmatory evidence of the inability to remember even quite short single utterances verbatim was specially commissioned from Professor Brian Clifford and presented at the 2003 'Glasgow Ice Cream Wars' Appeal. This was used to challenge successfully the claim of police officers that they had independently remembered, some of them for over an hour, verbatim and identically, utterances made by the accused at the time of arrest. Clifford's experiment tested the ability to remember a short, 24-word utterance and found that, even when such a small stretch of language was involved, most people were able to recall verbatim no more than 30 to 40 percent of what they had heard. (for details see <http://news.bbc.co.uk/1/hi/scotland/3494401.stm>)

Thus, the only way in which these two Molloy extracts could have come to share so much vocabulary and phrasing would be if one had been derived from the other or both from a third text. Sadly, it was not possible for me to test the acceptability and persuasiveness of these arguments in court, as the Crown conceded the appeal shortly before the due date, when compelling new evidence from document and handwriting analysts emerged to convince the judges of the unsafeness of the conviction.

### **9 Problems with lexical and grammatical cohesion in a disputed confession**

In the same Bridgewater Four case there was secondary, supporting linguistic evidence of a different kind to reinforce the opinion that the interview record was falsified and to demonstrate that it was derived from the statement. If we assume that the police officers had indeed, as Molloy claimed, set out to create a dialogue based on the monologue statement, they would have faced the major problem of what questions to invent in order to link forward and apparently elicit the actually pre-existing candidate answers, which they had derived from the statement. In this scenario one would expect there to be occasions when a question did not fit successfully into the text into which it had been embedded – and indeed there are.

In a developing interview, a police question usually links backwards lexically, often repeating word(s) from the previous answer. However, in creating a question to fit a pre-existing answer, there is always the danger that the question will only link forward. I will give two examples. The original statement has a two-sentence sequence – (21) They all looked shocked and were shouting at each other. (22) I heard Jimmy say 'it went off by accident' – which appears word for word in the interview record, except that the two sentences are separated by the inserted question "Who said what?". However, in this context the word "said", although cataphorically unremarkable – "said" links with "say" –

is anaphorically odd, because the men have just been described as "shouting". One would therefore have expected an anaphorically cohesive follow-up question to be either 'What/Why were they *shouting*?' or 'Who was *shouting* (what)?'; one would certainly not predict "who *said* what?". The choice of "said" is a most unexpected choice – except, of course, for someone who knows that the next utterance will be "I heard Jimmy *say*..." – then "said" has an evident logic.

An example of a *grammatical* misfit is where the statement version "on the settee I saw the **body** of the **boy**. **He** had..." is transformed into "Q. Did you see **the boy's body**? Yes sir, **he** was on the settee". The statement version correctly uses the pronoun "He" because the referent is the "boy" in "the body of the boy", but the reformulated version in the police interview, "the boy's **body**", would be more likely to have elicited "**it**" as a referent.

We also find examples of *process* misfit: in the exchange reproduced below, the question "what happened" requires a report of an action or an event, but in fact the response is a description of two states:

P      What **happened** then?  
M      I **was appalled** and **felt sick**.

Had the reply been "I vomited", it would, of course, have been cohesive. Similar process misfits are:

P      What were the others **doing**?  
M      The three of them **were** still in the room.  
P      What were they **doing**?  
M      They all **looked shocked**

It is possible to continue in this vein, but these examples are sufficient to show that textual oddities like these support the opinion that the interview record was created from the pre-existing statement.

### **10. Uncovering underlying dialogue in a disputed statement**

Until the Police and Criminal Evidence Act was passed in 1984, requiring the tape-recording of significant interviews with suspects, the English police standardly recorded linguistic evidence in handwritten form. Under Judges Rules in force at the time, they were required to record what the suspect said verbatim. They could record two kinds of evidence, an interview, when they had to write down in full both their questions and the suspect's answers and a statement, when the suspect dictated his version of events, during which the police were not allowed to ask questions. Nevertheless, suspects not infrequently claimed the police did ask questions, but then recorded this dialogue as if it had been a monologue. This was confirmed by Detective Chief Inspector Hannam when giving evidence in the 1952 trial of Alfred Charles Whiteway. In the following example the accused was not responsible for a single word in the quoted sentence which was attributed to him.

I would say "Do you say on that Sunday you wore your shoes?" and he would say "Yes" and it would go down as "On that Sunday I wore my shoes" (Court transcript of Hannam's evidence, p 156)

In such cases, when it is important to establish that a supposed monologue statement was in fact a collaboratively produced document, there are occasionally linguistic clues to the underlying dialogue. In the case of Derek Bentley, accused and later found guilty of the murder of a policeman, although he was under arrest at the time the policeman was shot, it was the occurrence of negative clauses within the narrative, (for a detailed discussion, see Coulthard 2002).

We begin with the observation that narratives, particularly narratives of murder, are essentially accounts of what happened and to a lesser extent what was known or perceived and thus reports of what did **not** happen or was **not** known are rare and special - there is always an infinite number of things that did not happen and thus the teller needs to have some special justification for reporting any of them to the listener, i.e. there must be some evident or stated reason for them being newsworthy.

In the Bentley case one negative statement

'I did not know he was going to use the gun'

assumed major importance in the trial. In his summing up, the judge who, because of the importance of the case was the Lord Chief Justice of England, made great play with this sentence arguing that its positioning in Bentley's narrative of events, before the time when was a single policeman on the roof, coupled with the choice of "the gun" (as opposed to "a gun") must imply that Bentley knew that Craig had the gun well before it was used - in other words "the gun" in its position in the statement must be taken to mean "the gun I already knew at this point in the narrative that Craig had". However, this argument only holds good if the statement was indeed a single author monologue. If it was, as alleged, a jointly constructed narrative, the positioning would be determined by a police decision to ask a question at that point and the "the" would then have no incriminating significance, but merely be a consequence of collapsing a QA sequence into a single sentence. So, for example,

- Q Did you know he was going to use the gun?  
(i.e. the gun we now, after the event, know he had in his possession)
- A No (I didn't even know he has a gun)

Could easily be transformed into the actual 'I did not know he was going to use the gun'.

In Bentley's statement we see several negative clauses which have no narrative justification and therefore are most easily explained as the product of the collapsing of a negative response to a yes/no question; for example sentence (17) below:

- (16) Chris then climbed up the drainpipe to the roof and I followed.  
(17) Up to then **Chris had not said anything**.  
(18) We both got out on to the flat roof at the top.

Chris is not reported as beginning to talk once they have got out onto the roof, nor is his silence contrasted with anyone else's talking, nor is it made significant in any other way later in the narrative. A similarly unwarranted denial is:

- (26) [The policeman] caught hold of me and as we walked away Chris fired.
- (27) **There was nobody else** there at the time.
- (28) The policeman and I then went round a corner by a door.

None of the possible inferences from the denial seem to make narrative sense here - i.e. that as a result of there being no one else there a) it must be the policeman that Craig was firing at, nor b) that it must be Craig who was doing the firing, nor c) that immediately afterwards there would be more people on the roof. So the most reasonable conclusion is that at this point in the statement-taking a policeman, trying to clarify what happened, asked a question to which the answer was negative and the whole sequence was then recorded as a negative statement. The fact that, like (17) and (27) other sentences in the statement may have been elicited in this way becomes particularly important in relation to sentence (23) the one singled out by the judge as incriminating.

- (22) We were there waiting for about ten minutes.
- (23) **I did not know** he was going to use the gun.
- (24) A plain clothes man climbed up the drainpipe and on to the roof.

This sentence too would only make narrative sense if it were linked backwards or forwards to the use of a gun - in other words if it was placed immediately following or preceding the report of a shot and thus the most likely explanation for its actual form and placing is that it is a collapsed Q-A sequence.

In another case, that of Robert Brown, it was the grammatical oddity of an individual phrase that was conclusive. As part of my evidence I focussed on the two clauses

- (45) "I was covered in blood, (46) my jeans and a blue Parka coat and a shirt were full of blood". (the highlighting is mine)

It will be noticed that the phrasing of (46) is most unnatural - no one would refer to their own clothes with the indefinite article once they had begun with the possessive. The most likely use of 'a' in this context would be to distinguish between 'mine' and 'not-mine' - e.g. "I looked round the room and I saw my jeans and a blue Parka coat and a shirt - they were full of blood" - but that meaning, of course, didn't make any sense in this particular narrative - Brown has to be referring to his own clothes. The phrase "a blue Parka coat and a shirt" could have occurred quite naturally, of course, from a careless conversion of a sequence of short questions and answers into monologue form. Indeed one can see how it might have happened by looking at the following sequence taken from the record of an immediately preceding interview with Brown

- "What were you wearing?"
- "I had a blue shirt and a blue parka" (highlighting is mine)

In this context the use of the indefinite article is normal, because, as noted above, when items are introduced for the first time the indefinite article is the natural choice. Once the grammatical oddity of the phrase had been pointed out the judges were in a position to draw the same inference about underlying questions as I could.

### **Idiolect and the detection of plagiarism**

The academic linguist is now being asked more and more to help with plagiarism and any investigation of plagiarism is based, consciously or unconsciously, on a notion of *idiolect*. In other words, the underlying expectation is that any two writers writing on the same topic, even if intending to express very similar meanings, will choose an overlapping, but by no means identical, set of lexico-grammatical items to do so. It follows from this that, in any comparison of two texts, the more similar the set of items chosen, the greater the likelihood that one of the texts was derived, at least in part, from the other (or, of course, that both were derived from a third text), rather than having been composed independently

Johnson's (1997) solution to the detection of this kind of student plagiarism or *collusion*, was to move away from using strings or sequences of items as diagnostic features and to focus instead on the percentage of shared individual lexical types and tokens as a better measure of derivativeness. Intensive testing has shown that this measure of lexical overlap successfully separates those essays which share common vocabulary simply because they are writing on the same topic, from those which share much more vocabulary because one or more of them is derivative (see Woolls and Coulthard, 1998). For example, in Johnson's study, whereas three suspect essays shared 72 different lexical types in their first 500 words, a set of three other essays from the same batch, whose authors had not colluded, shared only 13 lexical types, most of which were central to the topic under discussion. Copycatch Gold, (Woolls 2002) is a computer program which applies these insights and enables rapid comparison of large numbers of student essays, provided they are submitted electronically. It has recently been used to compare 200 three thousand word essays, each with every other one, a total of some 19,800 comparisons which took less than a minute. Essays which share suspiciously high proportions of lexis are identified for closer inspection.

Further work, (Woolls 2003), has shown that the most significant evidence of collusion is not the mere quantity of shared lexis, but rather the fact that, both texts have not only selected the same item, but have also used them only once. As such 'once-only' items are, by definition, not central to the main concern of the text, otherwise they would have been used more frequently, the chances of two writers independently choosing several of the same words for single use are so remote as to be discountable.

If proof were needed of the distinctiveness and diagnostic power of words used once-only – *hapaxes* as they are technically labelled – it comes from successful internet searches in cases of suspected plagiarism. Experience confirms that the most economical method to use when checking the internet for suspected plagiarised text is to search for distinctive

collocates whose individual items occur only once in the text in question. I will exemplify with the opening of a story written by an 11-year old girl:

**The Soldiers** (all spelling as in the original)

Down in the country side an old couple husband and wife Brooklyn and Susan. When in one afternoon they were having tea they heard a drumming sound that was coming from down the lane. Brooklyn asks,

“What is that glorious sound which so thrills the ear?” when Susan replied in her o sweat voice

“Only the scarlet soldiers, dear,”

The soldiers are coming, The soldiers are coming. Brooklyn is confused he doesn't no what is happening.

Mr and Mrs Waters were still having their afternoon tea when suddenly a bright light was shinning trough the window.

“What is that bright light I see flashing so clear over the distance so brightly?” said

Brooklyn sounding so amazed but Susan soon reassured him when she replied .....

The first paragraph is unremarkable, but the second shifts dramatically, “*What is that glorious sound which so thrills the ear?*”. The story then moves back to the opening style, before shifting again to “*What is that bright light I see flashing so clear over the distance so brightly.*” It is hard to believe that an author so young could write in both styles and raises the question of whether the other borrowed text(s) might be available on the internet.

If one takes as search terms three pairs of collocated *hapaxes* ‘thrills/ear’, ‘flashing/clear’ and ‘distance/brightly’ one sees the distinctiveness of idiolectal co-selection; the single pairing ‘flashing – clear’ yields over half a million hits on Google, but the three pairings together a mere 360 hits, of which the first thirteen are all from W.H. Auden’s poem ‘O What is that sound’. The poem’s first line reads ‘O what is that sound which so **thrills the ear**’ while the beginning of the second verse is ‘O what is that light I see **flashing so clear** Over the **distance brightly**, brightly?’. If one adds a seventh word ‘so’ and searches for the phrase ‘flashing so clear’ with the other two pairs, every one of the hits is the Auden poem.

As a result of this work by Johnson and Woolls, by the time of the Bridgewater Four Appeal it was possible to provide extra evidence to support the claim that the interview record was derived from the statement. An analysis of the shared vocabulary in the two Molloy texts showed that the highlighting in the two extracts presented earlier understates the similarities between the two texts – a closer examination revealed that there was in fact not one single item in Molloy’s statement, neither lexical nor grammatical, which did not also occur in the interview record. We have only found that degree of overlap on one other occasion and that was when a Copycatch analysis of two student essays showed a shared vocabulary of 97 percent. Closer inspection of the two essays showed that in fact the students had submitted identical essays – the 3 percent difference was accounted for by mis-spellings which the program coded as different words..

**Closing Observation**

I have tried in the space available to give an idea of the diversity of the problems tackled and the techniques used by forensic linguists. Details of many other cases not referred to here can be found in Cotterill (2002), Coulthard (1994), Dumas (2002), Gibbons (1994, 2003), Levi (1994a, b), McMenamin (2002) and Shuy (1993, 1998, 2002), while an excellent article on the areas in which the forensic linguist is and is not currently allowed to give evidence in the United States, is Tiersma and Solan (2002).

## References

- Bates, E, Kintsch, W, Fletcher, C R, and Giulani, V 1980, 'The role of pronominalisation and ellipsis in texts: some memorisation experiments', *Journal of Experimental Psychology: Human Learning and Memory* 6, 676-691.
- Cotterill J (ed) 2002, *Language in the Legal Process*, London, Palgrave
- Coulthard, R M 1994 'Powerful evidence for the defence: an exercise in forensic discourse analysis', in J Gibbons (ed), 414-42.
- Coulthard, R M 2002 'Whose voice is it? Invented and concealed dialogue in written records of verbal evidence produced by the police' in J Cotterill (ed), 19-34.
- Coulthard, R M (in press) 'Author identification, idiolect and linguistic uniqueness', *Applied Linguistics*
- Dumas B 2002 'Reasonable doubt about reasonable doubt: assessing jury instruction adequacy in a capital case', in J Cotterill (ed), 246-259
- Gibbons J (ed) 1994 *Language and the Law*, London: Longman.
- Gibbons J 2003 *Forensic Linguistics: an Introduction to Language in the Justice System*, Oxford, Blackwell
- Grice H P 1975 'Logic and conversation', in P Cole and J Morgan (eds), *Syntax and Semantics III: Speech Acts*, New York: Academic Press, 41-58.
- Hjelmquist E 'Memory for conversations', *Discourse Processes*, 1984, 7:321-336
- Hjelmquist E and Gidlung A 'Free recall of conversations', *Text*, 1985, 3:169-186.
- Johnson A. 1997 'Textual kidnapping – a case of plagiarism among three student texts', *Forensic Linguistics*, 4 ii 210-25
- Kaplan J P, Green G M, Cunningham C D and Levi J N 1995, 'Bringing linguistics into judicial decision making: semantic analysis submitted to the US Supreme Court', *Forensic Linguistics*, 2, i.
- Keenan, J.M., MacWhinney, B, and Mayhew, D. 1977 'Pragmatics in Memory: a study of natural conversation', *Journal of Verbal Learning and Verbal Behavior* 16 549-560.
- Levi J N 1993 'Evaluating jury comprehension of the Illinois capital sentencing instructions', *American Speech* 68, I, 20-49
- Levi J N 1994a *Language and the Law: A Bibliographical Guide to Social Science Research in the USA*, Chicago: American Bar Association.
- Levi J N 1994b Language as evidence: the linguist as expert witness in North American Courts, *Forensic Linguistic*, 1, i, 1-26.
- Levi J N and Walker A G (eds) *Language in the Judicial Process*, London: Plenum 1990
- McMenamin G 2002 *Forensic Linguistics: Advances in Forensic Stylistics*, London, CRC Press
- Prince E 1981 'Language and the law: a case for linguistic pragmatics', *Working papers in Sociolinguistics*, Austin: Southwest Educational Development Laboratory, 112-160.
- Shuy R 1993 *Language Crimes: the Use and Abuse of Language Evidence in the Courtroom*, Cambridge MA, Blackwell
- Shuy R 1998 *The Language of Confession, Interrogation and Deception*, London, Sage
- Shuy R 2002 *Linguistic Battles in Trademark Disputes*, New York, Palgrave
- Slembrouck S. (1992) 'The parliamentary Hansard "verbatim" report: the written construction of spoken discourse', *Language and Literature* 1, 2, 101-19
- Solan L 1998 'Linguistic experts as semantic tour guides' *Forensic Linguistics*, 5 ii, 87-106
- Tiersma, P and Solan L 2002. The Linguist on the Witness Stand: Forensic Linguistics in American Courts, *Language* 78, 221-39.
- Woolls D 2002 *Copycatch Gold* a computerised plagiarism detection program for more details see [www.copycatchgold.com](http://www.copycatchgold.com).
- Woolls D 2003 'Better Tools for the Trade and how to Use them', *Forensic Linguistics*, 10 i 102-112.
- Woolls D and Coulthard R M 1998 'Tools for the Trade', *Forensic Linguistics*, 5 i 33-57.

**Key Words confession, expert, forensic, language, legal, linguistics, opinion, plagiarism, witness.**

## **Abstract**

This article illustrates the problems faced and the techniques used by the linguist when acting as an expert witness. Examples are drawn from a wide variety of cases, ranging from disputes about the meaning of individual morphemes in a trademark case and individual words in jury instructions, through grammatical complexity in a letter and a statute to the 'ownership' of particular words and phrases in two plagiarism cases and accusations of the fabrication of a whole text in a murder case. Linguists are seen to use evidence derived from corpora and questionnaires as well as insights drawn from morphology, grammar, lexis, pragmatics, semantics and discourse and text analysis to reach and support their opinions.

## **2<sup>nd</sup> Abstract**

This article illustrates the problems faced and the techniques used by the linguist when acting as an expert witness. Examples are drawn from a wide variety of cases. The article first exemplifies disputes about the meaning of individual morphemes in a trademark case, where the American burger chain McDonalds claimed ownership of the morpheme 'Mc' on the grounds that they had invented a 'McLanguage', and about the interpretation of individual words like 'sufficient', 'preclude' and 'impairment' in jury instructions and health insurance proposals, where convincing evidence is offered that co-operative readers would not have derived the meaning intended by the legal authors of the texts. The article then examines the contribution linguists made in two specific cases to resolving questions about the degree of grammatical complexity in a disputed letter and a statute whose interpretation had been appealed, before moving on to use the concept of linguistic uniqueness to help resolve the question of the 'ownership' of particular words and phrases in two cases of suspected plagiarism. The concepts used in the plagiarism cases are then used to resolve a dispute about whether a whole interview record had been fabricated by the police in a murder case. Throughout the article examples are provided of the wide range of techniques that forensic linguists have developed and now use to reach and support their opinions, ranging from evidence derived from corpora and questionnaires to insights drawn from morphology, grammar, lexis, pragmatics, semantics and discourse and text analysis.

## **Opening**

Discourse is a major instrument of power and control and Critical Discourse Analysts feel that it is part of their professional role to investigate, reveal and clarify how power and discriminatory value are inscribed in and mediated through the linguistic system' (Caldas-Coulthard and Coulthard 1996 Preface)

Forensic linguistics is an area where linguists interface with the real world. As part of their job they sometimes confront the linguistic inequalities imposed by big corporations and other power groups like the police and the law and occasionally redress the balance, as many of the examples below, taken from real cases, demonstrate.

**Ending**

Critical Discourse Analysis is essentially political in intent with its practitioners acting upon the world in order to transform it.

As is evident in the examples above, forensic linguists have had notable successes: they have forced insurance companies and law-makers to produce texts which are more transparent to the intended lay audience, they have devised ways to defend authors against raids from text thieves and they have successfully protected many people wrongfully accused from the consequences of falsified evidence.