

Author Posting. (c) 'Sage publications'. This is the author's version of the work. The definitive version was published in *Language and Literature*, 11, 4, 325-55, 2002.

Revisiting the Notion of Faithfulness in Discourse Presentation Using a Corpus Approach¹

*Mick Short and Elena Semino, Lancaster University and
Martin Wynne, University of Oxford, UK*

Abstract

A number of recent studies have argued that the notion of faithfulness to an original should be abandoned in models of discourse presentation, and particularly in accounts of *direct speech* presentation. This has coincided with a shift of attention in the study of discourse presentation from written to spoken data. This article discusses the arguments that have been made against the notion of faithfulness, and proposes a context-sensitive account of this notion, and of its relation to the various clines of discourse presentation and their categories. Our account is prompted partly by the results of a corpus-based approach to the study of discourse presentation, and partly by a qualitative analysis of a set of newspaper articles on a particular news story from outside this corpus, which we undertook to provide a check on the conclusions we had reached from our corpus study. We believe that if a general account of discourse presentation is to be reached, similarities and differences across a wide range of texts and text types need to be examined. Our corpus work, which involves careful and systematic comparison of a balanced set of written fictional, news and (auto)biographical narratives is offered as a contribution towards the general account referred to above. We also believe that if such a general account/theory is to be reached, scholars will need a clearer and more consistent application of the various descriptive terms which have been used in this area of study during the 20th century, in particular (a) 'discourse', (b) 'speech', 'thought' and 'writing'; and (c) 'report', 'presentation' and 'representation'.

Keywords: direct discourse/speech/thought/writing; discourse report/presentation/representation; faithfulness; indirect discourse/speech/thought/writing; verbatim quotation

1 Introduction

This article is concerned with the role of the notion of faithfulness to an original in the report/(re)presentation of other people's words and thoughts.² The issues we are going to discuss are perhaps best introduced by means of an example. The extract below is taken from a newspaper article discussing the closure in 1994 of a newly-opened children's theme park in Bare, a suburb of a seaside resort in Northern England called Morecambe:

A woman walking her Yorkshire terrier along the empty paths said: 'I'm very sad. This place would have brought a lot of people to Morecambe.' The town's newspaper said: 'Imagine how we look now. Morecambe should hang its head in shame.'

(*Independent on Sunday*, 'Bare ladies' protest puts end to Crinkley Bottom', 4/12/1994)

The extract contains two instances of what is traditionally known as 'direct speech'. The first ('I'm very sad. This place would have brought a lot of people to Morecambe.') relates to the words *uttered* by an unnamed local resident, whereas the second one ('Imagine how we look now. Morecambe should hang its head in shame.') reports the *written* words of what is described as 'the town's newspaper'. Following the most traditional view of this kind of reporting, one would have to conclude that, because quotation marks are used in both cases, both stretches of text are faithful word-by-word reproductions of their respective originals. On the other hand, some recent revisions of the traditional view (e.g. Fludernik 1993; Sternberg 1982a, 1982b; Tannen 1989) argue that, because direct speech can be used when faithful reproduction is neither possible nor relevant (e.g. in casual conversation), its presence can never be associated with faithfulness to an original. Within this view, both of the stretches within quotation marks in the above extract are seen as constructions on the part of the reporter.

Critiques of the importance of faithfulness in (the theory of) discourse presentation vary considerably in terms of data and perspective. Tannen's (1989) work concentrates more or less exclusively on spoken data, and emphasizes the constructed nature of direct reports in informal conversation. Sternberg (1982a, 1982b) and Fludernik (1993) primarily draw their examples from written fictional narratives, but also generalise their conclusions well outside fiction (e.g. Fludernik 1993: 434). Clark and Gerrig (1990) approach the phenomenon of reporting from the perspective of the philosophy of language, and argue that direct quotations should be seen as 'demonstrations' which are inevitably selective in what aspects of the original they include and what they miss out. Collins (2001) starts from a historical-pragmatic perspective, and demonstrates the constructed and creative nature of direct reports of speech in Medieval Russian trial transcripts. In introducing a collection of studies focusing on 'institutional discourse', Baynham and Slembrouck (1999) also reject the notion of faithfulness to an original in the theory and analysis of what they call 'speech representation', but recognize the importance of a range of contextual factors in the report/(re)presentation of the words of others:

The major conclusion to be drawn from this brief survey is undoubtedly that researchers need to develop modes of enquiry into speech representation which stress the situational dependency of the forms and functions of speech representation. (Baynham and Slembrouck 1999: 448)

In this article we argue that there is a middle way between the traditional, writing-dominated, faithfulness-oriented, position on the one hand, and the more recent tendency to dismiss the notion of faithfulness on the other. This middle way is based on a context-sensitive account of discourse presentation (as advocated by Baynham and Slembrouck 1999), and focuses on how contextual factors affect the relevance of the notion of faithfulness to the production and reception of (stretches of) discourse presentation.

Within this approach, the two instances of ‘direct speech’ in the example above are quite different from each other. The first one is a *written* report of *speech* in the direct speech (DS) form, although the source is an unnamed member of the public, whose *exact* original words are unrecoverable and are less important for the article than the general attitude that is attributed to her. On the other hand, the second instance is best described as ‘direct writing’ (DW), since the source, like the report itself, is in the written medium: a newspaper which, although unnamed, is identifiable to at least part of the article’s readers, who could in principle check the accuracy of the quotation. This checkability, we argue, will lead to stronger expectations on the part of writer and readers that the quotation will be faithful. Whereas the first reported extract would be very unlikely to cause the reporter legal problems (no identifiable individual is specified, and there is unlikely to be an independently verifiable way of checking what the woman with the terrier said), it is possible to imagine a situation where an inaccuracy in the second quotation (the source of which was specific, written and public) could result in a court case.

In general terms, then, we would want scholars to take account of, and to describe systematically, the variety and complexity of the phenomenon which is made to seem deceptively simple when it is referred to as ‘discourse report’. We agree with Clark and Gerrig (1990) that what they call the ‘verbatim assumption’ for direct discourse arises from what Linell (1982) has called the ‘written language bias’ of linguistics and philosophy (see also Collins 2001), and we welcome the present attempts to show in detail how direct presentations of discourse in spoken language are signalled and function (e.g. Tannen 1989, Clark and Gerrig 1990, McCarthy 1998, Myers 1999). But to notice that what we prefer to call the faithfulness assumption (see below) does not apply in some forms and contexts of direct discourse presentation does not imply that it should be abandoned in *all* forms and contexts. We wish to demonstrate that the use of the direct forms in some contexts carries a greater commitment to a reproduction of the original words than it does in others, both from the point of view of reporters, and from the point of view of those who hear or read the reports. Thus our position is essentially similar to that of Clark and Gerrig in the quotation below (see also Collins 2001).

Suppose you want to represent a person’s locutionary acts verbatim. If you use indirect quotation, your addressee cannot be sure what sentence really was uttered. The mapping from your words to your source sentence is undecidable. What you need is direct quotation. Even that isn’t enough, since the locutionary act need not be a depictive aspect of the quotation. You must also say or imply that your wording is verbatim. It is CONVENTIONALLY implied that the wording is verbatim in newspapers, law courts, and literary essays, but, as we later argue, not elsewhere. (Clark and Gerrig 1990: 792)

One of the reasons that Clark and Gerrig use the word ‘conventionally’ above is because they acknowledge that what they call verbatim quotations do not have to be an *exact* mimicry of an original locution *in all respects*. They point out, for example, an American newspaper report, in which direct speech is used for an English quotation from the Pope which, from the context, must have originally been in Italian. Clearly this is not complete mimicry of the original, otherwise it would have had to be reproduced in a language which most of the newspaper’s readers would not have been able to understand. But it is also clear that the writer was claiming a greater

faithfulness to the original than an indirect speech version would have suggested, and that the extent of the claim is made clear contextually. It can be seen from this example, as well as from that in relation to the Bare ladies above, that the notion of faithful quotation is complex and, in varying degrees, context-dependent.

As Clark and Gerrig say at the end of the above quotation, the assumption of a ‘verbatim’ quality in a report will vary from one situation to another. We try to spell out some aspects of this complexity towards the end of our article. However, we would prefer to use ‘faithful’ rather than ‘verbatim’. Both notions are complex and have been defined in different ways. Clark and Gerrig (1990: 795-6) show the range of variation in existing definitions of what is involved in a ‘verbatim’ report. In the strictest sense, though, verbatim quotation does mean an absolutely exact reproduction, including, for example, normal non-fluency features in direct speech presentation, which is not what is involved in the notion of faithfulness as we intend it here. For us, faithfulness in direct discourse concentrates on those factors which are relevant in specifying as accurately as is feasible in context the precise communicative content of the discourse being reported. Faithfulness thus centrally involves the reproduction of the lexical items and grammatical structures used in the anterior discourse, and, where communicatively relevant, contrastive stress and other speech/writing production factors. But it does not necessarily involve reproducing, except where it is communicatively important, every single linguistic characteristic of the utterance being reported. The concepts of faithfulness and verbatimness are thus not always substitutable for one another, in spite of the fact that discourse presentation analysts (including ourselves in the past) have sometimes used them interchangeably.

Below we begin by reviewing the arguments that have been put forward in critiques of the notion of faithfulness in discourse presentation. We then go on to explain why we think that the notion of faithfulness is nevertheless important for discourse presentation theory, and make what we believe are some crucial terminological and conceptual distinctions which are necessary to make sense of this very complex area. Next we introduce a corpus-based study of discourse presentation in narrative texts in order to show how its results contribute (in both quantitative and qualitative terms) to the faithfulness debate, and then examine faithfulness in a set of DW presentations of a particular newspaper article in a series of news reports which followed it. We finish by proposing the outlines of a context-sensitive approach to faithfulness in reporting.

2 The ‘no faithfulness’ approach to discourse report/(re)presentation: is the baby being thrown out with the bath-water?

At least four different kinds of argument have been put forward in order to critique the traditional role of the notion of faithfulness in models of discourse presentation (Fludernik 1993: 22). Below we sketch out each of these arguments in turn and indicate why, in general terms, we do not feel that they justify a complete rejection of this notion.

Firstly, it has been pointed out that the direct forms can be used in cases where there is no such thing as a straightforward ‘original’ which can be faithfully reproduced. This is the case, for example, (a) when one refers to a hypothetical utterance (e.g. ‘If she was here, she would say “Over my dead body!”’), (b) when a person’s non-verbal behaviour is ‘translated’ into verbal form (e.g. ‘The expression on his face said “What the hell is this?”’), (c) when the reporter spells out what is implied by someone else’s speech (e.g. ‘What she really meant was “Get lost!”’), (d) when

what is reported relates to multiple originals (e.g. ‘My students are always saying “The books are not in the library.”’) (see Fludernik 1993: 409-414; Sternberg 1982a, 1982b; Tannen 1989: 110-119). We accept, of course, that these phenomena need to be included in accounts of the forms and functions of discourse presentation (for a detailed discussion of our approach to hypothetical discourse see Semino *et al.* 1999). However, we do not see that this necessitates a complete rejection of the notion of faithfulness because such cases only represent a small minority of total cases (see the quantitative evidence from our corpus presented in section 6 below); and in this minority of cases it is usually made clear contextually that *reporting* is not involved in the *presentation*. Contextual clues induce the reader to cancel the conventional assumption that the direct discourse form relates to a specific stretch of anterior discourse.

Secondly, it has been noted that the direct discourse forms, in particular direct speech, may contain expressions which explicitly suggest that the reporter is missing out some elements of the original (e.g. ‘so-and-so’, ‘what’s-his-name’, ‘something or other’), or may be accompanied by expressions which undermine the reliability of the quotation (e.g. ‘She shouted something like “I am not prepared to put up with this!”’). Our views concerning this kind of phenomenon are in line with what we said about the lack of original utterances, namely (i) that the proportion of such cases is very small and (ii) it is made clear contextually that the claim of making a *report* faithful to the words and structures used in the anterior discourse is not being made. In this case, the claim being made is that of *presenting a rough approximation* vividly. In other words, although the linguistic form canonically associated with DS report is being used, any faithfulness assumption that might be associated with it has been explicitly cancelled. Hence this kind of example does not undermine the relevance of the notion of faithfulness in all cases and in all contexts.

Thirdly, it has been pointed out that many features of original utterances are lost in *any* report, including direct report, especially when the anterior discourse is in one medium and the posterior discourse in another. So, for example, the use of DS in written texts inevitably fails to represent a range of aspects of spoken utterances, such as voice quality, accent, intonation, paralinguistic features, and so on. Similarly, it may not be possible, particularly in spoken contexts, to reproduce exactly all the characteristics of a written original, including layout, typeface and orthographic features (Sternberg 1982a, 1982b; Baynham and Slembrouck 1999). These are, of course, important considerations in assessing the extent to which a particular report can even begin to reproduce an original. It is also true, however, that not all aspects of an original utterance are equally important in assessing faithfulness to an original, and different aspects of the original will not be equally important in different reporting contexts. Given our focus on lexis and grammatical structures in defining ‘faithfulness’, we do not feel that taking these issues into account necessarily involves seeing all cases of direct discourse presentation as equally constructed. It is only with the modern technologies of the computer, the tape recorder and the photocopying machine that it has become possible to reproduce exact versions of an original when reporting it (though it is arguable that even with tapes of speech or photocopies of orthographic originals minute variations occur in the recording/copying process). Even if the discourse medium is held constant in the anterior and posterior contexts, variations are allowable within the concept of faithfulness. Hence, if someone re-says exactly what someone else said we would not expect the voice quality to be the same, simply because we know that the original speaker and the reporter are not the same person. Similarly, no-one would quibble with a direct writing presentation of

something in the Times font being reproduced in the New York or Chicago font, except in very particular circumstances indeed. This is because we all automatically make adjustments in accordance with the type/token distinction when we compare one thing with another. In our view, then, a direct quotation in English of something said by a Chinese politician in Chinese, although it is not verbatim, precisely because it is not in Chinese, does not mean that the direct speech string cannot be faithful to the original. It merely needs to be an accurate translation of the words originally spoken, and in a way that an indirect speech report of the same original utterance does not need to be. Similarly, for discourse presentations which are in a different medium from the original, we would not expect an accurate representation of features which cannot survive the medium change. What can be common to speech and writing are the words and grammatical structures used to produce some propositional structure, and so it is not surprising that traditional distinctions between the direct and indirect forms refer specifically to these two linguistic levels, even though other linguistic levels are sometimes involved (e.g. the use of italics to indicate contrastive stress in speech).

Last but not least, as we indicated in 1 above, it has been pointed out that in conversational contexts, where speech is the reporting medium, it is simply not possible, due to the limitations of human memory, for speakers to reproduce previous utterances word-for-word, even when they adopt a form of reporting which is the oral equivalent of direct speech presentation (Fludernik 1993: 414ff.). Our reaction here, as before, is to accept the factual evidence but to give it a different weighting in the theoretical discussion. In reports of previous speech in very informal settings, the notion of faithfulness as we have defined it does not normally apply. This is not simply because of the factors mentioned above, but also because speech has no straightforward equivalent for quotation marks, which, in a number of written contexts, is the main indication that a faithfulness claim is being made. It is interesting to note that Tannen (1989) uses inverted commas to mark off direct discourse presentation in transcripts of speech, even though such punctuation devices, being writing devices, did not exist in her original data. This indicates a tacit assumption that cases of direct speech in spoken and written language are automatically and directly equivalent, an assumption that is by no means obvious. Others, for example Myers (1999), are more sparing in their use of such orthographic markings, indicating a more careful attitude to the issue of cross-medium 'equivalence'.

The arguments we have reviewed so far demonstrate that the notion of faithfulness is largely irrelevant in the use of the direct form of presentation in a number of contexts, notably informal conversation. However, as we have already pointed out, this does not mean that the role of the faithfulness criterion should be dismissed across *all* the discourse presentation (speech, thought, writing) scales and across all contexts. While we do not want to support a 'naive' notion of faithfulness, we do feel that some critiques of faithfulness have ended up underestimating an important dimension of discourse presentation, thereby throwing the baby out with the bath-water, as it were.

While those arguing against the importance of faithfulness have tended to concentrate on spoken interaction and fictional dialogue, in this article we will examine the use of the direct forms in a corpus of written narratives (fiction, new reports, and (auto)biographies). This will enable us to show that the applicability of the notion of faithfulness is context-dependent, and that, in some contexts, faithfulness strongly affects the behaviour of writers and the likely expectations of readers. This is

particularly the case in *formal written reports of writing* using the direct presentational form. We examine an example of this kind of context in section 7 below.

Before introducing our data, it will be helpful to make it clear why we feel the criterion of faithfulness is too important to abandon (section 3) and to draw attention to some terminological and conceptual unclarities (some of which have been alluded to above) which have arisen in the discussion of discourse presentation theory (section 4).

3 Why is faithfulness important?

Given that it has been clearly demonstrated that faithfulness in direct discourse presentation does not always apply, particularly in ‘hypothetical’ and speech-reporting-speech situations, the reader may legitimately ask why we are so keen to retain this problematic notion in our approach to discourse presentation.

Firstly, outside linguistics, stylistics and critical discourse analysis, considerable store is set by the differing faithfulness claims associated with direct and indirect speech and writing. For example, (i) the laws of libel and slander are used precisely in cases where the exact words which someone uses, or is reported to have used are crucial; (ii) in educational and academic contexts unacknowledged use of what others have written or said constitutes the offence of plagiarism (and unacknowledged direct quotation is more heinous than unacknowledged indirect forms); (iii) in academic contexts, to misquote someone in order to attack their position more effectively would be seen as a grave offence and (iv) newspaper style books, manuals of style and grammars for lay people often make much of the faithfulness distinction between direct and indirect speech/writing presentation. We are told that in American newspapers, unlike those in the UK, free indirect speech (FIS) is not usually allowable as a form to be used by reporters, precisely because there is ambiguity over which words belong to the person quoted and which to the reporter (Don Freeman, personal communication). We have also reported in Short *et al.* (1999) evidence that writers of newspaper articles, although not entirely consistent in their quoting behaviour do, by and large, try to quote accurately and to acknowledge direct quotation, even when their source is an unattributed press release. All these kinds of behaviour would be difficult to account for without recourse to the notion of faithfulness. Similarly, the accounts within forensic linguistics of the illicit use by the police of inaccurate written DS representations of what is said in particular interviews (cf. Coulthard 1992, 1996) would not be a matter of such high interest if the assumptions concerning faithfulness were not commonly held.

Secondly, although it is technically possible to make the various distinctions among the speech, thought and writing (re)presentational categories on formal linguistic grounds alone (along with some additional contextual information in some cases), if we *only* take linguistic factors into account we have no *reason* for making the linguistic distinctions that we do. The reason for the various discourse presentation category distinctions, in our view, relates to the varying prototypical faithfulness claims of the different (re)presentational forms, even though the extent to which these claims apply varies depending on context (see Semino *et al.* 1997 for a discussion of this matter).

Thirdly, we need an explanation for how the presentational *effects* associated with the various categories come about. For example, the presentational qualities of directness, vividness etc. of DS and DW, and the more reserved and measured

qualities of IS and IW, themselves depend on the contrast between, minimally, the *impression* of faithfully reproducing another's words as opposed to that of recasting what was said in one's own words.

Finally, the examples used to support, or undermine, the relevance of the notion of faithfulness also need to be weighed against statistical trends in order for their representativeness to be assessed. We will report relevant statistical evidence from our corpus in section 6.

4 Some useful terminological and conceptual distinctions

It will be clear from the discussion in sections 2 and 3 that in order to arrive at well-balanced conclusions, it is necessary to be very clear concerning exactly what we are talking about. Unfortunately, the discussion of discourse presentation has suffered from a number of important terminological and conceptual confusions, confusions which we have ourselves been guilty of, as well as others. In this section we try to make clear some of this issues in order to create a firm conceptual base for what we want to say later.

4.1 Discourse

The first thing which we need to beware of is the increasingly frequent use of the term 'discourse' across a range of areas of study (e.g. stylistics, critical discourse analysis, literary theory). This has resulted in a range of possible meanings for the term which are not always carefully distinguished. For some writers 'discourse' means exclusively *spoken* language (cf. the well-known text vs. discourse distinction) but for others it means the whole of language:

Discourse - language beyond the sentence - is simply *language* as it occurs, in any context, (including the context of linguistic analysis), in any form (including two made-up sentences in sequence, a tape recorded conversation, meeting or interview; a novel or a play).

(Tannen 1989: 6)

It is presumably this latter view which gives rise to expressions like 'literary discourse', 'written discourse' 'political discourse' 'medical discourse', 'the discourse of AIDS', 'the discourse of ante-natal care', 'media discourse' 'advertising discourse' 'institutional discourse', 'colonial discourse' and so on. Different researchers are using the term 'discourse' in so many ways that it is becoming difficult to use it without ambiguity. We do not want to address this vexed issue in any detail here, but merely wish to alert our readers to the difficulties concerning the term 'discourse'.

Discourse presentation theory, unlike other uses of the term 'discourse', includes *thought presentation* as part of its remit. This is because much work in this area has concentrated on the analysis of fictional prose. The most straightforward use of 'discourse' in the field with which we are concerned, then, is as a generic term which has (re)presented speech, thought (and writing) as its hyponyms.³ However, it is easy for writers in this area to move from using it in this inclusive sense to using it as a synonym for one of its hyponyms:

Qua representational discourse, therefore, each act of quotation serves two masters. One is the original speech or thought that it represents, pulling in the direction of maximal accuracy. The other is the frame that encloses and regulates it, pulling in the direction of maximal efficacy. Reported discourse thus presents a classic case of divided allegiance, between original-oriented representation (with its face to the world) and frame-oriented communication (with its face to the reader). (Sternberg 1982a: 152)

There is much that we agree with in Sternberg's characterisation of the janus-faced character of the report of anterior speech, but note how the use of 'representational discourse' at the beginning of this quotation leads Sternberg to equate 'original speech' and 'original thought' in his 'the original speech or thought that it represents'. However, the thoughts of other people are never directly available to us, and even the 'quoting' of one's own anterior thoughts is problematic in that it is difficult to be sure what words, if any, were used to think those thoughts. The notion of faithfulness is therefore fundamentally problematic in relation to thought presentation, in a way that does not apply in the cases of speech or writing (re)presentation. Speech and thought (re)presentation are thus very different in important ways, which the use of the term 'discourse' here tends to disguise.⁴

For clarity's sake we believe it is important to distinguish systematically among speech, thought and writing, and only use 'discourse' when it is genuinely inclusive of its hyponyms. This latter practice is the one we adopt in this article, and as a consequence, following on from the systematic distinction between speech and thought presentation in Leech and Short (1981: Ch. 10) we distinguish, for example, direct speech (DS) from direct thought (DT) and direct writing (DW), and so on along the three separate, but related, clines of discourse presentation possibilities (see 2 below and section 5).

4.2 Speech

'Speech' is often used as a general term in discourse presentation theory, sometimes including thought (re)presentation and sometimes writing (re)presentation. A clear example would be Jones (1968), who uses 'speech presentation' as an inclusive term referring to the presentation of both speech and thought in Conrad's *The Secret Agent*. A more recent example is the 1999 special issue of *Text* (19, 4), which is called 'Speech Representation in Institutional Discourse' in spite of the fact that one of its articles is mainly about writing: how students refer to published works in their essays. In the introductory article to that issue (Baynham and Slembrouck 1999), the editors use the term 'speech representation' to include an example which they themselves say is 'rather ambiguous in status between being a logical deduction by the social worker on the basis of his own experience and a medical diagnosis which is reported by the social worker.' This generalising use of the word 'speech' presumably results from the fact that in ordinary language people often use locutions like 'I said' metaphorically when they are referring to what they have written. But if we are to be clearer about the differing functions, meanings and effects in discourse presentation we need to distinguish systematically between (i) speech, (ii) thought and (iii) writing with respect to (a) the anterior context and (b) the posterior context. For example we noticed in section 1 that the written report from *The Independent* represented both spoken and written anterior sources. We cannot assume, other things being equal, that

a written report of speech as direct speech (DS) will automatically have the same functions, meanings and effects as a spoken report of the same DS or a written report of the same content in the form of direct writing (DW). These remarks apply, *pari passu*, for the various combinations of speech, thought and writing in anterior and posterior discourse contexts (where they both apply - see 3 below) and across all the categories of the speech, thought and writing (re)presentation scales.

It is only when these various combinations have been adequately charted that it makes sense to examine 'mixed' cases, which might (re)present discourse using a mix of written and spoken means. Examples of such 'mixed' cases include multimedia web sites which use text, sound, video and pictures of documents, and TV news reporting, in which a reporter may summarise or quote the original utterance, as well as showing a sound or video recording or transcript of the original.

4.3 Fiction

We need to distinguish systematically between fictional and non-fictional discourse (re)presentation. Until recently, most of the work in discourse presentation theory was carried out on invented examples (in grammatical study) or fictional texts (in stylistics and literary theory). But fictional examples are fundamentally different from discourse presentation in non-fictional language use, given that they do not, by definition, refer back to an independent antecedent occasion when the language being re-presented was originally produced. When we read novels it would seem that if a reliable narrator uses a DS string we assume that the character whose speech is being presented used exactly those words. In other words, the presentation form is a pretty solid guarantee. Indeed, it is part of what is involved in the assumption that readers often have that the words of a character in DS are not being reported at all, but being witnessed within the world of the fiction *as they are being produced* for the first time. However the same cannot be said so easily of the more indirect forms of presentation, and this 'presentation' assumption, where 'discourse time' and 'event time' are the same, is easier with 3rd-person fictional narrations than with 1st- or 2nd-person narrations because the latter often do look back, reporting events which happened in their past. In fictions which report events that happened in a fictional past, we conventionally assume that the 'reports' are real within the fiction, and so use the canonical assumptions associated with the various categories on the discourse presentation scales analogically, and with effectively guaranteed faithfulness results. The over-use of invented and fictional examples has, in our view, led to (i) an unclear account of how discourse presentation works and (ii) some confusion over the meaning of the terms we discuss in 4 below. This is something which Baynham and Slembrouck (1999) also remark on. A fully explanatory account of discourse presentation in fictional texts clearly needs to be undertaken within a more general theory of discourse presentation, and indeed this was the main motivation for our creation of our corpus of fictional and non-fictional narratives described in 5 below.⁵

4.4 Report, presentation and representation

We need to be aware of the distinctions and overlaps among the terms 'presentation', 'representation' and 'report'. Because they have overlapping meanings they are

sometimes used interchangeably (indeed, to avoid over-repetition of the same word we have sometimes used them in this way in the discussion so far, although only when we thought they were properly interchangeable).

But the three terms emphasise different aspects of what is involved, and should only be used interchangeably with considerable care. Which of the terms is used as ‘basic’ also varies from one branch of linguistic study to another. The term ‘presentation’ is mainly used in stylistic analysis because stylisticians concentrate on fictional texts which, as we pointed out in 3 above, do not have an anterior discourse situation independent of the ‘reporting’ discourse. Representation and report, on the other hand, both involve an anterior discourse and a posterior discourse which reports/represents the anterior discourse. The term ‘representation’ is most often used in critical discourse analysis because such studies are often at pains to point out mismatches between the anterior discourse and its posterior representation (e.g. Fairclough 1988, Slembrouck 1992) which are important in relation to some socio-political criticism being made. The term ‘report’ appears to be the basic term in linguistic traditions where naturally-occurring texts and discourses are not the object of investigation, but where the analyst invents representative examples to argue theoretical positions (e.g. generative linguistics, logic, philosophy of language), and where it is therefore assumed that the relation between, for example DS and the speech in the anterior situation which it reports is unproblematic. Interestingly, Tannen (1989) uses the term ‘report’ even when she is undermining the assumption of identity between DS and the anterior speech which it relates to.

5 Our corpus and associated annotation system

In order to study systematically the forms and functions of speech, thought and writing (re)presentation in written English, we constructed an electronic corpus which contains approximately 260,000 words of contemporary written British English. The corpus was divided roughly equally among three different narrative genres: prose fiction, newspaper news stories, and (auto)biographies. Each of these three genres is represented by a set of 40 extracts of 2,000 words or more from 40 different texts, making a total of 120 text-specific samples in all (the length of the extracts varies slightly because we used ‘natural’ boundaries like section and chapter divisions to mark the beginning and end of each extract). Each genre contains 20 ‘serious’ and 20 ‘popular’ samples, in order to allow a comparison between texts produced for different audiences and/or purposes. To keep the serious/popular divide as clear as we could make it, we made a point of using what we took to be central examples of each sub-category. As far as newspapers were concerned, broadsheets (e.g. *The Independent*) were classified as serious and tabloids (e.g. *The Sun*) were classified as popular. When it came to fiction, we selected authors and texts which we thought could be uncontroversially included in one category or the other (e.g. Virginia Woolf’s *Night and Day* as opposed to Elizabeth Adler’s *Peach*), and also consulted a group of stylistics researchers at Lancaster University to check our own judgements. For biographies and autobiographies, we took into account the profession of the protagonist (e.g. politician vs. sports personality), the status of the author (e.g. historian vs. tabloid journalist), and the style of writing (high- vs. lowbrow). So, for example, the serious section contains Margaret Thatcher’s autobiography *The Downing Street Years* and Peter Ackroyd’s biography of T. S. Eliot, while the popular section contains the autobiography of Cilla Black, an English ex-pop singer and light

entertainment TV show host, and Alan Henry's biography of Damon Hill, the Formula 1 racing car driver.

The whole of our corpus was annotated (or 'tagged') manually for speech, thought and writing (re)presentation. All texts were checked and discussed by all three authors of this article. We initially adopted the categories of speech and thought presentation proposed by Leech and Short (1981), given that one of the aims of the project was to test the usefulness of the Leech and Short model. This model was developed and extended during the pilot phase of the project (see Short *et al.* 1996, Semino *et al.* 1997 and Leech *et al.* 1997), and further revised in the process of tagging our current corpus, during which we made our coding system conformant to the Standardised General Mark-up Language, or SGML (Wynne *et al.* 1998).

There is, of course, an issue as to whether our electronic corpus is large enough or well-enough designed to be properly representative. It is difficult to know how big corpora need to be for particular purposes, and this issue has had some discussion within the field (cf. Biber 1990, 1993). There are considerably bigger corpora of English available (e.g. the BNC [Lancaster] at 100 million words and the much larger 'Bank of English' [Birmingham], which now exceeds 300 million words), but there are no computer programs available to analyse discourse presentation automatically within those corpora, and, as far as we know, no-one has attempted to annotate a corpus for discourse presentation in the way that we have. Biber (1990) investigates text length and corpus design in order to test empirically Oostdijk's (1988) claim that the general linguistic corpora existing at that time (up to 1 million-words) were 'not suited' for the study of linguistic variation. Comparing figures for sub-parts of each of the corpora he examines (the 1 million word LOB corpus and the 435,000 word London-Lund corpus) with the rest of the corpora, he concludes that, for the specific range of factors he examines, 1,000-word extracts are representative for most purposes (our corpus uses sample sizes of at least 2,000 words) and that 120 individual text samples (the number in our corpus) replicates samples from twice or four times the number of texts to a high degree (though, not, of course, for every single factor). He concludes that 'the underlying parameters of text-based linguistic variation (as represented by the factorial structure) can be replicated in a relatively small corpus *if* that corpus represents the full range of variation' (Biber 1990: 269). All that said, the corpora Biber was investigating are 'generalist' in nature, and it is usually assumed that corpora designed for specific purposes do not have to be anything like so large to be representative. For example, Sampson's (1995) 'SUSANNE' corpus, which is also annotated by hand (for grammatical analysis) is only half the size of our corpus and yet is generally assumed to be adequate for its purpose.

Another way of approaching the problem is not to worry so much about overall corpus size and composition, but to consider the number of instances of the particular phenomenon under investigation. Two common (but to some degree conflicting) statistical rules of thumb reported to us from a discussion we instigated on a corpus linguistics discussion group are that one needs either (i) a total of 400 examples of the particular phenomenon under investigation, or (ii) a minimum overall total calculated from the following 'algorithm': 10 items per cell in any tabular array plus another 10 for every 4 cells in the table. This brief discussion suggests that our corpus can be claimed to be representative in many, but perhaps not all respects, and so we would welcome the checking of our findings by researchers using other corpora. As far as we are aware, there are, as yet, no equivalent corpora with which to compare.

Below is a list of our basic categories for speech and thought (re)presentation:⁶

NV = Narrator's Report of Voice
 NRSA = Narrator's Representation of Speech Act
 IS = Indirect Speech
 FIS = Free Indirect Speech
 DS = Direct Speech
 FDS = Free Direct Speech

NI = Narration of Internal States
 NRTA = Narrator's Representation of Thought Act
 IT = Indirect Thought
 FIT = Free Indirect Thought
 DT = Direct Thought
 FDT = Free Direct Thought

The above lists include the categories proposed by Leech and Short (1981) plus two new categories, NV and NI, which lie at the boundary between narration and, respectively, speech and thought presentation. NV (Narrator's report of Voice) captures minimal references to talk or speech events, where no indication is provided as to the words, content or even speech act value of the utterance (e.g. 'Voices fretted along the murmuring wire' in J. G. Ballard's *The Empire of the Sun*, or 'The winner [. . .] was yesterday being interviewed by officials' in *The Independent*, 12/12/1994). NI (Narration of Internal States) is similar to what Cohn (1978) has called 'psychonarration', in that it captures references to cognitive or emotional experiences that do not amount to the report of specific thoughts (e.g. 'My mind was racing.' in Michael Caine's autobiography). For a detailed discussion of NI, see Short *et al.* (1996) and Semino *et al.* (1997). It was in the process of annotating our data that we realised that some of the stretches of text which we had begun to code as speech presentation were in fact explicitly presented as relating to *written* texts rather than talk. We therefore decided to adopt a set of tags for writing presentation which were parallel to those for speech and thought presentation (see also Table 1 above):

NW = Narrator's Report of Writing
 NRWA = Narrator's Representation of Writing Act
 IW = Indirect Writing
 FIW = Free Indirect Writing
 DW = Direct Writing
 FDW = Free Direct Writing

All stretches of text which we felt did not (re)present speech, thought or writing were tagged as narration (N) and the reporting clauses of direct and indirect discourse were tagged separately as NRS (Narrator's Representation of Speech) in order to provide more accurate word count statistics in relation to the various categories. These tags are not included in the lists above as they are not strictly part of the discourse presentation scales. When we thought that a particular stretch of text was ambiguous between two (or more) categories, we used 'portmanteau' tags to include each of the categories involved (e.g. the portmanteau tag N-FIT indicates an

ambiguity between narration and free indirect thought). In addition, we added a number of further letters to our annotations to highlight specific variants of our main categories. For example, an ‘e’ was placed at the beginning of a tag if the relevant category was embedded within another category (e.g. an example of IS embedded within, say, DS was coded as ‘eIS’). We appended a ‘q’ to the end of a tag if a non-direct category contained a stretch of quotation embedded within it (e.g. a stretch of IS containing some words within quotation marks was coded as ISq). Finally, we highlighted, by adding ‘h’, those instances of speech, thought and writing (re)presentation (ST&WP) which were *not* presented as having occurred in the world of the text prior to the act of narration, but rather as parts of hypotheses, wishes, fantasies or predictions about the future (see section 2 above). We have discussed this phenomenon of hypothetical discourse in detail in Semino *et al.* (1999), where we have related it to possible-world theory. For more detail on our entire annotation system, see Wynne *et al.* (1998).

6 An analysis of the direct forms in our corpus

We will now show how an analysis of our corpus can help to arrive at a more context-sensitive approach to the notion of faithfulness in reporting. A preliminary example of this is the way in which, by adding the ‘h’ tag to what we thought were hypothetical instances of ST&WP, we have been able to gain some idea of the frequency of a phenomenon which, as we mentioned earlier, is often quoted as evidence against the applicability of the notion of faithfulness to direct reports. We have found that hypotheticals account for only 4% of all tags in our corpus, most of which (62%) are embedded inside other non-hypothetical categories of ST&WP. As for direct speech, only 1.5% of all instances turned out to be hypothetical. We used the ‘h’ tag to cover not just obvious straightforwardly hypothetical examples but a range of phenomena where it was made clear that there was no anterior discourse to be ‘reported’. But the total number of cases in the corpus was still very small. Hence, the hypothetical report phenomenon can be seen as an interesting (and well-defined) exception, but it cannot, in our view, be used as evidence against the general applicability of the notion of faithfulness.

One of the advantages of the availability of a tagged electronic corpus is easy access to statistical information about the different modes of discourse presentation. Given that the issue of faithfulness of reproduction relates particularly to the direct modes, below we provide figures for direct speech (Table 1), direct thought (Table 2) and direct writing (Table 3). The columns in each table relate to different sub-sections of our corpus. Moving from left to right, we first provide figures for the corpus as a whole and then figures for each of the three genres within the corpus, starting in each case with the serious section (in **bold**), followed by the popular section (in *italics*). For the sake of completeness, we have separated the direct and free direct forms, even though, as we mentioned earlier, we regard the free direct variants as sub-categories of the direct forms, rather than as categories in their own right. In each case we have provided raw figures for the number of occurrences of each category and, in brackets, percentages relating to the total number of tags in the relevant section of the corpus. For example, the cell immediately below ‘**Fiction Serious**’ in Table 1 indicates that there are 347 instances of DS in the serious fiction section of our corpus, which represent 12.9 % of the tags in that section of the corpus (percentages have been rounded up to the nearest decimal point).

[INSERT TABLES 1, 2, AND 3 ABOUT HERE]

In Table 1 (DS/FDS), the number of cases (2974) exceeds significantly the ‘400 items’ rule of thumb mentioned in 5 above and also the ‘algorithm’ rule of thumb, which suggests that as there are 12 sub-category cells in each of the tables (the left-hand column and the bottom row of cells merely indicate overall totals) we need a minimum of 150 items in each of Tables 2, 3 and 4. Table 2 (DW/FDW) has 141 cases, a figure which is well below 400 but close to the 150 cases suggested by the ‘algorithm;’ method. Table 3 (DT/FDT) has 107 cases which is significantly below the thresholds indicated by either of the rules of thumb mentioned. This would suggest that comparisons concerning DS/FDS from one sub-section of the corpus to another are safe in terms of representativeness, as would be general DS/FDS comparisons across the three (re)presentational modes, but that comparisons within DW/FDW are rather marginal and those within DT/FDT unsafe. Below we restrict ourselves to statistical comments which appear safely representative according to the parameters just outlined.

The overall totals in each table show that direct speech is a significant category in our corpus, representing 18% of all tags (note that the total number of tags includes the N tags for stretches of narration). On the other hand, direct writing and direct thought each represent less than 1% of all tags. This helps to explain why the word ‘speech’ is often used to refer to all the (re)presentational media when discourse presentation is discussed. It is also easy to notice that the direct (re)presentation of the three different media is not equally divided across the three genres in our corpus, nor across the serious and popular section in each genre (more on this below). A closer analysis of our data has also highlighted an important difference in the form of the free direct sub-categories between speech, thought and writing. As far as speech and writing are concerned, the free direct forms are mostly characterised by the absence of a reporting clause and the presence of quotation marks. On the other hand, free direct thought may or may not have a reporting clause, but usually has no quotation marks (which, incidentally, suggests that quotation marks are associated with the reporting of phenomena where the original can, in principle at least, be reproduced faithfully). All this highlights the importance of distinguishing not just speech presentation from thought presentation, but also writing presentation from the other two scales. As we mentioned earlier, the trend in the literature to subsume all three modes under the labels ‘speech presentation’ or ‘discourse presentation’, sometimes results in confusion, especially when claims are made about the notion of faithfulness. We shall now focus on each mode of (re)presentation in turn.

6.1 Direct speech

If we consider DS and FDS together, we can see that they are considerably more frequent in fiction than in the other two text-types. (Free) direct speech represents 27.4% of all tags in fiction, as opposed to 14.2% in the press and 11.8% in (auto)biography. This may be due to a number of factors. Direct speech helps to produce the effects of immediacy, drama and involvement that are particularly important in the telling of fictional stories. But this is not the only possible explanation. In spite of the fact that the anti-faithfulness argument has been advocated with most vigour by literary scholars, fiction is, in our view, one of the genres where the reader needs to assume that, other things being equal, direct speech *does indeed*

provide a faithful (re)production of a character's words (and our corpus suggests that things *are* equal in the vast majority of cases). This is because, as we pointed out above, in fiction the 'original' speech has no independent existence whatsoever, and is only accessible *via* the ('re')presentation itself. Indeed, it would be difficult to talk about some important aspects of the process of characterisation if it was never possible for readers to have the illusion of direct access to the words of fictional characters. In our view, therefore, the high frequency of direct speech in fiction can be seen as supporting rather than undermining the notion that the direct forms carry with them conventional expectations to do with faithful reproduction.

As for the other two genres, overall the newspaper data has slightly more (free) direct speech than (auto)biography, but the difference is not great (14.2% of all tags as opposed to 11.8%). What is interesting is that in all three genres, but particularly in (auto)biography and the press, the popular section has considerably more direct speech than the serious section. Again, this will, in part, be due to the greater wish for effects like dramatisation and immediacy in the popular texts. We would argue, however, that in serious non-fictional writing there may be a greater reluctance to use the direct forms where the original cannot be reproduced. Indeed, it is likely that readers will have different expectations of the status of material within quotation marks depending on their perception of the text-type they are reading.

While a quantitative analysis has enabled us to discuss general trends in our data, only a more qualitative focusing on specific examples will allow us to highlight the contextual factors which affect the issue of faithfulness in DS. Our first example is taken from the autobiography of the runner Linford Christie (co-written with Tony Ward), and specifically from a chapter entitled 'Harassment', where Christie narrates a series of racially motivated attacks on his family.

Dad was awoken by a telephone call about an hour later, around five-thirty.

'Mr Christie,' said a voice, 'you don't know me but I know you. We have met and I know that you are a good man and a religious man. I am therefore sorry to tell you that I am going to kill your son Russell.' Dad said nothing. He was too stunned

(Linford Christie and Tony Ward (1990) *Linford Christie: An Autobiography*, London: Arrow, p. 43).

Here DS is used in writing to represent an utterance which Linford Christie, the narrator, must have heard second-hand from his father, rather than experiencing it directly. It is also unreasonable to assume that Christie's father would have been able to remember word-for-word the threats he received in the middle of the night, and equally unlikely that Linford Christie would remember exactly what his father said to him (Christie's father may even have told the story of this incident without using direct speech in his narration of it). In any case, what matters here for the point being made is not the individual words which were uttered, but the general tone and content of the threats directed towards Christie's family. As a consequence, the choice of DS here cannot be related to a claim of faithful reproduction, but is presumably aimed at involving the reader and creating drama and suspense. It thus conforms to the view of DS that Tannen and Fludernik propose, as outlined sections in 1 and 2 above.

Quite different considerations apply, however, to the following extract from the autobiography of British politician Kenneth Baker, who is at this point describing

the period that led to Margaret Thatcher's resignation as Prime Minister of the UK government:

The BBC's intrepid political reporter, John Sergeant, who had been addressing the camera and had his back to Margaret, was alerted to her arrival and achieved his place in history by spinning round, defying Bernard Ingham and thrusting the BBC's microphone into her face, crying out, 'Here's the microphone Prime Minister.' Margaret then made her prepared statement, saying, 'I am naturally very pleased that I got more than half the Parliamentary Party, and disappointed that's not quite enough to win on the first ballot. So I confirm it is my intention to let my name go forward for the second ballot.'

(Kenneth Baker (1993) *The Turbulent Years*, London: Faber & Faber, p. 397)

Here there are two instances of DS. The first one relates to John Sergeant's invitation to Thatcher to speak into his microphone, 'Here's the microphone Prime Minister.' This utterance is clearly rather trivial, and its faithful reproduction is not crucial to the account. On the other hand, Sergeant's words were probably captured on camera, so it will have been easier for Baker to report them faithfully. The second instance of direct speech reports what Mrs Thatcher herself said. This is a crucial statement, both because it was produced by the Prime Minister, and because every word was important in conveying her attitude to the results of the first ballot of her Party, which was voting on whether she should continue as leader, and hence as Prime Minister. In addition, her statement was broadcast on numerous occasions on both radio and television. As a consequence, it is (a) important that her utterance was faithfully reproduced word by word, and (b) easy for the writer to do so. The fact that the written text cannot represent Thatcher's tone of voice, paralinguistic features and so on, does not detract from the high faithfulness stakes in this kind of context. Indeed, if it turned out that the quotation was not accurate in relation to the words and structures used, we would probably have to assume that the writer made a mistake or, if the discrepancy was significant, that he had the intention to misrepresent the original speaker and deceive his audience.

6.2 Direct writing

Table 2 shows that, in our corpus, DW and FDW are about four times more frequent in (auto)biography than anywhere else. This is actually due to the high frequency of direct writing in serious (auto)biography, which alone contains more than half of all instances of direct writing in the corpus (79 out of a total of 141). Assuming this extrapolation to be reasonably safe, its explanation appears to be related to a series of connected factors. Serious (auto)biographies are less concerned with the dramatisation of the lives of the protagonists than popular ones, and are much more punctilious about spelling out the documentary evidence on which claims are based (e.g. diaries, journals, letters). In addition, those whose (auto)biographies we classified as serious tend to be the kind of people who produce considerable amounts of written texts (writers, politicians, artists), whereas those whose (auto)biographies we classified as popular are less likely to have produced a great deal of written material (e.g. sports people, singers, TV presenters).

Consider the following example from Humphrey Carpenter's biography of W. H. Auden:

Auden had the experience, some time in the early 1930s, of being invited to dinner by Eliot and his first wife Vivien, of whose strangeness he was made aware on arrival: ‘I told Mrs T. S. E. that I was glad to be there and she said: “Well, Tom’s not glad”’. footnote: *Tribute*, p. 155)

(Humphrey Carpenter (1983), *W. H. Auden*, London: Unwin, p, 73.

Here a footnote provides a reference to the source of the written quotation, which amounts to a claim of accurate word-by-word reproduction. It is also worth noticing that the DS contained within the quotation marks (‘Well, Tom’s not glad’) also carries a fairly high faithfulness claim, since the whole point of the anecdote revolves around the unmitigated rudeness of the words used by Eliot’s wife, which were thus likely to be highly memorable. Unlike the case of the written quotation, however, there is no way of checking the accuracy of the DS report against the original.

The use of DW in (auto)biographies and the press is almost invariably accompanied by similar references to the source, so that the reader is entitled to assume faithfulness to the words and structure of the original (even though typographical characteristics are unlikely to be replicated). In fiction, on the other hand, there are cases where the accuracy of DW may be less likely and less relevant. The following extract is from the first-person narration novel *Behind the Scenes at the Museum* by Kate Atkinson:

Bunty holds up the dress against herself, in a sitting-position, like an invalid. She turns to me, ‘What do you think?’

I sigh and shake my head in envy and longing, ‘It’s lovely.’ (Extracts from Ruby Lennox’s school report, summer term, 1966 - *Ruby has a real talent for acting . . . Ruby was the star of the school play.*)

(Kate Atkinson (1995) *Behind the Scenes at the Museum* , London: Black Swan, p. 278)

Here we have to conclude that the stretch of text in italics is the narrator’s written DW representation of what she remembers a school report said about her performance in the dramatic activities of the school. More precisely, the DW is embedded inside the homodiegetic narrator’s free direct thought (FDT). As such, the chances that it is to be interpreted as a faithful reproduction of the report within the fictional world are fairly low. In this case, however, faithful reproduction is irrelevant even within the fictional world: the point of the quotation is to suggest that the narrator’s previous utterance (‘It’s lovely.’) was a lie, and that this lie was easily disguised as a sincere response, thanks to her acting talents.

Direct thought

There are too few cases of DT and FDT in the corpus for our findings to count as safely representative. Nonetheless, they are suggestive, and a number of qualitative observations can be made. With thought reporting there is no such thing as an accessible original in verbal form, so that, as we suggested earlier, the notion of faithfulness is strictly irrelevant. Table 3 shows that, not surprisingly, most of the cases of direct thought (re)presentation in our corpus occur in fiction, where the

reader's suspension of disbelief allows for the presence of omniscient narrators who have direct access to characters' minds. Provided that the narrator is not marked in some way as unreliable, thought presentation in fiction has a high degree of reliability, even if the notion of *exact reproduction* may not apply in the same way as it does in the report of speech or writing. Consider the following example:

She looked round the room; [. . .] 'It's homely,' she thought with satisfaction, 'It's what I like.'

(Graham Greene (1943 [1938]), *Brighton Rock*, London: Penguin, p. 69)

It has often been pointed out (e.g. Leech and Short 1981: 342-3) that the use of the direct form for thought presentation creates the impression of highly conscious and articulate thought, so that it is quite possible that readers will imagine that the character in the extract above 'says' the quoted words to herself in her own head. Even if this were not so, the novel's 3rd-person narration (and the reading convention of the omniscience and reliability associated with such narrators) would lead readers to the conclusion that the extract within quotation marks is an accurate verbalisation of the character's thoughts, within the context of the suspension of disbelief that the reading of a novel involves. In fact there is a double suspension of disbelief here – firstly, readers are likely to accept that the narrator is telling the truth about the fictional world in presenting the character as thinking something (when in fact we know it is all invented) and, secondly, they are likely to believe that the character is thinking in a verbal (and hence articulate and well-organised) way.

DT and FDT are much less frequent in the (auto)biography section of the corpus and rare in the press section (where only DT occurs). Indeed, all instances of DT in the press are embedded within another category, usually to do with speech presentation. In the following example from the *News of the World*, a direct thought report occurs within free direct speech. The article tells the story of a British aid worker who was lost in the jungle in Africa for several days:

Day five brought the hunger pains.

'I was very weak,' Don says. 'I found some berries which looked like small grapes, and there were ape droppings with bits of these berries in them.

'So I thought, "If they're good enough for them, they're good enough for me.'"

(*News of the World*, 'I was knee deep in mud, bullets and hippos ... and all I could think was "The missus isn't going to like this"', 4/12/1994)

All of the instances of DT in the press in our corpus are similar to the above example in that they involve someone reporting what *they themselves* thought. Exactly the same applies to our (auto)biography data, since all cases of direct thought presentation occur within autobiographies and relate to the thoughts of the protagonists themselves. So, even if we cannot assume that they originally produced their thoughts using exactly those words, nor that they thought in words at all, there seems to be a strong association between the use of the direct form and the most reliable possible source, namely the 'thinker' him/herself. This, again, suggests a connection between the direct forms and faithfulness.

7 A faithfulness analysis of reports of the article ‘Don’t scare tycoons away from politics’ by Lord Simon, *The Times*, 1 August 1997

In this section we discuss in detail a particular example of written reports of writing where we were able to have access both to the original text and to a number of subsequent direct reports of it. Our aim is to support our view that the notion of faithfulness should not be removed from discourse presentation theory, since it does affect the behaviour of reporters (and the likely expectations of the audience) in many contexts, including the language of the press.

The newspaper articles analysed below are not part of our corpus, but were collected later, because of its relevance to the topic of this article. The context is that of a political row in the UK in the summer of 1997. Lord Simon of Highbury, a prominent businessman, had joined Tony Blair’s new Labour government as Minister for Trade and Competitiveness. His appointment was clearly an attempt by the new government to show that it supported business, and the Conservative opposition mounted a smear campaign against him. This campaign eventually resulted in him resigning from the government, as a consequence of the fact that he had not divested himself of his £2 million share holding in BP, a company which he had chaired before joining the government. In the argumentative to and fro, Lord Simon wrote an article in *The Times* on 1 August 1997, defending his position. A number of newspapers picked up the story the same day, when the first edition of *The Times* came out, and quoted from it in their articles. *The Times* itself also carried a story about the article in the issues of the paper in which the article appeared. The original article was easily available for accurate report and the story was a politically important one. Hence we would predict a high degree of faithfulness. Below is a tabular account of the degree of faithfulness in Direct Writing quotation in the newspapers which carried the story and used direct quotation. All of the newspapers except one, *The Express*, are broadsheet newspapers, and so represent the serious end of the popular/serious divide in the British newspaper industry. *The Express* is a tabloid which was a broadsheet earlier in its history, and so is to the serious end of the tabloid spectrum.

[PUT TABLE 4 ABOUT HERE]

It will be clear from the above table that, although it is not absolute, faithfulness in DW in this context is being adhered to with considerable consistency. If we first of all consider all of the newspapers except *The Times*, we can see that all 300 words involved are quoted completely accurately. The only deviations from exact quotation relate to the fact that some material which is omitted from the original is not indicated as being omitted by the use of continuation marks (as one would expect in academic works, for example). Oddly enough, the only newspaper which alters the wording is *The Times*, and then in what are clearly three very minor cases involving a total of 7 words in the original and 11 out of 133 in the report. It is, of course, very strange that only *The Times* has these inaccuracies when this was the very newspaper which carried the original article. We are not privy to exactly what happened on the day that the article and accompanying story appeared, and have not been able to find a way of checking. However, we suspect that, given that the news story appeared in the first edition of *The Times*, the *Times* journalist probably based his first-edition article, which was repeated in later editions, on a pre-final draft of Lord Simon’s article, to which only the *Times* journalists would have had access. The discrepancies

would then, of course, be accidental, and connected to the rush of newspaper production. If this was the case, there would be no abrogation of the faithfulness assumption on the part of the journalist as he would have been quoting from a slightly different version of the article. And, in any case, the discrepancies between the Simon article and the report were very small.

This evidence from written sources containing direct writing presentation in a serious political context is thus contrary to the evidence produced by those arguing against faithfulness in direct discourse, and suggests that the removal of the faithfulness criterion from discourse presentation theory ignores the fact that faithfulness is a central factor in a number of contexts. Our view is that the variations can be explained on contextual grounds depending on (a) the medium of the anterior source, (b) the medium of the posterior report and (c) other factors, e.g. the casualness or seriousness of the discourse being reported and of the reporting situation. If this is the case, we effectively need a more complex notion of faithfulness than has been hitherto assumed, one where its assumed level of importance in discourse presentation is related to contextual circumstances and pragmatic assumptions. In the final section of this article, we sketch out how such a context-based model of the salience of faithfulness in discourse presentation might work.

8 Factors determining the salience of the faithfulness criterion in discourse presentation

Below we list a number of factors which need to be taken into account in determining the salience of the faithfulness criterion in particular discourse reporting contexts. This list is merely a set of suggestions we have arrived at in the light of the research reported in this article. We hope that they may stimulate discussion on the salience of the faithfulness criterion in different media and different contexts of (re)presentation. Further research would need to be carried out (a) to investigate our faithfulness salience proposals in general terms and (b) to arrive at a more complete list of relevant factors and develop a mechanism for weighting the factors appropriately.

1. Anterior discourse accessibility

In order to report an anterior discourse accurately, the reporter must have access to the anterior discourse. Writing and speech are both in principle accessible to the reporter, but speech is only accessible at the moment of utterance, and so memory limitations intervene unless a recording is available. Writing, on the other hand, does not degrade rapidly over time, and is therefore usually available for reporters (and others) to check. The thoughts of other people are never accessible directly, and, even when people report their own previous thoughts, there must be considerable doubt as to whether an ‘anterior thought discourse’ actually existed in linguistic form or even if it can be sensibly described as a discourse. The accuracy of 1st-person thought report, like speech report, will be affected by memory limitations, and because of the private nature of thought, that accuracy can never be checked.

In order for someone other than the reporter to check whether a direct discourse presentation is accurate, access to the anterior discourse is needed after the posterior discourse has been produced, and the factors described in the previous paragraph apply in a parallel way.

Reporters are thus normally able to achieve accurate writing report more easily than accurate speech report (this is one of the factors which helps to explain the

differences between Tannen's (1989) findings for speech report in conversation and ours for writing report in the press). However, it is not clear whether the faithfulness factor can be properly applied to thought report at all (though the parallels between the thought presentation scale and the other scales may well result in interpretative analogies with those scales on the part of readers/hearers). Similar remarks relate to any pressure reporters may feel as a result of assessing whether others can check up on the accuracy of their reports. The net result of the anterior discourse accessibility factor can be graphically represented on a scale as follows (see also Sternberg 1982a for a similar scale of decreasing order of 'reproducibility'):

Anterior discourse accessible

Anterior discourse inaccessible

Writing<----->Speech<----->Thought

2. *Posterior discourse accessibility*

For someone, either the reporter or someone else, to be able to check the accuracy of a discourse presentation the posterior discourse also has to be available. Again, this is easier when the reporting medium is writing rather than speech (and so is another factor in the differences between Tannen's findings and ours in 7 above), and the notion of a posterior discourse in thought mode does not properly apply - unless the reporter can manage extra-sensory perception! For the posterior discourse accessibility factor we thus have a scale parallel to that for anterior discourse accessibility:

Posterior discourse accessible

Posterior discourse inaccessible

Writing<----->Speech<----->Thought

3. *The importance of (the wording of) what is being reported*

It seems reasonable to assume that, other things being equal, the importance and sensitivity of the original utterance(s) will affect a reporter's felt need to be accurate in report, and the expectations of those interpreting the reports. Thus the views someone has declared about the weather as (part of) a phatic conversational move will usually be less important and less sensitive than views in relation to a political debate or an academic argument, and so we would normally expect the former to be reported less punctiliously than the latter. This factor would lead to reports of the words of government ministers about matters of state being more accurately reported in DS than the comments of media personalities about fashion. Another example would be Mrs Thatcher's remarks about her political future vs. the words of John Sergeant concerning the location of the microphone in the quotation in 6.1.

More specifically, faithfulness demands and expectations are at their strongest where the *exact wording* of the original is of paramount importance. We wrote the final draft of this article in the immediate aftermath of the terrorist attacks on the United States on 11 September 2001. At this particular point, a great deal of attention was being given in the media to the exact words which were uttered by Heads of State and Defence Ministers, in order to speculate on their implications for future action and to identify signs of convergence or divergence within the emerging international coalition against terrorism. President Bush's use of the words 'war' and 'crusade' came under particular scrutiny, while much controversy was sparked by a statement by Italian Prime

Minister Silvio Berlusconi on the alleged superiority of Western civilisation over Islamic cultures. The sensitivity of the situation thus greatly increased the faithfulness stakes, so that the accuracy of quotations was sometimes a topic of discussion in its own right. On 28 September 2001, for example, an interviewee on BBC Radio 4 claimed that Berlusconi had been misquoted on the BBC's website. This person's expectations of faithfulness in direct reporting was not apparently affected by the fact that Berlusconi's original statement was in Italian and the quotation was in English. On a lighter note, faithfulness is also crucial where the wording of the original utterance is important for other reasons, such as in the case of reports of puns, plays on words, extracts from poems, and so on.

4. The memorability of the original

Particular instances of speech or writing may be particularly memorable, for example because of the unusual context in which they occurred or the particular form of the language used (cf. our discussion in 6.2 of the report of the remark made by T. S. Eliot's wife, and our reference to jokes and poetry under point 3 above). The issue of memory is, of course, very complex and has received considerable discussion in Psychology. This factor thus needs very careful consideration.

5. The status, social role and personality of the producer of the original discourse

Faithfulness is also likely to correlate with the perceived status and social role of the original speaker. The difference in status between Mrs Thatcher and John Sergeant is thus another factor in the quotation from 6.1 we have just referred to in 3 above. Similarly, we may be more punctilious about reporting carefully the words of someone who has power over us, has a nit-picking personality, or is easily hurt.

6. The social role, personality and attitude of the reporter

The role of the reporter in the reporting context can also affect faithfulness and expectations of faithfulness. A newscaster reading out a direct quotation on radio or television will be expected to be faithful to the wording of the original utterance in a way that the same person using direct speech in an informal conversation will not be. In addition, other things being equal, we might expect a reporter who is sympathetic to what is being reported, or to who said or wrote it, to be more careful in relation to faithfulness than someone who is antipathetic to what was 'discoursed' or who produced it. However, there may also be circumstances where a sympathetic reporter may try to improve on the original, and thus be unfaithful to the original. Vice versa, someone may decide to quote faithfully an utterance they strongly disapprove of in order to cast maximum blame on the original speaker. The 'punctiliousness rating' of the character of the reporter could also be a factor.

7. Text-type or speech context

Although we do not have specific examples to illustrate the point, we suspect that, unless legal proceedings are thought possible, popular magazines and newspapers are likely to be less careful about faithfulness than more serious journalistic organs. Similar remarks apply to what is said in a court of law vs. what is said in a conversation in a bar, or to what is written in an informal letter vs. what is written in an academic paper. To some degree, this factor overlaps with factor 3 above as, for example, popular journalistic media present more trivial content than the serious media. Fictional texts are a special case with respect to this factor because, as we

mentioned earlier, in fictions no anterior discourse actually occurred. The writer creates the entire fiction, and, if the narrator is deemed to be reliable (and depending on the exact narratorial circumstances), readers either assume they are witnessing fictional world discourse ‘live’ (as in most 3rd-person narrations), or, if narratorial report is clearly indicated (as in 1st-person narrators talking about their past) they pretend to themselves, in line with the fictional world’s narratorial structure, that the narrator was a witness to the ‘anterior discourse’ being ‘reported’.

8. The part of text in which reporting occurs

It may well be that position in a text might affect the faithfulness quotient. For example, there is some evidence that headlines in newspaper news reports are less restricted in terms of faithfulness for text in quotation marks than the main body of reports (cf. Short 1988).

We have characterised the faithfulness factors outlined above in terms of how strongly a reporter will feel the need to be faithful in report. We would argue that, because these factors are understandable to all, listeners and readers would apply the same kind of reasoning in assessing whether to assume that the reporter is trying to report faithfully or merely dramatise the presentation of what is being discussed. This is an issue that could be investigated by means of informant testing.

9 Conclusion

In summary, we have argued that, although those who have questioned the relevance of faithfulness in direct discourse presentation have provided some important data which needs to be taken into account, the removal of faithfulness as a factor in discourse presentation is an unwarranted conclusion, and has some unfortunate consequences in the attempt to arrive at a comprehensive theory of discourse presentation. We also believe that to come to firmer conclusions about faithfulness in particular, and discourse presentation in general, scholars will need to distinguish systematically among speech, thought and writing presentation, and think of them as three separate, but related, phenomena in discourse presentation. We have argued that, in terms of medium, the notion of faithfulness applies most easily to writing presentation, and that readers’ interpretations of the meanings and effects of presentation categories in other media are likely to derive from canonical assumptions related to writing presentation, the scale of discourse most easily open to observation. In spoken DS report, as Tannen and others have shown, faithfulness is often of marginal importance, particularly in informal and non-crucial talk. The thought presentation scale is very different from the speech and writing scales, precisely because faithfulness does not sensibly apply to this scale at all. Nonetheless, the meanings and effects of the various thought presentation categories can, in our view, only be properly accounted for by relating that scale analogously to the other scales.

The distinctions we have made here, and our text-based and corpus-based research more generally, have led us to propose a set of factors which need to be taken into account when the salience of the criterion of faithfulness is assessed in accounts of discourse presentation. We do not expect our list of ‘faithfulness factors’ to be exhaustive. However, we would want to suggest that such factors will be important in developing a context-sensitive account of the role of faithfulness in discourse presentation. We hope that other scholars will help us to take up the challenge of

developing a fuller set of factors, and investigating (a) how these factors interrelate with one another in different reporting/presentation circumstances and (b) what counts as 'faithful' in different reporting contexts. It is only when this has been achieved that we can hope to arrive at a truly general, yet discriminating, account of discourse report/(re)presentation and how it operates in different circumstances.

Notes

- 1 The research project referred to in this article was supported by the British Academy's Humanities Research Board (grant BA LRG M-AN2314/AON3489) and by pump-priming and project-completion grants from the Social Sciences Faculty and the Committee for Research of Lancaster University. We would like to thank John Heywood, Katie Wales and two anonymous reviewers for their very perceptive comments on an earlier draft of this article. A yet earlier version was presented at the Poetics and Linguistics Association conference in Potchefstroom, South Africa in 1999 and is reproduced in the conference proceedings (Biermann and Combrink 2001: 484-509).
- 2 Throughout this article, we use 'discourse presentation' as the most general term, in place of the more cumbersome 'discourse report/(re)presentation'. We discuss the different emphases of the terms 'report', 'presentation' and 'representation' in section 4.4 of this article.
- 3 However, writing is usually disregarded, or treated in the same way as speech. In our corpus-based project, writing (re)presentation has been treated throughout as a separate discourse (re)presentation cline, enabling the different forms and functions associated with writing (re)presentation to be separately identified and clarified.
- 4 For an account of a quotation in Fludernik (1993) which has similar meaning shifts in relation to 'discourse' see Short *et al.* (1997).
- 5 It is worth noting that the terms 'direct speech' and 'indirect speech' were in use well before the invention of equipment to record speech, when the chances of checking up on the faithfulness of speech report were very small indeed. The earliest instances quoted in the *OED* come from the mid-19th century, but are, of course part of a tradition which Fludernik (1993: 26ff) has traced back to Roman and Greek times. Indeed, the terms *oratio recta* and *oratio obliqua* are, as Fludernik notes, obvious antecedents of direct and indirect speech. From the earliest times examples of writing and speech, and history and fiction, were used together when exemplifying or discussing discourse categories.
- 6 The direct/free-direct distinction was applied throughout, even though we already thought when we began the process of annotation that the latter was probably a minor variant of the former. This was in order to allow us to gather data to test our hypothesis. This issue is discussed in detail in Semino *et al.* (1997).

References

- Baynham, M. and Slembrouck, S. (1999) 'Speech Representation and Institutional Discourse', *Text* **19**, 4: 61-81.
- Biber, D. (1990) 'Methodological Issues Regarding Corpus-based Analyses of Linguistic Variation', *Literary and Linguistic Computing* **5**, 1: 257-69.
- Biber, D. (1993) 'Representativeness in Corpus Design', *Literary and Linguistic Computing* **8**, 4: 243-58.

- Biermann, I. and Combrink, A. (eds) (2001) *Poetics, Linguistics and History: Discourses of War and Conflict: Poetics and Linguistics Association Conference Papers 1999*. Potchefstroom, South Africa: Potchefstroom University.
- Clark, H. H. and Gerrig, R. J. (1990) 'Quotation as Demonstration', *Language* **66**: 784-805.
- Cohn, D. (1978) *Transparent Minds: Narrative Modes for Presenting Consciousness in Fiction*. Princeton NJ: Princeton University Press.
- Collins, D. (2001) *Reanimated Voices: Speech Reporting in a Historical-pragmatic Perspective*. Amsterdam and Philadelphia: John Benjamins.
- Coulthard, M. (1992) 'Forensic Discourse Analysis', in M. Coulthard (ed) *Advances in Spoken Discourse Analysis*, pp. 242-58. London and New York: Routledge.
- Coulthard, M. (1996) 'The Official Version: Audience Manipulation in Police Records of Interviews with Suspects', in C. R. Caldas-Coulthard and M. Coulthard (eds) *Texts and Practices: Readings in Critical Discourse Analysis*, pp. 166-78. London and New York: Routledge.
- Fairclough, N. (1988) 'Discourse Representation in Media Discourse', *Sociolinguistics*, **17**, 125-39.
- Fludernik, M. (1993) *The Fictions of Language and the Languages of Fiction*. London and New York: Routledge.
- Jones, C. (1968) 'Varieties of Speech Presentation in Conrad's *The Secret Agent*', *Lingua* **20**: 2, 162-76.
- Leech, G. N. and Short, M. (1981) *Style in Fiction*. London: Longman.
- Leech, G. N., McEnery, A. M. and Wynne, M. (1997) 'Further Levels of Annotation', in R. Garside, G. N. Leech, and A. M. McEnery (eds) *Corpus Annotation*, pp. 85-101. London: Longman.
- Linell, P. (1982) *The Written Language Bias in Linguistics*. Linköping, Sweden: University of Linköping Press.
- Oostdijk, N. (1988) 'A Corpus Linguistic Approach to Linguistic Variation', *Literary and Linguistic Computing* **3**, 1: 12-25.
- McCarthy, M. (1998) *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.
- Myers, G. (1999) 'Functions of Reported Speech in Group Discussions', *Applied Linguistics* **20**: 3, 376-401.
- Sampson, G. (1995) *English for the computer: The SUSANNE Corpus and Analytic Scheme*. Oxford: The Clarendon Press.
- Semino, E., Short, M. and Culpeper, J. (1997) 'Using a Computer Corpus to Test a Model of Speech and Thought Presentation', *Poetics* **25**: 17-43.
- Semino, E., Short, M. and Wynne, M. (1999) 'Hypothetical Words and Thoughts in Contemporary British Narratives', *Narrative* **7**, 3: 307-34.
- Short, M. (1988) 'Speech Presentation, the Novel and the Press', in W. van Peer (ed) *The Taming of The Text*, pp. 61-81. London: Routledge.
- Short, M. (1996) *Exploring the Language of Poems, Plays and Prose*. London: Longman.
- Short, M., Semino, E. and Culpeper, J. (1996) 'Using a Corpus for Stylistics Research: Speech and Thought Presentation', in J. Thomas and M. Short (eds) *Using Corpora in Language Research*, pp. 110-31. London: Longman.
- Short, M., Semino, E. and Wynne, M. (1997) 'A (Free Direct) Reply to Paul Simpson's Discourse', *Journal of Literary Semantics*, **XXVI**, 3: 1997, 219-28.

- Short, M., Semino, E. and Wynne, M. (1999) 'Reading Reports: Discourse Presentation in a Corpus of Narratives, with Special Reference to News Reports', in H-J. Diller and E. O-G. Stratmann, (eds) *English Via Various Media*, pp. 39-65. Heidelberg: Winter.
- Slembrouck, S. (1992) 'The Parliamentary Hansard "verbatim" Report: The Written Construction of Spoken Discourse', *Language and Literature* **1**, 2: 101-119.
- Sternberg, M. (1982a) 'Proteus in Quotation-land: Mimesis and the Forms of Reported Speech', *Poetics Today*, **3**, 2: 107-56.
- Sternberg, M. (1982b) 'Point of View and the Indirections of Direct Speech', *Language and Style*, **15**, 2: 67-117.
- Tannen, D. (1989) *Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse*. Cambridge University Press: Cambridge.
- Wynne, M., Short, M. and Semino, E. (1998) 'A Corpus-based Investigation of Speech, Thought and Writing Presentation in English Narrative Texts', in A. Renouf (ed) *Explorations in Corpus Linguistics*, pp. 231-45. Amsterdam: Rodopi.

Table 1 - Direct Speech and Free Direct Speech

	Whole corpus	Fiction Serious	Fiction Popular	Press Serious	Press Popular	Biog Serious	Biog Popular
DS	2047 (12.4%)	347 (12.9%)	485 (16.2%)	286 (10.5%)	398 (14.8%)	121 (4.4 %)	410 (15.3%)
FDS	927 (5.6 %)	282 (10.5%)	455 (15.2%)	28 (1 %)	58 (2 %)	24 (0.9 %)	80 (2.9 %)
Total	2974 (18%)	629 (23.4%)	940 (31.4%)	314 (11.5%)	456 (16.8%)	145 (5.3 %)	490 (18.2%)

Table 2 - Direct Writing and Free Direct Writing

	Whole corpus	Fiction Serious	Fiction Popular	Press Serious	Press Popular	Biog Serious	Biog Popular
DW	109 (0.6 %)	6 (0.2 %)	5 (0.2 %)	11 (0.4 %)	11 (0.4 %)	60 (2.2 %)	16 (0.6 %)
FDW	32 (0.2 %)	4 (0.1 %)	4 (0.1 %)	0 (0.0 %)	2 (0.1 %)	19 (0.7 %)	3 (0.1 %)
Total	141 (0.8 %)	10 (0.3 %)	9 (0.3 %)	11 (0.4 %)	13 (0.5 %)	79 (2.9 %)	19 (0.7 %)

Table 3 - Direct Thought and Free Direct Thought

	Whole corpus	Fiction Serious	Fiction Popular	Press Serious	Press Popular	Biog Serious	Biog Popular
DT	38 (0.2 %)	12 (0.4 %)	7 (0.2 %)	2 (0.1 %)	5 (0.2 %)	0 (0.0 %)	12 (0.4 %)
FDT	69 (0.4 %)	28 (1 %)	30 (1 %)	0 (0.0 %)	0 (0.0 %)	5 (0.2 %)	6 (0.2 %)
Total	107 (0.6 %)	40 (1.4%)	37 (1.2 %)	2 (0.1 %)	5 (0.2 %)	5 (0.2 %)	18 (0.6 %)

Table 4 - Faithfulness in the Lord Simon Report

Newspaper	Total words in DW from original article	Comments on the character of any non-faithful aspects of the DW reports
<i>The Times</i>	133	'It was decided' becomes 'It was therefore decided', 'next January' becomes 'in January next year' and 'Until then' becomes 'During this period'.
<i>The Daily Telegraph</i>	73	All accurate, but 'In the meantime' is missing from the beginning of a lead sentence in one quotation, with no indication that it is missing. This is also the case with an intervening sentence in the middle of one quotation
<i>The Guardian</i>	29	All accurate
<i>The Independent</i>	76	All accurate
<i>The Express</i>	122	All accurate, but one intervening sentence in the middle of a quotation is not indicated as missing. There is also one added paragraph boundary in the middle of another quotation.
TOTAL	433	