

Susan Blackwell¹

WHY FORENSIC LINGUISTICS NEEDS CORPUS LINGUISTICS

ABSTRACT: While corpus linguistics has existed since the 1960s, Forensic Linguistics is a relatively new discipline, involving both linguistic evidence in court and wider applications of linguistics to legal texts and discourses. Computer corpora of natural language may be marked up in various ways, grammatically tagged, parsed, lemmatised and analysed with concordance, collocation and other specialist software. In the relatively short history of forensic linguistics, its exponents have often employed corpus linguistics techniques in order to throw light on questions like disputed authorship. However, the corpora employed have been general ones such as the Cobuild “Bank of English”, rather than purpose-built databases of language used in legal contexts, with the result that such research sometimes raises more questions than it answers. Conversely, corpus linguists have from time to time incorporated data from legal settings into their collections; but they have tended to use these resources as the basis for sociolinguistic or historical linguistic research rather than as a means of exploring topics in language and law.

This paper makes a plea for these two fields, which are both already cross-disciplinary, to join forces and create a purpose-built corpus for forensic linguistics. It illustrates how corpus techniques may be successfully applied to questions of disputed authorship, citing both hypothetical and actual examples. It ends with an outline of the kinds of texts which a proposed new corpus for Forensic Linguistics should contain and the tools required to exploit it effectively.

KEY WORDS: corpus, linguistic, forensic linguistic

1 Some definitions

1.1 Forensic linguistics

The term “Forensic Linguistics” was probably first coined by Svartvik (1968), but acquired currency in the 1990s with a series of seminars and the establishment of the International Association for Forensic Linguists (IAFL)

¹ University of Birmingham, UK, s.a.blackwell@bham.ac.uk

and the journal *Forensic Linguistics: the Journal of Speech, Language and the Law*². The term can be said to have both a narrow and a broad definition. The former covers the use of linguistic evidence in court, concerning for example disputed confessions (Coulthard 1994), trademark disputes (Okawara, 2006 and forthcoming), threats and attempts at extortion (Shuy 1993), taped conversations allegedly offering bribes (Shuy 2005), suicide notes (Shapero, forthcoming), disputed authorship and alleged plagiarism (Kniffka 2000). The broader definition covers all areas of overlap between language and law, including courtroom interpreting (Berk-Seligson 1990), courtroom discourse (Solan 1993; Tiersma 1999), linguistic minorities in the legal process (Eades 1994) and children in the legal process (Walker 1999). These lists are by no means exhaustive but serve to give a flavour of the wide range of research currently being undertaken in Forensic Linguistics (henceforth FL).

1.2 Corpus Linguistics

The most concise definition available is probably that of Renouf: “The term ‘corpus’ will be used to refer to a collection of texts, of the written or spoken word, which is stored and processed on computer for the purposes of linguistic research.” (Renouf, 1987:1)

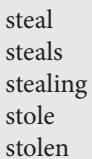
The first corpora in this sense were the Brown corpus (Kucera and Francis 1967) and the Lancaster-Oslo-Bergen (LOB) corpus (Garside, Leech and Sampson ed.s, 1987). These consisted of 1 million words of US and British English, respectively, from published sources in the year 1961. Since then a number of general corpora have been built including the COBUILD Corpus, now known as the “Bank of English” (Sinclair ed., 1987); the British National Corpus (Burnard, L. ed., 1995) and the International Corpus of English (Greenbaum ed., 1996). As well as general linguistic research aimed at achieving a more accurate description of natural language, larger corpora have often been used for lexicographic purposes, most famously the COBUILD range of dictionaries and grammar books such as the *Collins Cobuild English Dictionary for Advanced Learners* (2001).

² The journal was founded in 1994 as *Forensic Linguistics* but the title changed in 2003 to *The International Journal of Speech, Language and the Law* to reflect a broadening of academic coverage and readership.

The field of corpus linguistics has also spawned a plethora of specialised corpora, including the International Corpus of Learner English (Granger, 1994); the CHILDES database (really a collection of sub-corpora) of child language (MacWhinney 1995); the Bergen Corpus of London Teenage Language (COLT) (Stenström et al. 1998), the Leeds Corpus of English Dialects (Klemola and Jones 1999) and the Helsinki corpus of English texts (Kytö 1994). Specialist corpora may be used to study a language at a particular period in time or in a particular region, or to examine the linguistic patterns in a particular author or text type.

Corpora rarely consist of plain text, although Sinclair’s concept of a “monitor corpus” (Sinclair 1982) envisaged almost-raw text flowing through a series of software filters to extract information from it and then being discarded. More commonly, corpora are marked up with various kinds of information such as the sex of the speaker or the date of the text; less trivially they may incorporate part-of-speech tagging or higher constituent tagging (syntactic parsing). It may be considered desirable to lemmatise the text, in order to enable the linguist or lexicographer to retrieve all forms of a particular word in a single search expression; indeed, in highly-inflected languages such as Hungarian lemmatisation and consequent morphological mark-up are virtually essential (Pajzs 1991). In the case of parallel corpora, “hooks” into the translation equivalents in another language are embedded into the text (Botley et al. 2000). Markup may be carried out automatically or manually: usually some combination of the two is employed. Figure 1 illustrates the various forms related to the lemma “steal” while Figure 2 shows a fairly basic form of mark-up, “COCOA” tagging for the now-superseded Oxford Concordance Program (Hockey and Martin 1987), applied to the official court transcript of an English trial to label the speaker of each utterance and to mark certain text as “comment” to be excluded from any processing.

Figure 1: Lemma “steal”



steal
steals
stealing
stole
stolen

Figure 2: Corpus data with mark-up

```
[ Wednesday, 23rd January, 1985 ]
[RANSFORD NEDRICK: Sworn.
Examined by Lord Hooson, Q.C. ]
<P 01>
<S LH> Mr. Nedrick, when I ask you questions will you please face the jury as much
as you are now. What is your full name?
<S RN> Ransford Delroy Nedrick.
<S LH> Your normal address?
<S RN> 40 Hallett Drive.
<S LH> That is a council flat in Wolverhampton?
<S RN> A council maisonette.
<S LH> Your age?
<S RN> 49.
```

Once a corpus has been created and marked up ready for exploitation, specialist software can be used to analyse it to produce wordlists, concordances, collocation sets and more: the CLAWS suite of programs for the LOB corpus and the CLAN suite for CHILDES are two well-known examples. Kirk (1994) gives a good overview of the various types of corpora, annotation and processing software.

2. Forensic linguistics meets Corpus linguistics

2.1 Hypothetical scenario: The case of the disputed confession

Disputed confessions are a fairly frequent phenomenon in FL as narrowly defined. A typical problem involves a suspect denying that part or all of an incriminating statement consists of his/her own words. A forensic linguist tasked with evaluating the plausibility of such a claim will usually seek to obtain an undisputed sample from the suspect for comparison, along with an undisputed sample from anyone suspected of being the real source of the text, such as a police officer.

Let us examine a fictitious example of this “genre”, in which the text in question contains 4 instances of a relatively rare lexical item, such as “vehicle”, which does not appear at all in the defendant’s undisputed statement. However, it appears 5 times in the police officer’s witness statement, as shown in Table 1.

Table 1: Hypothetical example: “car” vs. “vehicle”

	Disputed statement	Undisputed statement	Police witness statement
“car”	2	6	2
“vehicle”	4	0	5

The linguist’s native-speaker knowledge of English tells her that “vehicle” is a rarer word than “car” and moreover belongs to a more formal register, likely to be favoured by officers of the law. It is tempting, faced with the evidence in Table 1, to conclude that the confession is the work of the police officer. However, such a conclusion would be unwarranted without taking into account the relative text size, as shown in Table 2.

Table 2: Hypothetical example: “car” vs. “vehicle” with text size

	Disputed statement	Undisputed statement	Police witness statement
“car”	2	6	2
“vehicle”	4	0	5
Total words	3,000	2,000	3,500

It now appears that the absence of the word “vehicle” in the undisputed statement may be due to its shorter length and not to the rarity of the word. The fact that “car” occurs here with a higher frequency than in the longer, disputed statement would seem to reinforce our original suspicions, but nonetheless it is hard to be sure. This is the kind of situation where a reliable corpus can be a godsend.

Table 3: Frequencies of “car” and “vehicle” in the Bank of English

	BBC radio	USA radio	General spoken	Total spoken
“car”	2,535	1,896	1,205	5,636
“vehicle”	618	269	121	1,008
Text size	20 m	10 m	4 m	34 m

The data in Table 3 are actual and not hypothetical. They indicate that for both British and US English the lexical item “car” is 4-5 times as frequent as “vehicle” in radio broadcasts, while in general spoken discourse it is 10 times as frequent. The forensic linguist can be confident, after all, that there

is something distinctly odd about a text in which “vehicle” appears more frequently than “car”, at least if its origins are supposed to be in speech rather than writing.

2.2 The Google “corpus” - a quick and dirty solution?

Some linguists now use search engines such as Google as a tool for checking the relative frequency of contrasting words in modern English, as a rough-and-ready general corpus; or as a means of demonstrating that phrases one might think were common are in fact quite unique to specific texts or speakers/writers. Subjecting “car” and “vehicle” to a Google enquiry for domain names ending in “.co.uk” and “.com”, on the assumption that these will yield British and US data respectively, is likely to produce statistics like those in Table 4.

Table 4: Results of a search for “car” and “vehicle” using Google

	.co.uk	.com
“car”	95,900,000	953,000,000
“vehicle”	17,300,000	244,000,000

It is gratifying to find that the general proportions of the lexical items under scrutiny are confirmed by a trawl of the World-Wide Web. However, we have no way of knowing the total “text size” of the pages from which these figures were returned. There are many questions which one can ask of a corpus but not of a search engine, such as “Do men and women use this word equally?” (requiring mark-up for the sex of the speaker/writer); “Is this feature more common in speech or writing?” (requiring control of the corpus collection), and “What words appear most frequently two places to the left of the key word?” (requiring collocation software). With part-of-speech tagging it is even possible to interrogate a corpus about particular usages of a syntactically ambiguous word, requesting all occurrences of, for instance, “judge” used as a verb but not a noun. None of this is possible with an Internet search engine.

Two case studies

3.1 Daniel Raphaie

Mr. Raphaie came from Iran to Britain in 1978 as a student. His first language is Farsi (Persian). He remained in the country, married, had a child, took various jobs and fell into bad company. In 1988 the police raided the flat of his former wife, where he was staying at the time. He was charged with dealing in drugs and stolen goods although no hard drugs were found in the flat: he was convicted solely on the statements of the police that at the time of the search he had admitted having just flushed a quantity of heroin down the toilet.

In a linguistic examination of the alleged incriminating statements by Raphaie, it was noted that these included the following:

“*Look* I didn’t want to get caught holding it.”

“*Look* I might get six or seven grammes, maybe more, every two days.”

There is no evidence from much longer, undisputed samples of Raphaie’s speech that he ever uses *look* as a discourse marker; yet in this supposedly contemporaneous transcript of the search, which contains just 232 words attributed to him, he is supposed to have used it twice.

The Cobuild Bank of English was searched for instances of the word “look” used as a discourse marker: the results are shown in Table 5.

Table 5: Discourse marker “Look” in the Bank of English

Total instances of “look” as discourse marker: Of these:	949
Total primary occurrences:	318
Total secondary occurrences:	631
Total corpus size: (spoken British English)	20,181,050

It was found that the use of discourse marker “look” could be divided into primary and secondary usages (Blackwell, 2000). Primary occurrences were instances of the speaker using “look” directly in addressing the hearer, as in “*Look*, you’ve got to be here on Sunday.” Secondary occurrences, by contrast, were examples of quoted speech in which use of the feature was

being attributed to someone else, or to the speaker at some previous time: “Yeah but the ANC are saying *look* equality for blacks.” Table 5 shows that secondary usage is twice as frequent as primary usage: in other words, “*look*” is twice as likely to appear in a reconstruction of someone’s purported speech than in their actual original words.

This in itself might not be sufficient to support Mr. Raphaie’s allegations that the supposed contemporaneous transcription of his speech at the time of the police raid was nothing of the kind. However, there is other evidence that the introduction of discourse markers is one way of giving a veneer of authenticity to texts which are not contemporaneous (Coulthard 1996). One may note, moreover, that “*look*” is a confrontational item, unlikely to be used by a suspect to a police officer. Finally, Lindsay and O’Connell (1995) have observed that transcribers tend to omit all discourse markers due to pressures of real-time writing and the lack of psychological saliency of such items for the hearer. The sum total of this linguistic and metalinguistic evidence was considered sufficient to discredit the police claims that the interview with Raphaie had been transcribed contemporaneously at the time of the search. This does not mean, of course, that the content was a total fabrication: it may have been based on a real speech event but written up some time afterwards. In that case, however, one is justified in asking why the police wrote up the alleged interaction in the Exhibits Book, claiming that this was the only book available to write in at the time the transcription was made.

This analysis of discourse markers was made available to Mr. Raphaie’s legal team and submitted as part of the evidence to the Court of Appeal. It is believed to be the first occasion when an appeal was heard in the English courts on the grounds of linguistic evidence. In the event, Mr. Raphaie’s appeal was allowed on legal grounds without the linguistic evidence being put before the court.

Eddie Gilfoyle

Eddie and Paula Gilfoyle were a married couple living in Upton, Wirral, England. On 4th June 1992 Paula Gilfoyle’s body was found hanging in the garage of her home. She was eight and a half months pregnant. Despite the fact that a suicide note was found in Paula’s handwriting, Eddie was prosecuted for her murder. The prosecution claimed that Eddie, a hospital nurse, had tricked Paula into writing the note and then murdered her, in effect using the

note as his alibi. The jury believed this and convicted him. He and his family and friends are still protesting his innocence.

Goutsos (1995) compared the language of the problematic suicide note with samples of Eddie’s writing and found a number of apparently incriminating phrases which were common to both, including “rebuild your life”, “turn back the clock” and “if I could, I would”. There was also a tendency in both texts to use couplets such as “cheated and lied”, “family and friends”, “pain and suffering” / “suffering and pain”, “hurt and suffering” and “pain and heartache”. It is tempting to conclude from this that Eddie was indeed the originator of the “suicide” note and had justly been convicted. However, as Table 6 shows, the Bank of English reveals that some of these phrases are common collocations in general use.

Table 6: Phrases from the “suicide note” in the Bank of English

rebuild + life	7
turn the clock back	16
could + would	125
pain and suffering	24
suffering and pain	3
pain and heartache	1

(from Goutsos, 1995)

Worse was to come. Goutsos recalls the investigations of the Birmingham University Forensic Linguistics group:

“We found that the surprising phrase *Goodnight and God bless* which appeared in the closing off section of the disputed suicide texts is in fact a common feature of death announcements in the press of the area where the texts originated.” (Goutsos 1995:108)

Further problems emerged when the nature of the texts being compared with each other was taken into account:

“One major problem is that our corpora were significantly skewed. The texts involved were not alike with regard to almost any parameter among the components of speech events as formulated by Hymes (1974) “... To achieve register objectivity, we would have to refer to comparable corpora with different variables such as a corpus of letters written by other people or a corpus of suicide notes.” (Goutsos, 1995:107)

Thus, although at first sight there had appeared to be a number of incriminating similarities between Eddie's language and that of the suicide note, on closer investigation this conclusion was not justified. In the first place, the phrases are not so unusual in colloquial English; secondly, it has to be borne in mind that people who live together intimately probably tend to converge in their language use; and thirdly, the linguists were not comparing like with like and did not have corpora which would enable them to do so.

The outcome for Mr. Gilfoyle was a less happy one than for Mr. Raphaie: his two appeals against conviction were unsuccessful and he remained in jail, vehemently protesting his innocence.

4. General vs. Specialised Corpora for Forensic Linguistics

While general corpora such as the British National Corpus or Bank of English may be adequate for some FL purposes, as in the Raphaie case, it is clear from Goutsos' remarks cited above that such corpora cannot answer questions such as whether or not "Goodnight and God bless" is a likely way to end a suicide note. There is a need for a specialised database of texts which can be used to research issues in language and law. A wide range of issues could be investigated with such a resource, such as differences between the language of prosecution and defence lawyers, or between expert witnesses and eye-witnesses. The language of judges, which has already been the focus of examination (Solan 1993), could be studied more effectively if a machine-readable, marked-up corpus were available to researchers.

Admittedly some legal language is already available in corpus form, most notably the proceedings of the Old Bailey from 1674-1834 which have recently been placed online. However, to date this material has been used mainly for historical linguistic and sociolinguistic research, as a rich source of information on variation and change at the interface of early modern and modern English in London. Dr Magnus Huber of the University of Giessen, for instance, is exploiting the Old Bailey data to analyse differences correlating with the social parameters of age, gender, place of origin and social status (Huber 2007). Similarly, the International Corpus of English (ICE) contains ten 2,000-word texts of "legal presentations" in each category (Nelson 1996), but the main purpose of this is to compare the language of such presentations across various parts of the world in which English is spoken rather than to investigate the language of the courtroom *per se*.

It is to be hoped that some of the existing collections of legal texts can be incorporated into the proposed corpus for forensic linguistic research.

Table 6 offers a tentative list of text types which might usefully be included in a specialist FL corpus.

Table 6: A corpus for Forensic Linguistics: Text types

suspects' statements to police, contested
suspects' statements to police, uncontested
suspects' statements to solicitors
suspects' statements to linguists
witnesses' statements to police, contested
witnesses' statements to police, uncontested
witnesses' statements to solicitors
witnesses' statements to linguists
police statements used in court
police reports for internal consumption
transcripts of court proceedings
threatening/extortion letters
transcripts of threatening/extortion telephone calls
suicide notes
"martyrdom" videos and letters

Table 7 indicates some of the features which it would be desirable to include in the mark-up of such a corpus.

Table 7: A corpus for Forensic Linguistics: annotation

age of text originator ²
sex of text originator
role of text originator (e.g. suspect, eye witness, police officer)
first language of text originator
other languages spoken/used by text originator
how long text originator has been resident in the country concerned

³ I have used the term "text originator" to indicate the person responsible for the language of the text in question. This is not necessarily the same as the person who writes it, as can be seen from the Gilfoyle case where a suicide note was allegedly dictated by Eddie, the text originator and alleged perpetrator, to Paula, the scribe and alleged victim.

text type (statement, court proceedings etc.)
whether text was written by text originator or transcribed by someone else
if transcribed: whether from speech or tape recording age, sex and role of transcriber(s)
if written: details of writing (e.g. handwritten; typewritten; word-processed).
if word-processed, were spell- or style-checkers available?
purpose of the text (e.g. evidence in court, background info. for solicitor)
whether other texts from same text originator exist in the corpus

5. Conclusion

This paper has attempted to provide an overview of the kinds of problems which face the forensic linguist. While some questions may be resolved satisfactorily by reference to language data readily available on the Internet or in a general machine-readable corpus, there remain thorny issues which can only be discussed properly when a specialist corpus for language and law becomes available. We cannot state with any degree of confidence that a disputed suicide note is a forgery until we have an idea of what a “normal” suicide note looks like. We may be sure that a particular utterance, such as “I then proceeded to exit the vehicle”, is so formal that it can only have been produced by a police officer; but a cross-examining lawyer is likely to put it to any expert witness stating this that perhaps in the formal setting of a police interview, suspects (especially seasoned ones with previous experience of such speech events) are likely to accommodate their language to that of the interviewing officer. When a person’s liberty and reputation are at stake, mere linguistic intuition is not good enough: there needs to be a solid basis on which to draw reasonable conclusions, preferably supported by quantitative data which can be subjected to statistical tests. The construction of a purpose-built corpus for research in language and law will not only nurture the academic curiosity of linguists, but should serve the wider interests of justice.

Bibliography

Berk-Seligson, S., 1990, *The Bilingual Courtroom: Court Interpreters in the Judicial Process*. Chicago: University of Chicago Press.

- Blackwell, 2000, "Looking up *look*: Discourse Markers in the Bank of English". In Kirk, J. ed., *Corpora Galore: Analyses and Techniques in Describing English*. Amsterdam: Rodopi.
- Botley, S., McEnery, A. and Wilson, A. (eds). 2000, *Multilingual Corpora: Teaching and Research*. Amsterdam: Rodopi.
- Burnard, L. ed., 1995, *Users' Reference Guide to the British National Corpus*. Oxford: Oxford University Computing Service.
- Collins Cobuild English Dictionary for Advanced Learners, 2001. London: Harper Collins.
- Coulthard, M., 1994, "Powerful evidence for the defence: an exercise in forensic discourse analysis". In Gibbons, ed., *Language and the Law*. London and New York: Longman.
- Coulthard, M., 1996, "The official version: Audience manipulation in police records of interviews with suspects". In Caldas-Coulthard and Coulthard (ed.s), *Texts and Practices: Readings in Critical Discourse Analysis*. London: Routledge.
- Eades, D., 1994, "A case of communicative clash: Aboriginal English and the legal system." in Gibbons, ed., *Language and the Law*, London: Longman.
- Fox, G., 1993, "A Comparison of 'Policeseak' and 'Normalspeak': a Preliminary Study". In Sinclair, Hoey and Fox (ed.s), *Techniques of Description: Spoken and Written Discourse. A festschrift for Malcolm Coulthard*, London: Routledge.
- Garside, Leech and Sampson ed.s, 1987, *The computational analysis of English: a corpus-based approach*. London: Longman.
- Goutsos, D., 1995, Review article: "Forensic Stylistics". In *Forensic Linguistics* vol. 2 no. 1.
- Granger, S. 1994, "The Learner Corpus: A Revolution in Applied Linguistics". *English Today* 39(10/3), pp. 25-29.
- Greenbaum, S., ed., 1996, *Comparing English worldwide: the International Corpus of English*. Oxford: Clarendon Press.
- Hockey, S. and Martin, J., 1987, "The Oxford Concordance Program Version 2", in *Literary and Linguistic Computing*, 2, pp. 125-131.
- Huber, M., 2007, 'Playing tag with Old Bailey. Creating a corpus of 18th-century spoken English'. In Meurman-Solin, A. and Nurmi, A. (eds.): *eVARIENG 1: Proceedings of the workshop on corpus annotation at ICAME 27, Helsinki, 24-28 May 2006*. Helsinki: VARIENG.
- Hymes, D., 1974, "Ways of Speaking", in R. Bauman and J. Sherzer (ed.s), *Explorations in the Ethnography of Speaking*. Cambridge: C.U.P.

- Ihalainen, O., M. Kytö and M. Rissanen, 1987, "The Helsinki Corpus of English Texts: Diachronic and Dialectal: Report on work in progress", in *Corpus Linguistics and Beyond. Proceedings of the Seventh International Conference on English Language Research on Computerized Corpora*, ed. W. Meijs. Amsterdam: Rodopi.
- Kirk, J., 1994, "Taking a Byte at Corpus Linguistics", in *Entering Text*, edited by Flowerdew, L. and Tong, A.K.K., pp. 18-43.
- Klemola, J. and Jones, M.J., 1999, "The Leeds corpus of English dialects – project". In *Leeds Studies in English* 30: 17-30.
- Kniffka, H., 2000, "Anonymous Authorship Analysis without Comparison Data? A Case Study with methodological impact". In: *Linguistische Berichte* 182, 179-198.
- Kucera, H. and Francis, W.N., 1967, *Computational analysis of present-day American English*. Providence, R.I.: Brown Univ. Press.
- Kytö, M. 1994, *Manual to the Diachronic Part of the Helsinki Corpus of English Texts: Coding Conventions and Lists of Source Texts*, 2nd ed. Helsinki: Helsinki University Press for Department of English, University of Helsinki.
- Lindsay, J. and O'Connell, D.C., 1995, "How do transcribers deal with audio recordings of spoken discourse?", in *Journal of Psycholinguistic Research*, 2:101-115.
- MacWhinney, B., 1995, *The CHILDES-Project: Tools for Analyzing Talk*. Second edition. Hillsdale, NJ: Lawrence Erlbaum
- Nelson, Gerald 1996, The Design of the Corpus. In S. Greenbaum (ed.), pp. 27-35.
- Okawara, M.H., 2006, *A linguistic analysis of Some Japanese trademark cases*. Ph.D. thesis, University of Sydney.
- Okawara, M.H., forthcoming, "Legal Japanese viewed through the Unfair Competition Prevention Law". In Gibbons, Prakasham and Tirumalesh, ed.s, *Justice and Language*. Delhi: Longman Orient.
- Pajzs, J., 1991, "The Use of a Lemmatized Corpus for Compiling the Dictionary of Hungarian", in *Using Corpora: proceedings of the 7th annual conference of the UW Centre for the New OED and Text Research*. Waterloo: UW Centre for the New OED and Text Research.
- Renouf, A., 1987, "Corpus Development". In Sinclair, ed.
- Shapero, J.J., forthcoming, Ph.D. thesis, University of Birmingham.
- Shuy, R., 1993, *Language Crimes: the Use and Abuse of Language Evidence in the Courtroom*. Cambridge, MA and Oxford: Blackwell.
- Shuy, R., 2005, *Creating Language Crimes: How Law Enforcement Uses (and Misuses) Language*. Oxford: Oxford University Press.
- Sinclair, J. 1982. "Reflections on computer corpora in English language research." In

- Computer corpora in English language research*, ed. Johansson, S.: 1-6. Bergen.
- Sinclair, J.M., ed., 1987, *Looking Up: An Account of the COBUILD Project in lexical computing*. London: HarperCollins.
- Solan, L., 1993, *The Language of Judges*. Chicago: U. of Chicago Press.
- Stenström, Anna-Brita, Andersen, G., Hasund, K., Monstad, K. and Aas, H., 1998, *User's manual to accompany the Bergen Corpus of London Teenage Language (COLT)*. Department of English, University of Bergen, Norway.
- Svartvik, J. 1968, *The Evans Statements: A case for forensic linguistics*, Gothenburg: Gothenburg University Press
- Tiersma, P.M., 1999, *Legal language*. Chicago: University of Chicago Press.
- Walker, A.G., 1999, *Handbook on Questioning Children: a Linguistic Perspective*. Washington: American Bar Association.