

**Eötvös Loránd Tudományegyetem
Bölcsészettudományi Kar
Nyelvtudományi Doktori Iskola
Angol Nyelvészet Program**

Ferenczy Judit

**Grammatical and Textual Features of Radio News
(Rádiós hírek: nyelvtani és szövegszerkezeti jellemzők)**

Témavezető:

Dr. Kontra Miklós

a nyelvtudomány kandidátusa

Budapest

2002

Table of Contents

INTRODUCTION.....		9
1	APPROACHES TO MEDIA DISCOURSE.....	14
	1.1 Non-linguistic media studies.....	14
	1.2 Linguistic non-media studies.....	20
	1.2.1 Systemic-functional grammar.....	20
	1.2.2 Conversation analysis.....	27
	1.3 Linguistic media studies.....	31
	1.3.1 Critical linguistics.....	31
	1.3.2 Critical discourse analysis.....	36
	1.3.3 Critiques of critical media studies.....	41
	1.3.4 Non-critical linguistic media studies.....	46
	1.4 Summary.....	49
2	APPROACHES TO SPOKEN AND WRITTEN LANGUAGE.....	51
	2.1 Cultural-philosophical approaches.....	51
	2.2 Approaches from communication studies.....	53
	2.3 Linguistic approaches.....	55
	2.3.1 Typical speech vs typical writing.....	55
	2.3.2 Planned vs unplanned discourse.....	55
	2.3.3 Fragmentation vs integration and involvement vs detachment.....	56
	2.3.4 The multi-feature/multi-dimensional approach.....	58
	2.3.5 The mode-switching model.....	62
	2.4 Media language – spoken or written?.....	64
	2.5 Summary.....	66
3	THE GRAMMATICAL CHARACTERISTICS OF PRINTED NEWS.....	68
	3.1 Basic grammatical characteristics of printed news texts.....	69
	3.2 Nominal elements in printed news texts.....	71
	3.2.1 Simple noun phrases and reference.....	71

3.2.2	Complex noun phrases: premodification	72
3.2.3	Complex noun phrases: postmodification.....	73
3.3	Verbal elements in printed news texts	75
3.3.1	Verbs	75
3.3.2	Variations in the verb phrase.....	76
3.3.3	Modal auxiliaries.....	78
3.4	Adjectives in printed news texts	79
3.5	Adverbs in printed news texts	81
3.6	Clause grammar in printed news texts.....	81
3.6.1	Complement clauses in printed news texts	82
3.6.2	Adverbials in printed news texts	86
3.7	Summary	92
4	THE GRAMMATICAL CHARACTERISTICS OF CONVERSATIONAL DISCOURSE	94
4.1	Basic grammatical characteristics of conversations.....	94
4.2	Nominal elements in conversations.....	95
4.2.1	Simple noun phrases and reference.....	95
4.2.2	Complex noun phrases: premodification	95
4.2.3	Complex noun phrases: postmodification.....	96
4.3	Verbal elements in conversations.....	97
4.3.1	Verbs	97
4.3.2	Variations in the verb phrase.....	98
4.3.3	Modal auxiliaries.....	100
4.4	Adjectives in conversations.....	100
4.5	Adverbs in conversations	102
4.6	Clause grammar in conversations.....	103
4.6.1	Complement clauses in conversations	103
4.6.2	Adverbials in conversations	105
4.7	Summary	109

5	THE TEXTUAL CHARACTERISTICS OF PRINTED NEWS	113
	5.1 The sociocognitive approach	114
	5.1.1 The thematic structure of printed news	114
	5.1.2 The schematic structure of printed news	117
	5.2 The structural discourse analytical approach	121
	5.2.1 News stories and personal narratives	121
	5.2.2 The discourse structure of news stories	123
	5.2.3 The event structure of news stories	126
	5.3 A comparison of the two approaches.....	128
	5.4 Summary	129
6	RESEARCH DESIGN	131
	6.1 Quantitative research questions.....	131
	6.2 Qualitative research questions	132
	6.3 The data.....	133
	6.3.1 Data collection	134
	6.3.2 Representativity.....	136
	6.4 Quantitative methods	139
	6.4.1 Data processing	139
	6.4.2 Problematic tagging cases	141
	6.4.3 Further procedures	145
	6.5 Qualitative methods.....	146
	6.6 Summary	150
7	GRAMMATICAL FEATURES OF RADIO NEWS: RESULTS AND DISCUSSION	152
	7.1 Frequency counts for word classes	152
	7.2 Phrase-grammar	158
	7.3 Clause grammar	160
	7.4 Multi-feature/multi-dimensional results	162
	7.5 News length	166

	7.6 Summary of findings	169
8	TEXTUAL FEATURES OF RADIO NEWS: RESULTS AND DISCUSSION	172
	8.1 Topics in the BBC corpus	172
	8.2 Topics and positions	179
	8.3 Textual structure	191
	8.4 Lead/Main Event	196
	8.5 Attributions	201
	8.6 Headlines	204
	8.7 Transformations	210
	8.8 Summary of findings	222
9	A SUPPLEMENTARY ANALYTICAL FRAMEWORK FOR RADIO NEWS	225
	9.1 What is a text?	225
	9.2 What makes a text?	229
	9.3 The concept of PANTEXT	231
	9.4 Future applications.....	240
	9.5 Summary	241
	CONCLUSIONS.....	242
	APPENDICES.....	249
	REFERENCES.....	267

List of Figures and Tables

Figure 1.1 The process-and-participant matrix of a newspaper story.....	34
Figure 5.1 Hypothetical structure of a news schema	118
Figure 5.2 Outline model of news texts	123
Figure 5.3 The time structure of a news story.....	127
Figure 6.1 The hypothetical topic structure of a topic area in the BBC corpus.....	146
Figure 7.1 The distribution of word classes in radio news	153
Figure 7.2 The lexical density of radio news.....	154
Figure 7.3 The distribution of lexical words in radio news	155
Figure 7.4 The distribution of function words in radio news.....	156
Figure 7.5 The average length of news items in the BBC corpus.....	168
Figure 7.6 The position of reports in the BBC corpus.....	169
Figure 8.1 Topic areas and positions in the BBC corpus.....	181
Figure 8.2 Legal issues – topics and positions.....	182
Figure 8.3 Conflicts – topic groups and positions.....	182
Figure 8.4 Economic issues – topics and positions.....	184
Figure 8.5 Human interest stories – topic groups and positions.....	184
Figure 8.6 Military issues – topic groups and positions.....	185
Figure 8.7 Nature-related issues – topic groups and positions.....	185
Figure 8.8 Immigration issues – topics and positions.....	187
Figure 8.9 Protests – topics and positions.....	187
Figure 8.10 Media-related issues – topics and positions.....	187
Figure 8.11 Elections – topic groups and positions.....	188
Figure 8.12 Anniversaries – topic groups and positions.....	188
Figure 8.13 Scientific achievements – topic groups and positions.....	188
Figure 8.14 The textual structure of radio news	192
Figure 8.15 The Headline information continuum in the BBC corpus.....	207
Figure 8.16 The transformation tendencies in the BBC corpus.....	213
Table 4.1 A contrastive presentation of the rate of certain grammatical elements in printed news and conversations.....	111
Table 4.2 A contrastive presentation of certain grammatical features in printed news and conversations.....	112

Table 6.1 The standard deviation of word-class frequencies according to the days of the week in the BBC corpus.....	138
Table 7.1 Descriptive statistics for the frequency count of linguistic features in Biber 1988 and the BBC corpus.....	163
Table 7.2 The average length of different types of news items in the BBC corpus.....	167
Table 7.3 The average number of news items in different editions of BBC news.....	167
Table 8.1 Topic areas in the BBC corpus and their rate of coverage.....	175
Table 8.2 The coverage rate of topic groups/topics in the BBC corpus.....	177
Table 8.3 Topic structure of the Human interest stories topic area.....	178
Table 8.4 The rate of textual categories in the BBC corpus.....	195
Table 9.1 A tentative classification of possible pantexts.....	241

Acknowledgements

First of all, I express my deep gratitude to my supervisor, Dr Kontra Miklós, whose competence, attitude and work ethic not only helped me professionally at all times, but also gave me a pattern to follow.

Thanks are due to my teachers at Eötvös Loránd University, especially to the director of the English Linguistics PhD programme, Dr Varga László, whose help and patience meant a lot to me, and Dr Kövecses Zoltán, whose understanding kept me going on with my PhD studies. I also think thankfully of my linguistics teachers at Babeş-Bolyai University, Kolozsvár, especially Dr Szilágyi N. Sándor and Dr Péntek János, who instilled in me a respect and love for language.

I am grateful to Dr Allan Bell for his useful comments on my dissertation proposal, and to Dr John Heritage for helpfully sending me eye-opening reading materials.

I am thankful to the Hungarian section staff of the British Broadcasting Corporation for sharing valuable information with me.

I owe thanks to the leadership of Kodolányi János University College – especially to Dr Stephanides Éva –, who contributed to making my work smoother; and to all my dear colleagues, who did a lot of hard thinking and careful listening to bits and pieces of the BBC news with me, and tolerated my usurping the department's best computer. I am especially grateful to those colleagues, who read earlier versions of my work, or helped me in other ways: Angeli Zsuzsa, Keith Hardwick, Kiszely Zoltán, Nagy Károly, Dr Sárdi Csilla, Dr Stephanides Éva, Szemere Pál, and Szitó Judit.

Special thanks to Bodor György Csaba, virtuoso of the computer, who not only made my life incomparably easier, but also shared excellent ideas with me.

Thanks go to the staff of the English section of the Hungarian Radio, Budapest, and of Vörösmarty Rádió, Székesfehérvár, who gave me an opportunity to gain some experience in radio work.

I am grateful to my parents, friends and fellow PhD students, who trusted me, encouraged me, cheered me up, or prayed for me in dire straits.

Importantly, thanks to my loving husband, Tibor, who supported me in all my ups and downs. Finally, thank you, Tatár, for the company at walks between two sentences.

And, above all, thanks to the Ultimate Linguist.

FJ

Agárd
July, 2002

Introduction

We spend a high proportion of our time as media-consumers, making use of it either as a source of information or of entertainment. This alone is enough of a reason for a student to get interested in the workings of the media. A PhD course in English linguistics, some personal experience as a media worker, and the stage is set for a doctoral dissertation on media language. But why radio, and why the news?

First, much research has been done on the news, but little on *radio* news. Second, there are a number of studies on radio language, but little on radio *news*. It is printed news that has received most professional attention. Then comes television as the next most covered topic in the media literature. Radio has received less consideration, with some, albeit significant, studies on talk radio or radio interviews. As Scannell (1998:266) notes about radio: “that beautiful medium so largely overlooked by current preoccupations with television”. This perceived neglect, together with some time spent on the other side of the microphone, has driven me to focus on radio news.

The media has become the subject of intellectual scrutiny within many academic fields, such as cultural studies, media studies, semiotics, and film studies. The groundbreaking work done in two academic centres, Glasgow University and the University of Birmingham, prepared the way for linguistic media studies. By now, linguistic media research has come to constitute a circumscribable area of discourse analysis. However, as Cotter (2001:418) points out, work in the non-linguistic domains is referred to by media discourse researchers perhaps more than in any other topical area of the discipline. This is probably due to the manifold character of the subject matter: the media, which— in Bell’s (1995:23) terms – is an important linguistic as well as a powerful social institution. And a cultural one, we may add. Moreover, the multimodality of its codes – verbal, aural, visual – results in the media easily lending itself well to many kinds of studies.

Most of media discourse analysis has been, and still is, done in a critical framework: first under the auspices of critical linguistics, and recently under the umbrella of critical discourse analysis. The critical stance manifests itself in an interest in revealing the ways in which (media) discourse produces and reproduces unequal power relations in society. The periodical *Discourse in Society* was especially founded by van Dijk twelve years ago to open a forum for discourse analysts sharing this

ideology-sensitive perspective. The validity of such an approach is unquestionable, albeit much debated. Alternative studies have been done on cognitive aspects of news production and comprehension, generic features of news texts, practices of news production etc., often in a non-critical manner, but critical discourse analysis is currently the most influential approach to media texts.

The early critical work focused on potentially value-laden grammatical structures in a functional linguistic framework, such as transitivity, modality, nominalisation, passivisation, agent deletion. But, as Bell (1995:31) observes, the focus of these analyses is the sociopolitical significance of these grammatical forms rather than a concentration on the nature of the language structure itself. Recently, both in critical discourse analysis and elsewhere, the interest has centred round the discourse level of news. All in all, language structure below the level of the sentence has received considerably less attention than the textual level of media language. Some exceptions are studies which deal with the use of reported speech in the news (e.g. Nir and Roeh 1987), or with phonological variables in newsreaders' language use (e.g. Bell 1984).

The present study is gap-filling in three ways:

First, it focuses on grammatical features of radio news language, that is, on the neglected area of levels below the sentence. The relative novelty of the study stands in its corpus-based character: the findings are grounded on a not too big but solid database of approximately 49,000 words. A corpus-based approach allows researchers to support their observations with indisputable facts, which may or may not verify initial intuitions, but are surely apt to reveal unique characteristics of a register.

Second, the study addresses a much debated issue in the linguistic literature: that of orality and literacy. According to the most accepted view, speech vs writing are not two opposing categories but the ends of a continuum (e.g. Tannen 1982), but there are studies which question the validity of the terms altogether, and propose that a whole system of features and dimensions has to be introduced to capture variation in English (cf. Biber 1988). Media language represents a challenge for extant conceptions of orality vs literacy, but even studies that have attempted to place media language along this variation (e.g. Vagle 1991) argue that broadcast news represents a written type of language in an oral media, such as radio. The present study seriously challenges this view, and seeks for evidence that radio news displays in the form of certain grammatical features that approximate it more to spoken genres than written modes of discourse.

Third, the discourse-level studies on the textual configuration of the news have again focused on printed news, with the insightful but slightly casual remark that a model which accounts for the discourse structure of press news should be valid for broadcast news as well. This intuition is basically true, but only on the surface: certain press text categories may prove superfluous, and new textual units may be necessary, to describe accurately the make-up of radio news. Here as well, the corpus-based approach displays its advantages in an ability to also indicate to what degree the textual categories identified actually appear in radio news.

Thus, the aim of the present study is also threefold: First, it is intended to reveal the main grammatical features of radio news. Second, it seeks to challenge the assumption that radio news has an absolutely literate character, and to offer numerical data as counter-evidence. Third, it is meant to shed light on the topics and discourse structure of radio news vis-à-vis its press counterpart, also paying attention to how news items on the same topic are constructed in terms of textual units. The study will offer a new model for the topic structure of radio news, a modified model for its discourse structure, and will present a supplementary framework for analysing news on the same topic. The above objectives are projected on the subgenres of radio news as well.

The study does not have a critical stance, not because it questions the validity of research in this vein. Rather, this standpoint is based on the reasoning that exploring the distinctive characteristics of radio news may lead to a better understanding of that genre. This may, for its part, serve as a firmer basis for ideologically sensitive research, just as van Dijk's (1988a) 'sociocognitive' framework for the textual structure of printed news was implemented as a theoretical tool to reveal the ideological nature of certain news coverages (1988b, 1991).

I have judged that the objectives outlined above can best be attained through a combination of quantitative and qualitative methods. The grammatical features of radio news and its spoken vs written character have been analysed via frequency counts of word classes, phrasal and clausal elements, and other relevant features. This decision was made because only exact numbers can be compared with the results of earlier studies (Biber 1988, Biber et al. 1999) for spoken and written genres. The analyses of topics and textual structure have been done qualitatively, but supplemented by calculations akin to content analysis, in a way similar to what Bell (1991:213-214) envisages: content analysis remains a basic part of many studies, but it is best when supplemented by other methods or a researcher's own qualitative insights. The

examination of textual features has also been done in a contrastive manner: comparing the discourse structure of radio news to extant models of press stories (van Dijk 1985a, 1988a; Bell 1991, 1998).

The structure of the study is the following:

The first part of the work, entitled 'Theoretical background and previous works' is made up of five chapters. Chapter 1 is a general overview of the approaches to media language, either from a linguistic or non-linguistic perspective. Chapter 2 reviews the literature on orality vs literacy, with special attention to media language. This chapter is concerned with works which set out from the mode of discourse (spoken vs written), and try to detect grammatical features that differentiate between the two modes. Chapter 3 and 4 take a reverse route: they depart from grammatical features, and seek in what ways they are employed in oral and literate registers. These two chapters are based on a comprehensive corpus-based grammar, Biber et al. 1999, and summarise the findings of the manual on printed news and conversations. In this way, chapters 3 and 4 provide concrete numerical data of the two endpoints that the quantities in radio news can best be compared to: press news being the written, and conversations the spoken, pole. (A section in Chapter 2 also offers numerical data against which radio news can be viewed.) Chapter 5 reviews the two most wide-spread frameworks (van Dijk's and Bell's) for the textual structure of press news, also serving as bases for comparison.

The second part of the study presents the research done in four chapters. Chapter 6 presents the research design: the quantitative and qualitative research questions are laid out explicitly, and the data gathering and processing methods are described in detail. Chapter 7 provides the answers for the quantitative questions, revealing the basic grammatical nature of radio news. Chapter 8 responds to the qualitative questions, establishing the topics of the corpus, outlining the main textual characteristics of radio news, and testing the above-mentioned discourse structure frameworks on their validity for radio news. It is here that the two models are offered: a new one for the topic structure, and another for the discourse make-up of radio news (the latter is a modified version of existing frameworks for printed news). The last chapter tentatively introduces a new concept, that of 'pantext', and delineates a supplementary framework for the textual analysis of radio news, a model that is potentially generalisable for other text types as well. The study is concluded by a summary of the results obtained and an assessment of the methods applied.

Finally, some comments on the title: It was on purpose that a definite article has been avoided, in particular because I would not like to give the impression that this work will cover (all) *the* grammatical and textual features of radio news. This account is inevitably partial, and intended to signal some characteristics which I believe are central to that media genre. But certainly, there is much more to it than can be extensively treated within the scope of such a study.

1 Approaches to media discourse

This chapter is aimed at encompassing the multifarious work in media research. The account will inevitably not include all related studies, since they are great in number and varied in goals, methods and the discipline which hosts them. I propose the following division of media research: (1) media studies which are not rooted in linguistics, but in cultural or social research areas; (2) linguistic work not mainly aimed at the analysis of the media, but later substantially utilised in media-focused studies; and (3) frameworks within the domain of linguistics, particularly conceived for the analysis of media texts. The sections in the present chapter follow this categorisation.

1.1 Non-linguistic media studies

Among the earliest types of analyses that media texts were subjected to is content analysis, a quantitative sociological method which was successful in identifying some of the patterns in media texts, such as what is most extensively written about in the papers. However, it was not until the second half of the 1970's that the media were paid systematic attention to.

The two most important intellectual workshops, both in the United Kingdom, where thorough media studies were first carried out, were the Glasgow University Media Group (GUMG), and the Centre for Contemporary Cultural Studies (CCCS) in Birmingham. The *Bad news* books of GUMG set the stage for media research. In this series, lexical choices, information structure and the use of quotations were subjected to content analysis to detect bias in industrial reporting in the British press. In another volume which comprises the most important articles in this framework (*Language, Image, Media*, edited by Davis and Walton, 1983), the editors' lay out their objective of research into the language of the contemporary broadcast media as documenting bias or systematic distortion in news reporting. ('Language' is interpreted broadly to mean visual as well as verbal messages.) Davis and Walton (1983a:1), establishers of GUMG, label the work done at Glasgow University as 'communications studies', which draws from psychology, sociology, linguistics, and more practical subjects such as photography, art and design. They (1983b:9) observe that, even in the West, the range of broadcast information seeks its security within a middle-of-the-road state-set agenda,

with a ‘closure’ against certain points of view instead of ensuring balance. In a series of papers, the authors of the volume set out to detect unilateral presentations in media texts through analyses of linguistic and ideological transformations (Kress 1983), membership categorisations of newsactors and the use of metonymies in the news (Jalbert 1983) etc. The area of research, however, was not delimited to media texts only: one of the contributions, for example, analyses of topic transformation and the institutional voice in the Judiciary Committee transcripts of the Nixon White House conversations (Lerman 1983).

A similar collection of papers was published by CCCS (*Culture, Media, Language*, edited by Hall et al., 1980), a volume which is meant to show how the Centre redefined work on the media in the broader framework of ‘cultural studies’. As Hall puts it (1980a:117-118), CCCS media studies broke with behaviourist stimulus–response models, in which media content served as a trigger, into a framework which drew much more on the ideological role of the media. CCCS also challenged the notions of media texts as transparent bearers of meaning, and gave much greater attention than had been the case in traditional forms of content analysis to their linguistic and ideological structure. They also broke with a passive and undifferentiated conception of the audience, replacing it with a more active idea of readership. Finally, they paid attention to the role which the media play in the circulation and securing of dominant ideological definitions and representations.

Work done at CCCS includes further research into ideological dimensions of media messages (Heck 1980), but also challenges the terms of ‘bias’ and ‘distortion’ (Connell 1980). Hall’s (1980b) famous concept of ‘encoding/decoding’, which replaced the linear, sender/message/receiver model of media communication with a dynamically conceptualised process, was also conceived within this framework. In the model proposed by Hall, a raw historical event is first signified, that is, encoded, within the aural–visual forms of, say, television, before it can become a meaningful discourse. This is then appropriated as such, and meaningfully decoded by the audience. It is only in this moment that the ‘effects’, the ‘uses and gratifications’ identified in positivistic research, can take place.

Van Dijk (1985b:v, 1988a:13) acknowledges the value of the work done in the Centre by pointing out that, instead of the surface units of quantitative content analysis, as represented by GUMG, here there is an account of underlying meanings or processes of signification. Nevertheless, he finds that the contributions of the Glasgow University

Media Group and the Centre for Contemporary Cultural Studies, cannot be regarded as a systematic discourse analysis, though they are first steps towards a full-fledged discourse analytical approach.

Parallel to the work done in the United Kingdom, media research in the United States took a more sociological direction. An important token of this work is Fishman's *Manufacturing the News* (1980), which is mainly concerned with examining the routine work of reporters to discover how news gets constructed out of an amorphous world of happenings. The main finding of the book (1980:140) is that news is a practical organisational accomplishment, and that news reporters heavily rely on the definition of the phenomena by governmental agencies and corporate bureaucracies. The author (1980:16) has reached this finding through examining the four stages of the news production process: news detection, interpreting occurrences as meaningful events, investigating their factual character, and assembling them into stories. The method of research has been extensive participant observation of newswork on a single newspaper in California. The most instructive part of the book is where the reader gets an insight to the journalists' so-called beats (1980:28), that is, objects of reporting consisting of a domain of activities occurring outside the newsroom, such as the city hall beat, the police beat or the justice beat. On a day-to-day beat round, the reporter meets people who provide him/her with valuable information which can later serve as news material. Fishman (1980:155) concludes that routine journalism communicates an ideological view of the world: what newswriters report is not what actually happens, not what is actually experienced by participants or observers of news events. It is a story woven around hard data, based on the bureaucratic categories and definitions of events that agency officials mean to happen.

In a review of the media literature, van Dijk (1985c:6) thinks that the main point of these types of studies is the recognition that, along the production phases identified by Fishman, the final news items are the results of a complex sequence of textual processing stages. However, van Dijk calls for a better understanding of which linguistic, cognitive and social factors impinge on this text processing. One way to this understanding, he proposes, is to see how cognitive models account for the strategies used by journalists in these text processing stages (see also van Dijk and Kintsch 1983, Section 5.1.1).

Another American accomplishment is Herman and Chomsky's *Manufacturing Consent* (1988), which however, in Cotter's (2001:431) view, has not had the same

academic impact as the Glasgow Media Group, for example. Herman and Chomsky challenge the myth of a free press, and propose a 'propaganda model', in which raw news go through a set of 'news filters' (1988:2) until they become fit to print. These filters marginalise dissent, and allow the government and dominant private interests to get their messages across to the public. The filters (1988:3-35) are the following: (1) the size, ownership and profit orientation of the mass media firms; (2) advertising as the primary income source of the media; (3) the reliance of the media on information provided by government, businesses and 'experts'; (4) 'flak' as a means of disciplining the media; and (5) anticommunism as a national religion and control-mechanism.¹ As Herman and Chomsky (1988:31) put it, the five filters narrow the range of news that passes through the gates and even more sharply limit what can become 'big news', subject to sustained news campaigns. They (1988:298) conclude that the societal purpose the media serve is not that of enabling the public to assert meaningful control over the political process by providing them with the information needed, but to inculcate and defend the economic, social, and political agenda of privileged groups that dominate the domestic society and the state. However, Herman and Chomsky (1988:302) are among the first to realise that this is not necessarily a conscious process: the system of presuppositions and principles which constitute an elite consensus are so powerful that they are internalised largely without awareness, sometimes resulting in a self-censorship on the part of the journalists.

The book reviewed above has an undeniable left perspective, as do other works, such as the volume *American Media and Mass Culture*, edited by Lazere (1987). Some of the main ideas expressed in the book are the following: news reproduces itself as a historical given by being strongly associated with, and spatially and temporally embedded in, the activities of legitimated institutions (Tuchman 1987); the media have come to accept a corporate version of reality (Dreier 1987); the media promote an image-centred view as an alternative to the age-old, traditional language-centred view of the world, which may be harmful to the youth (Postman 1987) etc.

Less sociological and more communications-oriented are the works of Hartley and Fiske. *Reading Television*, by Fiske and Hartley (1978) was born of the feeling that the tools of traditional literary criticism do not quite fit the television discourse. The book has a semiotic perspective, and undertakes to interpret the signs that television

¹ Now that the Cold War is over, this anticommunist filter may be substituted by an 'anti-Middle-East' one.

uses to encode its messages. After affirming (1978:21) that the starting point of any study of television must be with what is actually there on the screen, best established by content analysis, the authors cautiously warn (1978:24) that television does not represent the manifest actuality of our society, but rather, reflects symbolically the structure of values and relationships beneath the surface. To reveal what is in the depth, the authors borrow Barthes' concept of the orders of signification. In the first order we find denotations, representations; in the second are the connotations, the cultural meanings, the myths. The authors (1978:46) add a third order of signification, the locus of ideologies, that is, the broad principles by which a culture organises and interprets the reality with which it has to cope. Translated to television (1978:82), this means that a culture communicates with itself via second-level professional communicators who are manifested as the first-level encoded messages on the screen.

The most widely quoted part of the book is the concept of television as the culture's modern bard. According to this (1978:88), TV performs seven functions that the bard performed in a traditional society: (1) to articulate the established cultural consensus about reality; (2) to implicate the members of a culture into its dominant value systems; (3) to celebrate, explain, interpret and justify the doings of the culture's representatives; (4) to assure the culture of its practical adequacy by affirming and confirming its ideologies; (5) to expose, conversely, any practical inadequacies in the culture's sense of itself; (6) to convince the audience that their status and identity is guaranteed; and (7) to transmit a sense of cultural membership.

Fiske and Hartley (1978:105) also give the audience an active role in stating that, although a preferred meaning of the media message emerges at the connotative level, the same message can be decoded according to different codes, corresponding to the social experience of the decoder, and yet remain meaningful for all groups.

A development of the above views is to be found in Hartley's *Understanding News* (1982) and Fiske's more general *Introduction to Communication Studies* (second edition, 1990). Hartley continues his semiotic approach in further decomposing the sign-system of the news media, concentrating on the press, but mainly on TV, to find out how we become 'news-literate', that is, how we acquire to interpret the world in terms of the codes we have learnt from the news. Hartley (1982:21) proceeds on a structuralist route, through a concern for syntagmatic and paradigmatic axes of signs, for example, by showing why the syntagm 'the terrorists liberated sg' is unlikely because the two elements belong to two opposing discourses: *terrorist* is a boo-word,

and *liberate* is a hurray-word. An important insight is Hartley's (1982:118) proposal that the news and TV fiction (e.g. police series) are structured in the same way. News is more than an aggregate of distinct bits; there is the organising level of narrative, and the integrating level of newsreaders and, whilst you concentrate on one element, you are simultaneously encouraged by the narrative structure to use that element to make sense of the opposing one. Thus, news stories often cohere into a pattern by collectively signifying particular themes, issues and meanings.

This semiotic line of research represented by Hartley is continued on the columns of the journal *Media, Culture and Society* founded, among others, by Scannell in 1979. As Bell (1991:215) sees it, the semiotic approach differs from content analysis in being qualitative rather than quantitative, in focusing on subsurface rather than overt content, and in concentrating on the likelihood of different audiences giving different readings to the same text.

Fiske's introductory book is not only concerned with the media, but with communications in general. Starting from an overview of the communication models available in the literature – including general ones such as Jakobson's and models especially worked out for the mass media –, Fiske goes into a thorough discussion of codes and signification. His book (1990:133) is meant to show how structuralism teaches us to look for the deep structures that underlie all cultural and communicational systems, and how it can be utilised for a better understanding of the workings of the media. (For a communications-oriented study of the radio as a medium, see Crisell 1994.)

This section will be concluded by reference to a volume edited by van Dijk, entitled *Discourse and Communication* (1985). The book originated from the editor's (van Dijk 1985b:v) recognition that the study of mass communication and discourse analysis seems to ignore each other, a small link being content analysis. Thus, the book was intended to bridge this gap of "happy single-mindedness in the two disciplines" (*loc cit*). Indeed, the book contains an impressive collection of papers, some more communications-related (Davis 1985, Gerbner 1985,) others with a more linguistic/discourse analytical orientation (van Dijk 1985a, Connell and Mills 1985), and again others representing a transition between the two disciplines (e.g. Bentele 1985).

The present section has offered a short overview of the most commonly quoted non-linguistic media literature. Some of the most significant work in communication

studies and sociologically-based media studies have been reviewed, together with evaluations from authors who are now in the centre of media studies. As Cotter (2001:418) points out, work in the non-linguistic domains is referred to by media discourse researchers perhaps more than in any other topical area of discourse analysis. This explains why these studies, which served as points of orientation for linguists interested in the language of the media, have been paid attention to here as well.

1.2 Linguistic non-media studies

This section will take a reverse route: here, some important linguistic works will be reviewed, works which did not originate with an intention to provide tools for media analyses, but which were later utilised as such. Again, the overview will be short, focusing on just the general principles of the studies.

1.2.1 Systemic-functional grammar

Systemic-functional grammar is one of the most influential trends in present-day linguistics. The approach rests upon the Prague school and the Firthian tradition in its concept of ‘system’, namely, that a text is never simply a product, it is a process coming out of language as a system. One of the most important representatives of this approach is Halliday, who intended his *Introduction to Functional Grammar* (1985a) to be a reaction against grammar traditionally being the grammar of written language. In a systemic approach, the distinction between speech and writing dissolves: despite our every-day experience that “writing exists, whereas speech happens” (1985a:xxiii), we realise that both derive from the same system. (For a more detailed review on speech vs writing, see Chapter 2.)

Halliday (1985a:53) postulates three kinds of meaning that are embodied in human language: (1) ideational, that is, the representation of experience; (2) interpersonal, i.e. meaning as a form of action; and (3) textual, that is, relevance to context. If we translate this to the meaning of a clause, the ideational function is that of representing ‘processes’; the interactional function is that of exchanging roles (statements, questions, offers and commands), and the textual meaning is that of constructing a message. We will have a look at how these meanings are realised one by one.

To say that a clause has an ideational function, that it represents a process, is to attend to the grammatical system of transitivity, which, in Halliday's (1985a:101) terms, specifies the different types of processes that are recognised in the language. A process potentially consists of three components: the process itself, participants and circumstances. Processes can be material (actions and events), mental (perception, affection, and cognition) and relational (attribution and identification) or, additionally, behavioural, verbal and existential. Material processes involve Actors and potentially Goals, behavioural processes Behavers, mental processes display Sensers and Phenomena, verbal processes activate Sayers and Targets, existential processes are about Existents, and the participants in relational processes are Tokens and Values, Carriers and Attributes, and finally, Identifiers and Identified entities. Additionally, there may be Beneficiaries (Recipients, who are given to, or Clients, whom services are done for) and Ranges (the scope of the process). Some of the principal types of circumstantial elements are Extent and Location (in time and space), Manner (means, quality and comparison), and Cause (reason, purpose and behalf).

Turning to the interpersonal function of language, Halliday (1985a:68) considers that the two most fundamental types of speech roles are giving and demanding. These define the four primary speech functions: when one is willing to give goods and services, that is an offer, when one demands them, that is a command; when one gives information, that is a statement, and when one demands some, that is a question. Aspects of the interpersonal function of a clause include tense, mood, modality and polarity (whether it is positive or negative).

As for the textual function of a clause, the basic form of message organisation in a clause is through Theme and Rheme. Following the Prague school terminology, Halliday (1985a:38) defines the Theme as the element which serves as the point of departure of the message; it is that with which the clause is concerned. The remainder of the message, the Rheme, is the part in which the Theme is developed. Whatever is chosen as the Theme is put first in the clause. Themes can be unmarked if they coincide with the grammatical subject of the clause, and marked, if an adjunct or a complement moves to front position. If there are conjunctive adjuncts (such as *that is, or rather, briefly, likewise, nevertheless* etc.) or modal adjuncts (such as *probably, sometimes, in my opinion, frankly, luckily, strictly speaking* etc.), these tend to be thematic. Simple themes do not, but multiple themes have a further, internal structure of their own (including beside the grammatical subject, for example, vocatives). Theme and Rheme

are not synonymous with Topic and Comment, since, in Halliday's (1985a:56) terms, the Theme of any clause extends up to, and includes, the topical Theme, which is the first element in the clause that has some function in the ideational structure, that is, in transitivity.

Halliday's other significant contribution to the linguistic literature is on cohesion, the canonical work being Halliday and Hasan's *Cohesion in English* (1976). This work (1976:1-2) defines text as a unit of language in use, which is not a grammatical unit, nor is it defined by its size; it is a semantic unit. This unity with respect to its environment can be captured in the concept of 'texture' – the property of being a text –, provided by the cohesive relations between two interdependent elements in a text which are coreferential. One occurrence of a pair of related elements is called a cohesive tie. The concept of cohesion is not a structural, but a semantic one: it refers to relations of meaning that exist within the text. There are different kinds of cohesive ties: reference (situational and textual, a semantic relation), substitution (by pro-forms, a grammatical relation), ellipsis (also grammatical), conjunction (a semantic relation, expressing certain meanings which presuppose the presence of other components in the discourse), and lexical cohesion (achieved by the selection of vocabulary, that is, repetition of words, the use of synonyms or near-synonyms and collocations).

Texts cannot be interpreted without their contexts. For describing how the context of the situation determines the kinds of meaning that are expressed, Halliday and Hasan (1976:22) offer three general concepts (which follow from some of their earlier research): field, mode and tenor. The field is the total event, in which the text is functioning, together with the purposive activity of the speaker or writer; it thus includes the subject-matter as one element in it. The mode is the function of the text in the event, including therefore both the channel taken by the language – spoken or written, extempore or prepared – and its genre, or rhetorical mode, as narrative, didactic, persuasive, 'phatic communion' and so on. The tenor refers to the type of role interaction, the set of relevant social relations, permanent and temporary, among the participants involved. Or, in another formulation (Downes 1994:3510), field is the social activity performed; tenor is social roles enacted; and mode is the role of language in the situation. The linguistic features which are typically associated with a configuration of situational features – with particularly values of the field, mode and tenor – constitute a register. (For a brief review of how functional grammar can be used to interpret discourse, see Halliday 1985b.)

Other linguists, such as Martinet (for a review, see Martinet 1994) or Martin (1992), have taken different routes, or modified Halliday's functional grammar. Martinet's model is a very theoretical system, based on the concept of 'moneme', the minimal linguistic unit combining an acoustically perceptible form and a reference to some aspect of experience. Monemes are the same as units referred to in other grammars as morphemes, but Martinet avoids this term because it suggests form rather than meaning.

While Halliday's understanding of cohesion as non-structural textual resources remained on the level of lexicogrammar, Martin (1992) reworks the concept on a more abstract level of discourse semantics, focusing on text-size rather than clause-size meanings. As Martin puts it (1992:19), his book sets up a level of discourse semantics stratified with respect to lexicogrammar on a content plane, which will permit generalisations across structural and non-structural textual relations. In a review article, Martin (2001:39) points out the strength of his model, namely that it is stratified and, in this way, it can amount to the study of patterns of interaction among discourse semantics, lexicogrammar, and phonology or graphology in realisation. Martin (1992:494) refers to the system of tenor, field and mode collectively as register (a meaning slightly different from Halliday and Hasan's understanding), and deploys an additional level of context, above and beyond tenor, field and mode, referred to as genre. Genre comprises the ways in which field, mode and tenor variables are phased together in a text. In other words (Martin 2001:46), register is a pattern of linguistic choices, and genre a pattern of register choices.

Systemic-functional grammar may be utilised as a descriptive tool, but it can be put to 'critical' uses as well (e.g. in critical linguistics, see Section 1.3.1). Martin (1992:2) states that *English Text* has been written as a contribution to the linguistics envisioned by Halliday as "an ideologically committed form of social action".

Several concepts used in systemic functional grammar outlined above have been projected to media discourse. Such an enterprise is van Leeuwen 1987, which examines generic strategies in press journalism. Van Leeuwen posits a system of generic stages, which are derived from configurations of transitivity, mood, theme, conjunction, reference and tense. What van Leeuwen (1987:201) observes is that the social purposes (that is, the generic affiliation) of the stages which make up schematic structures can be derived from the functions of the linguistic signifiers that realise them. For example, a stage in a text which has predominantly temporal conjunction, declarative mood, people

as themes, specific reference, past tense and material, mental and verbal processes serves to relate past events, and is therefore a narrative stage. A stage which has predominantly logical and elaborative conjunction, declarative mood, topics or issues as themes, generic reference, universal present and relational processes serves to explain the world, and is therefore an expository stage. Van Leeuwen (1987:202) emphasises that he uses these categories to identify stages in a text, not to label whole texts as belonging to a narrative or expository genre.

The stages then give rise to generic structures. According to this structure, journalism is double-faced: interpretations are attributed to authorities or experts, the narration is done by the journalist. Some of the generic strategies that journalists apply to reach the generic structure are: adhortation, tension (whether or not two points of view are opposed in an article), description, and exposition. Van Leeuwen's approach to genre analysis provides a grammatical basis for recognising generic stages, and is able to show context-specific generic strategies which realise the social purposes of journalism (such as entertainment, social control etc.).

Van Leeuwen extends his model and places it under the umbrella of critical discourse analysis in van Leeuwen 1993. He (1993:193) propounds that discourse has a dual aspect: it is an instrument of power and control as well as an instrument of the social construction of reality. To bring out this dual aspect and apply it to multimodal texts, van Leeuwen (1993:194) proposes that discourse has not only a generic but a field structure as well. Generic structure is the syntagmatic structure of discourse, its 'beginning–middle–end' structure, which also, at the same time, realises discourse as social practice. Field structure is more dispersed through the text and realises the knowledge of some field as it is constructed in the context of a given institutional domain: e.g. the knowledge of politics as it is constructed in the mass media. In other words (1993:203), generic structure is the structure *of* the text, field structure is a structure *used* in the text, and its traces can be dispersed through the text in a multitude of ways. Van Leeuwen then offers an analysis of field structure, through which, for example, we may see whether a participant in the social practice is or is not referred to explicitly, whether (s)he is presented as human or not, whether (s)he is called by their name or not, backgrounded or not etc. Such an analysis allows the comparison of certain discourses with others, since it is able to detect the differences in the ways the same field is recontextualised in various discourses. (For a more recent analysis of multimodal texts, see Kress and van Leeuwen 1998.)

Another rethinking of systemic-functional categories, applied to media texts, is Östman 1999. While van Leeuwen's generic and field structure are social concepts, Östman introduces the term 'discourse pattern' as a cognitive construct. He (1999:78) thinks of genres and text types as forming a dichotomy, with genre zooming in on the external relations that a text/discourse displays, and text type focusing on its internal relations. For example, a genre like 'novel' can be written in the form of a narrative text type, but can also be in the form of, say, an argumentative text type. Östman (1999:80) continues that, as we need a 'meaning' filter between the traditional categories of form and function, we also need a filter between genre and text-type descriptions: that of discourse pattern. If text type is a structural and genre a socio-cultural category, then discourse pattern is a cognitive one. Östman (1999:86) accounts for why 'discourse pattern' is needed by saying that, for example, contact ads in British, American and Finnish papers are all of the same genre and the same descriptive text type, but their discourse pattern is different. English contact ads follow either a Human interest pattern (a text with a number of golden nuggets that will draw the reader's attention) or a Plot pattern (written in the form of a story with a plot). Finnish contact ads are very much different: they prototypically have the same structure and form as ordinary advertisements for cars or washing powder, that is, they have an Ad discourse pattern. Östman (1999:91-96) goes on to show that the concept of discourse pattern can successfully be used in an analysis of printed news: articles on the same topic may follow a totally different discourse pattern. Östman (1999:97) argues that discourse patterns play an important role in understanding and coherence, which receives support from research indicating the importance of scripts and cognitive schemata (see Lakoff 1987).

The concept of discourse pattern is put to test by Halmari and Östman (2001), where news coverage of a controversial nature is analysed. It is proposed that, if non-prototypical content turns out to require a change in discourse pattern while the text type and genre remain the same, the explanatory power of the notion is diminished. In an analysis of the coverage of a highly irregular execution, where the inmate was non-prototypical in being a young, attractive and religious woman, the authors (2001:820) found that the discourse pattern 'facts–reason of execution–description of crime–quotation of inmate's last words–quotes from victim's family witnessing execution' was still adhered to. A crucial question Östman (1999:89) asks in his earlier paper, that of how many discourse patterns are there, is answered here: there is no need to establish a

new pattern each time the existing ones cannot account for the empirical data. Rather, discourse patterns are dynamic and flexible.

Other studies conceived in a systemic-functional vein are based not on the concepts of genre and field, but on theme and topic. An example is Moya and Albentosa 2001, which, first of all, presents a critique of the Hallidayan concept of topical theme, and posits that topic and theme are two distinct concepts which should be defined from two different perspectives. Theme is a structural category whose main function is to determine the point of departure of the message, by its location in the clause. Following van Dijk (1977, van Dijk and Kintsch 1983), Moya and Albentosa view topic as a pragmatic and cognitive category that expresses what the message is about (see a discussion of van Dijk's similar concept in Section 5.1.1). This 'aboutness' comprises the communicative purpose of the writer, knowledge shared by reader and writer, and the linguistic and extralinguistic context. On a sentence level, the entity about which information is given is called 'local topic'.

Moya and Albentosa (2001:355-363) apply the notions described above to news items and tourist brochures to establish differences in topical and thematic organisation. They find that tourist brochures often introduce topics for the first time in the discourse in sentence-final positions, to create an atmosphere of great expectation. In 35% of the sentences in tourist brochures, theme and topic are realised by different clause elements. This rate is only 16% in the case of news, where the introductory topic is usually located in thematic position. This relates to the basic informative function of news. The authors (2001:365) conclude that, while news follows the unmarked pattern, in which the local entity about which information is given is clearly activated from the outset, tourist brochures are structured according to their persuasive function via placing local topics further from thematic positions to make the reader take interest. These relevant genre differences show that topic is a discursive and contextual function which goes beyond the rigid syntactic order of the English clause.

Finally, the cohesive means of reference has also been examined in media language. An example is Kronrod and Engel's study (2001), which examined referring expressions in newspaper headlines, also in the light of Ariel's accessibility theory of antecedents. In this theory (referred to by Kronrod and Engel 2001:684), high accessibility is accredited to personal pronouns as referring expressions, intermediate accessibility is attributed to the use of either first or last names with up to two-word long definitions, and low accessibility is ascribed to full names with modifiers or

definitions longer than two words. In an analysis of street newspapers and subscription ones, the authors (2001:691) found that genre does not influence referring style; both papers contained mainly intermediate accessibility markers due to a compromise between two conflicting tendencies: to be clear in referring and to save space and attract attention by omitting central pieces of information and using opaque language.

We have seen that systemic-functional grammar has been an extremely fruitful area of research, seminal for media studies as well. Concepts such as genre, field, theme, topic and reference have been interpreted and utilised in many works, with enlightening results. A significant intertwining of functional grammar and media research is represented by the emergence of critical linguistics, which will be discussed in Section 1.3.1.

1.2.2 Conversation analysis

Conversation analysis (hereafter CA) is a trend in linguistics with roots in both pragmatics and sociology, that is, it studies language-in-use in a primary locus of social interaction: conversations. The main objectives of CA, as seen by Drew (1994:749), is to uncover the sociolinguistic competences which underlie the production and interpretation of talk in social interaction. The principal focus of CA is ordinary conversation, following the aims of the sociological trend called ethnomethodology, represented by Garfinkel, Sacks, Schegloff, Jefferson etc. As interpreted by a practitioner (Psathas 1994:1161), ethnomethodology studies the methods actually practiced by members of society for accomplishing whatever it is they are doing, including their ways of talking about whatever it is they are doing. Translated to CA, this means that CA also aims at explicating the ways in which coparticipants make sense of one another's talk.

The fact that the analysed data are always ordinary conversations is absolutely central to CA. Heritage (1995:394) accounts for this by saying that natural conversation is the fundamental domain of interaction and indeed a primordial form of human sociality, since it is the primary medium of communication to which the child is exposed and through which socialisation proceeds. CA practitioners are deeply concerned also with methodology, first of all, with ways of transcribing oral communication. The transcription systems used by conversation analysts are not phonetic, but they do contain more than a mere transcription in terms of words. As

Drew (1994:750) informs us, CA transcription enables many essential characteristics of speech delivery to be conveyed, including some morphological and phonetic alternations, lexical choice, speech perturbation, and aspects of prosody such as stress and sound lengthening. More simply put, CA transcripts capture those aspects of talk which are interactionally salient.

What CA is mostly interested in is how conversations are systematically built up, how utterances follow from each other. The main conversational units they identify in conversations are the so-called adjacency pairs, that is, “utterances produced by two successive speakers such that the second utterance is identified as related to the first as an expected follow-up” (Richards and Schmidt 1983:128), openings and closings, topic selection and introduction, and norms of turn taking, that is, conventions which determine who talks, when, and for how long. (For a detailed discussion of CA, see e.g. Psathas 1995.)

Conversation analysis has fertilised media studies as well, resulting in a wide range of works using broadcast media conversations as data. Some of the main linguists who performed such work are Heritage, Greatbatch and Hutchby. Heritage (1985), for example, examined the differences between natural conversations and news interviews. He (1985:112) sees news interviews as a functionally specialised form of social interaction produced for an overhearing audience and restricted by institutionalised conventions. Heritage (1985:99-100) found that alignment work, which is common in natural conversations in the form of third-turn receipts such as *Oh, really?*, are lacking in news interviews. What is there, however, is ‘formulation’, that is, ways in which the interviewer summarises, glosses or develops the gist of the answerer’s earlier statements. This is specifically done for the sake of the overhearing audience, not for the interviewee, therefore lack of alignment with the informant is supplemented by an overall, if tacit, orientation to the news audience. Thus, the interviewer acts as the elicitor, but not the primary recipient, of the talk.

Heritage and Greatbatch (1991) aim at capturing the institutional character of news interviews. It is again found that the configuration of normative conventions in news interviews is distinctive from ordinary conversation, but also from other institutional forms of interaction (Heritage and Greatbatch 1991:130). The turn-taking system follows the simple rule that one party, the interviewer, will ask questions, and the other, the interviewee, will answer them. This system is associated with a central task and a core exogenous constraint of the news interview: the elicitation of talk

produced for the overhearing audience by an interviewer who should properly maintain a formally neutral or 'neutralistic' posture.

Greatbatch (1998) examines just how this neutralistic stance is established by interviewers. 'Neutralistic', in his terms (1998:167) describes a manner of style of interviewing, which refers to patterns of interviewer conduct that can escape formal charges of bias. The most important means by which this neutralistic stance is maintained are the following: interviewers produce utterances that are at least minimally recognisable as questions, and they refrain themselves from evaluative statements by attributing them to third parties. Interviewees usually collaborate in the maintenance of the neutralistic stance in that they do not generally treat the interviewer's utterances as expressing or indexing personal standpoints, but as soliciting their viewpoint on the issues raised. However, interviewees may also challenge, sometimes even openly characterise interviewers as having expressed a point of view on their own behalf.

Heritage and Roth (1995) go into how exactly the question-answer system of news interviews is organised. The paper is also distinctive for its supplementing the basically qualitative methods customary in CA with quantitative data analysis. It is found (1995:4) that interviewers normally attempt to question interviewees, who, for their part, produce responses, but departures from the norm are often specially accountable or the objects of protest or conflict. It is also found that a strictly grammatical coding of the data proved inadequate, given the interactional nature and the institutional context of news interviews. Therefore, the grammatical core was augmented to include instances where questioning was accomplished through forms other than questions, such as directives, declaratives that invoke sentiments or opinions, and phrases functioning as questions etc. The reverse may also happen, when a question does not do questioning at all. However, the final result (1995:52) is that interviewers massively restrict themselves to turns that take a questioning form. (For a summary of news interview research from a CA perspective, see Heritage 1998.)

News interviews are not the only subgenres to be analysed with CA methods. Hutchby is known for his studies of talk radio programmes. In his 1996 book, Hutchby analyses arguments, asymmetries and power on talk radio. The work brings a new perspective to CA often criticised for remaining agnostic on the issue of power: Hutchby (1996:109) successfully shows that CA can in fact be used to address that issue in a highly contextually and interactionally sensitive way. He demonstrates that interactional asymmetries abound even on talk radio, which is regarded as a democratic

forum. The opening of the call is not only designed to set up an environment in which callers introduce their topics, but it also places the participant on significantly asymmetrical footings (in a Goffmanian sense) with respect to those topics. The fact that callers are required to go first by expressing a point of view puts the host into a position from which (s)he can criticise or attack the caller's line simply by exhibiting scepticism about its claims.

Hutchby continues the Goffmanian line in his 1999a paper, in which he further analyses frame attunement and footing in the organisation of talk radio openings. He argues (1999a:41) that institutions do not define the kind of talk produced within them; rather, participants' ways of designing their talk actually constructs the institutionality of such settings. On talk radio, this takes place in not more than the first three seconds of each call. Hutchby (1999a:60) observes that the initial caller-identificatory announcement sees the host on a dual footing, addressing both the overhearing audience and the next selected caller. The host is thus an institutional representative at the interface between the 'private' sphere from which the caller is moving and the 'public' sphere of the broadcast colloquy. Callers, for their part, exhibit and sustain their identity as 'lay' contributors entering into the institutional frame of the broadcast.

This section has undertaken to show that conversation analysis has been influential in media studies, despite Scannell's (1998:258) view that CA came to the study of media, if not accidentally, at least tangentially. CA has been widely criticised for the small lens it uses, namely that it looks at individual interactions only, not being able to encompass broader relationships. Fairclough (1995b:23) and van Dijk (1994a:163), for example, reproach CA with being resistant to linking properties of talk with higher-level features of society and culture, such as ideologies or cultural values.² However, Garrett and Bell (1998:14), in a comment on a CA-based analysis (Greatbatch 1998), say that, by CA paying meticulous attention to the details of conversations, a CA analysis shows how much we can learn through just such a close reading of news interviews.

² More details on the complaints against CA follow in section 1.3.3.

1.3 Linguistic media studies

What follows in this section is a brief review of linguistics-based studies worked out especially for the study of media language. However, the analytic system they offer is more than just a toolbox for the better understanding of media texts; it is also a philosophy of language, which may not be radically new, but which certainly has a more polarised focus than the linguistic approaches reviewed in the previous section.

1.3.1 Critical linguistics

Critical linguistics is an approach which took up the critical, ideology-sensitive stance of non-linguistic studies such as those of the Glasgow Media Group and the linguistic implementation of Hallidayan systemic-functional grammar. The principal representatives of the approach are Kress, Hodge, and Fowler.

The theoretical foundation of critical linguistics is laid down in Kress and Hodge's *Language as Ideology* (1979), with the aim of showing that the study of discourse does not fall outside the scope of theoretical linguistics, which, in the sixties, meant a pure Chomskyan syntactic theory. However, critical linguistics continues both a Chomskyan and a Whorfian tradition in seeing language as very closely connected to the workings of the mind. Kress and Hodge (1979:1) envisage a linguistics which has an ambiguously wide scope: equally concerned with the relation between language and the psyche and with the connection between society and language. Language, as they (1979:13) see it, is an externalisation of a society's consciousness.

The relativistic tradition is, of course, much stronger in the theoretical basis of critical linguistics than the Chomskyan line. Kress and Hodge (1979:5) remind their readers that there is no 'pure' act of perception, no seeing without thinking, that we interpret the flux of experience through means of interpretative schemata, largely pre-packaged by the categories of language. In the linguists' witty formulation (*loc cit*):

Whatever has a name can become familiar, and is easier to classify and remember. Only what has a name can be shared. Communicable perception has to be coded in language. So language, which is given by society, determines which perceptions are potentially social ones.

This is how the psychological aspect intermingles with the social area. Language provides the categories and rules based on fundamental principles and assumptions about the world, which are not determined by thought, they *are* thought. But language is also the practical consciousness of a society, which is a partial and false consciousness, that is, an ideology. Thus, language, as Kress and Hodge (1979:6) conclude, is an instrument not only of communication but also of control. Linguistic forms allow significance to be conveyed and to be distorted. In this way, hearers can be both manipulated and informed, preferably manipulated while they suppose they are being informed.

This is what, from a critical linguistic stance, news is all about. Distortion is done through the grammar of a language, which is its “theory of reality” (1979:7). This is where Hallidayan functional grammar comes into the picture: language is a set of categories and processes, of models which describe the interrelation of objects and events. The first group of models is actional, in which one model is transactive, involving two entities related by a process: the causer and the affected of an action, such as *The batsman struck the ball*. The other model is non-transactive, relating one entity and a process, such as *The batsman runs*. The other group of models is relational, comprising equative models, such as *John is Lear*, and attributive ones, such as *John’s Lear is stunning* (Kress and Hodge’s examples). Transactive is not identical with transitive, since *Bill resembles his father* is transitive, however, it is not actional but relational. At this point, Kress and Hodge (1979:9) borrow Chomsky’s term of ‘transformations’ in a partly different sense: transformations are a set of operations on basic forms, deleting, substituting, combining or reordering a syntagm or its elements. However, they are not seen as being innocent, like in a generative-transformational theory: it is proposed that, in actual discourse, transformations alter the meaning of the basic form in not only making the linguistic form shorter, but also distorting its meaning.

A linguistic checklist of items, or categories of structure that figure frequently in critical practices is offered by Fowler (1985:68-74). The most important items are reproduced here:

- Lexical processes
- Transitivity
- Nominalisation
- Modality

Lexical processes refer to instances of lexicalisation, that is, the provision of a term for a concept. If many words are available for a concept, that is called 'overlexicalisation'; if there is lack of a term that would neatly encode a concept, that is 'underlexicalisation'. Overlexicalisation indicates the prominence of a concept in a community's beliefs and interests; underlexicalisation often results in laborious circumlocution. Transitivity refers to the kinds of processes and participants in clauses, as discussed above. Nominalisation is a rendering of the content of a verb in the form of a noun, which has two ideologically practical consequences: first, they are the source of new nouns, codings of experience that can be transmitted to social groups by propaganda; and second, they permit deletion of both agency and modality, thus making the participants mysterious. (Passives are less obscure, since they only delete agency but not modality.) Modality subsumes a range of devices that indicate speakers' attitudes to the propositions they utter, such as validity, predictability, desirability, obligation, permission, etc. These attitudes are signified by linguistic forms such as modal auxiliaries, sentence adverbs such as *probably*, *certainly*, *regrettably*; adjectives such as *necessary*, *unfortunate*, and some verbs and nominalisations, such as *predict*, *prove*, *likelihood*, etc.

These linguistic devices may indicate ideological beliefs and power, that is, the ability of people and institutions to control the behaviour and material lives of others (Fowler 1985:61). Fowler (1985:75) admits that it is sometimes difficult to relate textual structure to social context, since the relationship is indirect and not invariant. Nevertheless, as Kress (1985:30-31) points out in another chapter of the same volume, the presence of a linguistic feature in a text is always the sign of the presence of one term from a discursive and ideological system appearing in the copresence of other terms from that system. Choice of a lexical item or a syntactic construction is an ideological selection which is iconic: it signals and expresses the content of that ideological choice.

For an exemplification of critical linguistic practice, I will refer to a paper in the collection of studies entitled *Language and Control* (edited by Fowler et al., 1979). Trew (1979) takes a close look at the coverage in English national dailies of the 1977 Notting Hill Carnival, which ended in a conflict between the police and a group of young people. After a trawl through the headlines of various papers, Trew turns to the entire coverage in two newspapers, the *Morning Star*, with official links to the Communist Party, and the *Sun*, with no formal relationship with any party. Trew

(1979:123-125) sorts the terms of the texts into categories of process and participant, to abstract the distribution of agency and interaction among participants. The results are presented in a matrix, such as in Figure 1.1, which reproduces Trew's analysis of the participants and processes in a *Sun* article:

	uniden- tified	people	police	...	organ- isers	equip- ment	weapons
Unidentified	NNNNTN	TTTTT	TTT		T		
People		NNNNNN NNNNNN	T				
Police	TT		NNNNN	T		TT	
...		TTT	TTTT	NNNN	T	TT	TT
Organisers				T	NNNN		
Equipment			T	T		N	
Weapons							

Figure 1.1 The process-and-participant matrix of a newspaper story (after Trew 1979:130)

In Figure 1.1, the terms on the side of the matrix represent the causers or agents in transactive processes, the ones along the top the affected participants. The occurrence of a transactive process is recorded with a T in the square against the causer and under the affected participant. The occurrence of a non-transactive clause is recorded with an N on the diagonal for the participant involved, that is, on the square which has the involved participant both at its side and above it. 'Weapons' comprises bricks, bottles, cans, etc.; 'equipment' refers to goggles, truncheons, riot shields, etc. The three dots denote the group of participants for whom there is no term common to both the *Sun* and the *Morning Star* reports: the *Sun* contains 'black youths', 'gangs of youths', 'mob of black youths', 'thugs' (a case of overlexicalisation), and the *Morning Star* 'groups of youngsters'. Such a matrix is able to account for the differences in formulation in three

types of clauses potentially referring to the same event: e.g. *The youths stoned the police*, *The police were stoned* and *Stones hailed down on the police*. The above matrix is accompanied by a similar one for the *Morning Star* report, and a comparison shows that there are many more T's, that is, transactive clauses in the *Sun*, and the category '...' is the most active participant. Therefore, the *Sun* presents the event as an explicit interaction between identified groups of people and objects, while the *Morning Star* focuses more on processes.

Trew (1979:154-156) concludes that a linguistic theory in which the categories of process and causation are central can serve as a basis for a linguistic analysis aimed at presenting the ideological character of discourse. Trew adds that ideology is best understood by an analysis of both given texts and the processes of which texts are a part, that is, sequences of texts and related discourses, representing discourse in progress. However, Trew recognises as a limitation of this kind of work that it is not so easy to establish which combinations of forces and institutions find expression of their views in newspapers, which have the power to make their terms stick etc. Newspapers are only a part of the ideological institutions of a society; therefore, the analysis of the language contained in them provides only a partial insight in the ideological workings of the society. (For a full description of critical linguistics, assessment of its significance and detailed exemplification, see Fowler 1991.)

The shortcomings of critical linguistics have been widely discussed. In a state-of-the-art article, Bell (1995:24) judges that the pioneering analyses of media discourse conducted in the critical linguistics framework were stimulating but less than satisfactory. Fairclough (1995b:27-28) misses from the approach an audience aspect: he finds that texts tend to be interpreted by the analyst without reference to the interpretative practices of audiences. Further, critical linguistics tends to present a rather monolithic view of the role of media in ideological reproduction, which understates the extent of diversity and change in media practices. Another drawback is that the linguistic analysis is very much focused upon clauses, with little attention to higher-level organisation properties of whole texts.

Critical linguistics is now found to be too strong-viewed. Bell (1991:214) thinks that this approach is at its strongest in the direct comparison of different media accounts of the same event. However, Bell continues, critical linguistics has two main deficiencies: first, it presumes that there is a clearly definable relation between any given linguistic choice and a specific ideology; and second, it imputes to newswriters a

far more deliberate ideological intervention in news than is supported by other research on news production (e.g. Bell's own, *op cit*). Indeed, it has become controversial in media studies whether media discourse implies wilful ideological bias. Leitner (1997:199) finds that claims of deliberate acts of ideological distortion have been softened and, instead of 'bias', the term 'mediation' has come into use (e.g. Fowler 1991). The role of the audience has been augmented (e.g. Bell 1991). Nevertheless, the critical edge has also been preserved in a more and more extensively practised linguistic approach: critical discourse analysis. This is what the next section will be about.

1.3.2 Critical discourse analysis

Critical discourse analysis (henceforth CDA) is seen by Bell (1995:24) as increasingly the standard approach to media texts within European linguistics (though less familiar in the US). What was felt to be lacking from both conversation analysis and critical linguistics, namely, a search for the relationship between discourse and the higher-level social reality, is paid extensive attention to in CDA. However, CDA remains faithful to the critical stance of studies by Fowler, Kress and their associates in an interest for the workings of power in and through discourse. The paradigm has two distinctive trends: the 'sociocognitive' approach, practised mainly by van Dijk, and the 'social practice' approach, principally represented by Fairclough.

The leading Dutch theorist of discourse analysis, van Dijk, has worked out an outstanding framework for the analysis of media texts. He calls his paradigm 'sociocognitive', because the system he proposes is able to simultaneously account for news values, ideologies and other social constructs on the one hand, and cognitive processes of text production and comprehension on the other. The first section of Chapter 5 will provide a detailed presentation of van Dijk's framework (along with a similar model of Bell's); therefore, this section will focus on the theoretical and moral grounding of the approach.

Van Dijk (2001:352) sees CDA as a type of discourse analytical research that primarily studies the way social power abuse, dominance, and inequality are enacted, reproduced, and resisted by text and talk in the social and political context. It is not so much a school next to the many other approaches in discourse studies; rather it aims to offer a different mode or perspective of theorising, analysis, and application throughout the whole field. If language use, discourse, and verbal interaction belong to the

microlevel, and power, dominance, and inequality between social groups to the macrolevel of the social order, then – van Dijk (2001:354) continues – CDA has to theoretically bridge the gap between micro and macro approaches.

To encourage such enterprises, van Dijk founded the quarterly *Discourse and Society* in 1990, to provide a forum for those analysts who share this critical perspective. In an article proclaiming the principles of CDA, van Dijk (1993b:250) postulates that his concept of social cognition is the necessary theoretical and empirical interface, if not the ‘missing link’, between discourse and dominance. Social cognition, according to the linguist (1993b:257), is represented by the socially shared representations of societal arrangements, groups and relations, as well as mental operations such as interpretation, thinking and arguing, inferencing and learning, etc. These evaluative social representations have a schematic form.³ Ideologies are also social cognitions that reflect the basic aims, interests and values of groups; this is how the social aspect is linked to the cognitive one. Concrete text production and interpretation are also based on models, that is, mental representations of experiences, events or situations, as well as the opinions we have about them. These models may be shaped by other source texts, existing knowledge, and more or less variable or shared general attitudes and ideologies; this is how the cognitive aspect is linked back to the social one. (For a more explicit account of how underlying ideologies monitor semantic structures of discourse, see van Dijk 1995.)

But how does this theoretical account translate to linguistic analysis? Van Dijk suggests that, in the case of an individual text, we first need to look at its context: the access to socially valued resources and power of the originator(s), the setting in which the text is communicated, the genre, participant position(s) and role(s), and the kinds of speech acts that are used. After this, the analysts can take a closer look at the text itself: the topics of the text (its ‘macrostructure’), the structural organisation of the text (its ‘superstructure’), and on lower levels of analysis, local coherence and style (variations of syntax, lexicon and sound). (A more detailed presentation of the semantic and organisational structure of a text in van Dijk’s framework follows in Section 5.1. For a vast exemplification of the analytic toolbox, see van Dijk 1991.)

As for the moral stance of critical discourse analysts, van Dijk (1993a:131) suggests that, while the bulk of contemporary research has been decidedly ‘uncritical’ if

³ These schemes are similar to Schank and Abelson’s scripts, or cognitive linguists’ Idealised Cognitive Models.

not 'apolitical', CDA does not profess cynicism, distance and unwarranted relativism, attitudes that often are indicative of the indifference of academia vis-a-vis the remaining inequalities in contemporary societies. And, since any critique by definition presupposes an applied ethics, Van Dijk proclaims the moral standpoint of CDA: there cannot be an aloof, let alone a 'neutral', position of critical scholars (1993b:253); the ultimate scholarly aim is not only understanding the world, but also changing it (1994b:436).

In 1994, van Dijk (1994b:436) still felt that critical research, in common with some 'applied' research, is not only marginal and marginalised, but also less prestigious than 'pure' theory and analysis, and he admitted that, theoretically and methodologically, CDA is still in its infancy. Since then, the situation has changed: although 'applied' linguistics may still not be considered 'real' linguistics in some circles, by and large, their respect has increased. *Discourse and Society* has become the meeting point of the 'great names' in the field (see Kress 1993, Fairclough 1993, Bell 1994, van Leeuwen 1995), hosting a wide range of studies conceived in the critical perspective, e.g. on the following topics: the social representation of homosexuals (Meyers 1994); the intertwining of informative and entertaining elements in television news (Graber 1994); the media story-line built up for the Chernobyl accident in the Italian press (Triandafyllidou 1995); the ideological construction of Africa in the British press (Brookes 1996), etc.

The other paradigm in critical discourse analysis is the 'social practice' approach conceived by Fairclough, and put forward in Fairclough 1995a and b. Fairclough 1995a is a comprehensive collection of Fairclough's work in the field, from the early 80's to recent times. The first section of the book focuses upon linguistic details of texts; the second integrates discourse analysis with a social analysis of sociocultural change, with illustrations from media discourse; and the third section contains previously unpublished work on textual analysis in social research, with a concern for intertextual analysis, discourse strategies, etc. Fairclough 1995b is an application of the linguist's social practice approach to media discourse.

In an earlier book, Fairclough (1989:7-9) accounts for the birth of critical language study by building up the following development line: Mainstream linguistics, is an asocial way of studying language. Sociolinguistics has emerged as a reaction to the neglect by 'linguistics proper' of socially conditioned variation in language; however, it is strong on *what* questions, but weak on *why* and *how* and, in being so, it legitimates

social relations from a position that the sociolinguist cannot change reality. According to Fairclough's alternative scenario, the facts of the existing sociolinguistic order are lines of tension, contemporary configurations of the current balance of power relations, and critical language study shows the contingency of social facts and indirectly points to ways of changing them. Pragmatics, the science of language use, points at language being a form of action, but this action is atomistic, emanating wholly from the individual. While 'uttering as action' is a central idea of critical language study, too, Fairclough calls for a theory of social practice accounting for both the determining effect of conventions and the strategic creativity of individual speakers.

It can be seen that, while van Dijk lays special emphasis on the cognitive aspect of news production and comprehension, in Fairclough's model, the focus is on the social ties of news discourse. According to Fairclough (1995b:56), the analysis of any particular type of discourse involves an alternation between twin, complementary focuses: the analysis of communicative events on the one hand and the order of discourse on the other, in the following way:

A communicative event (1995b:57-58) has three facets: text, discourse practice and sociocultural practice. The analysis of texts covers traditional forms of linguistic analysis – aspects of vocabulary, syntax, etc. – and the analysis of textual organisation above the sentence. Fairclough works with a multifunctional view of text, in the Hallidayan terms of ideational, interpersonal and textual functions, re-termed by Fairclough as 'representations', 'relations' and 'identities'. The analysis of discourse practice (1995b:58-60) involves attention to the ways in which texts are produced and consumed, which are realised in features of texts (institutional routines, such as editorial procedures in producing media texts, transformations which texts undergo in production and consumption etc.). On the borderline between text and discourse practice in the analytical framework is the intertextual analysis of texts (1995b:61): looking at text from the perspective of discourse practice, at the traces of the discourse practice in the text. Intertextual analysis looks for the genres and discourses that were drawn upon in producing the text. The analysis of the sociocultural practice dimension of a communicative event (1995b:62) involves its immediate situational context, its wider context of institutional practices, or the yet wider frame of the society and the culture.

The order of discourse (1995b:62-66) is analysed in terms of configurations of genres and discourses, where 'genre' (1995b:56) means a use of language associated with, and constituting part of, some particular social practice – such as interviewing

people or advertising commodities –, and ‘discourse’ (loc cit) is the language used in representing a given social practice from a particular point of view – for example, the social practice of politics is differently signified in liberal or in Marxist discourses. Shifts within the order of discourse and its relationship to other socially adjacent orders of discourse are also relevant issues, and a discourse-order analysis is also able to account for the relations of discourse to power and ideology.

In the media order of discourse, Fairclough observes two tensions affecting contemporary media language: between information and entertainment, and between public and private. This results in two tendencies: the conversationalisation of public affairs media and the marketisation of the media represented by the move in the direction of entertainment. Conversationalisation is, for example, shown by lexical choice in favour of colloquial expressions. Marketisation may manifest itself in fast-moving, attention-grabbing forms of presentation in television, where the relationship of newswriters and viewers changes from knowledgeable reporter/interested audience to an entertainer/consumer model.

Fairclough intends his social practice approach to step over two limitations he sees in van Dijk’s sociocognitive model: first, from Fairclough’s (1995b:30) point of view, van Dijk’s approach focuses on representations, whereas social relations and identities in news discourse receive little attention. Second, texts are analysed linguistically but not intertextually, in terms of their constitution through configurations of discourses and genres.

What Fairclough’s model shares with van Dijk’s approach – besides its critical attitude – is its ethical stance. In Fairclough and Wodak’s (1997:258) view, CDA is similar to other discourse analytical approaches in that it analyses real instances of social interaction which take a linguistic form. However, CDA is different from other approaches in two ways: it sees discourse as a form of social practice, both constitutive of, and socially shaped by, the situation it occurs in; and sees itself not as dispassionate and objective, but as an engaged and committed social science, a form of intervention in social practice on the side of dominated and oppressed groups and against dominating ones.

As we have seen, CDA is a powerful and well-established approach to media texts. Despite the fact that work in and on English predominates (cf. Bell 1991:24), it is this line of research that has found its way in Hungarian media studies. An example for such work *in* English but *on* Hungarian is *Framing the Issues*, a collection of conference

papers edited by Cunningham and Horváth (1997). In this volume, Barát, for example, analyses how the meaning of the term ‘alien’ is discursively constructed in the Hungarian political discourse. Through an analysis of the definitions of the term in four Hungarian dailies, Barát (1997:145) found that the term is constructed at the intersection of various discourses of discrimination. In left wing papers (*Népszabadság*, *Magyar Hírlap*), ‘alien’ is invested to represent a threat for the reader; in the right wing press (*Magyar Nemzet*, *Új Magyarország*), the category is constructed as one that is not a threat for the reader but for others. Another paper in the collection, Elek 1997, compares the role of news in Western and Eastern European societies. He (1997:9) proposes that news in the west is a more or less well manufactured industrial product to be bought and consumed by people creating a demand on the media market (see Fairclough 1995b, reviewed above), whereas news in the east is a magic spiritual tool to work the complicated mechanism of the society, something you get offered. In Eastern Europe, journalism was for a long time listed more under the heading of ‘arts’ and less under the name of ‘crafts’. Unlike in the Anglo-Saxon news industry, for a long historical period in Hungary, the opinion was always king served by well-selected facts to fit whatever theory the opinion was supporting. Text-wise, this means that the Hungarian tradition still permits the media story-teller to constantly comment on the story as it unfolds. Elek (1997:11) finds that this tradition is so strong in Hungary that, despite the political changes having occurred, it made an abrupt turn from opinion to facts almost impossible.

CDA has been largely criticised on theoretical and moral grounds. These critiques will constitute the topic of the next section.

1.3.3 Critiques of critical media studies

The first claim of critical media studies that has led to controversy addresses deliberate distortion in the news media. As Scannell (1998:267) points out from a non-linguistic media study angle, the relationship between media, language and world can be thought along the axis of power/knowledge *and* that of understanding/truth. These two ontologies, each with their particular hermeneutic (distrust/trust), should not be thought in an either/or fashion. To privilege one at the expense or exclusion of the other is to distort the manifold reality of the world that we are in.

As has already been mentioned, Bell criticises the aura of a ‘conspiracy theory’ (1991:214) that can be found in early critical studies attributing conscious manipulation to newswriters. Later, this machination theory has loosened in critical studies as well. Kress (1985:31) notes that selection of an ideologically value-laden linguistic form may not be a ‘live’ process for the individual speaker: if ideological discourses exist in an already established repertoire in a social group, then the individual speaker will not be creating but merely reproducing the discursive forms. Herman and Chomsky (1988:xii) – as mentioned in Section 1.1 – give the issue a more sociopolitical treatment in saying that most biased choices in the media arise from the preselection of right-thinking people, internalised preconceptions, and the adaptation of personnel to the constraints of ownership, organisation, market, and political power. Van Dijk (1993b:254) writes that mind management is not always bluntly manipulative. On the contrary, dominance may be enacted and reproduced by subtle, routine, everyday forms of text and talk that appear ‘natural’ and quite ‘acceptable’. The possibility of unintentional bias thus allowed for, critique in this line has moved to offering evidence for this inadvertence.

One way to prove this is what Bell (1991, Chapters 3 and 4) takes in presenting in detail the so-called ‘news assembly line’ (1991:44), that is, the way of news from news source through chief reporter, journalist, news editor, subeditors and the editor to the media audience. A ‘journo-linguist’ (Cotter 2001:419), that is, a linguist with newsroom experience, Bell is able to give first-hand information on news production. As he (1991:57) observes, news sources are varied: press releases, press conferences, public addresses, news agency copies, interviews (face-to-face or telephone), organisationally produced documents, etc. Information from such input is integrated within the text, which then goes through various transformations as ‘many hands make tight work’ (1991:34), such as embedding, during which one speech event is incorporated into another, or editing, which involves deletion of details found unnecessary, lexical substitutions, and syntactic transformations. The functions of copy editing are: to cut, to clarify, to maximise news value, and to standardise language, maximising news value being the primary function. During this complicated process many mishaps might occur: events may be misreported, news agency data may be falsified, over-asserted, or refocused, due to time-constraints or news value maximisation, but not necessarily in the form of deliberate distortion. With so much editing involved, and all these persons working on a single piece of news, it is

problematic just who exactly could be held responsible for value-laden linguistic or textual presentation.

In a review of Bell 1991, DeFelice (1996:121) suggests that Bell's relationship to Critical Language Awareness – roughly the same as CDA – is inconsistent. While disparaging critical analysts on the grounds that choices made in the media are not conscious and newsmakers' interventions are not ideological, Bell unintentionally bolsters critical claims that conditions, if not people, conspire to influence the media.

We have seen Bell's account of the actual practices of newsmaking, which do not point to deliberate distortion. Another way of indicating that journalists are not necessarily the mouthpieces of certain ideologies is to show how they resist the expression of these beliefs. Geert (1999:254) chooses this route to provide evidence that the orthodox critical stress on determinacy – with news managers of various organisations telling newsmakers what to write – is incomplete. News management takes the initiative with the news rather than wait for the journalists' inquiries, thus, first of all, the newsmakers' agenda-setting is anticipated by that of the news managers. Through paying systematic attention to the structures of linguistic transformation between news management and newsmakers, Geert (1999:256) finds that the ideology is not always confirmed, it is often contested and reworded. For example, a euphoric description in a press release often gets substantially toned down. But writers of press releases are themselves indoctrinated to expect an attitude of distrust from journalists, therefore, ideological counter-tendencies may already have been built into press releases. Geert (1999:259) concludes that there is certainly no unilateral process here; instead, there seems to be a negotiation going on between newsmakers and news managers. At this point, Geert (*loc cit*) brings in Gramsci's term 'hegemony' not (just) in the static, deterministic sense, as a state of control and domination, but (also) as a process of consensus formation.

As we have seen, the first line of critique against critical media studies was connected with the wilfulness of bias on the part of newswriters. The second line of critique has to do with the theoretical/moral stance of critical discourse analysts. Since 1995, there has been an on-going debate on the columns of *Language and Literature* on the validity of critical discourse analysis, originally between Widdowson (1995, 1996) and Fairclough (1996), but later joined by other linguists as well. From a stylist's perspective, Widdowson (e.g. 2002:163) argues against the idea that there is a definite meaning to be discovered in texts; all texts give rise to diverse discourse realisations

and, in all cases, interpretation is bound to be partial. Therefore, it is misleading to claim that critical analytic procedures can discover the hidden significance of a text, since there is no such thing, significance is a function of the interpretative process. Widdowson's objection against critical discourse analysts is that they so often use their authority to privilege partial interpretations. Taking context into account, Widdowson argues, does not remedy partiality, because contextual features are, of course, just as prone to partial interpretation as are textual ones. What is typically done in CDA, he continues, is to look for textual evidence (with or without reference to context) for the contours of public discourse, of communal socio-political values, thereby necessarily disregarding the private residue, which, in Widdowson's opinion, is 'irreducible' (2002:164).

In Weber's (2002:155-158) understanding, Widdowson presents the following line of reasoning: non-literary texts are referential in referring to a specific context; literary texts are representational in that readers have to create the context for themselves; and, by decontextualising texts, CDA practitioners incorrectly treat non-literary texts as if they were literary. Close analysis is merely justified to explain the aesthetic effect of texts; since non-literary texts have no aesthetic effects, we should simply cooperate with such texts; what CDA practitioners do is taking up a deliberately non-cooperative position. Weber (2002:159) argues that this is not unique for CDA: many critics do the same with literary texts (e.g. postcolonial or feminist criticism).

While admitting to being much more in favour of CDA than against it, Toolan (1997) offers a fairly balanced view of CDA. On the one hand, he (1997:84) suspects that some of the objections from stylisticians have emerged for the very particular reason that provocative sociological CDA seems to threaten to make stylistics obsolete. On the other hand, he (1997:87) also gets the impression that critical discourse analysts are attempting to garner kudos for themselves for being the first to 'really see and address' the workings of power in discourse. Toolan (1997:88) suggests that the preferable CDA attitude is to concede that you cannot analyse or write about power, hegemony and dominance without yourself potentially being implicated and compromised by the powerful and hegemonising turns of your own discourse. Toolan (1997:94) has two additional concerns with regard to CDA: first, that sometimes theoretisation and global contextualisations are making the business of critical reading seem much more difficult than it needs to be; and second, that CDA has already become so diverse, so multi-voiced, that a certain methodological fragmentation has already

happened. Toolan (1997:101) suggests that CDA needs to critique some of its own theoretical distinctions and discourse, that it can be kept simple and textual, and that it must strive for greater thoroughness and strength of evidence in its presentation and argumentation. As a sympathetic observer, however, Toolan (*loc cit*) concludes by saying that the premises upon which CDA is based are irrefutable: if it did not exist, we would have to invent it.

Outside the debate in *Language and Literature*, Cameron (2001:137) evaluates CDA as essentially a form of textual analysis, which typically involves finding a regular pattern in a (set of) text(s) and proposing an interpretation of the pattern, an account of its meaning and ideological significance. Whether the proposed pattern is actually there or not is fairly straightforward; it is the interpretations offered by CDA that have led to controversy. Critics see no evidence that these are the ‘correct’ interpretations and not just meanings which the analyst is ‘imposing’ on the data because of his/her own ideological commitment. What Cameron (2001:140) suggests is that the analysis is enriched, and the risk of making subjective or sweeping claims reduced, by going beyond the single text to examine other related texts, and by exploring the actual interpretations their recipients make of them. As if in continuation to Cameron’s points, Schröder (1994:2416) argues that a purely textual analysis can still be justified in generalising about the meaning structures found, as long as these generalisations are made on the behalf of a specific interpretative community whose sign universe makes these meanings plausible.

Another dispute that is taking place in the linguistic literature is a CA vs CDA debate. As we have seen in Section 1.2.2, CA is devoted to the immanent/internal analysis of texts, with no regard for contextual surroundings, political or social influences on texts. A founding father and main representative of CA, Schegloff, criticises CDA for having “a kind of theoretical imperialism involved (...), a kind of hegemony of the intellectuals, of the literati, of the academics” (in Mey 2001:610). He also argues that the theoretical interests and concerns of critical discourse analysts take precedence over the concerns of individuals whose discourse is being studied. Schegloff, however, suggests that CA and CDA are compatible, but only after studying the “endogenous constitution” of the discourse can we begin a critical approach. Mey’s (2001:611) riposte is that a CA-based CDA poses difficult methodological and ideological problems, thus, it is doubtful that Schegloff himself believes that the approaches can be reconciled without effort.

Defenders of CDA over CA say that the problem with conversation analysts is that they rarely raise their eyes from the next turn in the conversation. With regard to Schegloff's critique that CDA theorists import their preoccupations into the analysis, Wetherell (in Mey 2001:611) finds it ironic that the conversational fragment selected for analysis in CA is always selected by the analyst, who thereby (pre)defines its presupposed relevance for the participant. Billig (in Mey 2001:612) wonders by what method can analysts claim that they are not imposing categories but merely labelling what can be observed to exist. The terminology of CA ('ordinary conversation', 'participants', 'members') conveys commonality and equality. It conveys an essentially non-critical view of the social world, an image of sharing, participation, equal members and first-name informality. To take this image as a microcosm of the social world is to take a highly ideological step. It is doubly ideological when the step is taken as if it is itself non-ideological, to be contrasted with the ideological biases of other approaches.

Other linguists, however, suggest that the effort of reconciling CA with CDA is worth making. Pomerantz and Fehr (1997:65) are of the opinion that the analytic approach of CA is not limited to an explication of talk alone but is amenable to analyses of how conduct, practice, or praxis, in whatever form, is accomplished. Studies within this tradition which do focus on talk view talk, nonetheless, as social action. Hutchby (1996, 1999b) has illustrated that CA can be applied to detecting relations of power in discourse (see Section 1.2.2).

This is not the place to see justice done to CDA or any other linguistic media approach reviewed above. CDA remains a vigorous approach and, ironically, while it strives to fight against social hegemony, it is true what Garrett and Bell (1998:6) observe with regard to its scientific position: CDA is criticised, but it nevertheless holds a hegemonic position in the field of media discourse, such that other approaches tend to have to position and define themselves in relation to CDA.

1.3.4 Non-critical linguistic media studies

In the title of this section, I use the term 'non-critical' instead of 'acritical' to indicate that the authors of these works do not necessarily want to challenge or ignore a possible critical stance; rather, they employ a different focus. However, a choice for doing other kinds of research may be thought of as an indirect way of criticising CDA practice. Nevertheless, as we shall see, some of the studies falling into this category

may easily lend themselves to critical considerations. Again, the review will only point at some significant works.

Bell is best known for his micro-level sociolinguistic studies in a variationist framework. Contrary to what had been established in the sociolinguistic literature, namely, that language style choice depends on attention given to speech, social context or community networks, Bell suggested in a 1984 article that speakers may also adjust their language style to the audience they are addressing. Comparing the linguistic behaviour in terms of certain phonetic variables of the same newscaster speaking on several radio stations, Bell found that the newsreaders responded to the social status of the different target audiences. This he explains by considering it typical that speakers “shift their style to be more like that of the person they are talking to” (Bell 1991:105). This unconscious phenomenon he called ‘audience design’. In pursuing such studies, Bell also provides what some reviewers of early critical studies missed: an audience aspect (see Section 1.3.1). Another contribution of this kind is Bell’s differentiated concept of broadcast audience: Bell acknowledges that the audience is not a homogeneous mass of people, they have ‘multiple roles’ (1991:90-91). Conceived in the form of concentric rings, the innermost circle is represented by ‘addressees’, the second person, who is known, ratified and addressed by the speaker. Next are the ‘auditors’, third persons present but not directly addressed. Then follow the ‘overhearers’, whom the speaker knows to be there, but who are not ratified participants. Finally, in the furthest circle from the centre, are the ‘eavesdroppers’, whose presence is not even known let alone ratified. This model is generalisable for all kinds of audiences, and relevant for the broadcast audience as well.

Bell follows this tendency of concern with the audience in his 1994 article, in which he examines how the audience comprehends the scientific information on climate change available to them through the media. He (1994:57) finds that the people overestimate climate change scenarios, which derives from their degree of concern about it, reinforced by the media’s framing of the issue. (For another piece of research on media reception, see Richardson 1998.)

Another non-critical study of media language, also from a sociolinguistic perspective, is represented by Ferguson 1983, which analyses the language of sportscasting in terms of register variation. After locating the register by characterising its occasions of use, Ferguson (1983:158-168) identifies certain syntactic characteristics of sports announcer talk, such as simplification (deletion of copula and sentence-initial

nominals), inversions, heavy modifiers (appositional noun phrases, nonrestrictive relative clauses etc.), result expressions (*for* + noun, *to* + verb), and routines, that is, the non-creative choice of the same linguistic alternative on several occasions. Some of the characteristics identified are shared with other registers: simplification is similar to registers of economy type (e.g. headlines, note-taking), whereas inversions and heavy modifiers are shared with literary registers of English. Nevertheless, Ferguson (1983:169) concludes, all of these features fit together and interact in the construction of sportscasting discourse, and they may all be employed as conventionalised register-markers.

The next study reviewed here is of a textual/pragmatic and contrastive character. Following Sinclair's interactive and autonomous planes of discourse, Al-Shabab (1986) studies the overall structuring of radio news discourse in English and Arabic in a 73,000-word extended and a 38,000-word restricted corpus. He identifies three metatextual/interactive functions: 'discourse organising elements'(DOEs), attribution and 'sentential and nominal background information'. DOEs (1986:73) are metatextual utterances which refer to, and organise, a part or the whole of the news discourse, and whose main aim is to maintain the channel of communication by addressing the listeners directly. DOEs may be opening (e.g. *The news, read by Ashley Hudson*), medial (e.g. *You're listening to the news on Radio Damascus.*), and closing (e.g. *And that is the end of the news from London*). Salient grammatical elements in DOEs are deictic demonstratives, address pronouns and nouns referring to the news. Attribution (1986:123) is the use of news statements which are presented from the point of view of their original source, realised in reporting clauses, so that the editor's commitment to them is lessened to some degree (in BBC news) or given support (in Radio Damascus news). Nominal background information (1986:223) is provided by the news editor in relation to a proper name to enable the listeners to identify an aspect of that name (occupation, location etc.); while sentence background information is provided about the topic of the news item. The distribution of these metatextual elements is regular: attributions generally, and sentence background information solely, appear in news item details, nominal background information is a function of summaries and item-initial parts. Al-Shabab's study provides valuable cross-linguistic information as well.

Attributions, which constitute a substantial portion of current journalistic reporting, have been studied elsewhere as well. Nir and Roeh (1987), for example, examine the declarative verbs used in the Israel radio news on a 300-item sample. Their

aim is to reveal connections between the choice of declarative verb on the one hand, and the status of the source and the content of the statement on the other. They (1987:19) find that there is a ‘rhetoric of objectivity’, reflected by the prevalent use of neutral, non-modal verbs (in about two thirds of the cases in their sample). They analyse the use of modally marked verbs along three parameters: type of discourse, status of source and the reporter’s attitude towards the content of the statement. Nir and Roeh believe that their study might shed some light on the ideological infrastructure of the news on the Israel state-owned radio.

1.4 Summary

Media studies has become a distinctive sub-discipline of discourse analysis. After a fruitful period of non-linguistic work on the media, the language of mass communication came into the focus of an extensive amount of research with a linguistic orientation. If one approaches the linguistic media literature globally, the following recent tendencies of analysis stand out:

- from scattered to systematic;
- from micro to macro levels;
- from ‘deceptive’ to ‘well-meant’ language use;
- from product to process;
- from production to reception;
- from outsider to insider position;
- from the press to all kinds of media.

That is, linguistic media studies have become more large-scale in their database, turning from just a couple of media texts to large samples. They have also moved from lexico-grammatical and syntactic studies to examining the discourse levels of media texts. After an earlier tendency of attributing wilful manipulation to newswriters, analysts have now begun to seek for other patterns of (possibly) ideological content and miscommunication. Attention has moved from the media text by itself to the media text as a result of a complicated process on the one hand, and to the ways audiences make sense of them on the other. Many analysts are now journalists themselves, which allows

for a more participant-like perspective. Finally, after a period of almost exclusive attention to the printed media, there is now much concern for radio and especially television as well, giving rise to a multimodal conception of the media text. There is also a growing interest in other electronic media, too, such as computer-mediated communication, which has not been reviewed here. (Some related issues will be presented in Section 2.3.5.)

The tendencies presented above are not exclusive: critical, microlinguistic and product-oriented studies still largely prevail. Recent, state-of-the-art collections of studies (such as Bell and Garrett (eds.) 1998) signal this by featuring articles belonging to a wide range of approaches. Given the ubiquitous character of the media itself, this plurality in its research is nothing but natural.

2 Approaches to spoken and written language

One of the main concerns of the present study is whether radio news represents spoken or written language. This question addresses the issue which has long been of interest for linguists and other scholars: that of orality and literacy. It has been acknowledged for a long time that speaking and writing do not simply differ in the channel and mode of communication, that is, writing is typically not visible speech, and speech is normally not audible writing. Rather, there are certain linguistic features that tend to be associated with speech, and others with writing. Besides linguistic differences, orality and literacy are often thought of as referring to a cultural dichotomy between societies with or without a tradition of print. This chapter will provide an overview of the most important literature on speech vs writing, with special attention to radio language in general, and radio news in particular, from the point of view of orality and literacy.

The usage of the terms 'spoken/written' and 'oral/literate' in the literature is not uniform. Sometimes they are used interchangeably, 'spoken' being synonymous with 'oral' and 'written' with 'literate' (e.g. Tannen 1982, Biber 1988). On other occasions, 'oral' and 'written' are thought of as two different verbal strategies that can be applied to both spoken and written communication (e.g. R. T. Lakoff 1982, Holly 1995). This latter usage is applied in the present work as well.

2.1 Cultural-philosophical approaches

An important finding of scholars addressing the issue of oral vs literate cultures is that the medium is the message, in other words, that the physical form in which the information is transferred and preserved has an important effect on the information itself. A seminal work in this vein is McLuhan's *Understanding Media* (1964), in which the term 'media' is used in a broad sense, referring not only to the printed and electronic press, but also to writing systems, books, the telephone, the telegraph, the phonograph, photography, film, and even to money or roads. The media in this interpretation are the extensions of man's five senses. McLuhan (1964:22-23) divides the media into 'hot' and 'cool'. A medium is hot if it extends one single sense in 'high definition', that is, in the state of being well filled with data. Cool media are low in definition, in the sense

that they require high participation or completion by the audience. Radio, film and photographs are hot, while the telephone, television and cartoons are cool. As for orality and literacy, speech is cooler than writing in that so little is given and so much has to be filled in by the listener, but writing, especially hieroglyphic and ideogrammic alphabets are cooler than print. The hottest is the phonetic alphabet (1964:87), which involves the separation of signs and sounds from their semantic content. Besides technique, writing styles can also be hot and cool, McLuhan says, citing Francis Bacon (1964:31): writing in complete packages contrasts with writing in aphorisms or single observations, because the latter are incomplete and require participation in depth.

While acknowledging the unique power and tremendous values of literacy, McLuhan (1964:84-88) hints at the possibility of purchasing these values at too high a price. Literate man undergoes much separation of his imaginative, emotional and sense life, but he also wins his personal freedom to dissociate himself from clan and family. The appearance of hot media of the mechanical, uniform and repetitive kind, such as typography, which resulted in a speed-up of information, fragments the tribal structure in society. However, a very much greater speed-up, such as occurs with electricity, McLuhan suggests (1964:24), may serve to restore a tribal pattern of intense involvement. McLuhan attributes such an effect to radio, the first massive experience of electronic implosion (1964:297-301), which he sees as a new 'tribal drum'. While literacy has fostered an extreme of individualism, radio has done quite the opposite in reviving the ancient experience of kinship webs of deep tribal involvement. Radio has the power to turn the psyche and society into a single echo chamber, thus providing a reversal of the entire direction and meaning of literate Western civilisation.

As we could see, McLuhan's is a philosophical approach, fathoming the deep significance of orality and literacy, and their effect on the social and psychological state of mankind. Radio, in his conception, belongs to orality, but not so much in the character of the language used, but as a cultural artefact.

This cultural-philosophical line is taken up by R. T. Lakoff (1982:239-240). She sees the cultural history of the past three thousand years as one of a deadly combat between literacy and non-literacy, based on the unsupported assumption of Western culture that only with the preservation of literacy can we hope to preserve culture and civilisation. The appearance of non-literate media – earlier, comic books; currently, TV and movies – unleashed the fear that literacy is dying. R. T. Lakoff opposes this view, saying that loss of literacy is not the same as loss of culture. She points out that the shift

in our society from a literacy-based model of ideal human communication to one based on the oral mode of discourse is not a new tendency, that there have long been oral strategies in written communication. R. T. Lakoff (1982:244-251) enumerates certain techniques of transferring spoken discourse to writing, which represent a mingling of oral and literate traditions: representations of dialogues in novels, quotation marks, italics, capitalisation, nonfluencies etc.

Without wanting to denigrate literacy, R. T. Lakoff argues (1982:259) that all points of view which represent orality vs literacy in terms of value-judgements are dangerously misleading. She foresees that, with sophisticated information-processing and audio-visual technology, we will have achieved a sort of meeting of the fullest benefits of literate and non-literate forms of information-sharing. In what relates to the present study, radio may well be one of these ways of information-transfer, with recourse to both oral and literate strategies.

2.2 Approaches from communication studies

Radio as a medium has aroused the interest of scholars of communication studies as well. An example of such an approach is Crisell 1994, which characterises radio as the 'blind medium'. Radio's codes, he says (1994:5), are purely auditory, consisting of speech, music, sounds and silence. Since the ear is not the most 'intelligent' of our sense organs, the deployment of these codes has to be relatively simple. Out of the auditory codes, Crisell argues (1994:54), the primary one is linguistic, since words are required to contextualise all the other codes. We make sense of the world in a visual way, but we communicate about it verbally, and this, Crisell writes (1994:226), gives the epistemological significance of radio.

Crisell (1994:60-61) applies Jakobson's communication model to radio as sender, and categorises different radio programmes from the viewpoint of the communicative function they fulfil. Thus, news, documentaries, and commentaries on public events are referential; interviews and chat-shows have an emotive function; commercials and public service notices perform the conative task; plays, story-tellings and poetry-readings have an aesthetic value, while the function of phone-ins is phatic and metalingual. News, therefore, is mainly aimed at spreading information, and is quintessentially a verbal genre. This explains why, even in this visual age of television and teletext, news is part of radio's 'rump' (1994:125).

What concerns the spoken/written character of radio language, Crisell (1994:55-58) states that, since much radio talk is first written down, to that extent it has a literary nature. But even the seemingly spontaneous radio talk is premeditated; therefore, it is much more explicit than spontaneous speech in that it creates its own context or situation to a much greater extent. Further, it is more fluent, precise and orderly, less diffuse and tautological, than ordinary speech. Although much of radio reading disguises itself as spontaneous in order to minimise detachment from listeners, news is clearly announced as being read, because the words must carry an air of definitiveness and accuracy; they must seem to be 'authorless' – originated by the events themselves. The quasi-objective style of newsreaders also suggests that they are the mere mouthpiece of the words, and not their originator.

As for linguistic structure, Crisell (1994:58-59) acknowledges the need for radio language to be syntactically fairly simple, so that it can be comprehended through the ear alone. He claims that radio messages will therefore tend to have a high level of redundancy, for speech is notoriously evanescent. When turning to radio news in particular, Crisell reiterates this need for redundancy, or reinforcement through repetition, which is a characteristic of colloquial language. It is not clear what Crisell means by redundancy here: if he refers to the total or partial repetition of news items from hour to hour, news is, indeed, redundant, but if he has lexical or grammatical repetition in mind, news seems to be much less redundant than spoken language in general.⁴ Crisell (1994:97) sums up the elementary nature of radio news language in claiming a relative simplicity of sentence structure.

In sum, Crisell finds that radio news has a literary nature in being scripted, but an oral character in being syntactically simple and displaying redundancy. He also points out (1994:56) that the differences between orality and literacy seem a good deal less absolute than is commonly supposed.

Another communication-oriented media study with an orality vs literacy concern – albeit only marginally and reflected to television – is Fiske and Hartley's *Reading Television*. The authors (1978:124-125) set out in opposing pairs the characteristics of the oral vs literate modes: dramatic vs narrative; episodic vs sequential; mosaic vs linear; dynamic vs. static; active vs artefact; concrete vs abstract; ephemeral vs

⁴ Kontra (personal communication, 2002) has drawn my attention to a tendency in Kossuth Rádió, the main public service radio of Hungary, that the location of the reported event is given in the very first

permanent; social vs individual; metaphorical vs metonymic; rhetorical vs logical; and dialectical univocal/ consistent, respectively. Television as a medium – they conclude – takes on its final shape from the active tension between the oral and literate modes. This finding of Fiske and Hartley's can probably be generalised to all forms of the media.

2.3 Linguistic approaches

2.3.1 Typical speech vs typical writing

It has been acknowledged that oral vs literate are not a matter of either/or, but rather two endpoints of a continuum (Tannen 1982). One of the most important manuals on linguistic variation along this cline is Biber 1988, which uses the terms 'typical speech' and 'typical writing' to refer to the unmarked genre in each mode (op cit:36-45), since there is no linguistic or situational characterisation of speech and writing that is true of all spoken and written genres. Following this logic, face-to-face conversation represents typical speaking and expository prose stands for typical writing. In terms of situational characteristics, stereotypical speech is interactive, and dependent on shared space, time and background knowledge; stereotypical writing has the opposite characteristics. In terms of linguistic features, stereotypical speech is structurally simple, fragmented, concrete, and dependent on exophoric reference; stereotypical writing has the opposite characteristics.

2.3.2 Planned vs unplanned discourse

Ochs (1979) takes a different route and, instead of the spoken/written differentiation, she proposes another distinction, that of 'planned' vs 'unplanned' discourse. Along this line, spontaneous dialogues, stream-of-consciousness writing and casual letters are examples of unplanned discourse, while formal expository writing and presidential addresses represent planned communication. As can be seen, there are spoken and written forms of communication in both planned and unplanned discourse, although written discourse may generally be more plannable than spoken discourse (1979:58). In writing, the communicator has more time to think out what (s)he is going

word of the news item, and is repeated at the end. This is a quite unnatural, and utterly redundant, language use.

to say and how it will be said; (s)he can rewrite and reorganise the discourse a number of times before it is eventually communicated. Ochs (1979:55) defines unplanned discourse as one that lacks forethought and organisational preparation, and planned discourse as one that has been thought out and organised (designed) prior to its expression. She identifies three features which differ substantially in the two kinds of communication, namely, in relatively unplanned discourse more than in planned discourse (1979:62-70):

- (1) speakers rely on the immediate context to express propositions;
- (2) speakers rely on morphosyntactic structures acquired in the earlier stages of language development; and
- (3) speakers tend to repeat and replace lexical items in the expression of a proposition.

More precisely, unplanned discourse makes greater use of so-called left-dislocation (a transformation that moves a noun phrase out of and to the left of a clause, leaving in its place a co-referential pronoun, e.g. *my father oh he's he's fit to be tied*, Ochs 1979:65), demonstratives as modes of reference, less passive voice, and present tense in narrations, while planned discourse displays no left-dislocation, definite articles instead of demonstrative determiners and relative clauses as modes of reference, more passive voice and past tense for narration. In a final section of her article, Ochs (1979:77) even speaks of 'planned unplanned' discourse, in which a speaker or writer will intentionally produce discourse that appears unplanned. Examples for such communicators are novelists and skilled rhetoricians such as journalists, politicians, and even academics at times.

Ochs (1979:56) finds it an extreme of total discourse planning when a speech has been written down in advance and read aloud. In this framework, therefore, radio news is to be considered an example for the planned endpoint of the cline.

2.3.3 Fragmentation vs integration and involvement vs detachment

A much cited work in the literature is Chafe 1982, which focuses on two major differences between speech and writing: fragmentation as opposed to integration, and involvement in contrast to detachment. The data used in this study come from four styles of language: informal spoken language (from dinner-table conversations), formal spoken language (from lectures), informal written language (from letters), and formal

written language (from academic papers). Chafe's starting point (1982:36-38) is that speaking is faster than writing but slower than reading, therefore, the abnormal quickness of reading fits together with the abnormal slowness of writing to foster a kind of language in which ideas are combined to form more complex idea units and sentences (here, 'idea unit' stands for a spoken sentence). The result is a high level of integration, which refers to the packing of more information into an idea unit than the rapid pace of spoken language would normally allow (1982:39). Means of integration are: nominalisations, participles, attributive adjectives, conjoined phrases, series, sequences of prepositional phrases and relative clauses, which can be found in typical writing. Typical speech display signs of fragmentation, where a typical idea unit consists of a single clause, and idea units are often introduced with coordinating conjunctions or even strung together with no connectives at all.

Another basic difference Chafe identifies (1982:45-48) is that speakers interact with their audiences, writers do not. An important feature of typical writing, therefore, is detachment from its addressees, which manifests itself in devices such as passive voice and, again, nominalisations. The involvement of speakers, on the other hand, is shown by more frequent first (and second) person references, verbs expressing mental processes, colloquial expressions aimed at monitoring the communication channel, such as *well, I mean* and *you know*, emphatic particles, like *just* and *really*, hedges and direct quotes. In a final remark, Chafe, too, emphasises (1982:49) that these seemingly categorical statements about spoken and written language apply in fact to extremes on a continuum, and that there are other styles of speaking which are more in the direction of writing, and other styles of writing which are more like speech.

Radio news, then, would exemplify just this 'other style of speaking' which is more in the direction of writing, due to the simple fact that it is previously put down on paper. Chafe makes reference to some previous findings (1982: 36-37), presenting the following data: the average speed of spoken English, including pauses, is in the neighbourhood of 180 words per minute; writing is between one-tenth (handwriting) and one-third (copying by typing) the speed of speaking; while reading may be between 200 and 400 words per minute. Thus, while speaking and listening necessarily proceed together at the same speed, writing and reading deviate from that baseline in opposite directions, writing being much slower and reading somewhat faster.

Loud reading, of course, is slower than silent reading, but quicker than spontaneous speech. This is especially important in the case of radio news, where a

written type of language has to be understood by mere listening, that is, an integrated and detached style of language is offered to the ear, which is otherwise trained for fragmented and involved speech. This takes us back to Crisell (see Section 2.2), who finds that radio news language should be syntactically simple and lexically redundant. But, as we shall see in Chapter 6, this is not always the case.

2.3.4 The multi-feature/multi-dimensional approach

The multi-feature/multi-dimensional analysis has been developed by Biber (1985, 1986), and described in detail as well as applied to an extensive corpus in Biber 1988. The term 'dimension' is proposed because it captures the idea of continuums of variation rather than discrete poles. Biber's system is based on the assumption that strong co-occurrence patterns of certain linguistic features mark underlying functional dimensions, which means that no single feature characterises a dimension on its own, rather, dimensions are bundles of linguistic features that co-occur in texts because they work together to mark some common function.

Biber's notion of 'dimension' (1988:21-25) is different from earlier conceptualisations of spoken/written modes of language in three respects. First, while previous studies analyse linguistic variation in terms of a single parameter, such as interactive/non-interactive, formal/informal, Biber's study starts out from the conviction that linguistic variation is too complex to be analysed in terms of any single dimension. Second, while previous studies have treated linguistic variation in terms of dichotomous distinctions, dimensions are identified as continuous quantifiable scales of variation. Finally, while previous studies have not checked whether the linguistic features they have analysed actually co-occur in texts, this study is based on frequency counts that yield precise results to check co-occurrence. What is also new in Biber 1988 is its finding that there is no reason to assume that the situational difference between speech and writing constitutes a linguistic dimension in English, since no set of co-occurring linguistic features can be found to distinguish all written texts from all spoken genres. Therefore, Biber (1988:55) finds it necessary to look for the particular ways in which each text type is oral and the ways in which it is literate.

This multi-feature/multi-dimensional approach (MF/MD) is proposed also to solve certain disagreements that Biber (1988:49-51) has encountered in the previous literature regarding the distinctive characteristics of spoken/written genres. There is no

agreement, for example, on the mean length of sentences in speech and writing: some find it roughly the same, others find that it is longer in writing. This contradiction may have to do with the difficulty to establish 'sentences' in spoken English. Another problematic issue is that of subordination vs coordination: some have found that writing has a much higher degree of subordination than speech, reflecting its greater structural complexity; others say that there is no considerable difference between the two modes in this respect, again others have found a reversed rate of sub- and coordination in spoken vs written texts.⁵ Similar contradictory results occur in dealing with the frequency of passive constructions. The reason for such disagreements, Biber says (1988:54), is that the findings have been reached through the analysis of single features and not relations of co-occurring features, since the same linguistic phenomenon may serve different functions depending on which other features are associated to it.

Biber (1988:63-67) has analysed 481 texts of different types, mainly taken from computer-based text corpora (324 written texts of 15 genres from the LOB Corpus, 141 spoken texts of 6 genres from the London-Lund Corpus of Spoken English) and 16 letters of 2 types from the author's own collection. Computer programmes have been used to count the frequency of different linguistic features, and multivariate statistical techniques, especially factor analysis, have been employed to determine the co-occurrence relations among the features. These features (1988:73-75) are categorised in 16 groups: tense and aspect markers; place and time adverbials; pronouns and proverbs; questions; nominal forms; passives; stative forms; subordination features; prepositional phrases, adjectives and adverbs; lexical specificity; lexical classes; modals; specialised verb classes; reduced forms and dispreferred structures; coordination; negation. These groups comprise 67 different features. The frequency counts are normalised to a text length of 1,000 words.

Based on the co-occurrence patterns of the features, Biber (1988:115) identifies six dimensions along which all the text types used in the study can be characterised. Dimension 1 is labelled 'Informational vs Involved Production', along which discourse with interactional, affective, involved purposes, associated with strict real-time production and comprehension constraints are opposed to texts with highly informational purposes, which are carefully crafted and highly edited. Dimension 2 is

⁵ Some research done in and on Hungarian (see Wacha 1988) indicates that speech tends to use 'successive', linear, subordinated structuring, while writing prefers simple sentences which abound in 'simultaneous', complex phrases and left-embedding.

called 'Narrative vs Non-Narrative Concerns', which distinguishes discourse with primary narrative purposes from discourse with non-narrative purposes (expository, descriptive etc.). Dimension 3 is that of 'Explicit vs Situation-Dependent Reference', which contrasts discourse that identifies referents fully and explicitly through relativisation, and discourse that relies on non-specific deictics and reference to an external situation for identificational purposes. Dimension 4 is labelled 'Overt Expression of Persuasion', which identifies features associated with the speakers/writers' expression of their own point of view or with argumentative styles intended to persuade the addressee. Dimension 5 is called 'Abstract vs Non-Abstract Information', and distinguishes between texts with a highly abstract and technical informational focus and those with non-abstract focuses. Dimension 6 is that of 'On-line Informational Elaboration', and opposes informational discourse produced under highly constrained conditions, in which the information is presented in a relatively loose, fragmented manner, with highly integrated informational or non-informational discourse.

For the purposes of the present study, it is useful to see how media genres have scored along these six dimensions. Mass media genres in the corpus used by Biber include press reportage (political, sports, society, financial, cultural and spot news), and radio broadcasts (sports and non-sports). The analysis yielded the following results and interpretation (1988:129-160):

Broadcasts show a more informational than involved production, non-narrative concerns, and extremely situation-dependent reference. Broadcasts rated very low in an overt expression of persuasion, they present both abstract and non-abstract information, and they use fairly integrated language. Broadcasts, therefore, represent informational, strictly non-narrative and non-persuasive discourse. Since they report events actually in progress, they encourage direct reference to the physical and temporal situation of discourse.

Press reportage displays very informational production, and it has both narrative and non-narrative concerns. Printed press equally uses explicit and situation-dependent reference, and it scores fairly low in an overt expression of persuasion. It presents both abstract and non-abstract information, and uses fairly integrated language. Thus, press reportage is a highly informational and strictly non-persuasive kind of discourse, in which narratives are typically framed within some larger expository discourse, subordinate to a larger purpose. Its fairly frequent use of nominalisations makes it

elaborated and explicit in nominal reference. Press reportage rates intermediate between abstract and non-abstract information due to the twin purposes of the genre: reportage of events involving concrete, often human, referents, and abstract discussion of the implications of those events in conceptual terms.

Biber's claim that speech and writing do not represent two endpoints of a scale in linguistic variation seems to be supported by the findings concerning media genres: broadcasts (an oral genre) and press reportage (a literate genre) are similar in that both are informational, non-persuasive and integrated, and both express abstract and concrete information as well. However, they display smaller or larger differences in terms of reference and narrativity.

It has to be noted that radio news is not included – or, at least, not separately highlighted – in the broadcast subgenres. One may also find the sports/non-sports division of the subgenres a little arbitrary. Biber (1988:68) has adopted the genre categories from the distinctions used in the corpora serving as a source of data; however, there seems to be no reason for choosing this grouping; it might as well have been news/non-news. The high level of situation-dependent reference found in broadcasts, for example, is undoubtedly due to the high representation of sports commentaries in the data. Had news been given special attention to in the corpus, the results may have been different.

In sum, Biber's multi-feature/multi-dimensional analysis shows that there is no absolute difference between spoken and written texts in English. However, speech and writing are relatively well-distinguished along Dimensions 1, 3 and 5, which, therefore, can be considered oral/literate dimensions. Along Dimensions 2, 4 and 6, spoken and written texts overlap. Broadcasts rate as spoken along two, and literate along one of the oral/literate dimensions. Biber (1988:163) points out that, although both speech and writing have oral and literate characteristics, the most literate of the written genres (exposition) is considerably more literate than the most planned and informational forms of speech. That is, there seems to be a 'cognitive ceiling' on the frequency of certain syntactic constructions and lexical complexity in speech, since even the most carefully designed and informational spoken genres are produced and comprehended in real time. This means that, despite its being written first, radio news must not surpass this cognitive ceiling, that is, it has to be syntactically and lexically simpler than the written expository genres.

2.3.5 The mode-switching model

All of the above views on orality/literacy begin by assuming that a stretch of discourse is either written or oral. However, an analysis carried out by Murray (1988) indicates that this is not always the case; there may be instances where speakers/writers – consciously or not – choose between available modes of communication.

The study examined the way an IBM team made use of the various available modes and media to communicate with their colleagues. By ‘medium’ Murray (1988:353) means established methods of communication through language (e.g. face-to-face conversations), while ‘mode’ refers to specific communication types within a medium (e.g. lunchtime conversations, office conversations). The focus of the study was computer-mediated communication (CmC), a relatively recent phenomenon which, according to Murray (1988:354), appears to be a hybrid of both oral and written language. CmC offered the IBM community the following options of modes: computer messages (i.e. max. 120-character instant messages between minimum two participants simultaneously logged on to computer terminals); forums (i.e. mail-type files on one technical topic with a wide distribution list to which recipients can append another piece of information); e-mail; and documents.

By taking language use, or contextual factors rather than form, as a starting point, Murray (1988:352) found that medium represents part of the meaning potential available to language users. Murray (1988:368-369) offers a taxonomy of factors which influence the choice of mode and medium:

- Field
 - Topic (whether it is sensitive/open or simple/complex)
 - Organisation of topic (on parallel or episodic threads)
 - Focus of topic (on social cohesion, action, or unfocused)
 - Distance between language and activity
- Speaker/hearer
 - Knowledge of audience (who and how big is the audience, which mode they prefer)
 - Role relations (depending on power, affect and previous contact)
- Setting
 - Institutional conventions (e.g. contracts always appear on hard copy)

Space (distance between interactants, interactivity, permanence, planning, availability)

Time (time zone, time management, physical constraints)

For example, interactants may want to switch from electronic communication to face-to-face conversations if the topic is sensitive or confidential; they choose e-mail instead of an e-message if they need to retain a permanent copy of the information shared; or they prefer to use e-mail over e-messages across time zones, where no simultaneous connection is possible etc. Murray hypothesises (1988:370) that there is a hierarchical organisation of the taxonomy, e.g. in business settings the availability of the person to contact may take precedence over other features, but adds that this assumption needs further examination.

What is new in this model is the conception of mode/medium not being a stable, given feature of the situation but subject to change and choice. As Murray puts it (1988:367), mode and medium need to be thought of as variables rather than only as factors in the context. That is, mode/medium-switching is a phenomenon similar to code-switching (1988:368), with the difference that it seldom occurs within an utterance, and it is usually linguistically marked.

Computer-mediated communication is a challenge for linguists and, to cope with the challenge, Murray has introduced the concept of mode/medium as variables. This has important consequences regarding the idea of speech and writing. In Murray's view (1988:370), literacy and orality are not dichotomous, nor do they represent ends of a continuum. They are but one aspect of the communicative situation, and not the cause of characteristics such as integration or involvement. Murray's is an organic model (1988:371), where a change in one dimension results in change in other dimensions, that is, when a participant mode-switches, the switch is an indicator to the recipient that there is likely to be some associated change in the dimensions of field, speaker/hearer or setting.

The mode-switching model seems to agree well with the multi-feature/multi-dimensional one in that both see speech vs. writing as but one, and not even a very decisive, dimension of communication. In my interpretation, however, the two models reviewed in Sections 2.3.4 and 2.3.5 do not call into question the validity, just the absoluteness, of the oral/literate concept. Computer-mediated communication indeed

opens up new perspectives in the field but, for more traditional ways of communication, such as radio, the oral/literate issue is still a valid point.

2.4 Media language – spoken or written?

There are a few studies in the linguistics literature which have addressed the specific issue of media language being oral or literate. One of these is Holly 1995, which takes over Ong's famous term to characterise the modern electronic media: that of 'secondary orality' (Ong 1982), and sets out to see how it actually manifests itself. He examines how secondary orality can be detected on three levels: conditions of communication, verbal strategies and cultural form. In conditions of communication, the media are different from face-to-face interactions in four ways: channels, persons, space and time.

From the point of view of the present study, it is the difference in channels and verbal strategies that counts. According to the channels used, Holly (1995:341) speaks about media that are reduced to the visual channel (newspapers, telefax), those which use only the acoustic channel (telephone, radio), and those which have both channels available (television, film). As for broadcast texts, in many cases their production and reception are separated not only by space and time, but also by the channel used: during production the language is written, but during reception the language is read and heard. This means that, although the 'realisation' of these texts is written, the 'conception' should be oral.

This underlying change of focus is similar to what Crisell (1994) formulates as a need for radio news to be syntactically simple and lexically redundant to enhance comprehensibility (see Section 2.2). Holly (1995:348), however, states the fact that, despite the necessity for oral verbal strategies, in reality, radio news is not formulated with an eye to the spoken performance. The traditional news report also contains many features of a literate style, such as nominalisations, a large number of attributes in noun phrases, lexical variation etc.

From a methodological point of view, Holly's study is both similar to and different from Biber 1988 (Section 2.3.4). Holly also admits that orality and literacy are matters of degree (1995:347), and that the distinction involves more than one plane (1995:345). He determines six factors that have to be taken into consideration to establish the grade of orality of a certain text:

- (1) production: from well reflected to spontaneous;
- (2) performance: from obviously read through recited, pseudo-spontaneous (read from concealed script), semi-spontaneous (improvised with keywords from a cuecard) to spontaneous;
- (3) participants: from monologues to dialogues;
- (4) actuality: from live-transmission to non-live, recorded;
- (5) authenticity: from fictional to 'real-life' documentation;
- (6) formality: from official to casual.

However, contrary to Biber 1988, Holly (1995:347) thinks that the grade of orality and literacy is not measurable in an objective or statistical fashion. He finds that the customary method of illustrating both styles is to contrast extreme examples that contain some of the typical features (a procedure Biber 1988 in fact also uses while comparing conversations and academic papers as two poles).

To sum up: what is relevant to the present study from Holly's contribution to the literature is that the orality of the electronic media is secondary and, due to the channel-gap between written production and oral performance/reception, radio texts such as news typically display literate verbal strategies, which hinders comprehension.

Another study of importance from our point of view is Vagle's (1991), whose hypothesis is that radio language can be described as variation along a continuum from spoken to written language. The data was taken from two hours of morning programmes of the Norwegian Broadcasting Corporation. Vagle is another author who believes in the validity of statistical analysis: as she puts it (1991:122) it is probably impossible to find linguistic features which are present only in spoken or written language, rather, some characteristics are more common in speech or writing. In other words, the difference between them is more quantitative than qualitative, making frequency counts a suitable method of analysis when dealing with concrete texts.

Therefore, the hypothesis was tested by means of co-variance analyses and ranking, using certain linguistic expressions as operational categories, with the independent variable 'communication situation' represented by different text types (newsreading, weather forecasts and magazine programmes). Vagle takes over Chafe's terms of involvement/detachment and fragmentation/integration (see Section 2.3.3), and expresses them in the following way (1991:123-125): involvement/detachment is

represented by the rate of pragmatic particles (*I think, well, you know*), omission (*Don't like nice weather*), interjection macrosyntagma (*Oh, no!*), address macrosyntagma (*Peter,...*) and false starts (*I don't like nice weather, but...*). Fragmentation/integration is expressed by the frequency of short sentences (without stating, however, how long short sentences are), sentence fragments (*Nice weather today*), speech co-joining (*and, or, but*, often as macrosyntagma-starters), irregular macrosyntagma and extraposition.

Vagle's results (1991:126-127) show news and weather forecasts as located totally at the 'writing' end of the continuum, but they are clearly written to be spoken. Vagle compares her results to those of the Speech Study of Oslo (TAUS), and finds that even the radio text types placed at the 'speech' end of the continuum are not real spoken data, so she concludes that radio broadcasts contain little genuinely spontaneous spoken language (another proof for the 'secondary orality' of the media). Vagle's final conclusion (1991:128) is that her results give an ambiguous answer to the hypothesis that variation in radio language can be accounted for as variation along a speaking/writing continuum. Instead of her initial formulation that broadcasting language is 'a hybrid between spoken and written language, where the ratio varies substantially from text type to text type' (1991:122), she characterises radio language as 'written language transformed into spoken language'. However, one misses from the study a description of how exactly these transformations from written to spoken language are done. The question whether these transformations are merely physical (just 'sounding' the words) or they involve some kind of linguistic change remains unanswered.

In sum, Vagle's study reveals that radio language cannot be unambiguously characterised along a speech/writing continuum. It has been found that none of the radio text types, not even the most involved and fragmented kind, is real spoken language, but some radio genres, for example news, represent a written text type transformed into spoken language.

2.5 Summary

Chapter 2 has provided an overview of the literature on spoken vs written language, with particular interest in radio language and especially radio news from this point of view. There now seems to be an agreement between scholars that speech and writing are not two distinct, independent and contrasted categories, but the two

endpoints of a continuum. It has also been acknowledged that the categories 'oral' vs 'literate' are too simplistic, and that there is more to it than the medium used. Several complementary distinctions have been suggested, such as typical speech vs typical writing, unplanned vs planned discourse, fragmentation vs integration, involvement vs detachment etc. Some studies go as far as stating that speech and writing are not even valid distinctions, and that a whole system of features and dimensions has to be introduced to capture variation in English. As regards the electronic media, it has been claimed to bring about an age of new non-literacy, of secondary orality, in which the traditional speech vs writing distinction is even less suitable. Radio news, however, is said to represent a kind of language which – despite being intended for the ear – very closely resembles written text types in its lexical and syntactic texture.

3 The grammatical characteristics of printed news⁶

Biber et al.'s (1999) *Longman Grammar of Spoken and Written English* (henceforth LGSWE) is a recent, highly professional, large-scale endeavour. As stated in its introduction, LGSWE is "a new kind of descriptive and explanatory account of English grammar" (p. 5); descriptive in the sense that it describes the range of grammatical features in English, and explanatory in that it also seeks an answer to how these grammatical factors are actually used. In this way, LGSWE includes *structural description* of all major grammatical units and constructions, their paradigms, and the range of variants for each structure, as well as *a description of patterns of use*, that is, the frequency of different features, the distribution of grammatical features across registers, and the distribution of those features in relation to others (p. 6). For this purpose, LGSWE adopts a corpus-based approach, using a large sample of naturally occurring language, mounting up to more than 40 million words.

The handbook derives its data from four registers: conversation (CONV), fiction (FICT), newspaper language (NEWS) and academic prose (ACAD). The authors account for the choice of text types in the following manner: these registers have the virtue of being (a) important, highly productive varieties of the language, and (b) different enough from one another to represent a wide range of variation. These registers differ in:

- mode: CONV is spoken, the other three are written;
- the degree of interactiveness: CONV is interactive, NEWS and ACAD are not, FICT is interactive only in dialogical parts;
- whether the participants share an immediate situation: in CONV they do, in the rest they do not;
- the main communicative purpose: CONV involves personal communication, FICT is meant for pleasure reading, NEWS is aimed at giving information and evaluation, ACAD at giving information, argumentation and explanation;
- audience: CONV is addressed to individuals, ACAD to a specialist audience, while FICT and NEWS to the wide public;

⁶ In this and the next chapter all references, if otherwise stated, refer to Biber et al. 1999. Therefore, unlike the convention used throughout the whole work, in Chapters 3 and 4 the references will only contain page numbers.

- and, finally, dialect domain: CONV is local, NEWS is regional vs. national, while FICT and ACAD are global.

(pp. 15-16)

For the purpose of the present study, in this chapter I shall consider the grammatical observations that LGSWE makes about the nature of printed news texts. (LGSWE's results and discussions on conversations will be summarised in Chapter 4.) What follows, therefore, is a summary of LGSWE's findings, with a special focus on NEWS. The data of this register form a subcorpus of 5,432,800 words, gathered mostly from British English newspapers, five national and five regional⁷, complemented by a subcorpus of American English news texts for the sake of dialect comparisons. All frequency counts reported in LGSWE are normalised to a common basis, per million orthographic words of text, to allow a direct comparison of results for the features in different registers. However, to give an even better idea of the rate of certain features across registers, I have chosen to indicate percentages. In some cases, unfortunately, Biber et al. do not give exact percentages; in these cases, a rough approximation is given on the basis of the histograms found in LGSWE.

3.1 Basic grammatical characteristics of printed news texts

The relationship between the number of different word-forms, or 'types', and the number of running words, or 'tokens', is called the type-token ratio (or TTR, that is types/ tokens x 100). LGSWE finds that the TTR varies with the length of the text: longer texts have many more repeated words, and therefore a much lower TTR (this is true for all registers). Of all the registers, NEWS displays the highest TTR, which means that repetition is the lowest in NEWS. The high TTR in news reflects the extremely high density of different terms referring to a diverse range of people, places, objects, events, etc. (pp. 53-54).

LGSWE differentiates between three major word classes. The first is that of lexical words, the main carriers of meaning in a text, classified in four different word classes: nouns, verbs, adjectives and adverbs. The second class is that of function

⁷ National newspapers: *The Independent*, *The Guardian*, *Daily Mirror*, *Sunday People* and *Today*; regional newspapers: the *Liverpool Echo*, *The Belfast Telegraph*, *The East Anglian Daily Times*, *The*

words, which bind the text together, such as determiners, pronouns, auxiliaries, co- and subordinators, etc. The third is what Biber et al. call ‘inserts’ – a term that is criticised for being used idiosyncratically, with not enough justification (Schneider 2001:140). The elements designated by this term do not form an integral part of a syntactic structure; for example *hm, yeah, bye, cheers* (pp. 59-60). Concerning the frequency of the different word classes, it is found that lexical words account for almost half of the words in the Longman Spoken and Written English (LSWE) Corpus, and there is considerable variation among registers: CONV has by far the lowest, while NEWS the highest lexical density, i.e. the proportion of the text made up of lexical word tokens (p. 62).

Within lexical words, the LSWE Corpus shows the following breakdown: nouns make up by far the most frequent lexical word class, verbs are less frequent, followed by adjectives and adverbs. Lexical word classes also vary greatly across registers: nouns are most common in NEWS, and least common in CONV; adjectives are most common in ACAD, and least common in CONV, while verbs and adverbs are most common in CONV and FICT. The proportion of the lexical word classes varies with register: while in CONV nouns and verbs are about equally frequent, in NEWS and ACAD there are three to four nouns per verb. The preponderance of lexical words in news texts is due to the pronounced informational aspect of that register. The high ratio of nouns to verbs in NEWS corresponds to longer clauses and more complex phrases embedded in clauses (pp. 65-66).

Considering the distribution of function word classes, LGSWE finds that pronouns are most frequent in CONV, and relatively rare in NEWS and ACAD. Prepositions and determiners, however, are most frequent in NEWS and ACAD, and less common in CONV. Auxiliaries and adverbial particles are more common in CONV and FICT than in NEWS and ACAD. Co- and subordinators vary to a lesser extent across registers than the other function word classes. The above distribution is closely connected with that of lexical word classes: the high frequency of nouns in NEWS corresponds to a low density of pronouns. NEWS has the highest frequency of prepositions and determiners, because those serve as extensions or specifications of nouns, which are also very frequent in NEWS. The authors of LGSWE find it somewhat surprising that subordinators are relatively rare in all registers, but more common in

Northern Echo and *The Scotsman*, a sample varied from the point of view of political affiliation, regional distribution, readership levels and topics.

CONV and FICT than in ACAD and NEWS. This, Biber et al. say, contradicts the widely held belief that conversation is grammatically simpler than the written registers (p. 7), and also reflects that, in CONV, complexity resides at the level of clause combinations (cf. the high frequency of verbs), while, in NEWS and ACAD, complexity lies rather at the phrase level. (pp. 91-93).

3.2 Nominal elements in printed news texts

3.2.1 Simple noun phrases and reference

By ‘nominal elements’ Biber et al. (p. 230) mean not only phrases headed by a noun or a pronoun, but all constituents which appear in positions characteristic of noun-headed structures including, for example, prepositional phrases, phrases with adjectives used as noun phrase heads and nominal clauses (introduced by *that* or a *wh*-word). Due to their referential specification, the role of nominal elements in a text is to specify who and what the text is about. In NEWS, nominal elements make up about 80% of the text, as opposed to approximately 75% in ACAD, 70% in FICT, and 55% in CONV. However, the main difference in nominal elements in the four registers can be related to the complexity of their realisation, not their frequency. In the LSWE Corpus, nouns are many times more common than pronouns in NEWS and ACAD than in FICT and CONV. As has already been mentioned, the high frequency of nouns is due to the informational purposes of news texts (p. 235).

There are also striking differences in the distribution of anaphoric expressions across registers. Personal pronouns and definite noun phrases are the main anaphoric devices in all registers. Both are equally common in NEWS, but NEWS and ACAD display a higher number of anaphoric definite noun phrases than other registers. This can be attributed to a dense use of nouns, and hence a great deal of potential competition among referents, which requires more specific anaphoric devices. A definite noun phrase meets that requirement.

Another anaphoric device relevant for NEWS is a definite article+synonym (e.g. *Kylie Minogue* on first mention, and *the Aussie singer* on next). This device is used twice more often in NEWS than in FICT or ACAD (and more than eight times more frequently than in CONV), though more rarely than a definite noun phrase, which is more unambiguous than a synonym. Nevertheless, synonyms are relatively common in

the written registers, because lexical variety is valued in these genres. In addition, the use of a synonym makes it possible to draw attention to different aspects of a referent, therefore it is more informative than simply repeating the noun (pp. 237-238).

As for the distance of an anaphoric expression to the nearest previous mention in NEWS, the average number of intervening words between coreferents is: 20 in the case of personal pronouns, 30 in the case of definite article+synonym, and 40 in the case of definite article+repeated noun. This reflects a clear relationship between explicitness and anaphoric distance: the more explicit the anaphoric expression is, the larger span it allows between itself and its referent (pp. 239-240).

3.2.2 Complex noun phrases: premodification

Noun phrases come mainly with modifiers in NEWS and ACAD. In all registers, noun phrases with premodifiers are somewhat more common than those with postmodifiers. In the case of NEWS, the rate is the following: 45% of all noun phrases have no modifiers, 28% have premodifiers, 17% have postmodifiers, and 10% have both. (ACAD has by 10% more modified nouns, but in CONV only 15% of all noun phrases have any modifier) (p. 578). NEWS stands out in that, while normally proper nouns rarely take a modifier, 20% of all proper nouns appearing in NEWS take some kind of modifier (p. 580).

While both pre- and postmodification are common in written expository registers, NEWS shows a slightly greater preference for premodification. This could be due to the space limitations of news reportage, since premodifiers are generally shorter than postmodifiers (p. 579). Adjectives are the most common category of premodifier in all registers, followed by nouns, then *ed-* and *ing-* participles. In the case of NEWS the total number of premodifiers breaks down as follows: roughly 53% are adjectives, 39% are nouns, 5% are *ing-* participles and 3% are *ed-* adjectives. However, it is exactly the NEWS where the relative frequency of noun premodifiers is the highest, apparently because of the great need for brevity, since these constructions express a wide range of meaning-relationships in a succinct form, although the exact relationship is often not explicit (p. 589).

In NEWS, four premodifying nouns are extremely productive (combining with more than 100 different head nouns), and extremely frequent (occurring as a premodifier over 100 times per million words): *government*, *police*, *home* and *world*.

Twelve premodifying nouns are extremely productive, but only relatively frequent (occurring over 50 times per million words): *business, car, city, council, family, health, labour, market, party, record, security* and *TV*. An additional eleven premodifying nouns are extremely productive, but relatively rare (occurring less than 50 times per million words): *cash, club, community, computer, drug, food, future, safety, school, state, tax* and *water*. These words reflect the general concern of the news. It is also notable that certain premodifying nouns combine with different head nouns in the news than in conversation, e.g. NEWS: *water + authorities, bill, companies, industry*; CONV: *water + balloon, bottle, fight, leak* etc. Although noun+noun combinations are space-saving – a characteristic much valued in the news –, they place a heavy burden on readers, who must infer the intended logical relationship between the two nouns. LGSWE authors think that news relies primarily on premodifying nouns from the semantic domains most commonly associated with current events, because these are areas news writers can reasonably expect readers to have well-developed pragmatic knowledge of, and so be able to decode the constructions without too much difficulty.⁸ (pp. 590-594)

More restricted in distribution, plural nouns also appear as premodifiers, and, again, they are much more common in NEWS than in other registers.⁹ The high frequency of plural first-elements in noun+noun constructions fits into the general pattern news follows to reach its aim: to convey a maximum of information as concisely as possible (pp. 594-595). Multiple premodification is also to be found, but the four registers are quite similar in their proportional use of premodification by length: 70-80% of premodified noun phrases have only a single premodifier, about 20% have two-word premodification, and only about 2% have three- or four-word premodification. Yet, longer premodifier sequences are slightly more common in NEWS (p. 597).

3.2.3 Complex noun phrases: postmodification

Prepositional phrases are by far the most common type of postmodification in all registers, making up about 65-80% of all postmodifiers (and most common in ACAD). In the case of NEWS this rate is approximately 73%. Relative clauses make up more

⁸ It might also be true that readers get used to frequent combinations, which, in this way, come to be cognitively processed as single words. – FJ

than half of the remaining 27 per cent. The order of frequency of other types of postmodification is the following: appositive noun phrases (also common in ACAD, but less in other registers), *ed*-clauses, *ing*-clauses and, finally, *to*-clauses (p. 606).

Restrictive relative clauses are more common than non-restrictive relative clauses in all written registers, but the relative frequency of non-restrictive relative clauses is twice as high in NEWS than in other registers. Generally speaking, restrictive postmodification is used to help identify the reference of the head noun, while non-restrictive relative clauses have an elaborating function. NEWS makes good use of this property of non-restrictive clauses in that it uses them to add information of potential interest but not directly related to the news story (pp. 603-606). *Who* is the most common relativiser in news, followed by *which*, *that* and zero (but ACAD prefers *which*, FICT and CONV favour *that*). Since *who* occurs only with animate (usually personal) head nouns, its high frequency reflects the emphasis on humans in the communicative focus of NEWS (pp. 610-612).

Appositive noun phrases are common in NEWS and ACAD, both in terms of their absolute frequency and proportionally. In NEWS, over 90 per cent of all appositive constructions involve a proper noun, the majority denoting humans. Since an appositive noun phrase is a maximally abbreviated form of postmodifier, it is favoured in a register with a high informational density. This descriptive phrase is used primarily to provide background information about people (p. 639).

LGSWE differentiates between postmodifiers and noun complement clauses, which superficially look identical, but represent very different structural relations. Thus, (1) *a report that lay behind him on the cupboard top* and (2) *an industry trade group's report that its leading indicator fell in September* are two different kinds of clauses: (1) is a postmodifying relative clause identifying the reference of the head, and is structurally incomplete, while (2) is a noun complement clause, which presents the complete content of the head noun, and could structurally stand on its own as an independent clause. The distinction is important because, although noun complement clauses are less frequent than postmodifiers, they are relatively common in NEWS and ACAD.

In NEWS, the most preferred types of clausal postmodification are *to*-clauses, then *that*-clauses and, finally, *of+ing*-clauses and *wh*-clauses. While postmodifying

⁹ Here a dialectal difference can be noticed: plural premodifying nouns are more productive in British English than in American English.

clauses can occur with almost any head noun, noun complement clauses are controlled by a closed set of head nouns, which mark some abstract (nominalised) stance towards the proposition in the complement clause. The fact that *to*-clauses are especially common in NEWS relates to the set of nouns that control complement clauses. The most frequent abstract nouns with *to*-clauses are the following: *chance*, *attempt*, *decision*, *plan* and *bid*, occurring over 50 times per million words in NEWS, but much less often in other registers (pp. 644-652).

3.3 Verbal elements in printed news texts

3.3.1 Verbs

LGSWE differentiates between three major verb classes in English: lexical or full verbs (e.g. *run*, *eat*), primary auxiliaries (*be*, *do* and *have*) and modal verbs (e.g. *can*, *may*). The most common class of verbs in the LSWE Corpus is lexical verbs: more than one word in ten is a full verb. Primary auxiliaries and modals are also very common, but primary verbs are about twice as common as modals. The high frequency of lexical verbs as compared to auxiliaries reflects the fact that main verbs often occur in simple tense forms and nonfinite clauses. Primary verbs are more common than modal auxiliaries because they express basic grammatical meanings of aspect and voice. In addition, *do* is obligatory in many questions and negations. The verbs are not distributed evenly across registers. Lexical verbs are less common in NEWS (and even less in ACAD) than in FICT and CONV (pp. 358-359).

LGSWE classifies lexical verbs into seven groups, according to the semantic domains they express in their most common meaning: activity verbs (e.g. *bring*, *go*, *take*), communication verbs (e.g. *ask*, *say*, *tell*), mental verbs (e.g. *think*, *love*, *taste*), verbs of facilitation or causation (e.g. *allow*, *let*, *require*), verbs of simple occurrence (e.g. *become*, *happen*, *change*), verbs of existence or relationship (e.g. *be*, *exist*, *have*), and aspectual verbs (e.g. *begin*, *keep*, *stop*). Overall, LGSWE finds that activity verbs occur much more commonly than verbs from any other semantic domain, but mental and existence verbs are also notably common. As for their register-relatedness, activity verbs are most common in CONV and NEWS (as well as in FICT). Existence verbs, however, are more often used in NEWS than in CONV.

The twelve most common verbs in the whole LSWE Corpus are the following: *say, get, go, know, think, see, make, come, take, want, give* and *mean* – in decreasing order of frequency. However, the twelve most common LSWE verbs account for only about 17% of all lexical verbs in NEWS. The most common verbs in each semantic domain of NEWS are the following: *make, say, see, help, seem,* and *start/keep/begin/continue* (equally frequent). It is to be noted that, while *get* is generally very common, it is not so frequently used in the written registers, because many of its uses have strong causal overtones which are avoided by more careful writers of informational prose. *Make* and *take* are more common in NEWS than in other registers, but they often occur in idiomatic expressions: e.g. *make his mark in politics, take advantage of changes* (p. 376-377).

NEWS (and ACAD) shows a high proportion of verbs occurring with inanimate subjects. Although the core meanings of activity, communication and mental verbs usually entail human subjects, NEWS (and especially ACAD) resorts to an abstraction, with an inanimate entity functioning as the subject of these verbs (pp. 378-379).

3.3.2 Variations in the verb phrase

Tense and modality:

The authors of LGSWE state that finite verb phrases in English can be marked for either tense or modality, but not both. Thus the presence of a modal verb precludes tense marking, although modal verbs sometimes express time distinctions (p. 453). It is found that about 85% of all finite verb phrases in the LSWE Corpus are tensed, which means that tensed verb phrases are much more common than verb phrases with modals. Modals are used to express a speaker or writer's stance (the degree of certainty of the proposition, meanings such as permission, obligation etc.). However, in a sense the absence of a modal verb can itself be interpreted as a choice of stance, with the addressor attributing unquestioned validity to the proposition. Overall, present tense verbs are somewhat more common than past tense verbs, and their distribution varies considerably across registers. NEWS stands out in that, unlike other registers, it shows no preference for one or the other tense: it uses both to about the same extent (with slightly more present than past) (pp. 456-457). At this point, LGSWE offers no explanation why this balance characterises NEWS.

Aspect:

Aspect in verb phrases relates to considerations such as the completion or lack of completion of events or states described by a verb. The perfect aspect designates events or states taking place during a period leading up to the specified time, and the progressive aspect refers to events or states in progress, or continuing, at the indicated time (p. 460).

Verb phrases unmarked for aspect are overwhelmingly the most common in all four registers (about 90%). When some aspect is marked, progressive verb phrases are slightly less common than perfect ones. (Perfect progressive aspect is extremely rare: less than 0.5% of all verb phrases.) The large majority of perfect and progressive verb phrases in CONV, NEWS and ACAD are in the present tense. (FICT shows the opposite preference.) In verb phrases marked for aspect, NEWS shows a slightly stronger preference for the perfect aspect over the progressive than other registers, since the perfective is used to report events or states existing at an earlier time, and this focus on reporting is typical of NEWS (pp. 460-462).

The aspectual markedness of verb phrases also varies according to dialects. In British news the perfect aspect is much more common than in American news. This, LGSWE speculates, might be accounted for by the space-saving drive of American newspapers, which adhere (even more than American CONV) to the general trend in American English to use the nonperfective past tense in contexts where British English favours the present perfect. Also, in British news the perfect is about twice as common as the progressive. The frequency of the progressive aspect in NEWS is slightly higher in American than in British English (pp. 462-463).

Voice:

The average rate of active and passive voice in verb phrases in the LSWE Corpus is the following: active verb phrases represent 75 to 98% of all verb phrases, with passive ones ranging between 25 to 2%, respectively. Passives are by far most common in ACAD, but notably common in NEWS as well, where about 15% of the verb phrases are in the passive voice. The reasons for the use of the passive are somewhat different in the two registers: ACAD favours the passive because (1) demoting the agent of the verb gives topic status to the affected, and (2) the extensive use of passive constructions conveys an objective detachment from what is being described, as required by the Western scientific tradition. NEWS, in turn, has space-

saving reasons for using the passive (especially the short or agentless one), since often the focus of a story is an event involving an affected person or institution, and the agent may be easy to infer, uninteresting, already mentioned, unknown, or not mentioned because of legal reasons (pp. 476-477).

In addition to register, lexical factors are strongly related to the choice between active and passive forms, since there are some verbs which normally take the passive voice, while others rarely do so. In NEWS, most of the verbs in the passive report negative events that happened to someone, omitting mention of the person who performed the activity: e.g. *be + accused, arrested, charged, injured, killed released, shot* etc. (pp. 479-480).

English verb phrases can be marked for combinations of aspect and voice. Out of the possible combinations, the perfect passive is the only one which is moderately common. Its occurrence is not even across registers: it is used relatively often only in ACAD and NEWS. Given the high frequencies for both perfect aspect and passive voice in these two registers, it is not surprising that the two are often used together. Typically, this combination retains the time orientation ('past with present relevance') of the perfect aspect while demoting the agent through the use of the passive.

Progressive passives are only occasionally used, and if so, they occur in NEWS and ACAD (pp. 482-483).

3.3.3 Modal auxiliaries

In the LSWE Corpus, the three most common modal auxiliaries are *will, would* and *can* (in decreasing order of frequency), and the rarest is *shall*. As for their distribution across registers, it is found that modal (e.g. *may, must*) and semi-modal (e.g. *be going to, have got to*) auxiliaries are most common in CONV and least common in NEWS.¹⁰ Although modals are rare in NEWS, the most often used ones coincide with those most frequent overall: *will* and *would*. *Have to* is the only semi-modal that is relatively common in the written expository registers (pp. 486-489).

Modals with obligation/necessity meanings in the LSWE Corpus are lower in frequency than the ones with meanings related to permission/possibility/ability or

¹⁰ Biber et al. use the term 'semi-modal' in the same sense as Quirk et al. (1985:143), that is, a set of verb idioms which express modal or aspectual meaning and which are introduced by one of the primary verbs *have* and *be*.

volition/prediction (these are the meaning categories in which LGSWE groups modals). This reflects a general tendency to avoid the face-threatening force of expressions with an obligation meaning. In addition, semi-modals have become better established in this semantic domain, replacing modals to a greater extent (pp. 489-490).

While the majority of modals do not co-occur with marked aspect or voice, certain modals show different preferences for these combinations. The perfect aspect with modals, for example is most common in FICT and NEWS, particularly with permission/possibility *might* and *may*, and obligation/necessity *must* and *should*. The aspect markers in such expressions serve to provide a past time reference, while the modals typically refer to logical necessity. Although the progressive aspect with modals is generally rare, an exception to this tendency is volition/prediction modals in NEWS. It should be noted, for example, that *shall* is rare but, when it occurs, it often combines with the progressive aspect. Volition/prediction modals in NEWS mark future events or situations that will take place over (or after) a period of time: e.g. *We shall be meeting with all parties in the near future*. The simple aspect in such contexts would suggest a strong volitional meaning, which is not intended (pp. 499-501).

Some semi-modals can occur in series following another (semi-)modal. These sequences are by far most common with *have to* in all four registers. *Will + have to* is most common in NEWS. Sequences of modal + *need* are also relatively common, but restricted primarily to ACAD and NEWS (pp. 501-502).

3.4 Adjectives in printed news texts

Adjectives are relatively common in all registers, and their distribution mirrors that of nouns: since adjectives are frequently used to modify nouns, thus adding to the informational density of expository registers, where nouns are common, adjectives will also be common in NEWS and ACAD (p. 504).

The main syntactic roles of adjectives can be attributive, modifying nominal expressions, and predicative, following a copular verb. Adjectives in NEWS (and ACAD) are mostly used in an attributive role: in 84% of the cases. This reflects the heavy reliance on noun phrases to present information in these registers (p. 506).

LGSWE classifies adjectives in the following semantic groups:

A Descriptors

- colour: e.g. *black, white, dark, bright*
- size/quantity/extent: e.g. *big, short, wide*
- time: e.g. *annual, late, young*
- evaluative/emotive: e.g. *bad, beautiful, nice*
- miscellaneous descriptive: e.g. *appropriate, free, strange*

B Classifiers

- relational/classificational/restrictive: e.g. *average, initial, necessary, similar*
- affiliative: e.g. *American, French, United*
- topical/other: e.g. *chemical, phonetic, visual* (pp. 508-509).

The four registers use attributive adjectives from the main semantic domains in quite different ways. Descriptors are found in all registers, but to varying extents, corresponding to the general frequency of adjectives in the register in question. (FICT uses them in a greater number than other text types). Classifiers, on the other hand, are used primarily in the informational written registers. NEWS uses affiliative classifiers to a greater extent than the other registers. Size, time and evaluative adjectives are also important in NEWS. These correspond to the typical subject matter of this register (pp. 501-511).

The most common attributive adjectives in NEWS are: *new/old* (time descriptors) –most frequent; *big/long/small/great/high* (size descriptors), *good/best* (evaluative descriptors), *same/full/general/major/final* (relational classifiers) and *political/public/international/ national* (topical classifiers) – of about equal frequency. This wide range of adjectives serves the twin purposes of clearly identifying the referents of noun phrases and providing descriptive details about those referents. Many of the most common attributive adjectives in NEWS are words derived from, or closely related to, nouns (especially those ending in *-al*: e.g. *political, national*). Such adjectives tend to delimit the domain of the head noun (pp. 512-514). The most frequent predicative adjectives in NEWS are: *able* and *likely*, followed by *sure, right, good, difficult, hard, ready, available* and *unlikely*. *Due* is only common in NEWS. Most frequent predicative adjectives in NEWS take clausal or phrasal complements. Thus, they provide a frame for making judgements and supplying information – both

important functions of news reports. Many predicative adjectives in NEWS occur in quoted speech (pp. 516-518).

In the comparison of adjectives, the general tendency is that monosyllabic gradable adjectives take inflectional suffixes (*-er* and *-est*), and polysyllabic ones take phrasal markings (*more* and *most*), with disyllabic adjectives varying in occurrence with the two ways of comparison. In all registers, comparative degree adjectives are about twice as frequent as inflected superlatives. NEWS stands out in that it shows a much greater frequency of superlative forms, probably reflecting a focus on the extreme in the interests of readers. The two most frequent adjectives in a comparative or superlative form in NEWS are: *more likely* and *most important* (pp. 522-525).

3.5 Adverbs in printed news texts

Adverbs are generally common in the LSWE Corpus, but less frequent in the written registers than in conversation, given that verbs are also less common in these registers, and adverbs most commonly co-occur with verbs (p. 504).

In the use of adverbs, NEWS does not stand out in any important respect in the LSWE Corpus. (The greatest differences in adverb use can be seen in CONV vs ACAD.) LGSWE, therefore, does not concentrate on this register in its discussion of adverbs. (A more detailed discussion of adverbs will be found in Section 4.5, focusing on the use of adverbs in conversations.)

3.6 Clause grammar in printed news texts

While LGSWE is very comprehensive in its discussion of major word classes and the comparison of their usages, in its treatment of major clause elements, basic clause patterns, negation, concord and types of dependent clauses, LGSWE is less concerned with register comparisons. Text type comparisons only occur in the part dealing with the form and function of complement clauses. By 'complement' Biber et al. mean all clause elements that verbs can take: objects, predicatives and obligatory adverbials (p. 122).¹¹

¹¹ This is a broader sense of 'complement' than in Quirk et al. (1985:54-55); 'predicative' means exactly the same as Quirk et al.'s 'complement', that is, the obligatory post-verbal element in clause types such as *He is handsome* or *She finds him handsome*.

One does find, however, a comparison of two individual texts at the beginning of the clause grammar section of LGSWE: of a conversation sample and a piece of news text. It is found that the conversation text contains a good deal of non-clausal material, while all the phrases in the news text belong to clausal units. Further, there are more single-clause units in the conversation text (i.e. 15) than in the news article (i.e. 1). The conversational piece is also characterised by shorter clauses (made up of 6 words in average, as opposed to 20 words in the news text) and a lower degree of embedding, that is, the occurrence of one clause as a constituent of another. The number of top level clauses is 20 in the conversation text and 7 in the news text; that of 1st level embedding is 6 in CONV, 8 in NEWS; while that of 2nd level embedding is 1 in CONV and 3 in NEWS. LGSWE summarises its findings saying that there is less structural integration in the conversation text, where clauses also tend to be shorter than in the newspaper article (p. 121).

3.6.1 Complement clauses in printed news texts

Complement clauses are dependent clauses that complete the meaning relationship of an associated verb or adjective in a higher clause. In such cases, the main clause verb or adjective controls the complement clause. The four major structural types of complement clauses identified by LGSWE are: *that*-clauses, *wh*-clauses, *to*-infinitive clauses (or, sometimes bare infinitive clauses) and *ing*-clauses (pp. 658-659). The first two dependent clauses are finite, and they are most common in CONV; the last two complement clauses are nonfinite, which are predominantly features of written language (p. 749).

That-clauses:

That-clauses are typically used to convey the speech, thoughts, attitudes or emotions of humans. Over 80% of all *that*-clauses occur in a post-predicate position controlled by a verb. The most common verbs controlling *that*-clauses are: *think*, *say* and *know*. Post-predicate *that*-clauses are very common in NEWS, where *say* is by far the most frequent verb controlling the *that*-clause: about six times more often used than the next most common verb, *believe*. This reflects that the pronouncements of public figures are considered newsworthy (pp. 668-674).

Post-predicate *that*-clauses controlled by adjectives are much less common overall, and especially in NEWS. *That*-clauses in subject position are rare in all registers. Extraposed *that*-clauses, however, are moderately common in NEWS (and ACAD), with the copular verbs *be*, *seem* and *appear* taking most of them, and the passive constructions *be known*, *be assumed* etc. taking the others. Subject predicative *that*-clauses are moderately common in NEWS (and ACAD). Most of them describe the nature of a problem, or present reasons, results and conclusions: e.g. *The problem is that...* *The net result is that...* These types of clauses use an impersonal, non-human noun phrase as a subject, and a predicate representing a static relation or attribute, which makes them suitable for the purposes of written expository registers (pp. 674-675).

In NEWS (and ACAD), it is not typical for the complementiser *that* to be omitted (in fact, it is retained in about 70% of the cases). Its omission, when it does take place, is promoted by three factors: the use of *think* or *say* as the main clause verbs, the occurrence of co-referential subjects in the main clause and *that*-clause, and the occurrence of a personal pronoun subject in the *that*-clause. These three factors are instrumental in the deletion of *that* in all registers, but their influence is strongest in NEWS (pp. 680-682).

Wh-clauses:

By *wh*-clauses Biber et al. mean dependent clauses starting with *wh*-words. LGSWE also includes in this category clauses starting with *whether* or *if*, referred to in other grammars as *yes/no* interrogative clauses (cf. Quirk et al. 1985:1053-1054). *Wh*-clauses are found to be relatively rare in NEWS and ACAD (pp. 683-688).

Post-predicate position *wh*-clauses usually follow a verb or a verb + a noun phrase. The most common verb controlling a *wh*-clause is overwhelmingly *know*, followed by *see*, *tell*+NP, *wonder*, *ask* and *understand*. Activity verbs can also occur with *wh*-clauses, but the most frequent verbs with *wh*-clauses are from some of the same semantic domains as those with *that*-clauses, especially communication and mental verbs. The verb *decide* is common in controlling *whether*-clauses in NEWS (pp. 689-692).

Infinitive clauses:

Infinitive clauses can occur in either post-predicate or subject position, and in a variety of other structures, but they are by far most common in the post-predicate position (p. 693). Over 60% of all *to*-clauses occur in this position. Infinitives controlled by verbs are very common in NEWS. *Want* is the most common verb in this function, followed by *try*, *seem* and *begin* in this register (p. 711). Infinitive clauses in verb-controlled positions are typically used to report the activities, desires and other thoughts or emotions of human participants, therefore, they are common in NEWS and CONV, which focus on such concerns. *To*-clauses complementing adjectives are also frequent in NEWS, especially those describing a stance held by some person regarding their own activities: e.g. *Mr Beregovoy has been quoted as saying that he is not ashamed to be called a social democrat.* (p. 722-723).

Extraposited *to*-clauses following an adjective, and subject-predicative *to*-clauses are moderately common, while subject *to*-clauses are rare in NEWS (p. 722). When they do occur, they mainly mark the information in the *to*-clause as being part of the current topic, providing a direct anaphoric link to the preceding discourse, or present a topical progression of connected ideas, foregrounded in parallel structures for effect: e.g. ... *the law also allowed the seizure of all assets, not simply those related to a specific offence. ... To take away the profits of a crime is one thing; to seize assets that have no connection with crime is another* (p. 725).

There are two verbs that can control either a *to*-clause or a bare infinitive: *dare* and *help*, the latter being more common. The *help*+infinitive structure is found to be especially frequent in NEWS, used to describe a process facilitated by some other factors: e.g. *Key testimony has been broadcast live on television and radio to help keep the city fully abreast of the proceedings...* The bare infinitive is more frequent than the *to*-infinitive in both British and American English, particularly in the pattern verb+NP+infinitive. Besides dialect differences, the distribution of these structures also differs across registers: bare infinitives predominate in CONV, NEWS and FICT, while in ACAD they are responsible for ca. 55% of the *help*+infinitive constructions. When the verb *help* is itself used as a *to*-infinitive in NEWS, 80% of the following complement clauses are bare infinitives: e.g. *US military police units might be forced to stay in Panama for at least a year to help keep the peace* (pp. 735-737).

Ing-clauses:

LGSWE does not differentiate between gerunds and present participles, using the above umbrella term to denote all clauses containing verb forms ending in the *-ing* suffix. Biber et al. find that *ing*-clauses are most commonly used in conjunction with an aspectual verb in the main clause, but they are also used to report speech acts, cognitive states, perceptions, emotions etc. They are by far most common in post-predicate position, but they can function as subjects or subject predicatives. The six most common verbs controlling *ing*-clauses are (in decreasing order of productivity): *keep/start* (aspectual verbs), *see+NP* (perception verb), *go on/stop/begin* (aspectual verbs again) (pp. 739-745).

In NEWS, *ing*-clauses are more frequent than in CONV and ACAD. The two most frequent verbs coincide with those most common in taking *ing*-clauses overall: *start* and *keep*, denoting the beginning or the on-going manner of an action, which is not surprising of NEWS typically reporting on actions and events of the past. NEWS is the only register to use verbs of offence or punishment frequently with an *ing*-clause, reflecting the common theme of 'crime and punishment': *admit, be blamed for/accused of/charged with* (pp. 746-748).

In all registers, adjectival predicates are less frequent than verbs in controlling *ing*-clauses. The most common adjective followed by an *ing*-clause is *capable of*. Some of the next most common are: *afraid of, aware of, necessary for, useful for*. Most of these adjectives convey either an affective stance or some other evaluation. In NEWS, these adjective+*ing*-clause constructions are not especially common (they typically occur in ACAD, often in an impersonal use) (p. 749).

There are certain verbs which can control both *that*-clauses and nonfinite clauses as complements: e.g. *agree, hope, remember*. The grammatical conditions which require the use of *that*-clauses are: when the subject of the complement clause is not co-referential with the subject of the main clause (e.g. *I hope that you were happy while you were here*), or when the complement clause includes a modal verb (e.g. *Remember that fortune and misfortune should be left to heaven and natural law*). For verbs that can control both *to*- and *ing*-clauses but not *that*-clauses, semantic factors influence the choice. In general, a *to*-clause has a meaning that is more hypothetical or potential than the meaning of the *ing*-clause with the same verb. For example, ca. 75% of the

occurrences of *like+to*-clause in NEWS are preceded by *would*, which has a hypothetical meaning (pp. 756-757).

3.6.2 Adverbials in printed news texts

Adverbials are elements of clauses with three major functions: to add circumstantial information about the proposition in the clause ('circumstance adverbials'), to express speaker/writer stance towards the clause ('stance adverbials'¹²), or to link the clause (or some part of it) to some other unit of discourse ('linking adverbials') (p. 762). The classes differ in the extent to which they are integrated into the clause structure, and the amount of variability in the precise functions of the class (p. 763). Circumstance adverbials are the most varied class, as well as the most integrated into the clause structure. They add information about the action or state described in the clause, answering questions such as '*how, when, where, how much, and why?*'. Stance adverbials convey speakers' comments on what they are saying, or how they are saying it. They have scope over the entire clause, and they are less integrated in the clause structure than circumstance adverbials, partly because they are always optional. Linking adverbials have the most peripheral relationship with the rest of the clause, and they serve a connective function between two units of discourse on different levels (pp. 764-765).

Adverbials are very common in all four registers. They are most common in FICT and least common in NEWS and ACAD. In all four registers, circumstance adverbials are by far the most common class of adverbials. Stance and linking adverbials are not especially common in NEWS (p. 765).¹³

Circumstance adverbials:

LGSWE divides these adverbials into seven major semantic categories:¹⁴

- place: distance e.g. *A woman who fell 50 feet down a cliff ...* ; direction e.g. *... rebels advancing rapidly southwards*; position e.g. *A bus lay in a ditch*;

¹² Another term that critics say is idiosyncratic and lacks justification (Schneider 2001:140).

¹³ In the following discussion, the three dots in the examples indicate truncation.

¹⁴ Some of these categories are different in Quirk et al. 1985.

- time: position in time e.g. *I'll see you all tomorrow night*; duration e.g. *It lasted years*; frequency e.g. *Occasionally she would gaze out the window*; temporal relationship e.g. *After this the conversation sank for a while*;
- process: manner e.g. *I found myself writing slowly*; comparison e.g. *The lip curled like a snail's foot*; means e.g. *... shipments be made by air*; instrument e.g. *... she fed it with a teaspoon*; agent e.g. *... caught and scratched by thorns*;
- contingency: cause/reason e.g. *He... died of head injuries*; purpose e.g. *... the knives were just for show*; concession e.g. *... despite intimidation, no one has gone back to work*; condition e.g. *If she smiles it will be at your hairline*;
- extent/degree: *... rateable values would rise by about seven times*; amplifiers e.g. *She looked very much like her mother*; diminishers e.g. *He hardly dared to look...*;
- addition/restriction: additive e.g. *also* in *The tycoon, who is also chairman of...*; restrictive e.g. *especially* in *A heart born especially for me...*;
- recipient: e.g. *I'll write the check for you*.

Certainly, there are other, less clearly defined semantic categories as well, and there are many cases in which adverbials do not clearly fit into one of the seven major categories, but even then, they often have a primary meaning (pp. 776-783).

In all four registers, place, time, process and contingency are the four most common categories, however, their order of frequency varies across registers. In NEWS, time dominates, followed by place, then process. These adverbials help to provide the information typically considered important in news stories: when, how, why or where someone did something. Further, since news is particularly concerned with current events, time adverbials are especially valued to make clear the time relationship of an event. The lower frequency of place adverbials relative to time adverbials can also be attributed to the fact that information about location can be included in structures other than adverbials: e.g. *Manchester company*. Within the category of process adverbials, manner is by far the most common subcategory in all four registers. In NEWS, agentive adverbials are moderately common (cf. the relatively high frequency of passive verb structures in NEWS, see Section 3.3.2). Within the category of contingency, cause/reason, purpose and condition are the most common subcategories in all four registers, but the proportional use varies across registers. In NEWS, cause/reason and purpose dominate (pp. 783-786).

Since prepositional phrases are the most common syntactic realisations of circumstance adverbials in all registers, interesting insights are offered by a look at the length of prepositional phrases across registers. The longest phrases can be found in FICT, followed by NEWS, where 36.5% of the prepositional phrases are more than 6 words long, 10% are 5, 13.5% are 4, 21% are 3 and 19% are 2 words long. It can be seen that, in more than half of the cases, there are pre- and/or postnominal modifiers included in the prepositional phrase. This adds to the dense packing of information characteristic of the written expository registers (pp. 791-792).

The most common circumstance adverbials in NEWS are: *also* and *now* (single adverbs), as well as words for the days of the week (noun phrases). *Yesterday* is very common in British news, while the words for the days of the week in American news. *Also* serves to mark information added to previous knowledge. The most common time adverbs show the emphasis on current information. In British news reports, the days of the week tend to be more often incorporated into prepositional than noun phrases as adverbials (e.g. *on Sunday, for Tuesday's trip*), hence their lower frequency. In American news, the names of the days of the week often appear without the preposition *on* (e.g. *The president said Tuesday he would continue...*), hence their higher frequency. *Now*, in NEWS, generally marks a change from previous events, showing the latest newsworthy event or telling readers what they can expect next: e.g. *There is now talk of an interest rate cut some time this week* (pp. 796-801).

Circumstance adverbials can appear in a clausal form as well. In NEWS, time and purpose clauses are frequently used. Time clauses are useful in two ways: they can describe certain events in relation to others, or they allow a more complete description of a time than is possible with phrase-level units. Purpose clauses reflect the need for news stories to explain motivations behind events. Some of the less common semantic categories, such as reason, and condition, are also important in providing explanatory information or background material for a news story (pp. 823-824). In all four registers, finite clauses are by far the most common type of circumstance adverbial clause. In NEWS, the distribution of the syntactic forms of clausal circumstance adverbials is the following: 58% are finite clauses, 23% are *to*-clauses, 15% *ing*-clauses, 4% verbless, while *ed*-clauses are virtually non-existent in this register (p. 826).

As for what can occur as subordinators of adverbial clauses, reason and condition have the least diversity, with *because* dominating in reason clauses and *if* in conditionals. Time, in turn, shows the greatest diversity in commonly occurring

subordinators, with NEWS (and FICT) making most use of the range of conjunctions: *when, as, after, before, while, until* and *since*. In NEWS, *after* can have strong overtones of cause as well: e.g. *He abandoned the attack after the woman screamed...* For concession, *though* and *although* are equally common in NEWS. Certain subordinators have multiple semantic roles: *as*, for example, can express manner, reason and time; *since* reason and time, and *while*, concession/contrast and time. NEWS mostly prefers *as* for time, *since* also for time, but *while* is used for time in 65% of the cases, the remaining 35% refers to concession/contrast (pp. 841-850).

In conditional clauses with *had, were* and *should*, it is possible to leave out the subordinator *if* and mark the adverbial clause with subject-operator inversion. In all registers, conditionals with *had* and *were* are very rarely marked by inversion; in NEWS, however, conditionals with *should* are marked by inversion in slightly over half of all occurrences. This shows that subject-operator inversion in conditionals is restricted to formal writing (pp. 851-852).

Stance adverbials:

Stance adverbials have the function of commenting on the content of style of a (part of a) clause. They are of three types:

- epistemic: doubt and certainty e.g. *maybe, certainly*; actuality and reality e.g. *in fact, really, actually, truly*; source of knowledge e.g. *apparently, according to ...*; limitation e.g. *in most cases, mainly, typically*; viewpoint or perspective e.g. *in our view, from our perspective*; imprecision e.g. *like, sort of, roughly, so to speak*;
- attitude: e.g. *(un)fortunately, inevitably, astonishingly, importantly, hopefully*;
- style: e.g. *honestly, frankly, more simply put*. (pp. 854-857)

Epistemic adverbials comment on the content of the proposition, the truth value that can be attached to it. Attitude adverbials express the speaker's attitude towards, or evaluation of, the content of the proposition, as well as an assessment of expectations. Style adverbials convey a speaker's comment on how the speaker is speaking, or how the utterance should be understood (p. 764).

Of all registers examined by LGSWE, NEWS has the lowest frequency of stance adverbials. In all four registers, epistemic adverbials are much more common than attitude or style adverbials, but in NEWS (and ACAD), attitude adverbials are slightly

more common than in CONV (and FICT). This class of stance adverbials mainly occurs in reviews, where the point of the text is largely to convey attitudes (pp. 859-860).

Three adverbials are especially common across all registers: *of course*, *perhaps* and *probably*. Of the four registers examined, NEWS has the lowest frequency of the most common stance adverbials: *according to*+NP is the only adverbial used with a higher frequency than in the other registers (especially in American news). Thus, the most common stance adverbials in NEWS are *probably* and *according to*+NP. NEWS uses *probably* and *perhaps* with predictions, suppositions, explanations and interpretations that have not been clearly proven. (pp. 867-869). The common use of *according to*+NP reflects the singular emphasis in NEWS on giving the source of a piece of knowledge. The noun phrase in these stance adverbials ranges from specifically named people and publications to sources identified only by their location (e.g. *according to Mr K* vs *according to reliable Westminster sources*) (p. 871).

Linking adverbials:

The primary function of linking adverbials is to state the speaker or writer's perception of the relationship between two units of discourse. Since they explicitly signal the connections between passages of text, linking adverbials are important devices for creating textual cohesion, alongside coordinators and subordinators.

LGSWE identifies the following semantic categories of linking adverbials:

- enumeration and addition: e.g. *first(ly)/second(ly)/lastly*, *for one thing/for another*, *in addition*, *further*, *similarly*, *moreover*;
- summation: e.g. *in sum*, *to conclude*, *overall*, *all in all*;
- apposition: e.g. *which is to say*, *in other words*, *i.e.*, *that is*, *e.g.*, *for example*, *for instance*;
- result/inference: e.g. *therefore*, *consequently*, *thus*, *so*, *hence*, *as a result*, *then*;
- contrast/concession: e.g. *on the other hand*, *in contrast*, *alternatively*, *though*, *anyway*, *however*, *yet*, *on the contrary*, *nevertheless*, *still*, *after all*, *in any case*;
- transition: e.g. *by the way*, *incidentally*. (pp. 875-879)

The distribution of linking adverbials differs across registers. Linking adverbials are considerably more common in CONV (and ACAD) than in NEWS (and FICT). NEWS, in fact, represents an extreme in that it uses the lowest number of linking

adverbials (with ACAD on the other pole). The rates of the different semantic types in NEWS are as follows: contrast and result represent a similar approximation of about 40% each, the next most common is apposition with 14%, followed by enumeration and transition with 3% each. Thus, when the relationships between units of discourse are marked, they mostly show contrast/concession and result or, more rarely, apposition giving additional clarification of information. Much more commonly in NEWS, however, the relationships between ideas are left implicit or encoded in other ways besides linking adverbials. Since reports of events are more common and arguments rarer in news, fewer linking signals are needed than, for instance, in academic prose. Sequences of events are typically reported with respect to their relationship in time, not with linking adverbials. Moreover, resultive relationships are easily inferred based on common knowledge and experience, so, again, there is no need to overtly signal them (pp. 880-882).

LGSWE extends its analysis on linguistic aspects such as 'lexical bundles' (by which Biber et al. mean extended collocations), idiomatic phrases, coordinated binomial phrases etc. However, because of a need to limit the scope of its description, Biber et al. focus only on the two registers in the LSWE Corpus that show the most striking differences in language use: conversation and academic prose. Since news reportage is not dealt with from the above points of view, these topics will also be neglected in the present account, which aims at a contrastive analysis of news and conversation.

In a review of Biber et al. 1999, Schneider (2001) has criticised it for the following: a lack of synchronicity in the corpus (fiction includes much earlier data than other registers); a lack of homogeneity of the samples in terms of size (the fiction corpus is much larger than the others); no attention to internal variation in the corpora; some idiosyncratic terminology with no justification; and no explicit references to earlier scholarship. Some of this critique seems exaggerated, since Biber et al. explain why their choice fell on data 'unbalanced' in size (1999:26), and point out that all frequency counts have been normalised to a million words to neutralise the imbalance. They also describe their efforts spent on representativeness in internal variation (1999:28). Schneider mentions some strengths of the manual as well: e.g. its emphasis on real language use and on differences between varieties of English; and its innovative and corpus-based character, which makes it "a major step ahead into a new direction of grammar-writing" (2001:142).

What one could miss from Biber et al.'s otherwise invaluable handbook are: data on the average length of sentences (measured in number of words), the average number of clauses per sentences, the average length of words (measured in number of syllables or letters), or the dominant clause patterns across registers. Occurrences across text types of verbs that can take several clause patterns are discussed on pp. 380-392, but only with respect to certain individual verbs (e.g. *see, call, tell*), not in a comprehensive manner.

3.7 Summary

The grammatical features of printed news identified by Biber et al. (1999) point in the direction of two functional characteristics of the press: an endeavour to effectively treat the typical subject matters of the (printed) media, and to pack as much information as possible in the smallest possible space. In other words, the grammatical make-up of printed news is closely related to *what kind* of information needs to be communicated, and *how* this information is conveyed.

Thus, the first group of grammatical features is related to the function of the press to inform on current events, newly emerged states, and the recent doings of people. These characteristics are the following:

- the verbs used are mainly activity and existence verbs;
- the present and the past tense are equally preferred in verb phrases (activities in the past, states in the present);
- the adverbials used are mainly circumstance ones: preferably time, place, manner, reason and purpose, in this order of frequency; and
- more than half of the nouns are modified, mostly by affiliative classifying adjectives, and relative or appositive clauses, to clarify the identity of referents.

Features belonging to the second group, and resulting in an informationally dense text, are the following:

- the highest lexical density, with 80% of a typical news text made up of nominal elements;

- a high frequency of nouns, determiners, prepositions and adjectives;
- a high rate of agentless passives;
- a high percentage of attributive adjectives;
- long, complex sentences with a high degree of embedding and structural integration;
and
- a common use of nonfinite clauses.

A parallel summary of the main findings in LGSWE on NEWS and CONV will be presented in Table 4.1 and 4.2 at the end of Chapter 4.

4 The grammatical characteristics of conversational discourse

While the previous chapter was an overview of LGSWE's observations on printed news discourse, this part will be concerned with the findings and comments of the same book on another of the four registers examined: conversation (CONV). The part of the LSWE Corpus dedicated to conversation comprises a British English sample of 3,929,500 words, complemented by a 2,480,800-word large American English corpus for dialect comparisons.¹⁵ The set of informants represented the range of English speakers in both countries across age, sex, social group and regional spread. These volunteers tape-recorded all their conversational interactions over a period of a week.

4.1 Basic grammatical characteristics of conversations

The type-token ratio is lower in the conversation texts than in all written registers. Thus, repetition is found to be a characteristic of spoken language. TTR is low in conversation because CONV is less concerned with the transmission of information than writing. Moreover, conversation is spontaneously produced, with little time for planning and varying the choice of words. Repetition may also be used for emphasis, to help the speaker's planning, or to make sure that the message gets across to the hearer (p. 53).

As for the types of words occurring in the LSWE Corpus, we have seen that, while NEWS has the highest lexical density, CONV has the lowest proportion of lexical word tokens. Inside the class of lexical words: although nouns are most frequent on an overall basis, they are by far least common in CONV, where nouns and verbs are about equally frequent. CONV has the highest frequency of verbs, which entails the highest frequency of adverbs, too (pp. 62-65). Verbs are frequent in CONV, because clauses are characteristically short, thus more numerous, and – since a verb is required in each clause – there will also be more verbs. Also, the interpersonal focus of CONV results in the frequent use of verbs to narrate events and to present personal attitudes. As for function words, pronouns are by far the most frequent in CONV. Prepositions and determiners are much less common in CONV than in the written registers. The high frequency of pronouns in CONV compensates for the low frequency of nouns.

Auxiliaries are more frequent in CONV (and FICT) than in the other text types. This follows from the high frequency of verbs, which are usually accompanied by auxiliaries. In spite of being relatively rare in all registers, subordinators are somewhat more common in CONV and FICT than in NEWS and ACAD. This is due to the high frequency of verbs, hence also of clauses and clause combinations. LGSWE concludes again that, since the use of co- and subordinators differs less markedly across registers than the use of phrase-level function words, register differences are more connected with the build-up of phrases than with the connection of clauses (pp. 91-93).

4.2 Nominal elements in conversations

4.2.1 Simple noun phrases and reference

Concerning the heads a noun phrase can have, LGSWE repeatedly states that pronouns are slightly more common than nouns in CONV, which is explained by the shared situation and personal involvement of the participants: the on-line production and context-dependent circumstances make it more appropriate to use pronouns instead of nouns (p. 235).

Out of the two main anaphoric devices in all registers (personal pronouns and definite noun phrases), personal pronouns are predominant in CONV (and FICT). Since there are fewer nouns in CONV than in other registers, there is less competition between potential referents of a pronoun, so it will be sufficient for identification. (If there is uncertainty, the addressee can ask for clarification, as it is often the case.) In fact, 85 per cent of all anaphors in CONV are realised by personal pronouns. These function words are not explicit, therefore they do not allow an especially large span between themselves and their previous mention: the average number of intervening words is 15 (pp. 237-240).

4.2.2 Complex noun phrases: premodification

Noun phrases in CONV come with a low degree of modification. CONV represents an extreme in this respect: only 15 per cent of all noun phrases have any

¹⁵ The LSWE Corpus includes a supplementary register of British non-conversational speech, too, such as lectures, planned speeches and sermons.

modifier at all (with ACAD at the opposite extreme). LGSWE accounts for this absence by making reference to the shared situation and personal knowledge of the participants in a conversation: speakers use referring expressions with minimal modification because they know the listener will have no trouble identifying the intended referent (pp. 578-579). Comparing the frequency of premodified noun phrases in the two registers of interest for the present study, we find the following: a complex noun phrase occurs 75 times per million words in NEWS, but only 21 times per million words in CONV.

With regard to the structural types of premodification, while in NEWS nouns were almost as productive premodifiers as adjectives, in CONV adjectives are far more common than nouns in a pre-noun position: roughly 70 % of premodifiers are realised by adjectives, 24% by nouns, and only 6 % by *ed-* or *ing-* participles. The high frequency of adjectives in a premodifying status in all registers and especially in CONV undoubtedly relates to the fact that they come from many different semantic classes, including colour, size/extent, time/age/frequency, affective evaluation (p. 589). Nouns, on the other hand, are not productive as premodifiers in CONV; in fact there are only four that produce more than 50 combinations: *car*, *Christmas*, *school* and *water* (for differences in the head nouns they tend to take, see Section 3.2.2). As for plural nouns occurring as premodifiers – much more rarely than singulars –, in CONV there is only one noun that combines with more than 20 different head nouns: *police* (pp. 592-594). Registers do not differ pronouncedly in the relative frequency of premodifier chains: in CONV, just like in other registers, 70-80% of premodified noun phrases have only one premodifier (p. 597).

4.2.3 Complex noun phrases: postmodification

Since it is not very common for nouns to have modifiers in CONV, postmodifiers are not frequent either. When they do appear, they mostly take the form of a prepositional phrase (in about 85% of postmodified nouns). The majority of the remaining 15% is postmodification by relative clauses, followed by appositives, *ing-* clauses, and, rarely, *to-* and *ed-* clauses. However, the relative frequency of *to-* clauses (having a future orientation) is higher in CONV than in writing. Further, in CONV postmodifiers often have a role not seen in the written registers: they can occur outside the regular clause structure, modifying a dislocated noun phrase, and clarifying the

reference of some nearby pronoun, e.g. *Cos Brenda, whose horse I rode up at Bridley – I was telling her...* (pp. 606-607).

The most preferred relativiser in conversational relative clauses is *that*, since it is the most flexible in terms of gap positions it can occur in and the meanings it can carry. (By 'gap' LGSWE means the missing constituent of the relative clause, which corresponds in meaning to the head noun.) Once every four times, however, the relativiser is left out altogether, due to the colloquial nature of CONV (pp. 610, 620).¹⁶ Another postmodifier speciality in CONV is the use of relative clauses which do not have a gap at all, sometimes resulting in constructions that might be considered nonstandard or even a disfluency, e.g. *Usually they give you ... a thing that you don't want it*. This characteristic is associated with the difficulties of on-line production. In addition, relative clause constructions in conversation are often complex, with deeply embedded gap positions, e.g. *I mean something which I think as a committee we should be in agreement with*. LGSWE considers the existence of such constructions in conversation surprising, since structural complexities are stereotypically associated with written exposition rather than speech. In formal writing, however, relative clauses with embedded gaps would be considered awkward (pp. 622-623).

Appositive postmodification and noun complement clauses are not especially common in CONV (pp. 639, 642). When they do occur, *that*-complement clauses typically follow the nouns *fact* and *idea*. *To*-clauses tend to occur with the noun *chance* (pp. 648, 652).

4.3 Verbal elements in conversations

4.3.1 Verbs

As has been mentioned in Section 4.1, verbs are more frequent in CONV than in any other text type. This is attributed to the fact that conversational partners talk a lot about actions and events, in short clauses often expressing a single main idea, and requiring a lexical verb each. Further, conversation commonly uses full verbs to frame the personal stance of the speaker (e.g. *I know, I mean, I think*). Analysing the relative

¹⁶ Interestingly, it is not CONV where the zero relativiser is most common, but FICT. Also, the frequency of relativiser omission is about the same in CONV and ACAD. In all registers, relativisers tend to be left out when the subject of the relative clause is a pronoun.

frequency of the different types of verbs in texts, however, LGSWE finds that lexical verbs are most common in FICT. Yet, they occur very often in CONV, too (pp. 358-359). According to the semantic domains lexical verbs belong to, activity verbs are most common in CONV and NEWS (as well as FICT). Mental verbs, however, are more often used in CONV than in NEWS. In CONV, the most common verbs in each semantic domain are the following: *get/go/come* (equally frequent), *say, see/know/think/want/mean* (equally frequent), *let, happen, look*, and *start/keep* (equally frequent). These reflect the typical communicative purposes of the register: talking about what people have done, what they think or feel, or what they said. And, since in conversation speakers operate on-line, they tend to rely on relatively few verbs, repeating them frequently. CONV is special in that the most common lexical verbs in the whole LSWE Corpus (*say, get, go, know, think, see, make, come, take, want, give* and *mean* – in decreasing order of frequency) occur much more often in it than in other registers: they account for nearly 45% of all lexical verbs in CONV (pp. 360-373). CONV also differs from the other text types in that, while *say* is most common in the past tense in all registers, CONV commonly uses it in the present tense as well (35% of the time). This happens because *say* in CONV often reports past speech as if in the present, with an effect of immediacy and personal involvement (pp. 374-375).

In CONV (and FICT), the large majority of verbs, irrespective of semantic domains, occur with animate subjects. The core meaning of activity, communication and mental verbs usually entail human subjects, and these registers rely heavily on these core meanings (pp. 378-379).

4.3.2 Variations in the verb phrase

Tense and modality:

CONV is similar to other registers in that the majority of the verb phrases it uses are marked for tense rather than modality. Nevertheless, it does differ from the rest of the text types in that it shows a strong preference for present tense forms. (ACAD is similar, but FICT shows the opposite pattern, with a strong preference for the past.) However, the two 'present-oriented' registers differ in the reasons for which they prefer present tense verbs. In CONV, the reliance on present tense reflects the speakers' general focus on the immediate context, while, in ACAD, the present tense is used to convey the idea that the propositions are true, regardless of time (pp. 456-458).

Aspect:

The rate of simple verb phrases and verb phrases marked for aspect is similar in CONV to the overall findings: about 90 per cent of all verb phrases are neither perfect, nor progressive. But, while overall the progressive aspect is slightly less common than the perfective one, in CONV this relationship is reversed (pp. 460-462). LGSWE does not comment on this phenomenon, but it seems logical that the preference for the progressive aspect reflects the interlocutors' pronounced interest in on-going activities and processes, which is in line with the 'here-and-now' orientation of conversations.

The aspectual markedness of verb phrases also varies according to dialects. In American conversation, the progressive aspect is much more common than in British conversation. Also, in American conversation, the progressive is about twice as common as the perfect. The frequency of the perfect aspect in CONV is slightly higher in British than in American English. This finding is in accordance with the general trend in American English to use the nonperfective past tense in contexts where British English favours the present perfect (for example with *yet* or *already*) (pp. 462-463).

Voice:

CONV represents an extreme in its use of passives: only 2 per cent of the verb phrases are in the passive voice, which is the lowest rate across registers. The *get*-passive, which is extremely rare overall, only occurs in CONV, but even here it is very rare (accounting for only 0.1% of all verbs in this register, most commonly occurring in the construction *get married*). LGSWE explains this decreased use of passives in CONV by the human-centred concern of conversations, where the interest mostly lies in people's actions, thoughts and stances, therefore the subject – who is often the speaker – is not demoted (pp. 476-477). Although passives are generally rare in CONV, a few passive verbs are more frequent in CONV than in written expository registers, e.g. the fixed phrase *can't be bothered* (p. 480).

Complex combinations of aspect and voice are rare in CONV (p. 482).

4.3.3 Modal auxiliaries

As has been mentioned in Section 3.3.3, modals and semi-modals are most common in CONV and least common in NEWS. The register differences in the use of semi-modals are particularly striking: semi-modals are five times more common in CONV than they are in the written expository registers. The greater frequency of modals and semi-modals in CONV is understandable given that these forms mostly convey stance-type meanings. The predominance of semi-modals in CONV has a historical explanation: they developed much more recently than central modals. As usual, linguistic novelty establishes itself first in CONV, and then spreads to the written registers. Since modals are most common in CONV, it is not surprising that the most frequent modals in CONV coincide with those overall: *will*, *would* and *can*. There are semi-modals whose relative frequency in CONV is high, but which are virtually non-existent in the written expository registers: *(had) better* and *(have) got to*. *May*, however, is rarely used in CONV and, when it does occur, it typically marks logical possibility rather than permission. For this latter meaning, CONV prefers *can* (pp. 486-492).

It is not very common for modals to co-occur with marked aspect or voice, even less common in CONV. The only modal perfective which is frequent is *must*, marking logical necessity in the past. With the progressive, the obligation/necessity modals (especially the semi-modal *be supposed to*) are relatively common in CONV (pp. 497-500).

Modal+semi-modal sequences are on the whole less common in CONV than in the written registers, even though their semi-modal elements are generally more common in CONV. This could be due to the preference of conversation for shorter, less complex structures. The only complex modal combination that occurs commonly in CONV is the one that combines the two most common semi-modals: *be going to have to* (p. 502).

4.4 Adjectives in conversations

Since the primary role of adjectives is to modify nouns, and nouns are relatively few in CONV, adjectives are also rare. But, while in other registers the attributive role of the adjectives predominates over the predicative function, in CONV, the two roles

occur evenly. The roughly equal frequency of predicative and attributive adjectives in CONV is in keeping with the general reliance on a clausal rather than nominal presentation of information (pp. 505-507). Most adjectives used in CONV are descriptors, with evaluative descriptors having the highest frequency. LGSWE does not interpret this finding, although it seems to be in line with the tendency in CONV to deal with people's feelings about entities and events.

The most common attributive adjectives in CONV are: *big/little* (size descriptors), *new/old* (time descriptors), *good/nice* (evaluative descriptors) and *same* (relational classifier) – all of about the same frequency. They are mostly monosyllabic and simple, consistent with the generally less-complex structures in this register (pp. 511-513). As for predicative adjectives, those most frequent in CONV tend to be evaluative and emotive. These are: *able, sure, right, good and nice, followed by true, wrong, bad, fine and funny. Alright and lovely* are two adjectives which are almost totally restricted to CONV. Unlike many predicative adjectives in other registers, those in CONV typically lack complements. Even *sure, true* and *wrong*, which normally take clausal complements in the written registers, commonly occur without complements in CONV (pp. 516-517).

A syntactic role of adjectives that is most common in CONV (and dialogues in FICT) is that of exclamations: e.g. *Good! I like that.*

Regarding the comparison of adjectives, inflectional comparison (with *-er* and *-est*) is much more common than phrasal comparison (with *more* and *most*). This reflects the generally lower frequency of polysyllabic adjectives. CONV has very few instances with phrasal degree marking, which shows an even greater tendency in CONV to use monosyllabic adjectives. Only in CONV, adjectives are occasionally doubly marked for degree (another nonstandard feature): e.g. *It's much more warmer in there* or *This is the bestest one you can read* (pp. 524-525).

Derived adjectives are relatively rare in CONV. Here, as in all other registers, the most productive suffix is *-al*. Adjectival compounds are rare in CONV but, when they do occur, they are mostly of the reduplicative type: e.g. *super-duper, goody-goody* (pp. 531-535).

4.5 Adverbs in conversations

Adverbs are common in all registers, and especially frequent in CONV (and FICT). This is due to the fact that adverbs most commonly occur as clause elements, together with lexical verbs, adding information to the relatively short, and therefore frequent, clauses of CONV (and FICT) (p. 504). The patterns of adverb use in the LSWE Corpus differ significantly in two registers, one end-point of the scale being CONV, and the other ACAD. Here, the main focus will be on CONV.

The two most common syntactic uses that adverbs are put to in CONV are those of modifying adjectives and adverbs. In the first case, the most common collocations are: *pretty good*, *really good* in American conversation, and *very good*, *very nice* in the British variety. These modifiers belong to the same semantic field: degree. CONV does not show a great diversity in adverb + adjective combinations, which is typical of this register's less varied word choice in general. Adverbs modifying other adverbs are less common, and most of them show a concern for amounts or comparison: e.g. *very much*, *much better*. There is one syntactic phenomenon involving adverbs which is mostly to be found in CONV: adverbs standing alone as structurally unconnected elements (e.g. *Getting there. Slowly but surely.*) or even complete utterances, for example, questions (e.g. *You can still vote if you lost it – Really?*) (pp. 544-551).

LGSWE divides adverbs into the following main semantic categories:

- Place: e.g. *there*, *backward*, *far*;
 - Time: e.g. *now* (position), *often* (frequency), *always* (duration), *already* (relationship);
 - Manner: e.g. *happily*, *fast*, *well*;
 - Degree: e.g. *very* (intensifier), *somewhat* (downtoner);
 - Additive/restrictive: e.g. *too*, *also*, *only*;
 - Stance: e.g. *probably/actually/reportedly/mainly/approximately* (epistemic), *unfortunately/curiously* (attitude), *frankly/simply* (style);
 - Linking: e.g. *first/secondly* (enumeration), *altogether* (summation), *namely* (apposition), *therefore/thus* (result), *however* (contrast), *incidentally* (transition).
- (pp. 552-559)

In CONV, the majority of adverbs fall into three of the above semantic domains (with the most common ones of each in brackets): time (*now, then*), degree (*very, really, too, quite*) and stance (*really*). This is not surprising, since the meanings of time-adverbs are deictic, which corresponds to the situation-relatedness of CONV, while the others reflect opinions, which also represent a common topic of conversations. The degree adverbs CONV uses are generally informal ones not used in the written expository registers: e.g. *bloody, damn, terribly, incredibly, real* (cf. *relatively, fairly, slightly common* in ACAD.) (pp. 560-569).

4.6 Clause grammar in conversations

In Section 3.6 we have seen that a typical conversation text contains a good deal of non-clausal material and shorter clauses, while in a typical news article all phrases are integrated in clauses, which tend to be longer than in an informal talk. Further, conversation also shows less embedding. If there are embedded clauses, however, these can be nominal (complement) or adverbial clauses.

4.6.1 Complement clauses in conversations

That-clauses:

Overall, more than 80% of *that*-clauses occur in post-predicate position controlled by a verb. These types of clauses are most common in CONV, where the most frequent verbs controlling *that*-clauses are those which are most common overall: *think, say* and *know*. The extremely high frequency of the verb *think* with *that*-clauses in CONV is largely due to the use of the clause *I think* to report one's own personal thoughts (pp. 668-669). These clauses are used most commonly to report indirect speech or mental states, using a personal, human noun phrase as a subject. These features make verb-controlled post-predicate *that*-clauses particularly suited to conversation (p. 675).

Although generally post-predicate *that*-clauses controlled by adjectival predicates are much less common in the LSWE Corpus than those controlled by verbs, when they do occur, they are most common in CONV. The most frequent expression controlling these clauses is *be sure*. Extraposed *that*-clauses are rare in CONV (and FICT), while *that*-clauses in subject position are rare in all registers (pp. 670-675).

In CONV, the omission of the complementiser *that* is the norm. Its retention, however, when it does take place, is encouraged by a number of factors, the most important of which are: the use of coordinated *that*-clauses, and the use of passive voice in the main clause. These factors are conducive to *that*-retention in all registers, but mainly in CONV (pp. 680-682).

Wh-clauses:

Wh-clauses are more common in CONV than in other registers. Post-predicate *wh*-clauses usually follow a verb or a verb + a noun phrase. The most common verb controlling a *wh*-clause is overwhelmingly *know*, followed by *see*, *tell*+NP, *wonder*, *ask* and *understand*, plus, in CONV, *do*. Out of the interrogative clauses starting with *whether* or *if*, the second variant is more common in CONV than in other registers, while *whether*-clauses are fairly evenly spread. With *if*-clauses, *see* tends to be associated with the addressee's state of mind, while *wonder* and *know* with the speaker's mental state: e.g. *Dad, try this on and see if it fits* vs. *Oh I wonder if they will I wonder if they will close our school down...* (pp. 688-692).

Infinitive clauses:

Infinitive clauses can occur in either post-predicate or subject position, and in a variety of other structures, but they are by far most common in the post-predicate position in all text types examined by LGSWE (p. 693). Out of the four registers, CONV shows the lowest rate of *to*-clauses. The only position in which infinitive clauses can be considered common in CONV is the post-predicative one complementing a verb, the most frequent of which is *want*, occurring almost five times more often than the next most common verbs, *try* and *like* (p.711). Infinitive clauses in this position are typically used to report the activities, desires and other thoughts or emotions of human participants; therefore, they are common in CONV and NEWS, which focus on such concerns (p. 722).

Another position for infinitives relevant to CONV is the one complementing an adjective. The most productive adjective controlling infinitives in the LSWE Corpus is *(un)likely*, followed by others from different semantic categories: *(un)able/determined/ready/ (un)willing* (ability, willingness), *free/glad* (personal affective stance), and *difficult/easy/hard* (ease or difficulty).

There is a special alternation with infinitives which occurs only with the verb *try*: *try+to+verb* vs. *try+and+verb* (only when *try* is in its base form). While in all registers *try+to+verb* is much more common than *try+and+verb*, in CONV the difference in frequency between the two structures is not so pronounced. This is related to the colloquial character of *try+and+verb*. In the written expository registers, this construction is mainly used to avoid repetition when *try* appears in a *to*-infinitive form, too: e.g. *He had practised putting on his kitchen floor at home during the winter to try and prepare himself for the greens* (pp. 738-739).

Ing-clauses:

Ing-clauses in a post-predicative position are least common in CONV, but when they do occur, they are normally controlled by verbs. But, while in FICT, for example, there is a whole range of different verbs taking *ing*-clauses, in CONV these clauses most commonly follow *keep*, *start* and *stop*. *Keep+ing* is a kind of progressive marker, emphasising that the action described in the *ing*-clause is continuous or recurrent, while *start* and *stop* denote the beginning or end of an activity. LGSWE does not comment on its findings at this point, but it seems typical of CONV to use such verbs, since it often deals with the narration of actions and events.

4.6.2 Adverbials in conversations

Adverbials are very common in all four registers, being most common in FICT and least in NEWS and ACAD. CONV falls between the two extremes. Circumstance adverbials are by far the most common class in all four registers. Stance adverbials are more common in CONV than the other registers, although they account for less than 10% of all adverbials in that register. This higher frequency of stance adverbials relates to the personal, interactive nature of this register (pp. 765-766).

Circumstance adverbials:

As we have seen in Section 3.6.2, in all four registers, place, time, process and contingency are the four most common semantic categories, however, their order of frequency varies across registers. In CONV, time and place are almost equally the most common, which is not surprising given that CONV is concerned to a large extent with the actions of interlocutors or other people, and place and time adverbials provide

important information about the situations these actions are performed in. Within the category of process adverbials, manner is by far the most common subcategory in all four registers. Within the category of contingency, cause/reason, purpose and condition are the most common subcategories in all four registers, but the proportional use varies across registers. In CONV, cause/reason and condition dominate. This is also consistent with CONV's concern for participants' actions, the reasons and conditions of which are important aspects to be considered (pp. 783-786).

Most semantic roles favour non-clausal syntactic realisations (except for contingency adverbials). In CONV, prepositional phrases are mainly used for place information, adverbs for time and addition/restriction, and noun phrases predominantly for time adverbials. While prepositional phrases are the most common structural forms for circumstance adverbials in all registers, CONV has twice as many single adverbs and many more noun phrases used as circumstance adverbials than, for example, ACAD. This higher frequency of adverbs and noun phrases reflect CONV's preference for shorter adverbials, since single adverbs are only one word, and noun phrases average two words in length. Furthermore, single adverbs and noun phrases are often expressions which are understood from the context (e.g. *here*, *last time*) or vague expressions (e.g. *a little*, *a bit*), both characteristic of conversational discourse (pp. 788-790).

This preference of CONV for shorter syntactic forms is also apparent from a look at the average length of prepositional phrases across registers. In CONV, about 44% of the prepositional phrases used as circumstance adverbials are 2 words, and 34% are 3 words long. 12.5% are made up of 4, 3% of 5 and 6.5% of more than 6 words. This means that, in almost half of the prepositional phrases, the complement of the preposition is either a pronoun or a noun with no modifier (e.g. *for you*, *to college*). This is in line with the real-time production constraints of CONV, since lengthy prepositional phrases would be difficult for hearers to process (pp. 791-792).

The most common circumstance adverbials in CONV are: *just*, *there*, *then*, *here* and *now* (single adverbs), and *this morning* (noun phrases). *Just* is more common in American conversation than in British informal talk, the rest are roughly equally frequent. These adverbials correspond to distinctive characteristics of conversational discourse. *Just* is used in various functions: it fulfils its primary semantic purpose of restriction, but it can also focus on parts of the clause (e.g. *It's just crazy!*), or soften

what it is being said (*I'm just correcting her!*) The other extremely often used adverbials are used for deictic reference to the situation of the utterance (pp. 796-799).

When circumstance adverbials are realised in the form of a clause, in CONV, condition, reason and time are particularly common categories. Besides their common use to mark the conditions on the truth of what conversants are saying, conditionals can also be used when giving commands or making suggestions: e.g. *You can hold her if you want*, where they mitigate the force of the order or suggestion, making it (seem) the hearer's choice. The high frequency of reason adverbial clauses is not surprising given the proportion of conversational discourse spent explaining thoughts, feelings and actions. Further, since a major focus of conversation is who will do what when, time adverbial clauses are also very common in CONV (pp. 821-822). The syntactic realisation of circumstance adverbial clauses in CONV is the following: about 88% of them are finite clauses, 11% *to*-clauses, the remaining 1% is divided between *ing*- and *ed*-clauses, while verbless clauses are virtually non-existent (p. 826).

As for subordinators, CONV shows a marked tendency to use just one subordinator very frequently for each semantic category of adverbials, the most common being *when* for time, *because* for reason, and *if* for condition. Concessive clauses are uncommon in CONV, but when they do appear, they use *though* slightly more often than *although*, which has a formal tone (thus fitting the style of ACAD). *Though* can also be used as a linking adverbial, marking a contrastive relationship between two units of discourse: e.g. *You just have to try and accept it, I guess. – It's kind of hard sometimes, though, isn't it?* In fact, most occurrences of *though* in CONV are as a linking adverbial. Out of the subordinators with multiple semantic roles, *as* is used mostly for manner (and not reason or time), *since* mainly for time (and not reason), and *while*, again, for time (and not concession/contrast) in CONV (pp. 841-850).

In CONV, marking conditional clauses with subject-operator inversion rather than the subordinator only occasionally appears. This is related to the lack of time for planning (pp. 852-853).

Stance adverbials:

CONV contains by far the highest frequency of stance adverbials. This is consistent with the general communicative characteristics of conversation: the focus on interpersonal interactions and the conveying of subjective information. Out of the three types presented in Section 3.6.2, epistemic adverbials are the most common, but CONV

also uses more style adverbials than the other registers. Attitude adverbials are less common in CONV (and FICT) than in NEWS (and ACAD). This might seem surprising, since speakers in CONV are certainly concerned with expressing their attitudes and evaluations. The solution may lie in CONV's particularly high frequency of using *that*-clauses following verbs and adjectives with evaluative or emotional meanings. Thus, it seems that CONV tends to use a clausal means of expression rather than the more condensed form afforded by stance adverbials (pp. 859-860).

Out of the possible syntactic realisations of stance adverbials, like all registers, CONV prefers single adverbs but, unlike NEWS (and ACAD), CONV (and FICT) prefers finite clauses over prepositional phrases in the rest of the cases. This is largely due to the use of comment clauses, e.g. *I mean, I think, I suppose, I bet*, used explicitly to mark a proposition as the speaker's opinion, or to convey some level of personal doubt/certainty (pp. 862-865).

The most common stance adverbials in all registers are *of course, perhaps* and *probably*, but CONV also shows a particularly high frequency of adverbials marking actuality (*actually, really*) and imprecision (*sort of, like* and *kind of*; out of which *sort of* is preferred by British, and *kind of* and *like* by American conversation). The most commonly used stance adverbials correspond to one of their most important functions: to show the doubt of certainty of a proposition (pp. 867-868). *Like* in CONV can be used in other functions as well: as an approximator (e.g. *I looked for like three weeks*) or to introduce indirect speech (e.g. *We're like I wonder what he meant by that?*) (p. 871).

Linking adverbials:

LGSWE finds it surprising that linking adverbials, functioning as the statement of relationships between two units of discourse, are more common in CONV than FICT or NEWS. In large part, this higher frequency is due to the relatively high frequency of *so* and *then* as result/inference linking adverbials. This semantic group is responsible for ca. 70% of all occurrences of linking adverbials. The next most common group is contrast/concession with 26%, followed by transition and enumeration with 2% each. The higher proportion of contrast/concession is similarly due to a high frequency of two items: *anyway* and *though* (pp. 880-883).

The four most common linking adverbials play an important role in the development of conversational discourse. *So* is often used in narrative accounts to move the story along, making clear how an event follows from another. *Then* is often used

when one participant sums up an inference based on another speaker's utterance. *Though* generally marks contrast. *Anyway* is often used by speakers as they move to their main point, marking the preceding discourse as less important. With *so* and *anyway*, the functions of linking adverbials become closely connected with those of discourse markers, which have interactional functions but lose the lexical content of adverbials (pp. 884-889).

Conversations are further analysed in LGSWE from perspectives such as lexical bundles, idiomatic expressions, coordinated binomial phrases etc. Here, these topics will not be touched upon, because they do not offer the possibility for CONV to be compared to NEWS. Biber et al. also examine conversations from the point of view of performance phenomena, e.g. disfluency and error, and they attempt to work out a special grammar of conversation. Because of their highly register-specific nature, these aspects will also be ignored in the present discussion.

4.7 Summary

Just like in the case of printed news (Section 3.7), the grammatical features of conversations identified by Biber et al. (1999) are also connected with *what kind* of information is typically communicated, and *how* it is conveyed. Two additional factors that influence the characteristics of conversations are related to the real-life situations in which conversations normally take place.

When people converse, they regularly talk about on-going processes, and actions and characteristics of their own or of third parties. The colloquy often involves a personal viewpoint and an evaluative stance. The grammatical features related to this aspect are the following:

- a high frequency of (mostly activity and mental) verbs, modals and adverbs;
- circumstance adverbials denoting time, place, manner, cause and condition;
- a preference for the progressive in verb phrases marked for aspect;
- adjectives derived from many different semantic classes (primarily size, age and affective evaluation); and
- a high rate of stance adverbs and adverbials.

The second group of features is related to a tendency of conversations to keep stretches of talk short and simple, and to present information in a more linear manner. These features are:

- the lowest lexical density across registers;
- a low frequency of passive constructions;
- adjectives both in attributive and predicative functions;
- more but shorter clauses, and mainly simple sentences;
- a low rate of nonfinite clauses; and
- a common use of linking adverbials.

The last four features point to a preference in conversations for a clausal, rather than nominal, presentation of information.

The third group of features is connected with the interpersonal focus and situationality of conversations, namely, that colloquies normally take place in a shared situation, the entities talked about are often directly present and the referents mutually known by the participants. The features that belong to this group are:

- a high frequency of pronouns and deictic expressions;
- verbs preferably used in the present tense; and
- a low degree of noun modification.

Finally, the features connected with the online production of conversational talk are:

- a high degree of non-clausal material; and
- the occurrence of non-standard grammatical forms (such as double comparison of adjectives or relative clauses with no gap).

Tables 4.1 and 4.2 sum up the main findings from LGSWE on printed news and conversational discourse in a parallel manner. Comparative forms (e.g. *higher*) refer to the relationship between the two registers; superlative forms (e.g. *highest* - also

signalled by bold letters) indicate that the text type represents an extreme in a particular respect compared to the other three registers examined by LGSWE.

FREQUENCY OF...	NEWS	CONVERSATIONS
NOMINAL ELEMENTS	80%	55%
NOUNS	highest	lowest
VERBS	lower (3-4 nouns/verb)	highest (1 noun/verb)
PRONOUNS	low	highest
DETERMINERS	highest	low
PREPOSITIONS	highest	low
AUXILIARIES	low	highest
MODIFIED NOUN PHRASES	55%	15%
(SEMI-)MODALS	lowest	highest
ADJECTIVES	higher, mainly attributive	lower, equally attributive and predicative
SUPERLATIVES	highest	low
ADVERBS	lower	highest
FINITE SUBCLAUSES:	lowest	highest
<i>THAT</i> -clauses	high	highest
<i>WH</i> -clauses	low	higher
Circumstance adverbial clauses	58%	88%
NONFINITE SUBCLAUSES:	highest	lowest
<i>TO</i> -clauses	high	lowest
<i>ING</i> -clauses	higher	lowest
SUBORDINATORS	lower	higher

Table 4.1 A contrastive presentation of the rate of certain grammatical elements in printed news and conversations

FEATURES	NEWS	CONVERSATION
REPETITION	lowest	highest
LEXICAL DENSITY	highest	lowest
TENSE PREFERENCE	present and past	present
ASPECT PREFERENCE (IF NOT SIMPLE)	perfect	progressive
RATE OF PASSIVE VOICE	relatively high (15%)	lowest (2%)
LENGTH OF PREPOSITIONAL PHRASES	more than 3 words in 60%	less than 3 words in 78%
PRESENTATION OF INFORMATION	mainly nominal	mainly clausal
PHRASE-LEVEL COMPLEXITY	more characteristic	less characteristic
NON-CLAUSAL MATERIAL	not characteristic	highly characteristic
SINGLE-CLAUSE SENTENCES	not characteristic	more characteristic
DEGREE OF EMBEDDING	high	low
AVERAGE NUMBER OF WORDS/CLAUSE*	20	6

Table 4.2 A contrastive presentation of certain grammatical features in printed news and conversations

5 The textual characteristics of printed news

This chapter will give an overview of two researchers' ground-breaking and wide-spread contributions to the media literature, which will serve as a background and frame of reference to the textual analysis of radio news presented in Chapter 8. The first is van Dijk, whose work is here represented by three pieces of research: van Dijk 1985a is a straightforward introduction to the theoretical framework for the study of the printed press later elaborated in van Dijk 1988a. Van Dijk 1988b complements the theoretical work with case studies as exemplifications of the approach. The second is Bell, who, in his 1991 book, gives a comprehensive analysis of the way of news, that is, process-oriented examinations of what and how a piece of information makes it into the news, and a product-oriented analysis of what the final structure of the news is like. Bell 1998 further refines the picture of the make-up of news, supplementing it with a consideration of one-sentence stories.

Van Dijk (1988a:vii) sees his approach as a new, interdisciplinary theory of news in the press, which represents a qualitative alternative to traditional methods of content analysis. Besides the usual linguistic, grammatical analysis of news language, this approach goes on to examine concepts such as the topic and schema of news and, finally, its stylistic and rhetorical structures. The approach is interdisciplinary because, besides analysing the linguistic aspects of news discourse, it also attends to the complex relationship between news text and context (thus sharing the aims of discourse analysis in general). Van Dijk (1988a:2) calls his method 'sociocognitive' because, on the one hand, he focuses on the cognitive processing of news both in production and understanding and, on the other, he makes explicit the well-known role of news values and ideologies, thus building a bridge between the psychological and sociological studies of news.

Bell's manual is basically concerned with three themes (1991:8): the importance of the processes which produce media language, the notion of the news story, and the role of the media audience. Bell's approach to the make-up of news stories is characterised in Garrett and Bell (1998:4) as 'structural discourse analysis', because it is mainly concerned with the elements that build up a news story, starting off from a comparison of this journalistic genre with conversational narratives. Bell 1998 demonstrates that the structure identified in Bell 1991 is valid even for one-sentence

news stories, and complements the earlier methodology with a guide to revealing the ‘event structure’ of press news, with special attention to the temporal arrangement.

Since the present study has a product and not a process orientation, here, the news production and comprehension part of van Dijk’s manual will only be touched upon, with the main focus on the framework for the textual structure of news (Section 5.1). Similarly, the news production and reception parts of Bell’s work will not be concentrated on; the stress will be laid on the story structure (Section 5.2). The aim of the review is to outline a method which then can be applied to radio news in Chapter 8.

5.1 The sociocognitive approach

5.1.1 The thematic structure of printed news

The basic notion in a thematic analysis is that of ‘topic’, which for van Dijk (1988a:31) means “the summary, gist, upshot, or most important information” of a stretch of talk or text. Since topic is not a property of words or sentences but of larger linguistic units, this theoretical notion belongs to the global, macrolevel of discourse description.

The meaning of a stretch of discourse can be broken down to ‘propositions’, which, in van Dijk’s terms, are “the smallest, independent meaning constructs of language and thought” and, along the referential dimension, “also the smallest semantic units that can be true or false”. A proposition is made up of a predicate and one or more arguments, which may denote things, persons, or events. A discourse consists of a set of lower-level propositions – which roughly correspond to clauses –, and higher-level propositions, which are called ‘macropropositions’ (van Dijk 1988a:32), and can be derived from lower level ones by so-called ‘macrorules’. Macrorules are semantic mappings or transformations which we may intuitively call summarising. They are of three kinds:

- deletion, by which we leave out all the information that is no longer relevant;
- generalisation, by which we replace a sequence of propositions with a single, more general macroproposition; and
- construction, by which we replace usual components of an act with a single macroproposition that catches the act as a whole.

These three rules reduce information of a text to its topics. Macrorules are recursive, that is, they can apply again at higher levels, until we reach one or two macropropositions that resume the text as a whole. This way we get to the overall macrostructure or thematic organisation of a text, and grasp its global coherence as well, since topics guarantee that a text or talk has semantic unity.

Based on some previous research on strategies of discourse comprehension (see van Dijk and Kintsch 1983), van Dijk (1988a:33-34) draws attention to the fact that meanings are assigned to texts in processes of interpretation which may be subjective, therefore, topics too can be subjective, or dependent on individual interest and relevancies. Further, the cognitive assignment of topics does not wait until the reader has read the whole text, rather, (s)he starts to make expedient guesses about the most probable topic(s) of a text at each stage of reading. Later on, the assigned topic acts as a major control instance on the further interpretation of the rest of the text. Finally, macrorules do not only have a subjective side, but also an intersubjective one in that they require generalised knowledge of the world, such as frames or scripts¹⁷.

In short, van Dijk (1988a:34) sees a topic as a “strategically derived subjective macroproposition, which is assigned to sequences of propositions by macroprocesses on the basis of general world knowledge and personal beliefs and interests”. Such a topic is part of a hierarchical thematic structure of a text, expressing the gist or upshot of the discourse.

Van Dijk (1988a:35-41) then goes on to illustrate, test and refine the above theory of macrostructures for the specific text type of news in the press, and notes the first important feature of newspaper discourse: Topics may be signalled by headlines and leads, which routinely express the major actors involved, the macroaction or event, and possibly the location, or the main issue involved. However, van Dijk warns (1985:75), there is not just one topic or possible summary of a text, but several. Summarisation, then, may take place on a continuum, from leaving out a few less essential details on the one end, until leaving out all information except the most relevant, on the other, expressed by headlines. But since headlines are macrostructural from the point of view of the journalist, they may also be subjective, and indicate a preferred overall meaning of the text. And, because readers try to derive a provisional

¹⁷ Or ‘Idealized Cognitive Models’, as called in cognitive linguistics, see G. Lakoff 1987.

topic from the text as soon as possible, headlines and leads are important textual devices in journalists' hands to signal the assumed macrostructure of the news text.

Another conspicuous and typical feature of topic realisation in news discourse is its 'instalment' character (van Dijk 1988a:43-44), that is, each topic is delivered in parts, not as a whole. This is caused by the 'top-down principle' of relevance organisation in news, namely, that the most important information is put in the most prominent position, both in the text as a whole and in the individual sentences. This way, for each topic, the most important information is presented first. When the important information of all topics has been expressed, earlier topics are reintroduced with lower-level details. This top-down organisation allows editors to cut the final paragraphs of a news story any time if needed, without the loss of essential information. Van Dijk (1985a:82) proposes a hypothetical model of the strategy of news discourse production, as can be inferred from the structure of the product. The steps of the production process are the following:

- Activate the model of the actual situation (from earlier knowledge).
- Derive an overall thematic structure from this situation model.
- Decide which of the main themes are most relevant or important, given a system of news values.
- Start actual production by expressing the main theme as a headline and the rest of the top structure of themes as the lead of the news text.
- Main themes about main events are at a next lower level of macrostructure, formulated as the first sentences/paragraphs of the text.
- Each next paragraph deals with the next lower topic, according to the following writing strategies: important consequences come first; details of events or actors come after overall mentioning of events or persons; causes or conditions of events are mentioned after the event and its consequences; context and background information come last.

Van Dijk (1988a:43-44) suggests that the understanding and the writing of a news text are in fact similar processes, only in a reversed order: making sense of an article takes place along the lines of summarisation (macrostrategies) outlined at the

beginning of this section; realising a news text takes place by the application of inverse macrorules (which he calls ‘specification rules’), as outlined above.

Van Dijk (1985a:82-83) notes that these cognitive production strategies are rather different from the production of other stories. Everyday conversational stories, for example, follow a left to right principle: from causes, conditions, circumstances or setting to the actions of events themselves, with the results, consequences last. (News stories in popular mass newspapers in fact often apply this narrative organisation.) Typical news stories, however, are based on the ‘actuality’ principle (1985a:78), which means that last main events and consequences are more important and appear in a more front position in the news text than earlier events or causes.

Van Dijk (1988a:48) concludes that news discourse exhibits a thematical realisation structure that is summarised in the headline and lead, and then elaborated in a basically top-down, relevance/actuality controlled, and cyclical manner (in instalments).

5.1.2 The schematic structure of printed news

While the thematic structure is a formal representation of the global content of a news text, the schematic structure describes the overall form of a discourse. Van Dijk (1985a:69, 1988a:48) proposes the term ‘superstructure’ to refer to these fixed, conventional schemas, which organise thematic macrostructures much in the same way as the syntax of a sentence organises its meaning. Superstructures are made up of different categories typical (at least in part) of news discourse, which correspond to a specific sequence of propositions/sentences of the text (1985a:85-86). This overall syntax (1988a:49) defines the possible forms in which topics can be inserted and ordered in an actual text.

Figure 5.1 is a reproduction of van Dijk’s schema for news superstructures:

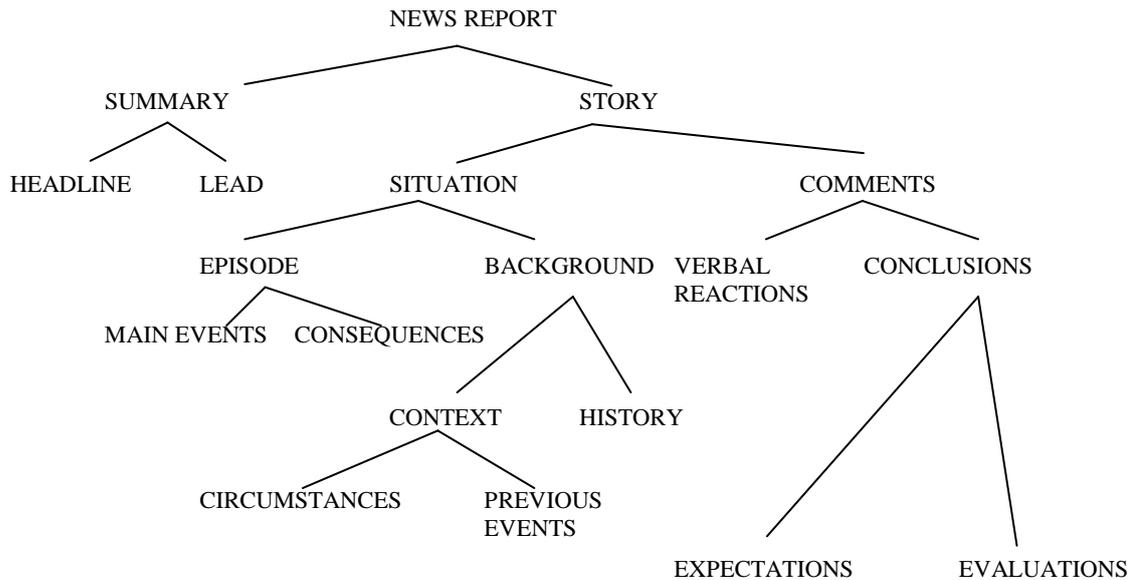


Figure 5.1 Hypothetical structure of a news schema (after van Dijk 1988a:55)

The above schema is a slightly modified version of van Dijk 1985a, where Background is dominated by Events, and Verbal reactions is placed under the node Consequences (1985a:86); here, they both occupy a more prominent position. The change must have been made to allow for Consequences to come earlier in the text than Background, since this ordering is probably empirically more grounded.

Figure 5.1 expresses the canonical order of news schema categories from left to right: **Headline**→ **Lead**→ **Main Events**→ **Consequences**→ **Circumstances**→ **Previous Events**→ **History**→ **Verbal Reactions**→ **Expectations**→ **Evaluations**. Our discussion, too, will follow this order:

The elementary rule is that **Headline** precedes **Lead**, and they both precede the rest of the news text. Together they express the major topics of the text, this is why van Dijk (1988a:53) introduces the category **Summary** to control both. Their distinguished character is indicated by graphic layout as well: they are printed in bold type and, possibly, across several columns.

Episode comprises the **Main Events** and **Consequences** of the reported happenings, often embedded in **Context** and **Background**. (There is a slight – and otherwise unimportant – imprecision in terminology here: it seems that van Dijk altered his 1985a structure, but in the 1988a explanation of the schema he still slides into the

earlier denominations.) As shown by its name, Main Events deals with the most important, current events of a situation, while Consequences is vital in that it gives causal coherence to news events. Sometimes, van Dijk (1988a:54) observes, Consequences are even more important than the main events themselves; in such cases the category may even move up into the headline.

Circumstances is often signalled by words such as “while” or “during” to express simultaneity. Previous Events reminds the reader of what has happened before (and what was probably reported earlier in the same newspaper). By History, van Dijk (*loc cit*) only means parts of a news text which deal with nonrecent past history of actual situations. In actual cases, van Dijk adds, it may sometimes be difficult to distinguish between Circumstances, Previous Events, and History. However, grammar may help out in such cases: the differences are often marked by different verbs, verb tenses, or temporal adverbs.

Verbal Reactions is a specific news schema category that may be seen as a special case of Consequences (as, indeed, appears in the 1985a schema). It is a common journalistic procedure to ask for the comments of important participants or prominent personalities; these views are reported in Verbal Reactions. The category is signalled by names and roles of news participants and by direct or indirect quotes of verbal utterances. The category routinely appears towards the end of the news text.

The category Conclusions contains the opinions of the newspaper staff itself. Van Dijk (1988a:56) warns that, although newsmakers share the ideological view that fact and opinion should not be mixed, this final Comments category often appears in the news, albeit sometimes in an indirect form. The two subcategories are Evaluations, which contains assessments made by the journalist, and Expectations, which comprises possible political or other consequences of the events or future predictions made by the newsmaker.

Van Dijk (1988a:56) concludes by saying that the superstructure represented in Figure 5.1 is theoretical in the sense that all categories are mentioned; in reality, many news texts display only some of these categories. Strictly speaking, only Headline and Main Events are obligatory categories in a minimally well-formed news discourse. Categories such as Backgrounds, Verbal Reactions and Comments are optional; others may be recursive, and even, the same text sequences may have several functions simultaneously.

Moreover, van Dijk adds (1985a:89-90; 1988a:57), the superstructure may undergo transformations, so that categories such as Verbal Reactions may move to higher positions in the tree-diagram.¹⁸ Constraints that may cause such transformations are recency, or relevance: if, for example, the comment of a distinguished personality is perceived to be even more important than the event (s)he comments on. Another problem that may cast doubt upon the validity of the schema is the discontinuous ordering of categories in news texts, so that elements of Main Events may appear separated from each other by parts of text belonging to other categories.

Van Dijk (*loc cit*) accounts for such seeming irregularities by saying that thematic and schematic structures of news texts are abstract, underlying structures, which in actuality may be realised in different ways. An actual headline may consist of several parts, such as a main headline, a superheadline, and a subheadline. Leads may be expressed separately and boldly printed, or may coincide with the first, thematical sentence of the text. Moreover, there may be cultural differences as well: news texts in other languages may display different realisations, but the categories are probably the same.

In sum, the news superstructure conceived by van Dijk proposes two main formal categories in a news text: Summary – made up of Headline and Lead – and Story – consisting of Situation and Comments, in this order. Not all categories are compulsory in a well-formed news article, and the actual realisation of texts may be considerably different from the abstract superstructure, due to constraints such as relevancy or recency.

Van Dijk 1988b contains an illustration of the theory in a concrete case study, which examines the international press coverage of a prominent world event: the assassination of president-elect Bechir Gemayel of Lebanon on September 14, 1982, followed by the invasion of West Beirut by the Israeli army. More than 700 articles from 138 newspapers were selected from a large sample of 250 newspapers from 100 countries, and were systematically analysed, both quantitatively and qualitatively (1988b:31). The extensive amount of empirical work succeeded in verifying the hypothetical structures offered in van Dijk 1988a, and the comparative analysis of the

¹⁸ It is interesting to note that the terms van Dijk uses here – “tree-like form schema”, “well-formed news discourse”, “obligatory” and “optional” elements, “transformations” etc. – remind one of the terminology of generative-transformational grammar. One could speculate about the possible reasons for this usage, such as an underlying wish to show that text linguistics or discourse analysis, traditionally regarded as non-mainstream linguistics, can be just as rigorous and formal as Chomskyan grammar.

accounts from many different countries, political systems and cultures suggests that the theory is generalisable.

Section 5.1 has given an outline of van Dijk's model of news in the press. This model proposes a semantic macrostructure of news discourse, and a macrosyntax that gives form to the overall content of the text. We have started out from a presentation of the macrorules that govern readers' understanding of news texts, that is, different ways of summarising information to derive subtopics and the main theme of a discourse. It has been suggested that news writing follows the same steps as understanding, just in a reversed order, in a top-down manner. This gives rise to a canonical system of news texts, represented in the superstructure offered by van Dijk. This superstructure may be altered by social constraints, such as news values, in the form of different principles at work (actuality, relevance) during news production. To close the circle and turn back to readers, van Dijk (1985a:91) suggests that this text scheme actually facilitates comprehension, storage and retrieval from memory. All these aspects considered, van Dijk's model indeed provides the link between psychological (cognitive) and sociological (ideological) studies of news in the press.

5.2 The structural discourse analytical approach

Bell (1991) strongly builds upon van Dijk's textual categories when proposing a model of his own for press stories. What he is also concerned with is the slightly misleading term that journalists use for press news: 'stories'. Bell analyses the ways in which news stories differ from more typical stories, that is, personal narratives.

5.2.1 News stories and personal narratives

Bell's starting point is the observation that press news – although they are called 'stories' – are much different from usual narratives. He takes Labov and Labov & Waletzky's famous analyses of personal story syntax (referred to in Bell 1991:148), and compares the six categories identified by Labov and his colleague to the structure of news stories (op cit:149-155). He finds the following:

The lead has precisely the same function in the news as the Abstract in personal narratives: it initially summarises the central action and establishes the point of the story. However, while Abstract is optional in a personal story, it is obligatory in news.

Orientation, that is, setting the scene, is just as important and compulsory in news as in personal narratives: *who*, *what*, *when* and *where* are basic facts which concentrate at the beginning of a news story, and may be expanded further down.

Evaluation is the means by which the significance of the story is established, the events are focused, and claiming the audience's attention is justified. Press news also contains evaluation but, while in personal narratives evaluative devices are typically concentrated near the end of the piece, in news it can be found at the beginning. The lead and the headline are nuclei of evaluation, because it forms the lens through which the remainder of the story is viewed.

Action, the central part of the story proper, is nonetheless present in news as well. However, while in personal stories the action is invariably told in the order in which it happened, news stories are seldom if ever told in chronological order (a point observed by van Dijk as well). Perceived news value overturns temporal sequence: results and outcomes are told earlier than processes or events, which are communicated prior to causes.

Resolution, the event that finally happened to conclude the sequence of events, is what a personal narrative moves to; news stories often do not present such clear-cut results. When they do, the results will appear in the lead, not at the end. Bell's witty formulation at this point (1991:154) goes like this: "News stories are not rounded off. They finish in mid-air."

In personal narratives, an optional Coda wraps up the action. No such element exists in a news story, because the functions it serves are not shared by news stories: it is not necessary to explicitly mark its finish, to return the floor to conversational partners, or to return the tense from narrative time to the present. Bell observes here (*loc cit*) that the Coda does have some parallel in broadcast news, where the end of a news bulletin is usually signalled by something like "That is the end of the news". In my view, however, since these announcements refer to the whole news programme and not to individual news items, they cannot be viewed as Codas, or else we regard the entire bulletin as one story.

Bell concludes that there are important similarities and differences between personal narratives and news stories, and adds that there is one characteristic that they share: a penchant for direct quotation, since "[t]he flavour of the eyewitness and colour of direct involvement is important to both forms" (1991:155).

5.2.2 The discourse structure of news stories

Bell (1991:169-174) offers a slightly different news structure from the one suggested by van Dijk 1988a. Figure 5.2 reproduces this model:

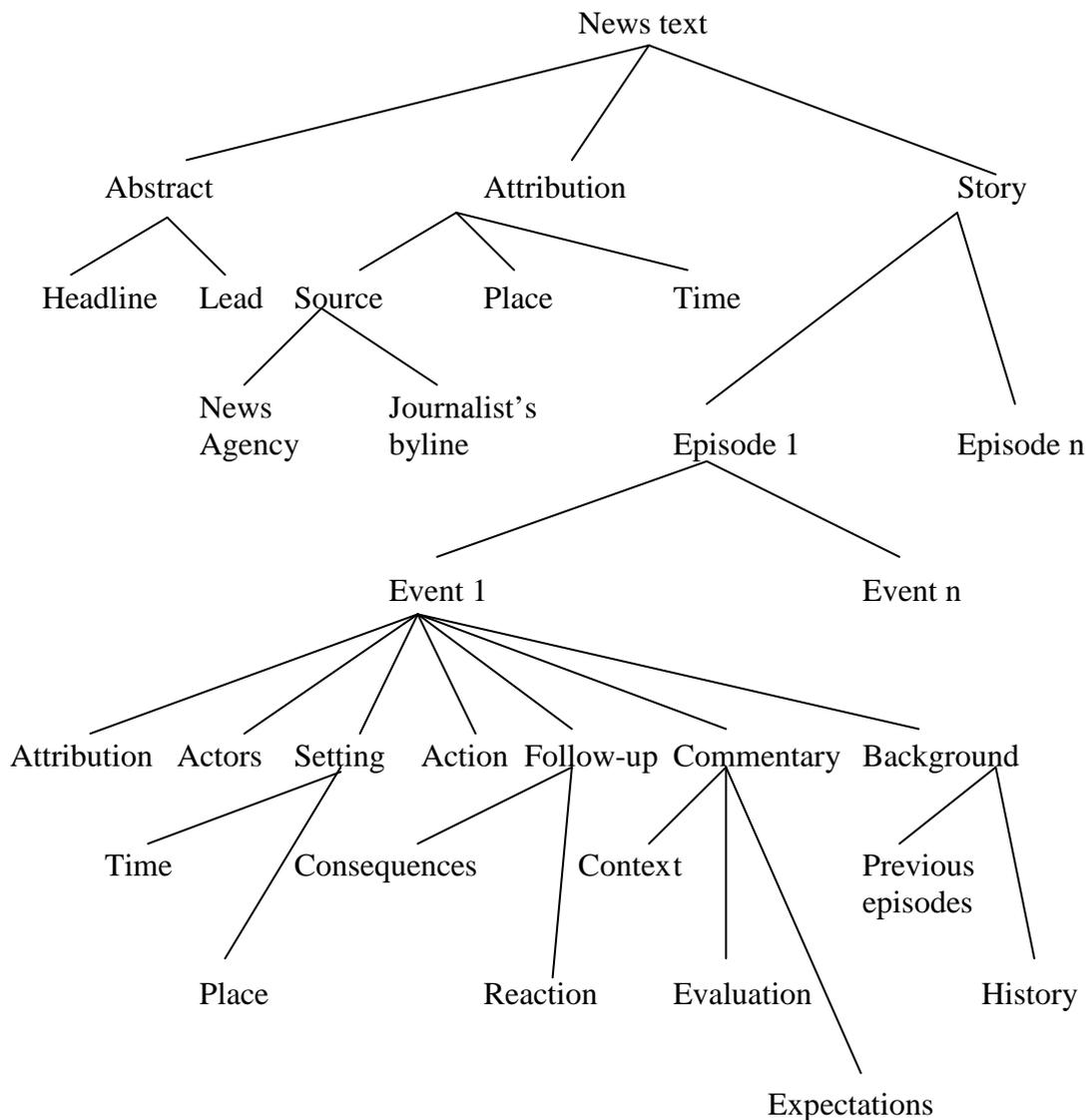


Figure 5.2 Outline model of news texts (after Bell 1991:171)

The news text, in Bell's conception, consists of Abstract, Attribution and the Story proper. The Abstract breaks down into Headline and Lead. Bell (1991:176) thinks that it is the lead, and not the headline, which is the most distinctive feature of news discourse. The lead is a micro-story: it comprises the main actors, the main event and

place of the reported happening. Another function of the lead is to concentrate the news value of the story: e.g. negativity, eliteness or facticity (more on news values can be read in Section 8.3.1). In Bell's (1991:178-179) ten examples, all but two of the leads start with the actor, which reflects the news value of personalisation. Bell (1991:180) points out that, although news writing lore has it that verbs should be active, passives are much more common than the prescription implies, because often this is the only way to get the main news actor at the front of the lead. The lead is a 'directional' (1991:183) summary of the story: a lens through which the point of the story is focused and its news value magnified. Besides being a summary, it must also provide a springboard for telling the whole story. Moreover, leads often consist of more than one event, and may also contain information which never resurfaces in the body of the story. Thus, Bell (1991:184) concludes, it is inadequate to describe the lead as simply a summary of the story; it is also part of the story itself.

Headlines, as Bell (1991:186-189) sees them, are the ultimate in the journalist's drive for summarising information. But, unlike leads, headlines are stand-alone units, they simply abstract the story, they do not have to begin it. Headlines are completely derivable from the story, most often from the lead alone. Headlines generally delete structure from the lead; for example, when a lead contains several events, the headline pares this down to a single main event. Thus, headlines usually contain only main events and no previous episodes, there is one such category that is obligatorily expressed in: performative speech acts. Otherwise, attributions are not usually expressed. Place is sometimes signalled in headlines, but never time. Finally, headlines are not just a summary, but part of news rhetoric as well: they display a distinctive monosyllabic lexicon, and they often use devices such as alliterations, punning etc.

Attribution (1991:169) is where the story comes from: agency credits, journalist's byline, optionally plus time and place: e.g. *Moscow, 31 March, NZPA-Reuter*. Besides attributions for the whole story, the press news regularly contains attributions within the body of the story as well (1991:190), to remind the audience that this is an account which originated with certain persons and organisations. News sources are of two broad kinds: suppliers of information or news actors whose own utterances – announcements, proposals, reactions etc. – have news value. The choice of news sources has an ideological side: news is what an authoritative source tells a journalist; unofficial news sources are little used; and alternative sources, such as

individuals, opposition parties, unions, minorities, fringe groups and the disadvantaged, tend to be totally ignored.

Attribution is made up of naming the source (1991:191-198) and using a speech verb (1991:204-210). Identifying the source can take place in different ways: giving the person's full title, or, if it is too long, using a generalised label on first mention and detailed affiliation on the next. Curiously enough, sources are often unnamed: e.g. *one top-ranking US official*. A common grammatical procedure in naming sources is determiner deletion: e.g. where *the Neighbours star Kylie Minogue* becomes *Neighbours star Kylie Minogue*. Bell notices an important switch in the semantics of such phrases: instead of the first NP being primary, and the second parenthetical, the name becomes the head, and the first, descriptive NP acquires the status of titleness. This way, Bell explains, the quasi-title embodies the claim of the ordinary person or the newly elite to news fame, plus the determiner deletion serves another goal of news writing: brevity.

The canonical neutral speech verb is *say*, along with *tell* and the useful nonfinite form *according to*. When talk itself is news, the verbs include *announce*, *declare*, *threaten*, *refuse* etc. Bell notes that journalists love the performatives of politics or diplomacy, because they constitute indisputable facts, and no other facts have to be verified. Speech verbs, however, can convey the stance of either speaker or reporter to the statement: *claim*, for example, is routinely used for 'noncredible' sources. The statements can be presented in two forms: direct quotations or indirect speech. Out of the two, indirect speech is the norm; direct speech is valued because it is an incontrovertible fact, it may be used to distance and absolve the journalist from endorsement of what the source said, or to add the story the flavour of the newsmaker's own words. Indirect speech is signalled by tense concord following a past tense speech verb.

A Story (1991:169) consists of one or more Episodes, which in turn consist of one or more Events. Events must contain actors and action, usually express setting (and may also have attributions). Episodes (1998:67) are clusters of events which share a common location or set of news actors, and need only be specified when a single story contains two or more clearly distinct sets of events.

Three additional categories (1991:170) can contribute to an event: Follow-up, Commentary and Background. Follow-up covers any action subsequent to the main action of an event. It can include verbal reaction or nonverbal consequences.

Commentary provides the journalist's or news actor's observations on the action. It can give context, explicit evaluation or expectations on how the situation could develop next. Background covers any events prior to the current action. These are classed as Previous Episodes if they are comparatively recent, or History, if they go beyond the near past. Follow-up and Background may have the character of episodes in their own right. Therefore, Episode is a recursive category, and can be embedded under consequences, reaction, history or background.

5.2.3 The event structure of news stories

Bell's (1998:66) most perceptive insight in the analysis of news stories is the deduction of an 'event structure' for a story. This involves reconstructing from the often fragmented information presented what the story says actually happened. It concentrates on the basics of story-telling, which in the news are encapsulated by the journalist's famous 'five Ws and an H': *who, what, when, where, why, and how*. For Bell (1998:94), event structure is the order in which events actually happened, while discourse structure is the order in which they are told in a story. From this it is apparent that, while event structure comprises three main components: actors, place and time (1998:89), time is found to be the most important dimension of news stories, and Bell (1998:93-97) devotes a separate discussion to it.

We have seen in Section 5.2.1 that news stories do not typically report the events in the order they occurred in; the make-up of news stories is dictated by other factors (e.g. news values). As Bell (1991:172) puts it: "In news, order is everything but chronology is nothing." Journalists follow a distinctive 'inverted pyramid' structure (1998:97), which means that the important information must come as early as possible for the story to be able to end at any sentence. This results in a radical discontinuity of time, with the story moving backwards and forwards in chronology, picking out different points or further details on each cycle. In a sample analysis of a single story, Bell (1998:94-96) perspicaciously numbers the time periods of the events, with Time Zero being the story present, minuses used to label times prior to this, and pluses for subsequent times. These labels are indicated one below the other, near every sentence of the news story. Figure 5.3 presents the result (my own illustration – FJ).

As can be seen, the time structure of this – typical – story is very complex, with nine points in time identified. The lead sentence alone covers three event times. Further

on, the story is taking us back in time, presenting the events in reverse order of actual occurrence, plus a couple of excursions into story future time. The earliest events are reported last of all.

Bell (1998:99) summarises his findings by saying that the canonical/chronological order of everyday story-telling is not used in modern news – though it used to be in the nineteenth century –, nor is the backwards pattern used in its pure form of telling an entire story from last event back to first. However, sections of news stories do run events in reverse order.

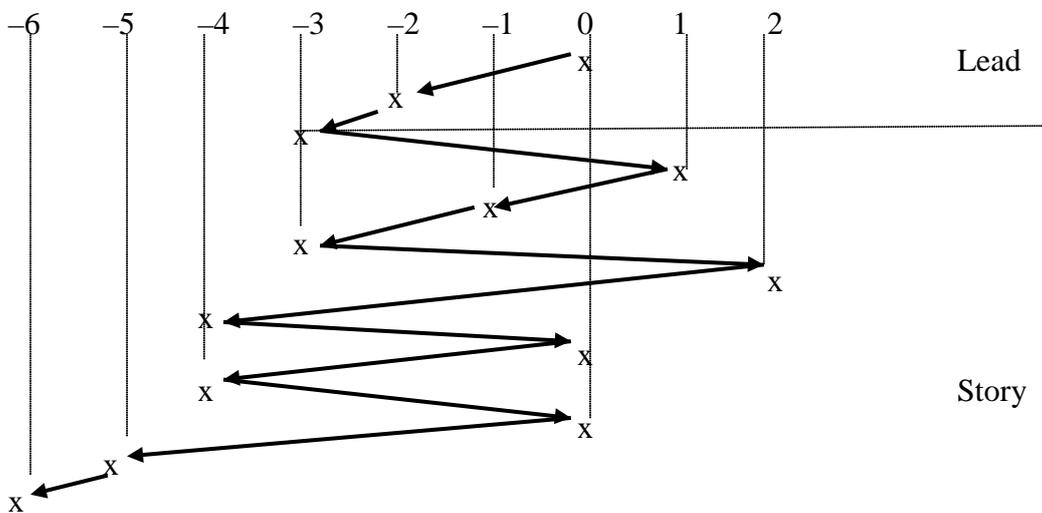


Figure 5.3 The time structure of a news story (analysed in Bell 1998:94-96)

Bell (1998:76-80) complements his framework with a full guide to analysing and interpreting the discourse (and event) structure of a news story. Bell finds it important that this method is suitable for revealing the ambiguities, unclarities or ideological bias of a story as well. Often, headlines are not valid representations of the story as a whole. The identity of news actors is not always clear, and sources are not always directly specified. Temporal and causal relationships between events are sometimes not justified. News stories often remain silent on places or times of the events. A lack of cohesion and specific reference across the story can also lead to confusion.

Bell (1991:174) points out that the minimal well-formed hard news story is just one sentence long. Many newspapers publish one-paragraph stories to fill odd corners or assembled in a column of news briefs. Bell (1998:69-75) provides a detailed analysis

of such single-sentence stories. He finds that as many as five events can be comprised in a one-sentence story, and the atemporal order of narration can be detected here as well. But the general structure seems to be: main event, news actors, place, and generally time. Thus single-sentence stories resemble leads in their structure, which is no surprise, because they probably originated as the lead paragraphs of longer stories from international agencies.

Bell (1998:101-103) sees the significance of his framework in that it can lead one closer to “what the story actually says happened”, it can ‘open up’ stories by revealing a lot about their character, indications of how they were made, manifestations of the news values behind them, and evidence of discrepancies and gaps. He adds that the analytical work involved is not small, but the demand is the true reflection of the real complexity of this journalistic subgenre, which sometimes appears deceptively simple.

5.3 A comparison of the two approaches

Van Dijk’s sociocognitive news schema and Bell’s discourse structure of press news have much in common. Bell (1991:169) finds that the categories van Dijk has identified in news discourse are needed, and sees his own framework as providing a few relabellings and additions. Indeed, what is a Summary in van Dijk’s framework is an Abstract in Bell’s. The straightforward category of Attribution for the whole story is absent from van Dijk’s schema, which van Dijk has probably not considered part of the news text itself. But other, more subtle differences can be detected as well. What is most striking is a formal difference, namely, van Dijk’s endeavour for each category to bifurcate,¹⁹ which has probably been found more elegant than having several branches from the same node; Bell does not seem to bother about this.

As for content, it stands out that Bell’s model specifies elements of events, such as actors, action and setting (time and place), while van Dijk does not go into details at this point. This comes from the fact that van Dijk’s starting points are propositions, and to separate actions from actors and setting would mean going below the level of propositions. This shows that, indeed, Bell’s schema is built upon the ‘five Ws and an H’ of journalistic writing.

¹⁹ This is another similarity to generative-transformational grammar, where all nodes in tree-diagrams tend to have two ‘daughters’.

Another important addition of Bell's to the van Dijk model is the category of Attribution referring to the individual events. This is a necessary category because, clearly, information source – where we find out about the action – is totally different from verbal reaction – an opinion about the action. Another difference is that what is Main Events in van Dijk's model is Event 1 to n in Bell's framework. This is more than just a terminological difference, because the numbering gives the analyst the opportunity to indicate the chronological order of events, while a single group of Main Events does not.

Context, in the Bell model, is part of a journalist's commentary on the action, while in van Dijk's framework it is part of the (background to) the situation, not so heavily attributed to the journalist. This may be an insignificant difference, given that the whole story is in fact the journalist's personal contribution.

The rest of the divergences are minor. Consequences and Reactions are fused in the same Follow-up category in Bell's model, while they are headed by different categories in van Dijk's schema (albeit in van Dijk 1985a they are arranged similarly to the Bell schema). The order of Background and Comments/Commentary are reversed in the two models.

A difference affecting the whole structure is that, in Bell's framework, the entire unit of Attribution+Actors+Action+Setting+Follow-up+Commentary+Background is repeatable while, in van Dijk's schema, the elements are independently recursive. This is a matter of principle; in practice, both models are capable of accounting for the actual news texts.

5.4 Summary

Chapter 5 has been a review of two prominent models of printed news discourse in the literature: van Dijk's sociocognitive and Bell's structural discourse analytical approaches. In the first two subchapters, the frameworks were introduced. In the last part of the chapter, the two approaches were compared. It has been found that they are in great consonance, the main difference being that Bell has complemented van Dijk's model with the category Attribution, for both the story as a whole and the individual pieces of information contained. Another advantage of the Bell model is that, through dividing Main Events into separate Events and numbering them, it can grasp the chronological order of the events, and can better illustrate the discontinuous time-planes

that make up a news story – a characteristic that otherwise has been noticed by van Dijk as well. Overall, while both frameworks are well-grounded, van Dijk's model has a more accentuated theoretical – and Bell's a more practical – character.

6 Research design

This chapter will set forth the research questions that the study is intended to answer, and describe the data collection and processing methods employed. The research questions are of two kinds: quantitative, aimed at finding out about the grammatical features of radio news, and qualitative, meant to discover the textual characteristics of broadcast news. The answers for the quantitative questions are to be found in Chapter 7, while the qualitative questions will be responded to in Chapter 8.

6.1 Quantitative research questions

One of the main concerns of the present study is whether radio news, as shown by the BBC corpus, is less literate in character than printed news, which is evidently a written genre. To find this out, the following research questions are asked:

1. What are the main grammatical features of radio news? Do any of these grammatical features approximate it more to spoken than written language?
2. Is there any grammatical feature in which radio news differs from all other media genres?
3. Do the subtypes of radio news identified in the present study differ in their spoken vs written character?
4. Do the subtypes of radio news identified in the present study differ in quantitative features such as length of items, number of sentences, length of sentences and position?

My hypotheses are that:

- there must be a quantitatively measurable difference between language use in radio news and the press, hence the common instruction given to would-be journalists at news editing courses: to write the news “for the mouth”;
- the subgenres of news must differ in their grammatical features such that the pieces sent by correspondents should display spoken characteristics (e.g. shorter sentences);

- short news must also differ from long items in both the number and the length of sentences, and possibly, in a smaller number of items per edition²⁰;

6.2 Qualitative research questions

Turning to textual aspects of radio news, the following questions are asked:

5. What are the topics of the BBC corpus?
6. Is there any relationship between the position and the topic of a news item?
7. How does the position of (partially or entirely) repeated news items change?
8. What are the textual units of the BBC corpus news? Is the discourse structure suggested by van Dijk (1985a, 1988a,b) and Bell (1991, 1998) for printed news valid for radio news as well?
9. What is the relationship of headlines to news items?
10. What are the transformation techniques of textual units when
 - i. a correspondent's report is later incorporated into a news item?
 - ii. a news item is partially repeated?
 - iii. a long news item becomes a short one?

I hypothesise that:

- the items dealing with the most important topics – in terms of amount of coverage – tend to be placed towards the beginning of the news programme;
- repeated items tend to be placed lower than their first mention since, not being the latest news anymore, they are less prominent as well;
- van Dijk's sociocognitive news schema and Bell's discourse structure are, by and large, suitable for capturing the textual nature of radio news, though to a different extent;
- the most important news items are headlined, and the order of headlines faithfully follows the order of news items;
- when news items later incorporate reports or are partially repeated, the main textual transformation type is deletion;

²⁰ An 'edition' is, for example, a World News programme between, say, 7 and 7:05 am, which is made up

- short news items, as suggested by the name of the programme type they occur in ('news summaries') summarise information of earlier versions.

6.3 The data

Linguistic sampling is a delicate issue. It poses both quality and quantity problems: on the one hand, it is difficult to collect data that would accurately represent a certain community of speakers/writers, and on the other, the analyst faces the problem of how much data from how many sources is neither too little nor too much. In Sebba and Fligelstone's view, perfect representativeness is, in theory, impossible, because any data is only perfectly representative of itself. However, selecting both a large enough and diverse enough sample to obtain a suitable corpus can maximise representativity. Where, Sebba and his associate (1994:769-771) say, the corpus represents what is taken to be a single text type (e.g. newspaper reports or parliamentary proceedings), this may not be problematic.

Leitner (2000) differentiates between General and Special Purpose Corpora, depending on whether they aim at representing a language as such, or just a slice of it. He (2000:175-176) considers it inconceivable to ever design a representative corpus if one understands by that a representation of the language at large, and suggests that one should tie the notion of representativity to that of the language experience of a *selected* group of language producers (Leitner's italics). The present subchapter will address the question of just how problematic the sampling and representativity of one-text-type corpora can also be.

Similarly to the problems of gathering linguistic data in general, sample-taking in linguistic media-research is also problematic, and varied in depth and amount. Bell offers a list of the kinds of samples gathered by a few researchers: the data range from a single radio programme to recordings of several months. Bell (1991:10) also points out the need for a delicate balance between limiting the amount of data to be gathered to manageable proportions and ensuring it remains representative. The tendency with media language, Bell says, is to collect too much, not too little, and so run the risk of drowning in data (1991:22). Content analysis researchers have tested the reliability of different sizes and designs of samples across different time periods. It has turned out

that, to represent a month of news, samples consisting of recordings from every fifth day (a total of six days) provided content proportions similar to those of the entire month. To represent a full year of news, those of 6, 12, 18, 24 and 48 days of a year were all adequate, and samples larger than 12 days provided little additional reliability. A frequent sampling pattern is to take every n^{th} day of a given period provided that n is not equal to 7, which would repeat the same weekday (1991:23).

Besides the quantity of media samples, quality issues are also to be addressed. When considering broadcast news as data, Bell draws attention to the fact that, if one was to study the core of scripted news as produced by the anchor person of a news programme or bulletin, materials such as commercials, opening or closing words, headlines, and even voice reports from journalists should be excluded (1991:16). Additionally, radio presents the problem of repeated news items. Bell judges that to include identical or almost identical repeats of a news item would unduly bias the quantitative analysis of some linguistic features if an unusual form happened to occur in a repeated item. Therefore, in the 35-hour sample of his 1977 sociolinguistic study of Auckland radio news, Bell included only the first occurrences of repeated items, which eliminated no less than a third of the total sample (1991:24).

6.3.1 Data collection

The data of the present study are news from the BBC World Service (hereafter BBC WS). There are three types of news programmes on the BBC WS: (1) five-minute World News, (2) two-minute news summaries, and (3) fifty-five-minute news programmes, containing fairly long correspondents' reports, news interviews, and news analyses revealing background information, plus a two-minute news summary in the middle of the programme. Five-minute news is broadcast every hour on the hour, except when replaced by news summaries. One-minute news summaries occur at 11:30 pm on weekdays, at 8 pm on Saturdays and at 4 pm, 6 pm and midnight on Sundays. The fifty-five-minute Newshours are on the air twice a day: at noon and at 9 in the evening.

For the purposes of the present study, the data were collected during a period of three weeks running: from 7 to 27 January, 2002. In contrast to the ' n^{th} day' type of sampling cited above, researchers found that consecutive weekdays over-represented certain kinds of content. However, the reason for choosing a sample of successive days is that one of the objectives was exactly related to content issues – news dealing with

the same topic –, and items far apart would not have been relevant to addressing those issues.

Every day, two to three morning news bulletins and the midday Newshour were recorded from the BBC WS available live through the Internet (for the Universal Resource Locator address, see References). Evening programmes were left out for convenience, on the grounds that (1) the most important news items are routinely included in morning broadcasts to allow for audience which do not listen to the radio during the night; and (2) there is nothing to indicate that evening news is significantly different from that of any other time of day. On five occasions, the Newshour recordings proved to be unfit for use because of server problems: the recording contained mere chunks of speech surrounded by a hissing sound, with a varied rate of intelligible speech versus noise. But this problem affected only the midday broadcasts; if any such difficulty occurred in the morning, there was always the possibility of recording the bulletin of the next hour.

The procedure described above resulted in a 21-hour sample. However, after leaving out the non-core news material from Newshours, such as interviews, analyses, long reports, advertisements, and announcements of subsequent programmes, the actual data decreased to a little more than 4.5 hours. This includes a total number of forty-nine 5-minute bulletins (broadcast between 6 and 10 am), and sixteen 2-minute news summaries (on the air at 12:30 in the afternoon).

Since the study does not aim at analysing phonological aspects and performance phenomena of news-reading, the recordings were transcribed only in terms of words. Although I am aware of the fact that intonation contributes to the textuality of a discourse (see Csölle 1999 on the role of intonation in text-cohesion), in this study intonation was only taken into consideration when separating news items, on the basis on pauses, and establishing sentence boundaries, on the basis of falls (Brazil 1985; Cruttenden 1981; Ladd 1980; Nevalainen 1992). Phenomena such as contracted negatives, slips of the tongue and repairs or bloopers were not represented in the transcriptions. The study deals exclusively with the verbal codes employed in radio news, ignoring suprasegmental elements, accent, pauses, and other aural codes such as sounds, noises, music and silence. This decision is based on the fact that news is exclusively prefabricated, written down and acknowledgedly read out loud. Stress and intonation in newsreading function only to clarify and disambiguate the meaning of the text, since newsreaders are trained to 'voice' the sentences with a tone of objectivity and

a minimum of expressiveness. Crisell points out that the primary code of radio is linguistic, since words are required to contextualise all the other codes (1994:54). Another relevant observation of Crisell's is that the newscaster is likely to speak with RP so that her/his reading will maximise the symbolic function – the meaningfulness – of the words while minimising her/his voice's function as an index of her/his personality. By this means it is suggested that s/he is the mere mouthpiece of the words (1994:58). The fact that different newscasters read out the same news item with contracting the negatives in different places proves that the clitics were not indicated in the news texts. To put it simply, the transcriptions in this study were aimed at restoring the scripts most probably used by newsreaders.

The obtained corpus contains roughly 49,000 words. The number is an approximation, because (1) in not more than 10 cases, and affecting less than 20 words, the speech remained unintelligible even after repeated listenings invoking native help, and (2) amounts were sometimes rendered by numbers, other times by letters, which all affect the total number of words.

These 49,000 words make up 452 news items. Out of these, only 28 were verbatim repetitions of previous items, though many more were cases of partial repetition. Unlike in the Bell study cited above, exact repetitions were not excluded, since these were responsible only for about 6% of the total corpus. Nor were the partial reiterations removed, because one objective of the study is to analyse what exactly had been left out from, or added to, previous broadcasts of the same item.

To sum up, the BBC WS data employed by this study consists of a 4-hour 30-minute recording of morning and afternoon news, which amounts to a 49,000-word sample with a little inner repetition characteristic of radio news discourse.

6.3.2 Representativity

At this point, the problem of representativity has to be addressed. It is not easy to tackle the representativeness of media samples, since, statistically speaking, the population is not delimited. If one wants to draw conclusions that apply to the population of Hungary, for example, there is a finite set of elements – roughly ten million people – to yield a sampling frame. Media output, on the other hand, is an endless mass. In tackling this problem, as we have seen, some studies resort to the production amounts of conventional time units such as a week, a month or a year. These

delimitations are in fact arbitrary, yet they are valid, since the workings of the media also follow these time units so deeply entrenched in culture.

Even if we take these time units as a population, the sample of the present study has not been randomly taken, nor is it large enough to allow for generalisability. (In mathematical terms, out of the roughly 730 hours of yearly news output on the BBC WS, the 4.5 hours considered in this study are only a drop in the ocean.) The method applied in this study at best approximates 'nonprobability purposive sampling', described by Babbie (1989:204) as based on selecting data according to one's knowledge of the population, its elements (i.e. BBC WS programmes and staff information), and the nature of research aims, based on one's judgement and the purpose of the study.

Statistically speaking, the data discussed here might at best be said to represent the January 2002 news output of the BBC WS. However, as Babbie (1989:188) points out, the representativeness of a sample is increased by certain factors, one of which is low degree of variation: a homogeneous population produces samples with smaller sampling errors than does a heterogeneous population. Therefore, if the news turn out to be largely homogeneous – as, indeed, is the case – the representativeness of the data may be considered higher. This statement is supported by the findings of Al-Shabab (1986:16), who worked with a corpus of 74,000 words, and found that the language of radio news is standard and elaborate, therefore, what is found in a small sample will be typical and highly representative of radio news broadcasts.

Another aspect of the BBC WS corpus is its variability according to producers, that is, news editors. An attempt was made to obtain information concerning the identity of news editors of the particular bulletins recorded, certainly without asking for exact names or confidential information, just to see whether the personnel were identical on particular days. Through personal communication with a leader of the BBC WS news staff, however, it was found that there is no way to retrieve the identity of the news editors from the archives. It was revealed, though, that there are approximately 80 editors and a similar number of newscasters working at the BBC WS newsroom, and there is no regularity in the way they are assigned shifts. The identity of newsreaders is public information, since they introduce themselves at the beginning of a newscast. In the BBC WS corpus, 32 newscasters were identified – 40 per cent of the total number of those employed by the BBC WS –, out of whom one person delivered news five times,

three people read them four times, six people three times, eight persons twice and fourteen newscasters worked only once (that is, only 10 per cent of the newscasters are over-represented in the sample, to the extent that they read news more than twice). Since the number of editors and newsreaders is roughly the same, through an analogy one may conclude that 40 per cent of the editors is probably represented in the three-week sample. Further, if there is no regularity regarding their working times, and so, theoretically, each person has a nearly equal chance of selection, the data could be taken as naturally randomised. All these considerations increase the representativeness of the sample in this study. Nevertheless, there is always the possibility of an editor having been ill or on holiday in January 2002, and thus being totally unrepresented in the sample.

Word-classes	Average percentage	Standard deviation from the mean
Adjectives	7.45	0.91
Adverbs	3.24	0.51
Conjunctions	3.84	0.78
Determiners	14.50	1.19
Gerunds	0.83	0.33
Lexical verbs	12.19	0.71
Modals	1.21	0.39
Nouns	30.85	1.02
Other	2.40	0.60
Prepositions	13.18	1.00
Primary auxiliaries	4.52	0.88
Primary main verbs	1.38	0.40
Pronouns	4.35	0.56

Table 6.1 The standard deviation of word-class frequencies according to the days of the week in the BBC corpus. The group ‘Other’ includes existential *there*, the negative particle *not*, the infinitive marker *to*, verbal pro-forms, foreign words and symbols.

Another variable that should be taken into consideration is the day of the week. A group of the quantitative results of the present study is in connection with word-class calculations, and these results have been analysed for standard deviation. As Table 6.1 shows, the values found for the different days are not scattered far around the average. (Only the numbers referring to the gerund show a slightly higher deviation, but this is due to the overall low frequency of gerunds; even one occurrence may cause a

considerable difference in deviation.) This means that the BBC January news is uniform in respect to the rate of word-classes. Since there is no reason to believe that other months would give considerably different results, the corpus of the present study may be considered to accurately represent the BBC news in general.

In sum, the three-week data can be considered to accurately represent news output on the BBC WS. The data collection most closely approximates purposive or judgmental sampling based on familiarity with the material to be examined, and the purposes of the study. The representativity of this small-scale data is boosted by the homogeneity of the population and a certain degree of natural randomisation.

6.4 Quantitative methods

6.4.1 Data processing

The data were divided into two groups: the so-called 'long news', that is, morning editions of the BBC World News, and 'short news', i.e. two-minute news summaries in the middle of afternoon Newshour programmes. The long news yielded 347 items, while the short news gave 105 pieces. Two subgroups were identified in the long news: 'reported news', that is, the ones containing news reports by correspondents, and 'non-reported news', i.e. those not involving correspondents. By nature, none of the short news contained reports.

After this grouping, news texts proper were separated from headlines and 'discourse organising elements', which, in Al-Shabab's (1986:72) definition, refer to, and organise, news discourse, and whose main aim is to maintain the channel of communication by addressing the listeners' directly. Such elements include openings, medial elements and closings such as *This is the BBC in London*, *You are listening to the news from the BBC* and *This was the news*. These utterances do not share the Hallidayan ideational function of the rest of the news, rather, they are interactive and metatextual and, as such, they cannot be considered news proper. Similarly, introductions to news reports from correspondents, such as *Geordine Kochland reports from The Hague (7/9)* have also been removed.²¹ Some hesitation was caused by

²¹The source of examples is given in the following way: Since the month is always January, the first number refers to the hour of broadcast, the second to the day. E.g. 7/9 means the 7 a.m. edition of the BBC WS news on January 9, 2002.

introductions which contain information other than the name and whereabouts of correspondents (e.g. *Our correspondent, John Leyne, has been travelling with Mr Powell, and sent us this report from Kabul 7/17*), but these were also excluded, partly because the extra information they contain does not refer to the events reported on, but the circumstances of the reporting act, and because such introductions represent only about 8% of the cases.

The news texts proper were then subjected to QTAG 3.01, a part-of-speech tagging programme available for academic research purposes, developed by Oliver Mason at the University of Birmingham (Mason 2001/1994). To each word of an input text, the programme attaches a grammatical label establishing word-class and containing additional information concerning, for example, the number or subclass of nouns, the morphological forms of verbs, the grade of adjectives or adverbs etc. The list of tags used by the programme, the Birmingham-Lancaster Tagset, a variant of the more common Brown/Penn-style tagsets, is to be found in Appendix 1/a. The tagset contains 70 tags, 58 for words and 12 for punctuation marks.

However, no such automatic programme is a hundred percent accurate, therefore, it needs careful semi-manual checking. Also, the tagset did not prove to be varied enough for some of the purposes of the present study. The following insufficiencies were experienced:

- The tagset separates the primary verbs *do*, *be* and *have* from other verbs, but does not differentiate between their usage as main verbs or aspectual, voice or emphatic auxiliaries.
- The programme divides conjunctions into two groups: coordinators and subordinators, but it does not distinguish between phrase-level, clause-level and sentence-initial coordinators.
- QTAG does not state whether an adjective is used attributively or predicatively.
- Modal auxiliaries originally have a single tag, which does not differentiate between present or past tense modals, and secondary or semi-modal auxiliaries, which can also occur in nonfinite forms.
- Genitives are not analysed for being classifying or specifying.
- There is a single tag for verbs in their base forms, which conceals their being finite or nonfinite.

- Verb forms ending in *-ing* have a single tag, unable to distinguish between gerunds and present participles.
- Pronouns have tags corresponding to some of their major classes (personal, reflexive), but pro-forms for phrases or clauses (e.g. *one, so, do*) are not identified. In the case of personal pronouns, there is no way to know from the original tagset which person they refer to.

All of the above problems have been solved by adding new tags to the tagset (Appendix 1/b), which in this way has grown to a list of 120 items. The attachment of tags has been checked word by word, and the new tags have been assigned by hand. Such an activity is in fact the first stage of analysis, in that it involves important decisions of parsing, which will be described in detail in the following subchapter.

6.4.2 Problematic tagging cases

In many cases the computer programme does not assign the right label to a word. Some of these errors are banal: such as parsing *Buenos Aires* as two plural proper nouns, on the basis that the two words start with capitals and end in *-s*. Others are less evident: e.g. the programme is unable to distinguish between homonyms, such as *close* used as a verb or a noun, *present* used as a verb, a noun, or an adjective etc. An illustrative example for such problems is *that*, which can behave in five different ways:

<i>The reports from both sides say <u>that</u> the Palestinian militants cut through ...</i>	(7/9)
	= conjunction
<i>... all seven marines on board a military plane <u>that</u> crashed in Pakistan ...</i>	(7/10)
	= relative pronoun
<i><u>That</u> hearing could be postponed ...</i>	(7/9)
	= demonstrative determiner
<i><u>That</u> was when Beijing made a defensive ruling ...</i>	(8/10)
	= demonstrative pronoun
<i>Decisions, however, will not wait <u>that</u> long.</i>	(6/15)
	= adverb

The analyst has to make other, more sophisticated choices as well. Such a problem may be the common/proper character of nouns. Naturally, the annotation of proper nouns consisting of two parts, such as the above-mentioned *Buenos Aires*, has

been reduced to one tag. But there was the question of titles, with or without personal names. Here, the following decisions have been made: if the titles occurred on their own: e.g. *president*, *prime minister*, the title was attached a single common-noun tag, on the basis that, although in a given time period the reference of such a noun is unique (there is one president in a country at a time), the nouns do not resist pluralisation and the usage of the indefinite article. In the case of titles accompanied by a personal name, the following two cases have been identified, and thus tagged in two ways:

<i>the</i>	<i>Zimbabwean</i>	<i>foreign minister,</i>	<i>Stanislav Mudenge</i> (7/9)
Determiner	Adjective	Common noun	Proper noun
<i>President Andres Pastrana ...</i>		(7/19)	
Proper noun			

In the first example, the proper noun has been considered to be an appositive postmodifier to the preceding head noun, while in the second, the three have been tagged as one common noun. The decisions have been determined by the presence or absence of the definite article so that, if the second phrase had read *the president of Colombia, Andres Pastrana* or *Andres Pastrana, the president*, it would have been analysed similarly to the first. The second, one-unit analysis has been retained even in the case of relatively long phrases if the definite article was missing: e.g. *Nigeria Labour Congress spokesman Adams Oshiomhole* (6/16).

Another common/proper-noun problem is the case of dates. Corresponding to the general view, days, months and festivals have been analysed as proper nouns. This analysis, however, has been extended to whole dates (e.g. *January 9, 1992*) or time expressions such as *7:37 GMT today, Monday 9:30 pm, May last year*, or even *the 1960's*, since all these expressions have unique reference (the last term resembling plural geographical names), and cannot take pluralisation or the indefinite article. Even more 'proper' is the expression *September 11*, which metonymically refers to a whole series of events, just like *World War II*.

Numbers have also been analysed in different ways. Typically they act as post-determiners (e.g. *1500 people* 7/9, *60 years old* 6/17), but they can also act as pro-forms (e.g. *Afghanistan needs 1.7 billion dollars for the first year, 15 billion for the next decade* 6/21). Special cases occur in the representation of ages; on such occasions the numbers have been tagged as nouns: e.g. *Mr Vance was 84* (8/13) or *at the age of 18*

(6/17), on the basis of their function as subject complements or complements to prepositions, which are all noun positions.

Titles of films, songs, names of organisations, however long, have been tagged as single proper nouns: e.g. *A Beautiful Mind* (9/21), *The Folks That Live on the Hill* (9/22), *the Popular Front for the Liberation of Palestine* (6/16). Elsewhere, elements have been tagged as common or proper depending on whether they are part of a name or not: cf. *federal judge* (8/12) and *Federal Reserve* (6/16); or *[microbes] living deep under the earth* (8/17) and *new life forms on Earth* (8/17).

Words which act as premodifiers of nouns, and are originally nouns themselves, have been tagged as nouns and not as adjectives, irrespective of the fact that they have lost some of their noun characteristics, for example, they cannot be pluralised: e.g. *passenger plane* (6/17) (for a detailed analysis of such syntactic constructions, see Marosán 1994, 2000). However, in some cases, it was exactly their syntactic position that decided on the tagging of certain elements: e.g. *military* has been tagged as an adjective if it functioned as a premodifier (e.g. *American military base* 6/17), or as a noun if it occurred in a noun position (e.g. *a senior officer in the United States military* 7/9).

Genitives can be functionally analysed as central determiners, since they do not allow the use of other such words. However, such an analysis would have concealed the fact that these 'determiners' are in fact nouns, which in effect would have disfavouredly modified the frequency counts for nouns. Therefore, genitives have been analysed as nouns functioning as premodifiers, and they have been given a refined tagging distinguishing between specifying (e.g. *Thailand's interior minister*) and classifying (e.g. *women's fashion*) genitives.

Gerunds have been separated from present participles on similar grounds: through their functions they increase the rate of nominal elements in a text, therefore, disregarding them would falsify the results. However, only real gerunds, and not verbal nouns, have been tagged as such: cf. *an end to kidnapping* (7/13) and *a series of killings, bombings and shootings* (8/13).

A special problem has been posed by quite widely used 'double' expressions, such as *rebel-held*, *Palestinian-controlled*, *Pakistani-backed* or *factory-closures*, *job-losses*, *weapon-smuggling* etc. These expressions are in fact entire clauses condensed in phrases ('held by rebels,' 'the closing of factories'). In such cases the tagging has been

done according to the head of the phrase, which is generally the second element of the compound.

When assigning tags to past form modal auxiliaries, attention has been given to whether the modals have real past tense reference or a tentative, hypothetical usage: cf. *Some analysts had predicted the peso would collapse* (7/12) and *Some school principals have welcomed the idea of mandatory testing civil liberty activists say would infringe students' rights* (8/11).

Semi-modal auxiliaries, such as *have to* or *seem to*, have been assigned separate tags. Some verbs, however, can act in different ways; in such cases the tagging has been decided according to whether they are followed by a full verb or not: e.g. *The UN Security Council is expected to vote ...* 6/15 – semi-modal; *Further moves are expected today ...* (6/15) – lexical verb.

Other, less important decisions have had to be made in some cases: for example, there were quite a number of words which form one unit: e.g. *a number of*, *a host of*, *a spate of* count as single determiners, *above all* is one adverbial, *in connection with* and *in advance of* are single prepositions, *most patriotic* forms one adjective etc. The same words may behave as single units on certain occasions and as separate ones on others: e.g. ... *out of the five candidates running for the post ...* (6/15) – one unit; ... *nothing will come out of the meeting ...* (7/14) – separate units. Other 'amphibians' include *one* as a numeral (e.g. *At least one person was killed ...* 6/16) or behaving more like a determiner (e.g. *Seeing is believing, says one official here ...* 6/16). Conversely, the indefinite article may also act as a numeral: e.g. *International donors are called on to provide over a billion dollars ...* (6/21).

Certain shortcomings of the programme have not been eliminated. For example, while first and second person pronouns have been identified, neither the original tagset, nor the new one can, for example, distinguish between third person pronouns, let alone the different usages of the pronoun *it*, as a third-person pronoun (e.g. *At 7:30 GMT today it was due to pass ...* 9/7) or a prop *it* in extraposition (e.g. ... *it is still possible the trial will be postponed* 8/9).

Nor is the programme able to tackle cases when certain elements have a double function. *Wh*-pronouns, for example, act as subordinators and also have functions in the subclause they introduce: e.g. in ... *the VSO asked more than a thousand adults what came to mind when they thought of the developing world* (9/7), the *wh*-element is a conjunct and a pronoun in subject function at the same time. These have been tagged as

wh-pronouns. The same tag has been attached to relative pronouns as well. Other interesting cases of double identity have been found: e.g. ... *the situation is still dangerous, though very slowly improving* (7/18), where *is* acts as a main verb followed by a complement (*dangerous*) on the one hand, and as a progressive auxiliary (to *improving*) on the other. The decisions in such cases have been made according to proximity: on the basis of the word's function in relation to the element closer to it.

As a result of the tagging procedure described above, the number of tags is smaller than that of words: the roughly 49,000 words have been assigned 40,644 tags.

6.4.3 Further procedures

After the tagging process, the beginnings and ends of noun and prepositional phrases were marked to establish the length of such phrases. This procedure was made difficult by the fact that such phrases are typically concentric, in the sense that a noun phrase may contain a prepositional phrase or another noun phrase, and a prepositional phrase always holds a noun phrase. It has been decided that, for the sake of simplicity, only the outer shells will be marked, so that *a hunger strike by asylum seekers held at a camp in central Australia* (6/23) has been considered one noun phrase, irrespective of the fact that there are others present (*asylum seekers, central Australia* etc.). This has been decided with the reservation that, this way, the value obtained for the average length of noun and prepositional phrases slides to the greater. Nevertheless, had all the noun phrases within prepositional ones been considered, the value would have been levelled.

Occasionally, the beginnings and ends of the phrases were dependent on interpretation. For example, in *there was now a real prospect of progress and stability under the new interim government* (8/8), it is not easy to decide whether the prepositional phrase is a postmodifier of the preceding noun or a separate unit acting as an adverbial. In most of the cases, however, the decisions were not difficult to make.

In sum, the collected BBC corpus shares some of the features Miller and Cann (1994:816) consider as characterising natural data. The advantages were the following: the analyst could not influence the text; and the corpus is useful for more than its original purpose. The inevitable disadvantages of such data were encountered here, too: the collection was time-consuming, the transcription laborious and it had to be coded for relevant properties. However, it was only the tagging and parsing procedure

described above which made calculations possible to answer the quantitative research questions.

6.5 Qualitative methods

The first step in this stage was to establish the topics of news items. The qualitative method of topic-assignment is the one offered by van Dijk 1988a, and detailed in Section 5.1: the news texts have been broken down to propositions, then the macrorules of summarising, such as deletion, generalisation, and construction, have been applied, until the topic, that is, “the summary, gist, upshot or most important information” (van Dijk 1988a:31) of the news item has been established. In this part of the analysis, repeated items have not been included, since they do not display any topical progression, and add nothing new to the analysis.

The next step has been to arrange these topics into units, which involves further generalisation, based on certain common aspects, or elements in the news items. Twelve large units have been identified, which have been called ‘topic areas’. Each topic area comprises smaller units, called ‘topic groups’, which then branch off into ‘topics’. The lowest level units are the ‘subtopics’, which, however, are not compulsory. The individual news item always belongs to the lowest level. That is, if a topic is treated in a single news item, then the topic directly defines this item. If, however, a topic is treated along several items, it is the subtopic that defines the individual item. Figure 6.1 shows the model obtained:

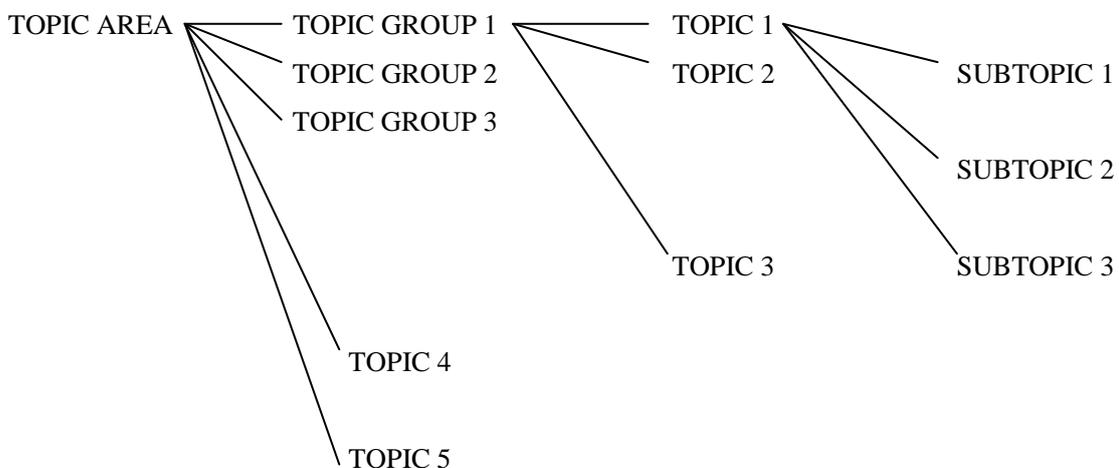


Figure 6.1 The hypothetical topic structure of a topic area in the BBC corpus

For each news item, its topic and position have been introduced into the database, to make retrieval possible for any correlation between the two variables.

The textual analyses have been done qualitatively: textual units have been identified according to van Dijk's and Bell's systems. The recognition of Main Events, Consequences, Expectations and Evaluations did not pose any problems. Context and Previous Events were differentiated on the basis of the tense used: present tense for Context and past for Previous Events. History and Previous Events were separated according to the time they spanned: Previous Events were considered to reach through a couple of days or weeks, History through months or years.

However, the task of categorising has not been without difficulties all the time. First, a matter of principle had to be attended to. Both researchers point out that several of the text categories may be recursive (Van Dijk 1988a:56; Bell 1991:170), but they also agree that news texts are typically delivered in instalments (Van Dijk 1988a:48; Bell 1991:165). The two statements have different presuppositions: if a category is recursive, we may think of *several different* instances of the same category occurring again; if a category is delivered in instalments, we may think of *the very same* instance of a category to be communicated discontinuously, divided into several parts. This may well be a matter of principle and, most of the time, it is not problematic, but sometimes it has serious practical consequences. Consider, for example, the following news items:

- (1) *A spokesman for Prince Charles, the heir to the British throne, has confirmed a newspaper report that Prince Charles's second son, Prince Harry, was involved last year in under-age drinking and smoked cannabis. The spokesman also acknowledged that Prince Harry, who was sixteen at the time, was taken by his father to a counselling session at a drug rehabilitation centre. There is no suggestion that Prince Harry himself needed treatment. He has since returned to school, and the spokesman said his family now considers the matter closed.(7/13)*
- (2) *The British Prime Minister Tony Blair has arrived in Islamabad on the latest stage of a mission to try to reduce the hostility between India and Pakistan. Mr Blair travelled from Delhi after talks with the Indian Prime Minister Atal Behari Vajpayee. Both men agreed that terrorism must be defeated before any solution could be found to the Kashmir dispute. India blames Pakistani-backed Kashmiri militants for last month's attack on its parliament. British officials say Mr Blair who is keeping in close touch with President Bush during his visit will put more pressure on Pakistan to stop the violence. A BBC correspondent travelling with*

the Prime Minister says that so far Mr Blair has nothing to show for his attempt to bridge the divide between India and Pakistan. (9/7)

- (3) *The computer software giant Microsoft has suffered a set-back in the US Courts after a federal judge rejected a proposal to settle more than 100 private cases against the company. Microsoft had offered a package of 1 billion dollars with their software and services to schools in order to settle claims that the company had practised unfair pricing.*

M. F. reports from Washington.

This deal had been proposed by one of the lawyers representing the plaintiffs. They've successfully argued that Microsoft had overcharged millions of people for using its software. The company would give computers, software, training and support to thousands of schools in underprivileged areas instead of paying what would often have been tiny damages to software users. But after negotiations, which ended on Thursday, the two sides failed to agree on the details. Passing throughout the proposed deal judge Frederick Motz said he felt the donation was probably too small to achieve what was promised. And perhaps more important that the deal would give Microsoft unfair access to the schools' market. (7/12)

All three texts have two underlined passages, both of which have been classed as Previous Events. But do they belong together? The following solutions are proposed: in example (1), the two clauses – although delivered discontinuously – form a single Previous Events category together, since the second part is a temporal and logical continuation of the first (the Main Event being that a report about Prince Harry's alcohol and drug consumption has been confirmed). In example (2), although the two sentences directly follow each-other, they are two separate Previous Events, since the second event is prior to that of the main clause in the first sentence (first the two men agreed, then Mr Blair left). In example (3), although the two sections show much cohesion and coherence (*unfair pricing* corresponds to *overcharging*), the two sentences have been analysed as two Previous Events on the grounds that they belong to different subtexts (represent different subgenres): the first is in the introductory part of the news item, and the second in the correspondent's report.

Occasionally, the hanging together of two textual pieces is overtly signalled, such as in example (4) below:

- (4) *The French Parliament is today debating a government move for overturning a controversial legal precedent that in effect established the right not to be born. In three cases French judges have ruled that families whose children were born with birth defects could sue the doctors who failed to spot them in the womb. The groups representing handicapped people say the rulings devalue their lives. And some specialists have stopped carrying out scans because of the risks of lawsuits*

and the pressure to recommend an abortion at the slightest sign of trouble. The government wants a change in the law so that no one can claim damages for having been born. (8/10)

Here, the conjunction *and* at the beginning of the second underlined sentence is an explicit linguistic link to the first, a connection that is born out by the content of the two statements. Thus, they have been analysed as forming one category of Previous Events. A similarly overt connection is shown by the following item:

- (5) *Researchers in California say they've developed a new technique that could help them diagnose Alzheimer's disease in living patients even in the early stages. At present the disease can only be medically confirmed by a post mortem examination. Using the technique a radioactive dye is injected to a vein and shows up lesions in the brain, symptomatic of damage caused by Alzheimer's. At present there's no cure for Alzheimer's, which is a fatal brain-wasting disease.* (7/10)

Here, both sentences start with *at present*, which signals their interrelatedness. Thus, the sentences have been regarded as two parts of the same Context category. On the other hand, *the [technique]* in the intervening sentence refers back to *a new technique* in the first, so this sentence classifies as a different, Details of Main Event category.

It was also necessary to differentiate between Verbal Reactions and Attributions. The second has been found indispensable for the analysis of the BBC corpus since, clearly, a common phrase such as *The BBC correspondent says* is not a Verbal Reaction, but an Attribution. In more general terms, when a person or institution acts as an information source, his/her contribution cannot be considered a Verbal Reaction, which expresses opinions or standpoints. Take, for example, item (6) below:

- (6) *Ceremonies are taking place today in the Northern Bosnian city of Banja Luka to mark the tenth anniversary of the Bosnian Serb Republic. A service is taking place in the town's main orthodox church and there will be a ceremony in the war cemetery. The current president of the republic Mirko Sarovic has described the commemoration of the founding of the republic on January 9, 1992 as an important jubilee. The BBC correspondent in Sarajevo says many Serbs view their republic as one of the most tangible gains of the war.* (7/9)

The above news item contains three Attributions, which have been indicated by underlining. Out of the three, however, only the first and the third have been regarded as

belonging to a Verbal Reaction. The second is an independent Attribution unit serving as information source.

As for the interrelatedness of Verbal Reactions, the decisions were much easier to make. Two or more successive instances of this category have been considered to form one unit if they are parts of a speech made by the same person, but different units if their origin is not identical. This is exemplified in (7) below:

- (7) *As 20 Taleban and al-Qaeda prisoners spent their first night in the United States detention camp in Cuba, the American Defence Secretary has said he does not consider them prisoners of war. Mr Rumsfeld said the detainees were illegal combatants and under that status were not protected by the Geneva Conventions. However, he insisted the prisoners would be treated humanely. Cuba has said it will cooperate with the United States plan to house the prisoners in Guantanamo Bay.* (7/12)

The above item contains three statements made by the American Defence Secretary: these form a single Verbal Reactions unit. However, the next statement, made by Cuban officials, is considered a separate one.

The comparison of the different subgenres of news has also been done qualitatively, involving a careful search for textual units, and for the transformations they undergo in the course of being partially or otherwise repeated. While headlines were excluded in the quantitative data analysis, in this part of the methodology they were re-introduced to allow for an analysis of their relationship to the items they accompany.

The reader is reminded here that topic-assignment, the application of macrorules and the derivation of macrostructures, as van Dijk (1988a:33) also warns, are subjective procedures. This means that other analysts might have arrived at different topics and groupings, and that, in this way, the results presented in Chapter 8 are nothing absolute. However, their analyses would also be subjective, so a guarantee of intersubjectivity – not of objectivity – could only lie with a research team.

6.6 Summary

The research questions in this study have been formulated so that they can find out whether radio news is any different from press news in its grammatical features and discourse structure, and if subtypes of radio news are dissimilar from each other in

grammatical features, size and textual make-up. The study also aims at discussing whether two frameworks for the textual structure of printed news (van Dijk's and Bell's models) are relevant for radio news as well. Additionally, it is asked whether there is any relationship between the topic, positioning and headlining of radio news items.

The data were gathered so that they provide the best chance of answering the research questions. For example, while to the quantitative/grammatical features it does not make a difference, the topic-related qualitative issues (e.g. textual transformations) could only be addressed through a database yielded by consecutive days of broadcast. Thus, the BBC corpus collected for the purposes of the present study consists of a 4-hour 30-minute sample of BBC World Service news from morning and afternoon programmes for three weeks running. The news was transcribed in terms of words, which resulted in a body of text made up of approximately 49,000 words (452 news items). The data are fairly representative of the BBC WS news output and, given the prestige and acknowledged exemplary character of the BBC, the findings are prone to be generalised to all Western-type public service radio news production.

The database was quantitatively tagged with the help of QTAG 3.01 part-of-speech tagger, and manually checked for errors and cases in need of clarification. Some shortcomings of the programme were also eliminated. The news was qualitatively analysed for topics and textual units, and the subgenres were compared in terms of discourse structure. Textual configurations obtained, the results could be viewed in the light of the two approaches referred to above, so that their relative applicability could be judged.

7 Grammatical features of radio news: Results and discussion

This chapter will provide the answers to the quantitative research questions asked in the previous chapter (Section 6.1), namely to find out about the grammatical features of radio news, with special attention to those which hint at the written vs spoken character of that media genre. To be able to compare radio news to written (press news) and spoken genres (conversation), we need numerical data on the use of word classes, phrases and clauses in the BBC corpus.

7.1 Frequency counts for word classes

The calculations have been done according to the number of tags, not words. However, for the sake of simplicity, in the following presentation of results I will speak about words, not tags. Most of the relevant features have been counted on an overall basis, and separately for short news and reports, which mean only the texts of correspondents, without the introduction by the newsreader. This has been done because short news and reports are intuitively felt to be different from news in general, short news in their summary-like and reports in their speech-like character. All the values presented in this subchapter have been rounded for convenience.

For the rate of different word classes, the following results have been obtained:

As can be seen in Figure 7.1, the most frequent word-classes are nouns, determiners, prepositions and lexical verbs. This roughly corresponds to Biber et al.'s findings about written news texts (1999:65, 92). Contrary to the intuition, there is no significant difference between the overall and the short news values (except for maybe nouns and prepositions), but the values for reports do show certain deviations. The most important difference can be seen in the value of primary main verbs, whose rate in reports is almost double the average. This can be interpreted in the following way: since *do* is very rare (2.5% of all primary main verbs), it is *be* and *have* which are responsible for this high frequency. Both these verbs are stative in meaning, which might suggest that, while studio-news mainly emphasises actions and events, reports contain much background information regarding the state of affairs.

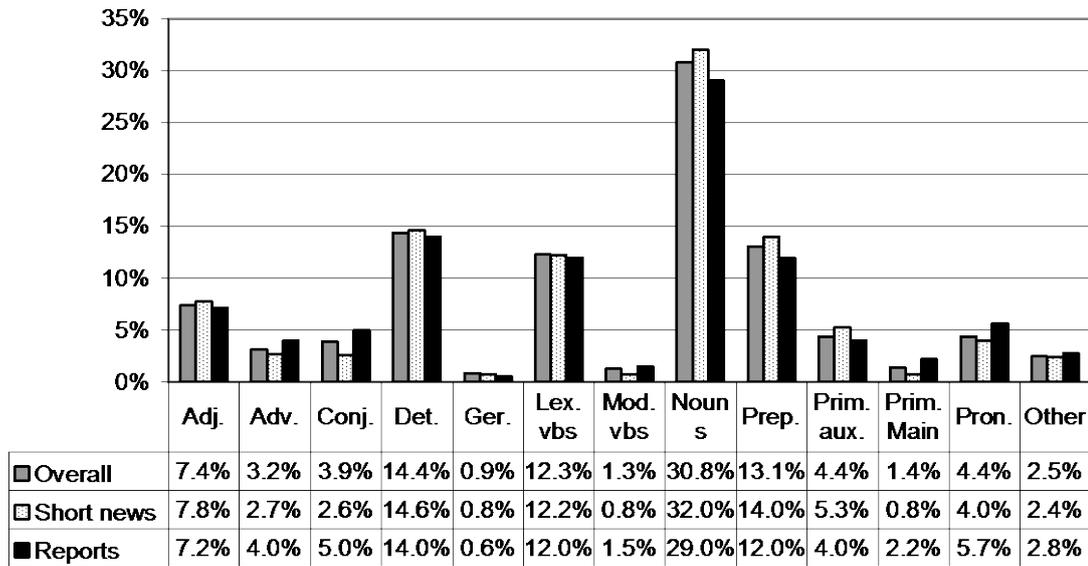


Figure 7.1 The distribution of word classes in radio news (as a percentage). Notation: Adj. = adjectives; Adv. = adverbs; Conj. = conjunctions; Det. = determiners; Ger. = gerunds, Lex.vbs. = lexical verbs; Mod.vbs. = modal verbs; Nouns = nouns; Prep. = prepositions; Prim.aux. = primary auxiliaries; Prim.Main = primary main verbs; Pron. = pronouns; Other = existential *there*, the negative particle *not*, the infinitive marker *to*, verbal pro-forms, foreign words and symbols.)

Another eye-catching difference between reports and short news is in the rate of pronouns (5.7 is higher than 4.0 by 42%) and conjunctions (5.0 is higher than 2.6 by 92%). The greater percentage of pronouns in reports goes along with a slightly lower frequency of nouns, which shows that reports are a little closer to conversations than radio news in general. The greater frequency of conjunctions must be due to more phrase-level coordination, since elsewhere the results show that there is not much difference in the number of clauses in reports and overall (see Section 7.3). Coordination makes the text of reports more linear and less integrated in nature than the average radio news.

While verbs are slightly less frequent in reports than overall, adverbs are more common (4 is higher than 3.2 by 25%). This is strange because generally these two word-classes tend to correlate in number. The results might show that reports are more concerned with the circumstances of events than the actions themselves, which is in harmony with the higher rate of stative verbs indicated above.

The lexical density of the texts, that is, the proportion of the texts made up of lexical word tokens as opposed to function words (Biber et al. 1999:62) shows the following results (Figure 7.2):

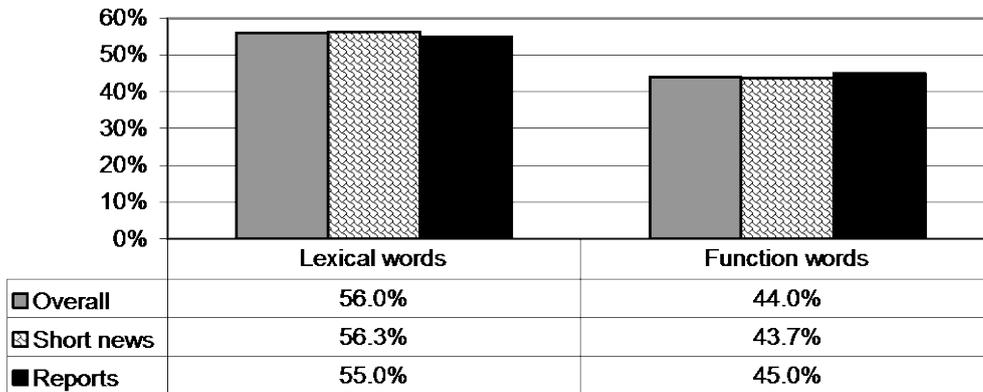


Figure 7.2 The lexical density of radio news (as a percentage)

It is apparent that, while Biber et al. (1999:65) found that the rate of lexical words is around 63% in written news texts, radio news presents a lower lexical density of only 55%, but not as low as conversations, whose lexical density is roughly 41%. A small difference between news types is also shown by the results: lexical density is slightly lower in reported news than overall, which again hints at reports being slightly more like conversations than other subgenres of radio news. Short news shows no significant difference from the average from this point of view.

Taking only lexical words into consideration, the following results occur (Figure 7.3):

The general tendency found by Biber et al. (1999) about the most and the least frequent word classes in all kinds of registers can be seen here as well. The differences between the three kinds of news in the BBC corpus are not considerable here either, but a slightly lower rate of nouns in reports than the average goes along with similarly fewer adjectives. The higher percentage of verbs on this figure (as opposed to a lower rate on Figure 7.1) is due to the fact that here primary verbs in main verb usage and lexical verbs have been added up to form the category of verbs. The noun/verb ratio, however, does show differences: overall, the rate is 2.3, in short news it is 2.5, and in reports it is only 2, that is, in reports the rate of verbs as opposed to nouns is 15% higher

than in average. The ratio distinguishes radio news, and especially reports, from written news texts, in which there are 3 to 4 nouns per lexical verb (Biber et al. 1999:65).

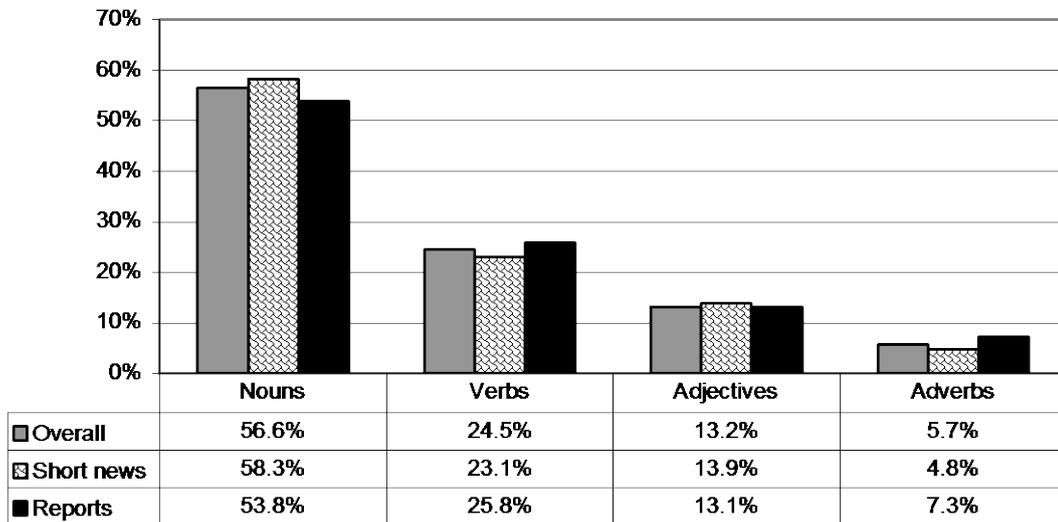


Figure 7.3 The distribution of lexical words in radio news (as a percentage)

The ten most frequent nouns in the BBC corpus are the following: *government, people, official* ('person who holds a public office'), *Afghanistan, United States, police, town, talk, leader* and *country*. This is in accordance with the most common topics of news, which is mainly concerned with politics and, more precisely – at least in January 2002 –, Middle-Eastern issues, and the role of the greatest world power in those issues. The ten most common verbs are: *say, kill, hold, make, tell, find, come, give, call* and *allow*. Three out of these correspond to the most commonly used lexical verbs in news reportage: *say, make, and come* (cf. Biber et al. 1999:375). Two of the most frequent verbs are communication verbs (*say, tell*), showing the main concern of news to report the words of important people (cf. *official* as the third most frequent noun). The meanings that the verb *hold* is mostly used in are divided between activity verbs (e.g. *hold back water*) and relationship verbs (*hold sy responsible*). *Make* is most commonly used in expressions (*make a decision, a mistake, an attempt* etc.). The typical meaning of *find* in the BBC corpus is dynamic (*find a solution, the way, a lost lottery ticket* etc.). *Come* is used either as a multi-word verb (*come up with, come in for*) or expressing occurrence (*The shootings come as ...*). *Give* and *call* stand for activities, while *kill* and *allow* express causation.

The ten most frequently used adjectives are: *Palestinian, American, Israeli, new, military, international, Afghan, Indian, other* and *interim*. Almost exclusively, the most common adjectives are classifiers, half of them affiliative, which are non-gradable. Even *new*, which may have gradable meanings, is mostly used in the sense of ‘recent’ (e.g. *new deadline, new bushfires* etc.). This shows that, in radio news, just like in printed news (Biber et al. 1999:512-3), the most typical adjectives are not descriptors but classifiers, expressing origin or topical affiliation. The high frequency of the adjective *interim* may seem strange, but this adjective is almost exclusively used to refer to the interim Afghan administration. This corresponds to *Afghanistan* being the fourth most common noun in the BBC corpus.

Among adverbs, the ten most common are: *today, now, also, about, almost, at least, earlier, later, still* and *just*. Six of these are directly connected to time (*today, now, earlier, later, still, just*), four in connection with the present. Three of the others are degree adverbs (*about, almost, at least*). This is connected to a high frequency of quantifiers in the BBC corpus (see below), to which these adverbs serve as approximations.

As far as function word classes are concerned, the differences between the types of news are more discernible (Figure 7.4):

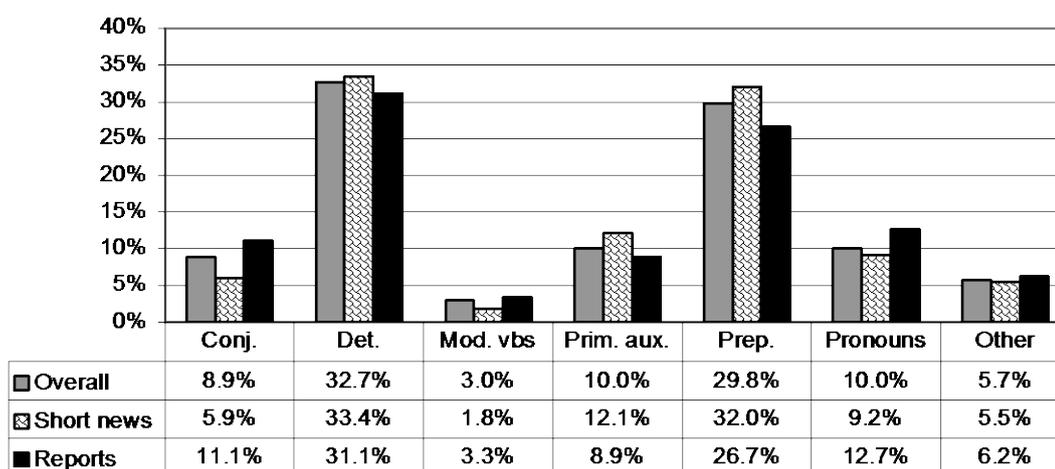


Figure 7.4 The distribution of function words in radio news (as a percentage)

As has been indicated above, conjunctions and pronouns are significantly more common in reports than in average, but prepositions are less frequently used. This latter

may be connected with a lower rate of nouns, since prepositional phrases are very often used as postmodifiers to nouns, so the low frequencies of the two presuppose each other.

Although function words are closed-system items, there may be some point in seeing which are the most frequent ones. Here are some notable findings: *and* is by far the most common conjunction, occurring almost three times more often than the next most frequent one, *that*. On the phrase-level, *and* is a means of integration, that is, to pack as much information as possible in a clause. On the clause-level, *and* is often used as an all-purposive connector, able to express temporal or causal relations as well, since, to be on the safe side, news may just vaguely suggest these relationship rather than clearly state them (e.g. *They removed him as a leader, and the government arranged for new local elections... 10/7*). Some of the next most common conjunctions are: *but*, *as*, *or*, *after* and *when*. *After* and *when* reflect the interest of news in relating events and their occurrence in time, while *as* is another all-purposive subordinator.

Naturally, the articles are the most frequent determiners, where *the* occurs more than three times more often than *a* and *an* together. This may be connected to the high degree of modification in noun phrases (see Section 7.2), which results in it becoming definite, since the news is very particular about clarifying the referents of its text. The next most common determiners are: *this*, *two*, *first* and *last*. *This* again functions as a clarifier, while *first* is often used to connect events to previous or subsequent happenings (e.g. *the first visit by a Western leader to the country... 8/8*). *Last* most commonly occurs as an indicator of the time when an event took place (e.g. *last week, last month, last July*).

As for pronouns, third person pronouns head the list, where singulars are twice as frequent as plurals, and masculines are more than ten times more common than feminines. This reflects the common subject matter of news to deal with the doings of prominent individuals and not so much of groups, and also shows the higher presence of men on the international political/military scene. Next come the relative pronouns *which*, *who* and *that*, again pointing at the efforts made in news to maximally clarify the identity and characterise referents. Quantifiers are found to be very common altogether, although none of them is frequent on its own. This high rate corresponds to the news' interest in numbers and amounts.

As for modals, *will* is by far the most common one, followed by *would*, *could* and the semi-modal *be expected to*.. Besides *can*, *should*, *must*, and *may*, two more

semi-modals, *be to* and *be due to*, are also included in the ten most common modals. The fact that the most frequent modals express prediction is again in consonance with news being interested in expectable events. The high frequency of modals roughly corresponds to the overall counts found by Biber et al. (1999:486), except for semi-modals: in the BBC corpus this subclass is extremely common: 24.5% of all modals are semi-modals. This frequency exceeds even the value found by Biber et al. (loc cit) to be characteristic of conversations, where only 18.5% are semi-modals (in printed news, the number is much lower: 8%). An interesting use of a semi-modal is that of the verb *fail*: out of twenty occurrences in the BBC corpus, eight are cases of main verb usage (e.g. *every attempt at reconciliation has failed* 6/16), nine are instances of semi-modal usage (e.g. *the policemen failed to hit the attackers* 6/22), while in three cases, the semi-modal fail is followed by a bare infinitive (e.g. [Cyrus Vance] *will be remembered for the time his diplomacy failed secure the release of American hostages in Tehran* 8/13). As could be seen in Section 3.6.1, the two verbs controlling both *to*- and bare infinitives in printed news are *dare* and *help*; the BBC corpus adds one more such verb to the list (although *dare* is not represented in it at all).

Among the prepositions, *of* is on top of the list, the next most common being *in*. The high rate of the preposition *of* is related to its genitive usage, be it in a possessive usage (*the level of violence*), a partitive meaning (*the south of the country*), in an appositive usage (e.g. *the town of Tulkarm, the aim of signing a ceasefire*), or following nominalisations (*the mediation of the United Nations*), again connected with ways of referent characterisation in news. The preposition *in* is mostly used to express physical location (*in Texas*) or, slightly less frequently, to refer to time (*in March*); the place and time of events being central in news.

7.2 Phrase-grammar

Turning to noun phrases, it has been found that 53.7% of all noun phrases is one or two, 46.3% is three or more words long, which means that slightly less than a half of the nouns are modified (with the reservation that, out of the two-word noun phrases, there is no way to know if they contain a determiner plus a head noun or a zero determiner, a premodifier and a head noun). This is slightly lower than Biber et al.'s (1999:578) results for written news where 55% of all the noun phrases were found to

display some kind of modification. The record in the BBC corpus is a 42-word long noun phrase (counted in orthographic words):

*[the resignation of the country's Prosecutor General, Shin Seungnam, who offered to step down on Sunday after the arrest of his younger brother on suspicion of receiving bribes from a businessman now facing trial on charges of embezzlement and stock price manipulation]*_{NP} (7/14)

These results point to the fact that radio news texts employ all kinds of methods to convey as much information as they can in the shortest time possible. The above noun phrase is typical in the sense that it contains a nominalisation (a verbal noun ending in *-tion*), which is one of the ways of making the text denser. Another frequent feature of radio news noun phrases is the use of genitives (the above example was one of an *of*-genitive), another means of integration. Consider the following examples of premodification by genitives:

the government's refugee policies (9/23)

an international donors' conference (8/25)

the weekend's volcanic eruption (12:30/22)

It can be observed that the three genitives are of different types: the first is a specifying genitive, the second is a classifying one, while the third is a special type expressing the time of the event marked by the following noun. This third type is evidently another means of integration, since it comprises a whole idea which could have been expressed by an entire separate clause (*the volcanic eruption which occurred last weekend*). In the BBC corpus, almost 10% of all 's-genitives are such 'temporal' ones.

As for prepositional phrases, 36.8% of all occurrences are maximum 3 words long, 63.2% are four or more words long, which means that roughly a little less than 40% is made of a preposition and a simple noun phrase, and a little more than 60% contains complex noun phrases. This, again, stands close to Biber et al.'s findings (1999:791) about printed news. (As a reminder: these prepositional phrases talked about are not used as postmodifiers, but as adverbials; those included in noun phrases have not been considered here.) The maximum length of prepositional phrases is 33 words, which occurs in four cases. An example is:

*[to an announcement by Palestinian police that they have arrested the man who heads Israel's list of most wanted men, the leader of the Popular Front for the Liberation of Palestine, Ahmed Saadat]*_{PP} (8/16)

Looking at the syntactic function of adjectives, it has been found that the great majority (92.7%) appear in attributive usage, with just 7.3% in predicative usage, which is even a stronger preference for premodifying adjectives than that found by Biber et al. (1999:506) for printed news, where the rate was 84%. The overwhelming majority of adjectives occur in their base form, with only 1% in comparative, and 4% in superlative forms. This usage also reflects the Biber et al.'s (1999:524) results for the higher frequency of superlatives over comparatives, probably reflecting a focus on the extreme of news in general.

As regards the grammar of the verb phrase, the tense usage is the following: of all the finite verb phrases, 45.5% are marked for the past, 54.5% for the present, so there is no preference for either one over the other. Very similar results have been found in Biber et al. for written news (1999:461). When the aspect is not simple, 77.5% of the aspectually marked verb phrases are in the perfect aspect, and 22.5% in the progressive, so there is a strong preference for the perfective, even stronger than in printed news, where the rate is 71.5% over 28.5%. Given that the perfective aspect expresses actions and events with a strong relation to the present, it is not surprising that radio news, the quickest medium, commonly uses this construction. (A more detailed discussion of this feature follows in Section 7.4.) Of all the verb phrases, 12% are marked for the passive. Biber et al. (1999:476) have found that, in written news, the rate of passives in finite verb phrases is 15%, but unfortunately, they do not indicate the rate of the passive in nonfinite verb phrases, which makes the results incomparable.

7.3 Clause grammar

The results in this part of the discussion are not exact, since the tagger has not been able to analyse the texts for clause constituents and the relationship between clauses. However, to get an idea of the clause grammar of radio news, the following simple calculations were done: The number of main verbs was counted to obtain the rough number of clauses. Then, this number was divided into finite and nonfinite clauses. The number of nonfinite clauses has been added up with that of subordinators

to get the total number of subclauses. This value has been extracted from the total number of clauses, which gave the number of main clauses.

Due to several factors, the calculations are not exact. For example, not all subclauses are introduced by subordinators. Further, clauses may also be verbless, which has not been marked, since nonexistent elements cannot be tagged. No separate tags have been used for the subordinator and relative pronoun function of *wh*-elements. Despite their nominal character, gerunds also have verbal features in producing clauses on some occasions, but acting just as premodifiers on others; similarly, present and past participles can act as modifiers or adverbial clauses etc. All these considered, the conclusions reached through the following numbers can at best be regarded as tentative.

A total of 5478 clauses have been identified in the BBC corpus, 65.5% of which are finite, and 34.5% non-finite. Of the latter, 38.5% are infinitival clauses, 32.5% are past participial clauses, 17% gerundial clauses, and 12% present participial ones. Out of the finite clauses, 56% have been found to be main clauses and 44% subclauses. If we compare the rate of finite and non-finite subclauses, we find that there are more finite than non-finite ones, which places radio news closer to conversations than printed news. The total number of subordinates – finite and non-finite together – considerably exceeds the number of main clauses, in fact, only one-third of the clauses are superordinate. The fact that there is a perceivably high rate of subordination in radio news also approximates it to conversations rather than written news, where grammatical complexity typically resides on phrase-levels.

The instances of coordination are roughly equally divided between phrase-level (52%) and clause-level (48%) coordination. This is similar to Biber et al.'s results for the use of *and* in two functions in written news (1999:81). Out of the clause-coordinators, however, only 81% occur between clauses, 19% are placed in a sentence-initial position. Using *and*, *or*, and *but* at the beginning of a sentence, even just in one-fifth of the cases, is an exclusively spoken characteristic (see their analysis as discourse connectives in Schiffrin 1987, Chapter 6). Out of the subordinators, the use of the complementiser *that* is worth taking into consideration: a hand-count shows that, in 65% of the cases, *that* is deleted. This is a strikingly high rate, since in printed news *that* is only deleted in 30% of the cases (Biber et al 1999:681).

The following are additional, albeit not too common, cases of marked 'spoken language' character in the BBC corpus:

- inverted word order in nominal relative clauses:

It is not clear what are the measures Mr Powell might suggest... (7/16)

- the use of hedges:

Mr Bush sort-of distanced himself from the company's chief executive...

(7/11)

- sentence fragments:

So no last minute deals or compromises. (7/16)

- the deletion of the preposition *for* in time expressions:

...the detainees have been here about a week now... (8/20)

Such cases are rare but existent in the news of the BBC World Service.

7.4 Multi-feature/multi-dimensional results

What follows is a comparison of the values yielded by the BBC corpus with those of Biber's multi-feature/multi-dimensional approach (see Section 2.3.4), taking into consideration the average results of the Biber 1988 corpus as a whole, of press reportage and broadcasts. Since Biber's data have been normalised to a text length of 1,000 words, the same operation has been done to the BBC data to ensure comparability. Because of limited software capabilities, not all the features identified by Biber 1988 could be counted in the present study. Table 7.1 presents the results in a parallel manner. The first three columns contain data obtained by Biber 1988, the first being an overall count and the second and third referring to two different genres. The fourth column comprises the results of the present study.

The main point of the multi-feature/multi-dimensional approach is that it is not the individual features but their co-occurrence that reveals the true nature of a text. Therefore, a lack of the results on the rest of the features and the inability to perform a factor analysis of the BBC corpus mean that the conclusions based on this section can only be tentative. However, the fact that most of the results coincide with those of Biber et al. 1999 enhances the generalisability of the conclusions. Table 7.1 indicates the following:

On the one hand, radio news approximates printed news in the use of the past tense, third person pronouns, modals, and gerunds. The first three features are connected with the narrative concern of radio news. The last one is a means of integrating as much information as possible in a text, which goes along with the severe time-constraints of radio news. The common use of gerunds normally goes along with

nominalisations, which may well be the case in radio news as well, but QTAG 3.01 has unfortunately not been able to count the frequency of these nouns separately.

	Mean	Press reportage	Broadcasts	Radio news
Past tense	40.1	45.1	18.5	40.2
Perfect aspect	8.6	8.0	6.4	19.6
Present tense	77.7	55.8	74.8	70.3
First person pronouns	27.2	9.5	11.8	0.5
Second person pronouns	9.9	1.1	2.7	0.0
Third person pronouns	29.9	28.7	31.7	24.3
Indefinite pronouns	1.4	0.6	0.5	0.7
Gerunds	7.0	7.8	3.8	8.0
Nouns	180.5	220.5	229.8	140.7
Passives	10.4	12.6	4.3	15.1
<i>Be</i> as main verb	28.3	20.7	21.9	10.9
Existential <i>there</i>	2.2	1.8	2.2	1.4
The complementiser <i>that</i>	3.6	3.4	1.2	5.8
<i>Wh</i> -complementisers	6.0	5.0	2.4	8.6
Infinitives	14.9	13.8	9.8	18.0
Present participial clauses	2.6	3.1	1.2	7.8
Past participial clauses	2.6	3.3	4.3	23.7
Prepositions	110.5	116.1	118.0	121.5
Attributive adjectives	60.7	64.5	61.1	63.7
Predicative adjectives	4.7	3.0	3.0	5.0
Adverbs	65.6	52.8	86.3	23.7
Modals	13.5	11.0	5.9	11.7
<i>That</i> -deletion	3.1	2.0	1.3	10.7
Phrasal coordination	3.4	4.1	1.3	11.5
Clausal coordination	4.5	2.3	8.9	10.6

Table 7.1 Descriptive statistics for the frequency count of linguistic features in Biber 1988 and the BBC corpus

On the other hand, radio news stands close to broadcasts in the use of the present tense, which Biber sees as dealing with topics and actions of immediate relevance (1988:224). This corresponds to radio's function to report on activities and events often

parallel to the broadcasting time. Further, the common usage of the present may be related to the high frequency of speech act verbs, which typically occur in the present tense.

More importantly, however, the frequency of certain features points to the similarity of radio news to conversations. The most striking is the rate of nouns: 140.7 as opposed to 220.5 in the press or 229.8 in broadcasts. Biber (1988:262) reports that the frequency of nouns in conversations is 137.4, which is just a little lower than in radio news. A similarly 'conversation-like' result is that of *that*-deletion, whose frequency in radio news is even higher than in face-to-face conversations, where the number is 9.5. It should be noted, however, that the rate of undeleted *that* complementisers is also high (5.8 as opposed to 4.1 in conversations), which goes along with the high frequency of speech act verbs ('public verbs' in Biber 1988:75), especially *say*, which mainly require a *that*-complement clause. And, the more *that*-slots there are, the more can be left out.

Also high is the result concerning predicative adjectives, which, as opposed to attributive adjectives, also indicate a fragmented, speech-like text. Biber has found press reportage to use 4.4%, and broadcasts 4.7%, of its adjectives in a predicative position. In the BBC news, 7.3% of the adjectives are used predicatively. (Here, there is a strong contradiction between Biber (1988:247) and Biber et al. (1999:506): in the latter, printed news has been found to use 16% of its adjectives predicatively.)

Radio news also rates high on clausal coordination, with its value being even higher than that of conversations (10.6 as compared to 9.5). However, radio news shows an unusually high degree of phrasal coordination as well (11.5 as opposed to the highest value in Biber 1988 of official documents, 7.3). That is, radio news presents a remarkably frequent usage of coordination in general, but no preference for its phrasal or clausal type. From this point of view, therefore, radio news rates intermediate between fragmented and integrated texts.

A feature that makes radio news unlike either the printed press, broadcasts or conversations is the extremely frequent use of the perfect aspect. None of the genres – spoken or written – examined by Biber 1988 has such high results in the use of perfect constructions as radio news. (The highest rate of perfect aspect verbs is found in mystery fiction, but even that value is lower than in radio news: 14.9 as compared to 19.6.) Perfect aspect forms are said to mark actions in past time with 'current relevance' (Quirk et al. 1985:189). It seems that the prominently common use of the perfect aspect

is a unique, distinguishing feature of radio news, which may be due to an aspiration of radio news to obtain an up-to-date character and a dramatic effect. Radio, indeed, is the quickest of all media, with the microphone being more mobile than the camera, and recording/broadcasting faster than the press. In this way, radio can afford to keep abreast of the events, which allows it to render actions as being strongly present-related. (Indeed, 83% of the perfective auxiliaries are in the present tense, with only 16% in the past and roughly 1% nonfinite.)

Another result that deserves attention is the use of passives, which is even more preferred in radio news than in the press. (another contradiction between Biber 1988 and Biber et al. 1999). Passives have been taken as one of the most important surface markers of the decontextualised or detached style that stereotypically characterises writing (Biber 1988:228). The high rate of passive constructions contributes to radio news being partly abstract, partly more concerned with the action/event than with the performer.

Interestingly enough, radio news presents an unusually low frequency of adverbs. Given that radio news is otherwise highly concerned with the circumstances of actions and events, the rarity of adverbs may be explained by the probable fact that these adverbials are mainly expressed by other construction, such as prepositional phrases, and not adverbs. The high rate of prepositions seems to justify this supposition. (The most frequent preposition in the BBC corpus is *of*, mainly occurring in genitive constructions, but almost equally frequent is *in*, followed by *to*, *on* and *for*, which, indeed, are able to express place and time.)

Radio news scores extremely low in the use of first and second person pronouns, which corresponds to its impersonal character and a total lack of overt persuasion. The pronoun *you* appears almost exclusively in the phrase *You're listening to the news from the BBC*, which shows 'involvement with the addressee' (Biber 1988:225), but which is not part of news proper. Radio news also shows a very rare use of *be* as a main verb, which is probably due to its more accentuated interest in actions and events than states. This orientation manifests itself in the low frequency of existential *there* constructions, too.

The strangely high rate of present and past participial clauses and the fairly high scores in *wh*-complementisers are probably due to a difference in the way of counting. This study has added up participles occurring as simple premodifiers (e.g. *continuing losses*; *alleged attempt*) as well as those occurring in as postmodifying clauses (e.g. *a*

correspondent travelling with the prime minister; an asteroid discovered just one month ago), adverbial clauses (e.g. *Tony Blair, who arrived in Islamabad from Delhi, trying to reduce tension; Israeli security sources, angered by what they say is a recent attempt by Palestinians to smuggle weapons*), or ‘condensed clauses’ (e.g. a *power-sharing package; Palestinian-administered territories*). Biber 1988 may have counted present and past participles differently, e.g. leaving out one-word participles, which may account for the remarkable differences. As for *wh*-complementisers, Biber 1988 is very precise in distinguishing between *wh*-conjuncts and *wh*-relatives in subject vs. object positions, or preceded by prepositions. QTAG 3.01 used in this study has not made it possible to apply these distinctions in the tagging process. However, if *wh*-clauses are indeed common in radio news, this is another feature that makes it resemble conversations (see Section 4.6.1).

To sum up, a multi-feature/multi-dimensional analysis of radio news, though partially performed here, indicates that radio news approximates printed news, or written language in its use of certain linguistic features (the past tense, third person pronouns, modals, gerunds, and the passive voice), but it also has features which make it similar to broadcasts (the present tense), or spoken language in general (rate of nouns, the conjunction *that* and especially its deletion, and predicative adjectives). Radio news also has some features rating it intermediate between spoken and written genres (phrasal vs clausal coordination). Radio news displays a distinctive linguistic feature typical of the genre: the high frequency of the perfect aspect. Although not all of the statistical calculations required by Biber 1988’s multi-feature/multi-dimensional model could be done, the results obtained here mostly coincide with those of the previous sections, which increases their reliability.

7.5 News length

The lengths of different news types in the BBC WS corpus have been compared. As a reminder: ‘Short’ refers to news summaries, ‘Long’ to average morning news, while a ‘Reported’ news item consists of an introduction and a ‘Report’ from a correspondent.

	Overall	Short	Long	Reported	Reports
1. Length of news items	96.14	46.23	112.57	173.87	113.79
2. Number of sentences	4.32	2.10	5.00	7.96	5.26
3. Length of sentences	22.21	21.96	22.25	21.83	21.61

Table 7.2 The average length of different types of news items in the BBC corpus. Lines 1 and 3 show the number of words (tags).

The results summarised in Table 7.2 show that reports roughly coincide in length with average morning news items (in the number of both words and sentences), and are preceded by an average introduction of 60 words, which make up a mean of 2.7 sentences. The average length of sentences is roughly 22 words per sentence, which is even more than Biber et al.'s result for printed news (20) yielded by a single, though typical, news text (1999:121). The mean length of sentences does not differ considerably in either type of news, which means that the main difference between short and long news arises from the number, and not the length, of sentences.

Table 7.3 below is a count of how many news items generally occur in the same edition. 'Short', again, refers to news summaries, and 'Long' to morning editions, which typically contain both items containing reports and pieces with no recordings from correspondents.

	Short	Long	Overall
Items per edition	7.00	6.78	6.83

Table 7.3 The average number of news items in different editions of BBC news

As can be seen, the editions do not differ to a great extent in this respect. The numbers range from a minimum of 6 items to maximum 8, but the average is around 7 in each type of news. This means that, if short news contains as many items as long news, the difference will stand in the length of items. This, indeed, is the case, as could be seen in Table 7.2.

Figure 7.5 demonstrates the relation of the average news length to the position of the item within a single edition.

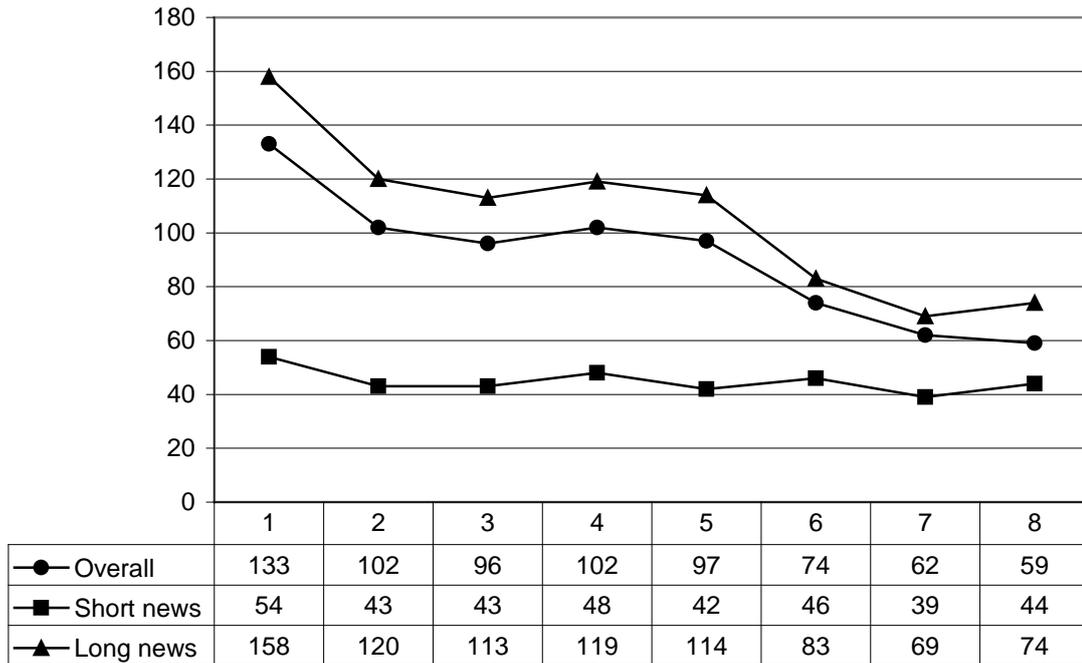


Figure 7.5 The average length of news items in the BBC corpus. The horizontal axis refers to the position of news items (from first to eighth); the vertical axis expresses the length in number of words (tags).

It is apparent that there is a general tendency of front-position news items to be longer, and end-position ones shorter, but this tendency manifests itself to various degrees in different news types: in short news, the number of words per item ranges between 39 and 54 but, in long news, the length varies between 69 and 158 words. At this point, the findings are worth comparing to those shown on Figure 7.6. It can be seen that the position of reports coincides with the length of news items, so the greater vacillation in length of long news items is due to the fact that these are the ones which do or do not contain reports.

The size of the news item might be in direct relationship with its perceived importance. Crisell (1994:85) thinks that, in a news item, longer is not necessarily the same as more important and, in any case, our sense of duration is less certain than our awareness of the spatial length of a press story. Crisell might be right in saying that an item is not perceived as important because it is long, but length may correlate with the importance of the topic an item is concerned with. This issue will be addressed in Chapter 8.

It may be of interest that, while generally the length of items is inversely proportional to its position, that is, the further back it is placed, the shorter it gets, position 4 is an exception: here, there is a jump in length. This is probably due to the fact that the place of medial discourse organising elements, such as *You are listening to the news from the BBC in London* is between positions 3 and 4. Seemingly, these discourse particles serve as attention-attractors, based on the assumption that, after three news items, the listener's attention might flag but, getting it back this way, the programme can 'afford' a longer item again. Or else, this jump in length may be due to the fact that a higher number of correspondent's reports occurs in position 4 than in place 3, as can be seen on Figure 7.6.

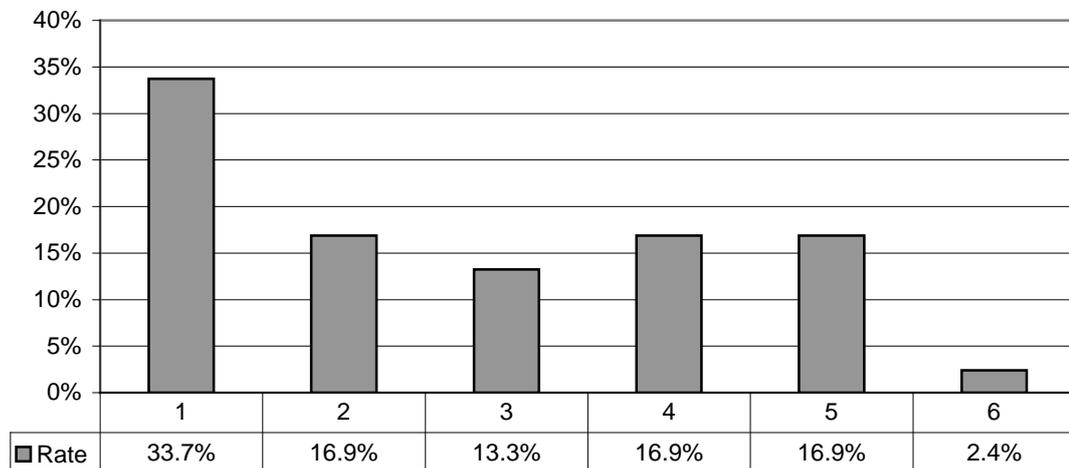


Figure 7.6 The position of reports in the BBC corpus. The horizontal axis refers to the order of news items in the same edition (from first to eighth); the vertical axis indicates what percentage of reports occurs in that particular position (repeated in the lowest row).

7.6 Summary of findings

The grammatical features of radio news are in a direct relationship with the general subject matter of news on the one hand, and the way information is presented in radio news on the other:

The fact that radio news is normally concerned with the doings of prominent personalities who have to be easily identified by the listeners, is shown by a high ratio of determiners and nouns modified by adjectives, prepositional phrases, participles and

relative clauses. The most frequent words in the corpus also reveal the concerns of news to give the circumstances of, and amounts involved in, current events (time and place expressions, quantifiers). The most frequent modals (expressing prediction) are also related to news dealing with expectable events.

Radio news also strives to present as much information as possible in the shortest possible time. Its grammatical features, therefore, display means of integration, such as long sentences, long noun and prepositional phrases, multi-purpose genitives, gerunds and a high number of nonfinite clauses.

Radio news has been found to present some features characteristic of written genres, namely, the length of noun and prepositional phrases, and aspectual marking (a preference for the perfective aspect over the progressive, when not simple). In these features, radio news is much similar to the printed media.

In other features, however, radio news differs considerably from literate language forms:

- In lexical density, radio news stands between written news texts and conversations.
- The average noun/verb ratio of radio news (2.3 nouns per verb) places it between news reportage (3 to 4 nouns per verb) and conversations (a 1 to 1 relationship).
- Clause grammar, that is, the rate of finite vs non-finite subclauses, coordination vs subordination and the retention vs omission of *that*, shows a more marked similarity to spoken than written language.
- The extremely common usage of semi-modal auxiliaries makes radio news especially ‘conversation-like’.

Therefore, the answer to Question 1, whether there are any grammatical features of radio news which approximate it more to spoken than written language (Section 6.1) is yes, quite a number. This alters considerably the picture drawn by previous works, which placed this radio genre totally at the writing end of the oral vs literate continuum.

The study has identified a distinctive feature of radio news, namely, a remarkably high rate of the perfective aspect. This must be in a direct relationship with the aim of radio news to be of an accentuated present relevance. This finding answers

Question 2, if radio news displays any distinctive features, in a positive way: the strong preference for the perfective proves to be a unique characteristic of radio news.

The hypothetical answers given to Question 3 and 4, whether the subgenres differ from each other in any important way, have only partly been justified: reports do not use shorter sentences than other subgenres; their spoken characteristic lies elsewhere. Neither does short news display shorter sentences: there is no difference between the news subtypes in this respect. Nevertheless, reports stand closer to conversations than to the press in their more frequent usage of pronouns, conjunctions and adverbs. The quantitative difference between long (reported or unreported) and short news stands in the number of sentences, not in their length, that is, short news displays fewer, but not shorter, sentences. There is no difference between the number of news items per edition either: the average number is seven both in the two-minute and the five-minute programmes.

In sum, the great majority of the hypotheses suggested in the previous chapter have proved to be true: radio news, indeed, shows appreciable similarities to spoken language both overall and, to a different extent, in its various subgenres.

8 Textual features of radio news: Results and discussion

The previous chapter has examined the grammatical aspects of radio news and what the rate of different word classes reveals about the spoken vs written nature of this journalistic genre. This chapter will turn to textual aspects, in an attempt to find the answers to the qualitative questions outlined in Section 6.2.

8.1 Topics in the BBC corpus

The topic-assignment has, in many cases, been a matter of compromise. As could be seen on Figure 6.1 sketching the general structure of a topic area in the BBC corpus (Section 6.5), when one attempts to capture such a divergent phenomenon as the topics covered by the news, several complications set in:

First, not every topic has subtopics, only those which are dealt with over several news items. In other words, if a topic is only discussed in a single news item, that topic is not divided into subtopics.

Second, not every topic belongs to a topic group; certain topics are directly controlled by the topic area. Some of these topics are insignificant in their coverage amount, and have nothing in common, except for the broad topic area they belong to. For the sake of simplicity, in a presentation of all topic areas – such as in Table 8.1 below – the category ‘Other’ has been introduced as a topic group, to comprise these small topics. In other cases, a topic is so extensively dealt with in the BBC corpus – such as the Afghan financial crisis – that it should occupy a prominent position in the system, this is why no topic group has been assigned to it. (For example, the Afghan and the Argentine financial crises could have been contracted into a single ‘Economic crises’ topic group, but this would have added nothing vital to the system.)

Third, there are many topics that could have been grouped under several headings. For example, a news item about a separatist Muslim leader, who faces trial in the Southern Philippines, belongs to the topic group Legal actions against terrorist suspects, but the next item, which reports on clashes between the protesting supporters of the accused and the police, could also classify as a member of the category Protest. Or an item about Argentinians marching because of the strict measures taken by the government to tackle the financial crisis could again rank as an item of the Protests

category. As an extreme example, the news dealing with Greenpeace activists who face trial because they tried to prevent a military missile test in protest could theoretically belong to four topic areas: Legal issues/Court cases, Protests, Environmental issues, and Military issues. In such cases, the dilemmas were resolved on the basis of a – partly subjective – judgement of which aspect of the event is more relevant, a relevance which can also be signalled by the formulation of the news item. Consider the texture of the first sentence of the item mentioned above:

(8) *Fifteen Greenpeace activists and two journalists are due to go on trial in the United States today for allegedly attempting to interfere with a test launch of a missile from a military base in California. (7/8)*

Here, the main new information – the rheme – of the sentence is that the activists and journalists are facing trial. The reason for the charges only appears in the continuation, the environmental aspect is only signalled by the affiliation of the activists, and the term ‘military’ only appears as an attribute of the location. But the legal case is explicitly formulated. This is why the item has been placed in the Legal issues topic area.

Fourth, further topic areas could have been formulated, such as Human rights issues, which may have included many topics grouped elsewhere. For example, news items connected with the detention conditions of Taleban and al-Qaeda terrorist suspects in a US military base in Cuba, or with hunger strikes in Australian refugee camps against asylum claim processing time, could have made part of such a category. But then, ultimately, one might as well formulate as many categories as many topics are assigned. Therefore, an attempt has been made to render the topic structure of the BBC corpus as simple as possible, while trying to retain the explanatory power of the model.

Even so, in two cases, the structure of the topic groups is not so neat as may be expected. First, in the case of the Crime topic group of the Legal issues topic area, two additional ‘topic subgroups’, Murder and Corruption, have been added to account for the similarity of certain topics. Second, in the Afghan financial crisis topic of the Economic issues area, the subtopic International Donors’ Conference for raising funds seems to govern several others, and this has been indicated in the table without the clumsy ‘sub-subtopic’ heading. As can be expected, these inconsistencies occur in the most extensively covered topic areas, whose structure, due to the wide range of topics that are touched upon, will be more complicated than that of less preferred topic areas,

and this complexity will find its way to the present representation as well. Such complexities are the true representations of the disparate threads of topics in the topic area.

All in all, Table 8.1 contains the topic areas, groups and/or topics that have been identified in the BBC corpus. The order of the topic areas follows a decreasing order of their coverage amount, based on the number of items dealing with the particular topic area, which yields the rate of its coverage presented in the right-most column of the table.

If we start off from the assumption that the relevance of a topic is in direct proportion with its amount of coverage in the media, it is apparent that Legal issues and Conflicts are the most prominent topic areas in the BBC corpus: together they make up for almost half of the whole database. Economic issues are also important, they are responsible for almost one-sixth of the news in the BBC corpus. Next in the row are Human interest stories, Military, Nature-related and Immigration issues: they are of roughly equal importance. Less significant topic areas are Protests, Media-related issues, Elections, Anniversaries, and Scientific achievements: their coverage rates are all below 5%. However, the coverage rate of the different topic groups/topics (without topic areas) is worth looking at, the order of which yields a slightly different picture.

Table 8.2 indicates that, for example, not all Legal issues topics are more important than Conflicts or Economic issues: certain topic groups belonging to lower categories are in fact more prominent in terms of coverage than higher category topics. It is proposed that Table 8.1 gives a just picture of the thematic interests of the BBC World Service news in general, while Table 8.2 reflects the thematic interests of the January 2002 news of the BBC World Service. The high percentage of the topic group Legal actions against terrorist suspects seems natural, four months after the September 11 events. Undoubtedly, a database collected in another time period would yield different results what concerns the topic groups/topics, but would probably display a similar representation of the topic areas in the news. That is, Table 8.2 would be different, but Table 8.1 similar in, say, January 2003.

TOPIC AREAS	TOPIC GROUPS / TOPICS	RATE (as a percentage)
1. Legal issues	Legal actions against terrorist suspects	9.3
	Crime	5.5
	New legislation in Zimbabwe	5.5
	Court cases	3.8
	<i>Total of legal issues</i>	<i>24.1</i>
2. Conflicts	Israeli–Palestinian conflict	8.7
	Conflict resolution attempts	6.0
	Colombian civil unrest	4.9
	Other conflicts	2.7
	Afghan–Pakistani conflict	0.8
	Iraqi–Iranian conflict	0.4
	<i>Total of conflicts</i>	<i>23.5</i>
3. Economic issues	The Afghan financial crisis	6.7
	The Argentine financial crisis	4.2
	The Enron scandal	2.9
	Other economic issues	1.9
	<i>Total of economic issues</i>	<i>15.7</i>
4. Human interest stories	Well-known personalities	4.0
	Significant events	2.2
	Other human interest stories	1.6
	<i>Total of human interest stories</i>	<i>7.8</i>
5. Military issues	Asian military issues	1.7
	Attacks on diplomatic outposts	1.7
	American military issues	1.6
	Russian military issues	1.6
	African military issues	0.5
	<i>Total of military issues</i>	<i>7.1</i>
6. Nature-related issues	Disasters	5.1
	Environmental issues	1.4
	<i>Total of nature-related issues</i>	<i>6.5</i>
7. Immigration issues	Australian refugee camp strike	3.0
	Other immigration issues	2.0
	<i>Total of immigration issues</i>	<i>5.0</i>
8. Protests	General strike in Nigeria	1.6
	Other protests	1.4
	<i>Total of protests</i>	<i>3.0</i>
9. Media-related issues	Other media-related issues	1.1
	Last independent Russian TV station	0.9
	<i>Total of media-related issues</i>	<i>2.0</i>
10. Elections	Election preparations in Sierra Leone	0.9
	Other elections	0.9
	<i>Total of elections</i>	<i>1.8</i>
11. Anniversaries	Second World War anniversaries	0.9
	Other anniversaries	0.9
	<i>Total of anniversaries</i>	<i>1.8</i>
12. Scientific achievements	Medical achievements	0.9
	Astronomical achievements	0.8
	<i>Total of scientific achievements</i>	<i>1.7</i>
Total		100

Table 8.1 Topic areas in the BBC corpus and their rate of coverage (as a percentage, in a decreasing order)

The verification of this hypothesis needs further research, but the assumption is supported by the consonance that the above topic areas bear with news values in general. Bell (1991:156-158) gives the following list of values in news actors and events:

- Negativity: the basic news value that makes a number of concepts such as damage, injury, death, disasters, accidents, conflict between people, political parties or nations;
- Recency: the best news is something which has only just happened, but at least within the past 24 hours;
- Proximity: geographical closeness can enhance news value;
- Consonance: compatibility with preconceptions about the social group or nation from which the news actors come;
- Unambiguity: the more clear-cut a story is in terms of facts, the more it is favoured;
- Unexpectedness: the unpredictable or rare is more newsworthy than the routine;
- Superlativeness: the biggest building, the most violent crime, the most destructive fire gets covered;
- Personalisation: something which can be pictured in personal terms is more newsworthy than a concept, a process, the generalised, or the mass;
- Eliteness: reference to elite news actors, such as politicians, film stars or nations make news;
- Attribution: the eliteness of a story's sources, e.g. a socially validated authority, can be crucial in its news chances; and finally
- Facticity: the high degree of facts and figures on which hard news strives, e.g. locations, names, sums of money, numbers of all kinds.²²

It shows that the topic areas preferred by the BBC corpus are of high news value: Legal actions, Conflicts, Nature-related issues, many Economic issues, and Protests deal with negative events, Human interest stories and Military issues deal with elite characters and nations. Elections and Anniversaries also reflect the media's interest

²² Bell (1991:14) defines 'hard news' as the core news product: reports of accidents, conflicts, crimes, announcements, discoveries, and other events which have occurred or come to light since the previous issue of a paper or programme.

in political and historical events. The place occupied by the Scientific achievements category is in concordance with what Bell has to say about it, namely, that science is a low-priority news area, but gains coverage when there is a breakthrough to report (1991:157).

Topic groups/Topics	Rate (%)
Legal actions against terrorist suspects	9.3
Israeli–Palestinian conflict	8.7
The Afghan financial crisis	6.7
Conflict resolution attempts	6.0
New legislation in Zimbabwe	5.5
Crime	5.5
Disasters	5.1
Colombian civil unrest	4.9
The Argentine financial crisis	4.2
Well-known personalities	4.0
Court cases	3.8
Australian refugee camp strike	3.0
The Enron scandal	2.9
Other conflicts	2.7
Significant events	2.2
Other immigration issues	2.0
Other economic issues	1.9
Asian military issues	1.7
Attacks on diplomatic outposts	1.7
Other human interest stories	1.6
American military issues	1.6
Russian military issues	1.6
General strike in Nigeria	1.6
Environmental issues	1.4
Other protests	1.4
Other media-related issues	1.1
Last independent Russian TV station	0.9
Election preparations in Sierra Leone	0.9
Other elections	0.9
World War Two anniversaries	0.9
Other anniversaries	0.9
Medical achievements	0.9
Afghan–Pakistani conflict	0.8
Astronomical achievements	0.8
African military issues	0.5
Iraqi–Iranian conflict	0.4
<i>Total</i>	<i>100</i>

Table 8.2 The coverage rate of topic groups/topics in the BBC corpus (as a percentage, in a decreasing order)

The topic areas that may be considered unpredictable given the general news values related above are the Immigration and the Media-related issues. The first category might be explained by a certain sense of threat caused by massive immigration that can be detected in First World societies, but increasingly in Second World countries as well (see the analysis of the Hungarian counterpart of the term ‘alien’ in Barát’s study reviewed in Section 1.3.2). The second area is legitimated by a sort of self-interpretation of the media: most news items belonging to Media-related issues are concerned with cases when the freedom of the press has been violated in some way. Giving news of such cases reinforces the BBC’s ‘free’ position in a democratic country. Thus, the newsworthiness of both categories must be connected to the news value of Relevance: relevance to the audience, in the case of Immigration issues, or relevance to

the newsmakers – and, indirectly, to the audience as well, because only a ‘free’ radio is worth listening to – in the case of Media-related issues.

Figure 6.1 in Section 6.5 has given a theoretical structure of a topic area. To show what this theoretical structure looks like in practice, the structure of the Human interest stories topic area is presented in Table 8.3. The topic structure of all other topic areas can be found in Appendices 2 to 13.

Topic area	Topic group	Topics	Subtopics
Human interest stories	Well-known personalities	President Bush has medical problems	Faints from pretzel in White House (1) Further checks after fainting spell (2)
		Saint Laurent retires (3) Prince Harry consumes drug and alcohol (3) Former US State Secretary dies (2) Muhammad Ali is 60 (2) Peggy Lee dies (2) British Chancellor’s daughter dies (1) Pavarotti performs after mother dies (1) Mariah Carey paid to end contract with EMI (1)	
	Significant events	Australian Open tennis championship winners announced	Capriati wins women’s finals (2) Johansson wins men’s finals (2)
		Golden Globe film festival winners announced (2) Detroit motor-show opens (2)	
		Bereaved Americans grieve with Afghans in mourning (2) US town borrows snow from another (1) New Zealand schoolboy excels in golf (1) Prostitution area in Venice assigned (1) Winning lottery ticket searched in Australian town rubbish tip (1) Thailand students to undergo drug tests (1) Malaysian university population to declare allegiance to authorities (1)	

Table 8.3 Topic structure of the *Human interest stories* topic area (topics in decreasing order of coverage amount; subtopics in order of broadcast). The numbers refer to the news items dealing with the particular (sub)topic.

Table 8.3 displays all the important characteristics of the BBC corpus topic areas: not all topics arch over several subtopics, and not all topics have topic groups that govern them. Here, the last two topics could have been assigned a separate topic group, such as Higher education issues, but this possibility has been rejected on the basis that neither of the two items are in fact concerned with educational problems, but rather with human rights and personal freedom. Moreover, its drug-relatedness might as well have paired the Thailand item with the one concerned with Prince Harry's drug consumption. In such cases, a simple common-sense principle has been followed in the analyses.

8.2 Topics and positions

This section seeks the answer to the question if there is any correlation between two variables in the BBC corpus: the topic of the item, as defined in Section 8.1, and the place it occupies within the news bulletin. We have seen in Table 7.3 in the previous chapter (Section 7.5) that the average number of items in a news edition is around 7, the minimum being 6 and the maximum 8. Figure 8.1 represents the position of news items according to the twelve topic areas identified in the previous section. Below the graph the exact number of items on the various topics are indicated.

The top line of the chart is straight, because six items are sure to be present in every bulletin, be it made up of short or long items. After position 6 the top line drops according to the decreasing number of editions consisting of seven or eight items. Figure 8.1 demonstrates the following:

Position 1 features five topic areas: Conflicts, Economic issues, Legal issues, Nature, and Military issues, in decreasing order of frequency. However, the numbers of items occurring in this position, ranging from 18 to 6, do not show as great a difference as in position 2, for example. Items belonging to the first three areas in first position are 18, 15 and 14 in number. This suggests that position 1 is always reserved for the most recent event that deserves attention, and not so much dependent on which topic it covers.

Position 2 shows greater polarisation: Legal issues and Conflicts appear with 25 and 19 items, respectively, but the next most common topic areas in this place, Economic and Military issues, are only present with 8 and 5 items, in the corresponding order. Starting from these, a dramatic fall is noticeable: Human interest stories, Nature-related, Media-related issues and Protests are only represented by 1 item each.

The variations in position 3 are slightly more levelled: the order is Legal issues (20 items), Conflicts (18), Economic issues (10) and Military issues (5). Two new topic areas pop up: Immigration issues (4 items) and Elections (2). The row closes up with two more areas: Human interest stories (2), and Nature-related issues (1).

Positions 4 to 7 already display all but one topic area, but this missing one is different in almost all four: Anniversaries in positions 4 and 7, Scientific issues in place 5, and Elections in position 6. Otherwise, the numbers of items on various topic areas show a fairly smooth decrease in positions 4 to 6. In position 4 this is: Conflicts (17), Legal issues (13), Economic issues (10), Military issues (5), Protests (4), Human interest, Immigration and Nature-related issues (3 each), Elections (2), Media and Scientific issues (1 each).

In position 5 the order of preference is: Legal issues (15), Economic issues (14), Conflicts (11), Human interest stories (6), Protests, Military and Immigration issues (3 each), Nature-, Media-related issues and Anniversaries (2 each) and finally Elections (1).

In position 6 we find: Legal issues (15), Conflicts (14), Economic issues (6), Nature-related issues (5), Military, Scientific issues and Anniversaries (4 each), Human interest and Immigration stories (3 each), and finally Media and Protest news (2 each).

Up to now, there is much consonance in that Legal, Economic issues and Conflicts are responsible for the greatest number of items. Position 7, however, shows a significant change: this is the first place where the most popular item is not hard news, but Human interest stories, represented by 15 items. After a considerable jump, the next most favoured topic area is Nature-related issues (6), then Legal and Economic issues (5 each), Conflicts (4), Elections, Media and Scientific issues (2 each), and Protests, Immigration and Military issues at the end of the row with a single item each.

Position 8 is similar to 1 in that the number of topic areas covered falls to five (note that position 8 is not compulsory to be filled in, while position 1 is essential). The leading position of Human interest stories is perceptible here as well (4 items). Four other topic areas are represented with one item each: Economic issues, Conflicts, Media- and Nature-related issues.

Figure 8.1 has shown the relationship between topic and place from the point of view of positions. Figures 8.2 to 8.13 represent this variation from the angle of topics, supplementing the information gained from Figure 8.1 with breaking the topic areas

down to topic groups. Topic areas are discussed in decreasing order of coverage amount.

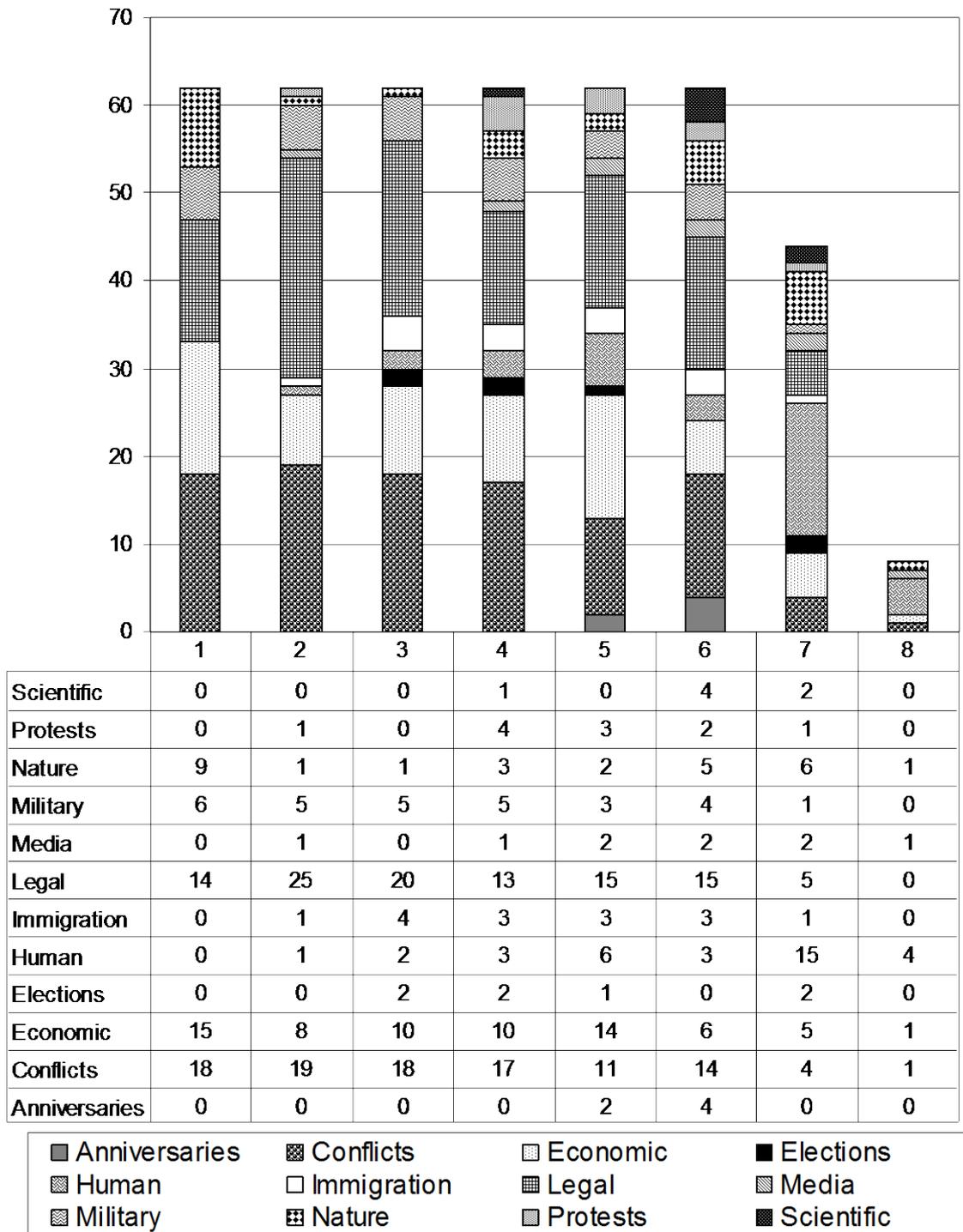


Figure 8.1 Topic areas and positions in the BBC corpus. The horizontal axis indicates the position of news items within the edition (from 1 to 8); the vertical axis shows the number of news items occurring in that particular position. Notation: *Scientific* = Scientific achievements; *Nature* = Nature-related issues; *Military* = Military issues; *Media* = Media issues; *Legal* = Legal issues; *Immigration* = Immigration issues; *Human* = Human interest stories; *Economic* = Economic issues; otherwise identical with content.

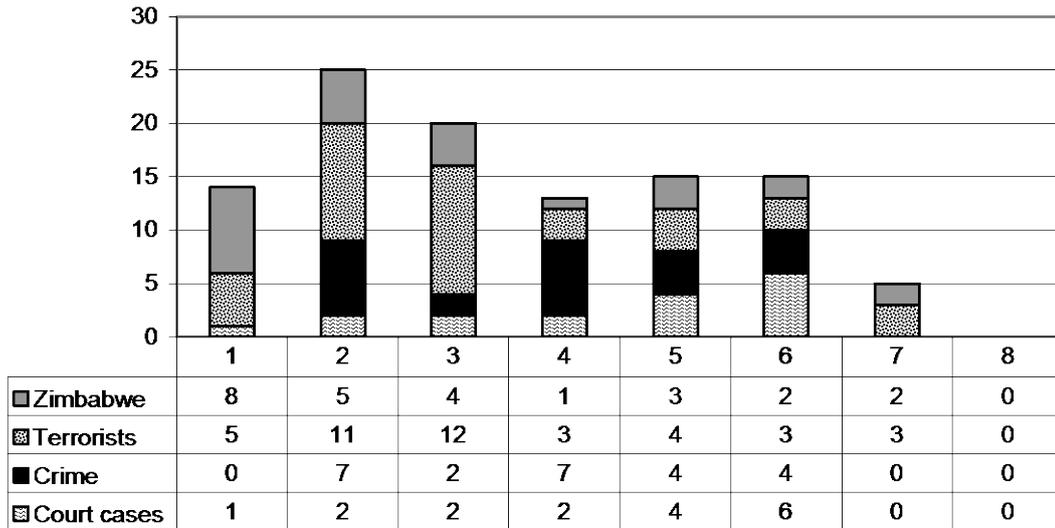


Figure 8.2 Legal issues – topic groups and positions. The horizontal axis indicates the position of news items within the edition (from 1 to 8); the vertical axis shows the number of news item occurring in that particular position. Notation: *Zimbabwe* = New legislation in Zimbabwe; *Terrorists* = Legal actions against terrorist suspects; otherwise identical with content.

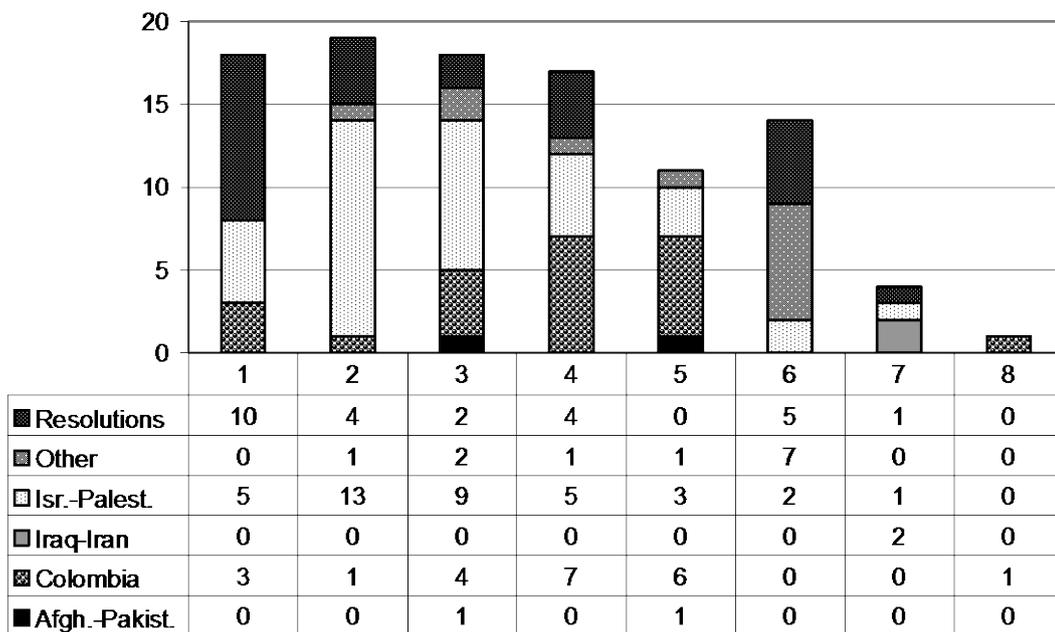


Figure 8.3 Conflicts – topic groups and positions. The horizontal axis indicates the position of news items within the edition (from 1 to 8); the vertical axis shows the number of news item occurring in that particular position. Notation: *Resolutions* = Conflict resolution attempts; *Isr.-Palest.* = Israeli–Palestinian conflict; *Iraq-Iran* = Iraqi–Iranian conflict; *Colombia* = Colombian civil unrest; *Afgh.-Pakist.* = Afghan–Palestinian conflict; *Other* = Other conflicts.

Figure 8.2 indicates the distribution of items on Legal issues across positions. The highest number of Legal news occurs in positions 2 and 3, but none is placed in position 8. The most common topic group in the area, Legal actions against terrorist suspects, arches through all positions from 1 to 7, and the same is true of the third most preferred topic: New legislation in Zimbabwe. Crimes, although high in number, are concentrated in mid-positions from 2 to 6. Court cases, although lower in number, are present in all positions except 7.

Figure 8.3 shows that, while Legal news tops in positions 2 and 3, Conflicts are much more evenly distributed. Almost the same amount of Conflict news occurs in the first four places. There is a slight drop in position 5, otherwise the decrease is smooth. Conflict is a topic area which is covered in all positions. The inner distribution of the topic area displays interesting aspects: although Conflict resolution attempts is the second most favoured topic group in the area, it is responsible for the great majority of items occurring in first position. The most popular topic group in the area, the Israeli–Palestinian conflict, tops positions 2 and 3. Colombian civil unrest, third in order, surfaces mostly in positions 4 and 5, and it is also responsible for the only item in position 8. Place 6 mainly features Other conflicts, while the Iraqi–Iranian conflict is only present in position 7. The Afghan–Palestinian conflict is rather arbitrarily placed 3 and 5.

Figure 8.4 is a representation of the places occupied by Economic issues. The two most prominent topics in the area, the Afghan and the Argentine financial crises, are roughly equally divided along positions 1 to 6, with the Aghan issues peaking in position 1 and the Argentine problems in 5. The handling of the Enron scandal is interesting: when it is present, it mostly appears in position 1, with another top in place 5. Other economic issues are placed towards the end of programmes, in positions 4 to 8, with the last two places reserved entirely for these various topics.

Figure 8.5 switches to Human interest stories. The change observed above is apparent on the bar-chart as well: these stories peak in place 7, and none of them appears in top position. It is the items dealing with the deeds of Well-known personalities that are present through positions 3 to 8, with two peaks in positions 5 and 7. News about Significant events is concentrated in places 5 to 8. Other human interest stories are quite arbitrarily dispersed.

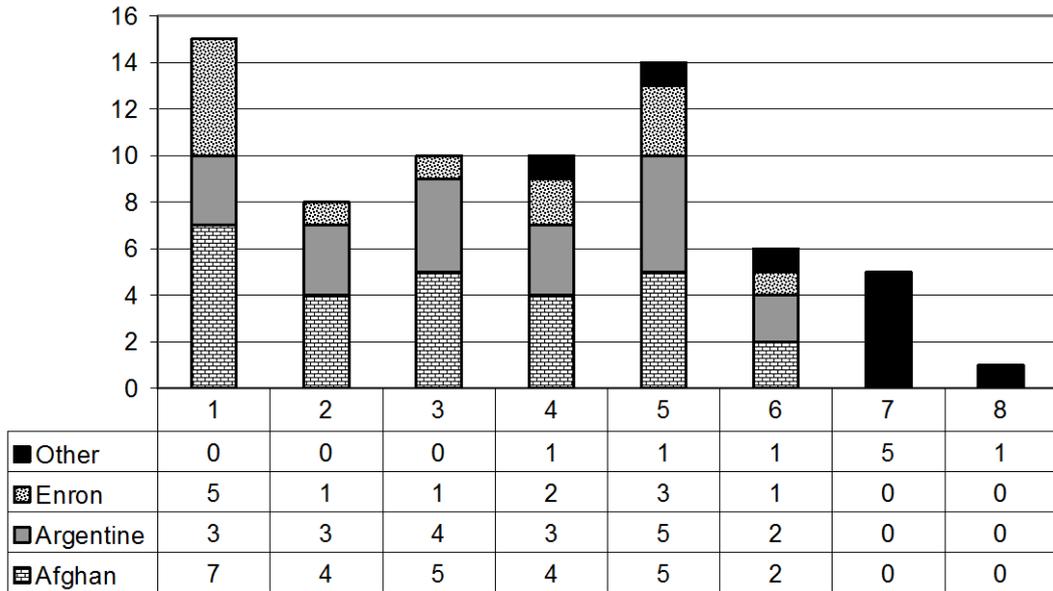


Figure 8.4 Economic issues – topics and positions. The horizontal axis indicates the position of news items within the edition (from 1 to 8); the vertical axis shows the number of news item occurring in that particular position. Notation: *Enron* = The Enron scandal; *Argentine* = The Argentine financial crisis; *Afghan* = The Afghan financial crisis; *Other* = Other economic issues.

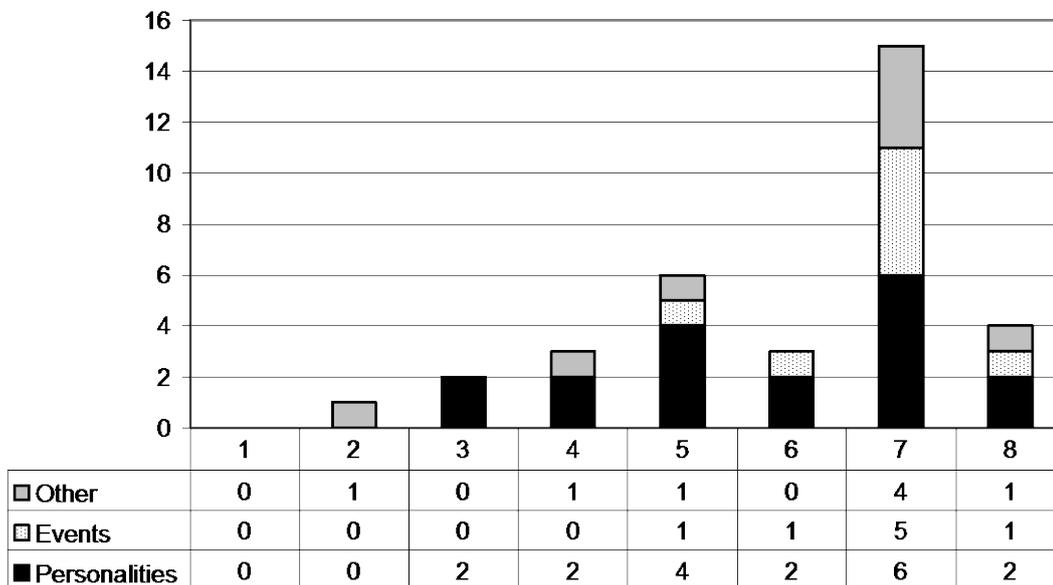


Figure 8.5 Human interest stories – topic groups and positions. The horizontal axis indicates the position of news items within the edition (from 1 to 8); the vertical axis shows the number of news item occurring in that particular position. Notation: *Events* = Significant events; *Personalities* = Well-known personalities; *Other* = Other human interest stories.

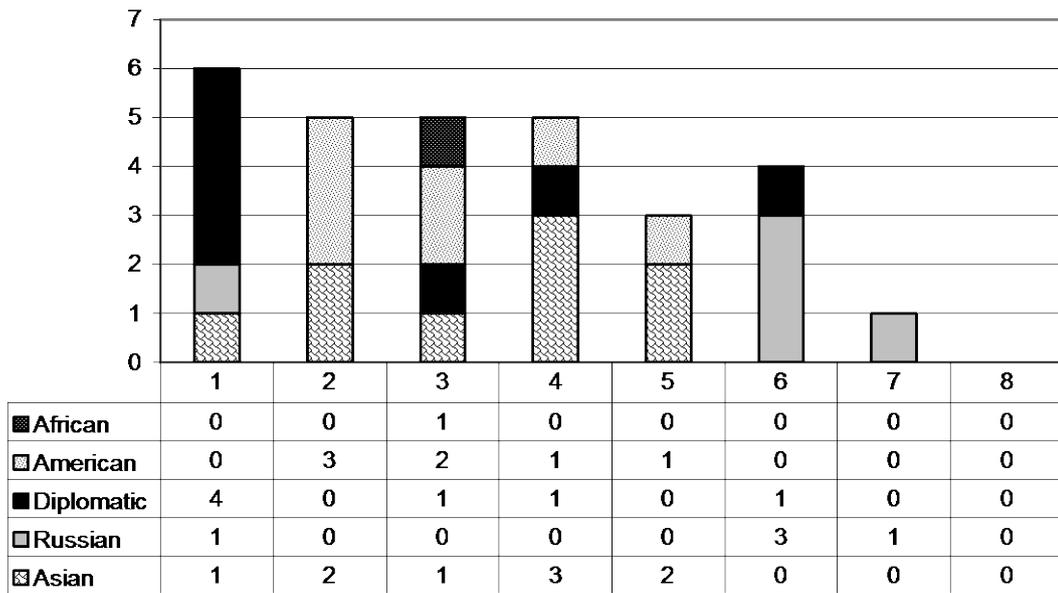


Figure 8.6 Military issues – topic groups and positions. The horizontal axis indicates the position of news items within the edition (from 1 to 8); the vertical axis shows the number of news items occurring in that particular position. Notation: *African* = African military issues; *American* = American military issues; *Diplomatic* = Attacks on diplomatic outposts; *Russian* = Russian military issues; *Asian* = Asian military issues.

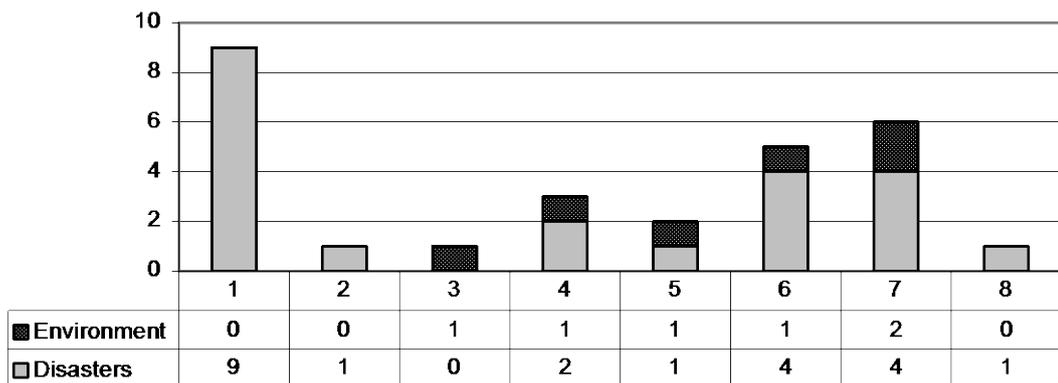


Figure 8.7 Nature-related issues – topic groups and positions. The horizontal axis indicates the position of news items within the edition (from 1 to 8); the vertical axis shows the number of news items occurring in that particular position. Notation: *Environment* = Environmental issues; otherwise identical with content.

Figure 8.6 turns to Military issues. This area is represented in the first position as well, and equally covered in positions 2 to 4. Interestingly enough, while the most popular topic group, Asian military issues, is fairly smoothly distributed through positions 1 to 5, the second most covered topic group, Attacks on diplomatic outposts, is responsible for the great majority of the items in first position, and represented by one item each in positions 3, 4 and 6. American military issues appear in places 2 to 5, while the Russian ones in 6 and 7, with a single one in first position. The single item on African military issues shows up in mid-position 3.

Figure 8.7 is concerned with Nature-related issues. This is another topic area that is present in all positions. It is eye-catching, however, that its first group, Disasters – present in all positions except 3 – displays almost half of its items in position 1. Thus, when the BBC World Service reports on calamities, they are likely to move up to the top of the bulletin. The other topic group, Environmental issues, is almost evenly distributed across positions 3 to 7.

Figure 8.8 deals with Immigration issues. The Australian refugee camp strikes are grouped in mid-positions, equally divided between positions 3 to 5, with an additional item in position 7. Other immigration issues are concentrated in position 6, with one item each in positions 2 and 3.

Figure 8.9 illustrates the positioning of Protest news. Similarly to Immigration issues, these again tend in mid-positions (4 to 6), but the different topics are more evenly dispersed. However, the more prominent topic, General strike in Nigeria, once moves up to position 2, while a less important topic from the category ‘Other’ slides to position 7.

Figure 8.10 shows the positions of Media-related issues. These are special in the sense that, although they rate low on the topic area priority list, the items are found in almost all positions (except 1 and 3), in the same amount. This omnipresence must be in connection with the fact that, in an indirect way, when the BBC World Service deals with media issues, it in fact deals with itself. The positioning is not systematic; the topics are dispersed rather casually.

Figures 8.11 to 8.13 feature the last three topic areas: Elections, Anniversaries and Scientific issues. None of them are privileged in the BBC corpus: besides their low coverage, their positions are not too prominent either. Only Elections shows a fairly even distribution over positions 3 to 7 (except 6), with the more important topic, Election preparations in Sierra Leone, occupying the higher places. Anniversaries are

exclusively placed in the less distinguished positions 5 and 6. Scientific issues are even more neglected, sliding to positions 6 and 7, with a single Medical issue in position 4.

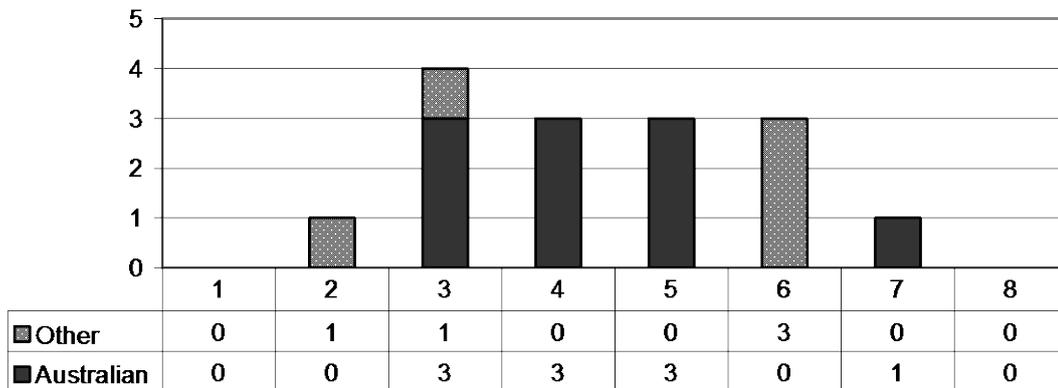


Figure 8.8 Immigration issues – topics and positions. The horizontal axis indicates the position of news items within the edition (from 1 to 8); the vertical axis shows the number of news item occurring in that particular position. Notation: *Australian* = Australian refugee camp strikes; *Other* = Other immigration issues.

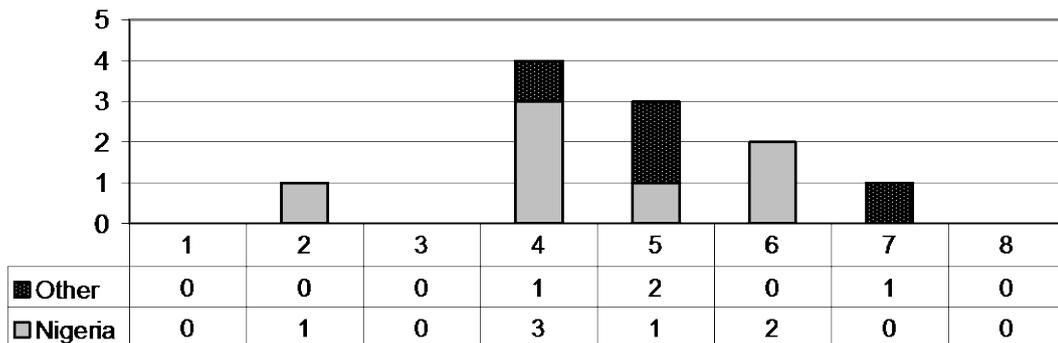


Figure 8.9 Protests – topics and positions. The horizontal axis indicates the position of news items within the edition (from 1 to 8); the vertical axis shows the number of news item occurring in that particular position. Notation: *Nigeria* = General strike in Nigeria; *Other* = Other protests.

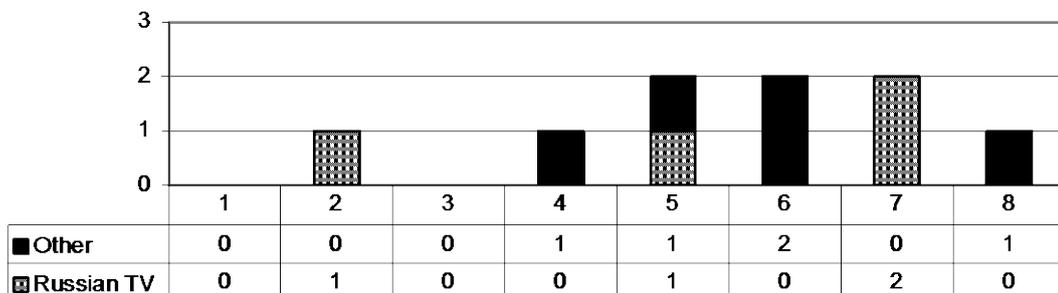


Figure 8.10 Media-related issues – topics and positions. The horizontal axis indicates the position of news items within the edition (from 1 to 8); the vertical axis shows the number of news item occurring in that particular position. Notation: *Russian TV* = The last independent Russian TV station; *Other* = Other media-related issues.

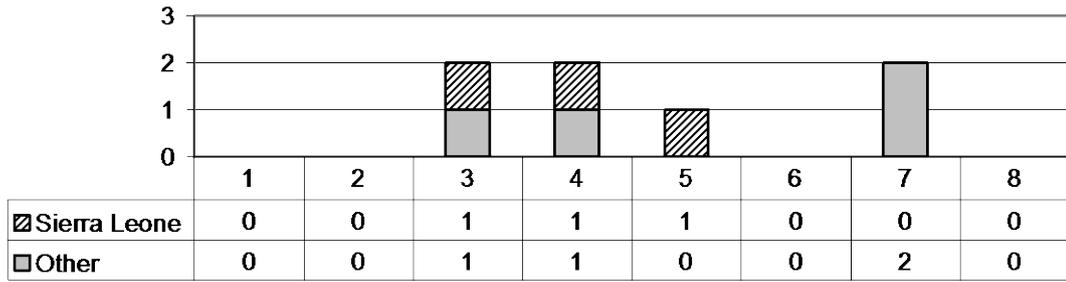


Figure 8.11 Elections – topic groups and positions. The horizontal axis indicates the position of news items within the edition (from 1 to 8); the vertical axis shows the number of news item occurring in that particular position. Notation: *Sierra Leone* = Election preparations in Sierra Leone; *Other* = Other elections.

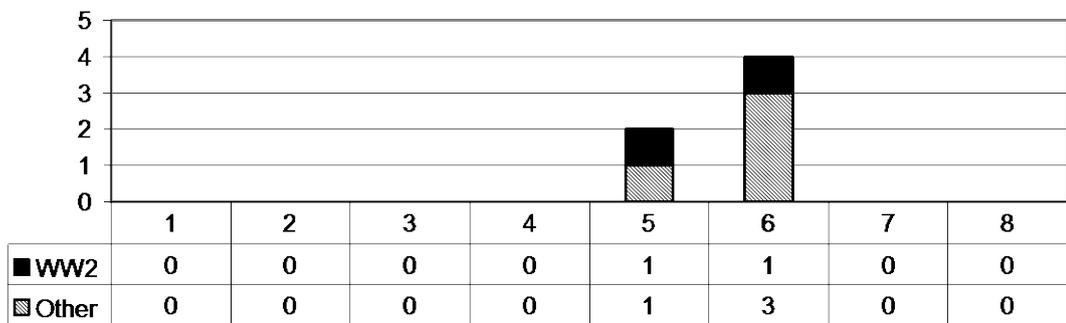


Figure 8.12 Anniversaries – topic groups and positions. The horizontal axis indicates the position of news items within the edition (from 1 to 8); the vertical axis shows the number of news item occurring in that particular position. Notation: *WW2* = Second World War anniversaries; *Other* = Other anniversaries.

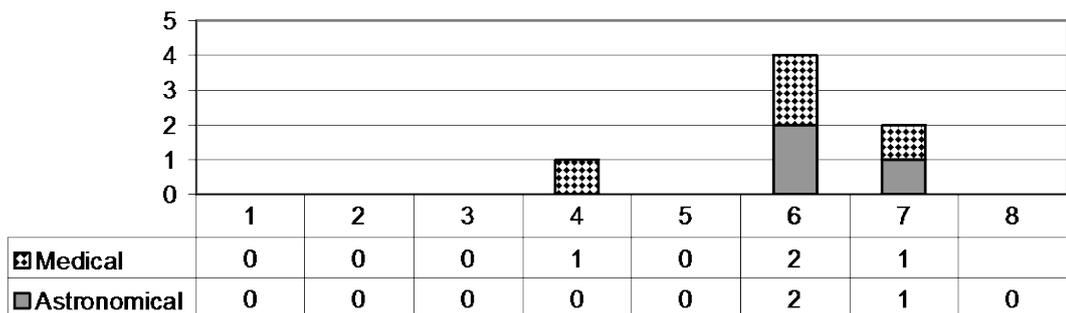


Figure 8.13 Scientific achievements – topic groups and positions. The horizontal axis indicates the position of news items within the edition (from 1 to 8); the vertical axis shows the number of news item occurring in that particular position. Notation: *Medical* = Medical achievements; *Astronomical* = Astronomical achievements.

We may conclude that, roughly speaking, the importance of the various topic areas in terms of coverage amount is reflected in their prominent positioning as well, according to the simple convention that more important topics tend to occur in higher positions, and vice versa. Positions 2 to 6 feature the most significant topics, while positions 7 and 8 are occupied by the so-called soft news. Position 1 does feature the first five topic areas, but the rate of topics is dictated more by the news value of recency. Only two of the topic areas are present in all positions: Conflicts and Nature-related issues; others tend to be more concentrated in certain positions. Despite the consonance between coverage amount and positioning, an analysis of news placing is not futile, since it may shed light on facts that coverage amount alone may not reveal. An example is the case of Military issues, where American and Russian ones are equal in coverage but dramatically different in positioning (2 to 5 vs 6 to 7, respectively). Another example is Media issues, which, despite their low coverage, are quite evenly dispersed in many positions. Therefore, one may draw the general conclusion that the perceived importance of a topic is reflected by a delicate balance between coverage amount and positioning.

Crisell (1994:86) points out that the news editor has no visual means of indicating that certain items are more important than others, there is therefore a risk that everything in radio news will assume an equal importance – or lack of it. In Crisell's view (1994:85), it is only the listener that is likely to assume that the most important items will come first, the less important later. Our findings suggest that this is just part of the story: ordering is indeed an instrument in editors' hands to indicate importance.

Another issue that may come up in connection with placing is the position change that – partly or totally – repeated items go through. The BBC corpus indicates that, in 48% of the cases, the position of a repeated item is the same as on first mention. This is normal if we take into consideration that the perceived importance of a topic is not likely to change in the course of a single day. The next most common tendency is for the repeated item to move down (in 32.4% of the cases). Normally, this means dropping one slot lower. This is again understandable: the editor might feel that an item that has lost its 'freshness' by having been heard before is relatively less important. In 19.6% of the cases, however, an opposite direction is observable: the repeated item moves upwards. This may be due to a previous item having been left out in the next edition, or to a possible negotiation on the part of the editor not being able to decide

which topic is more important, so they are presented in order *a–b* in an earlier edition, and *b–a* in a later bulletin.

If we break down the position changes to topic areas, we find some tendencies of individual topic areas to be different from the general ones. For example, Media issues, Protests, Human interest stories, Nature-related and Scientific issues, and finally Elections show a stronger tendency to move down than to keep their positions. It seems that topic areas that are less covered show their moderate importance also in that their second mention is even less important than the first.

The above account of position changes only signals some tendencies; the BBC corpus is too small to yield generalisable conclusions in this respect (in a database of 452 pieces of news, only 28 items are exactly and 176 partially repeated). Moreover, it might make a difference whether a piece of news takes position 6 in a six-item or in an eight-item edition. Possibly, an index calculated from the position of an item and the total number of items in the bulletin may lead to slightly different conclusions. This difference, however, may not be that significant, since seventh and eighth positions are probably time-filling ones, and their presence or absence does not affect the order of the preceding items so much.

A final remark on the ordering of news items within a programme: it is rare for two items belonging to the same topic area to be found in the same news bulletin. But if they do, they tend to be grouped together. For example, in (7/17), the item on the general strike in Nigeria caused by fuel price rises is directly followed by an item on similar measures in Indonesia. Or in (7/10), although Scientific issues are poorly covered in the BBC corpus, two such items are packed in the same edition, both medical. Therefore, there are indications that news programmes do show a certain degree of inner coherence, although, on the surface, they might seem as a random collection of texts. Another phenomenon giving a certain unity to adjacent news items is the fact that, quite often, the newscaster inserts an *and* between the last two items of a programme, thus signalling that what comes is the final item. This gives news programmes the character of an enumeration: “We inform you on this, this, this, and this.” A listing like this, similarly to an utterance such as *I bought a loaf of bread, some milk and a bar of chocolate*, does endow news programmes with a sense of unity, as ‘things that belong together’ (see Al-Shabab 1986 on the coherence of news programmes).

8.3 Textual structure

Bell (1991:175) notes that the structures of broadcast news still largely reflect its historical roots in printed journalism. Elsewhere (1991:148) he says that, because press stories are generally longer and carry much more detail than broadcast news, the structure of press stories is more complex, and that a framework which handles press news is likely to be adequate for the text of broadcast stories. This must basically be true, but it is not a reason for making analysis superfluous. However suitable press news schemas might be for capturing the structure of radio news, it is still worth seeing what exactly are the constraints that influence the actual manifestation of these models. Some categories may prove to be unnecessary, or new categories may be needed, to adequately describe radio news.

Both van Dijk's and Bell's models are surely based on extensive data and – in the case of Bell's – inner observation. Both researchers point out that not all the categories they have identified are compulsory. According to van Dijk (1988a:56), Headline and Main Events are obligatory in a minimally well-formed news discourse, all the rest are optional. In Bell's (1998:69) view, the smallest well-formed story is just one sentence long, and this is the Lead (plus the Headline, of course). A corpus-based analysis might reveal just how often this might happen and, if other categories are present, how frequent they are.

The present analysis has led to the following results: the categories Headline, Lead, Main Event, Previous Events, History, Circumstances, Context, Consequences, Verbal Reactions, Evaluations, Expectations and Attribution have been found to be useful and adequate for radio news as well. On the existence and name of most of these categories, there is agreement between the two linguists, but Attribution is unique to Bell's model. In the model suggested here for radio news, van Dijk's 'Main Event' (in the singular) has been chosen instead of Bell's 'Event 1 to n', leaving out the category Episode. This is because – due to its shortness in relation to the press – radio news very rarely has several main events (in as little as 2% of the cases did I encounter two main events). Therefore, only this Main Event category proved to be necessary, supplemented by a new unit of Details of Main Event to account for additions.

Another added category is Reasons, simply because it has been found illogical to have Consequences but nothing to relate motives. Reasons is a subcase of Previous Events which show a causal relationship to the Main Event.

Figure 8.14 is a model proposed for the textual structure of radio news. Most of the categories are based on van Dijk 1988a and Bell 1991, with the additions and alterations described above:

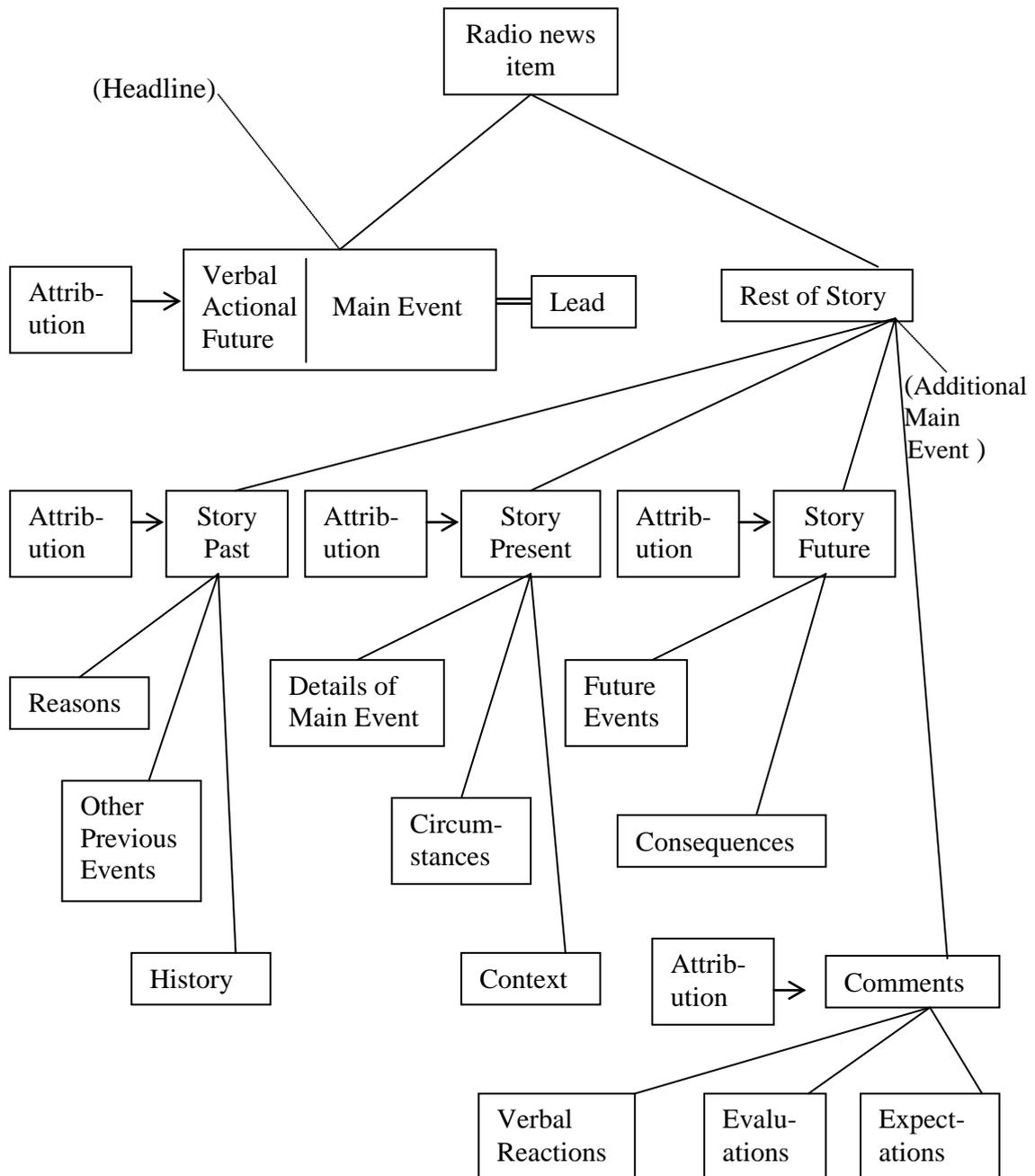


Figure 8.14 The textual structure of radio news

The model needs some explanation. A significant difference from earlier models is that the unit called Abstract (in van Dijk 1988a:55) or Summary (in Bell 1991:171), represented by Headline plus Lead, has changed into Main Event. This is because one

of the most striking differences between the press and radio news lies in the status of Leads. Van Dijk (1988a:53) regards Leads as forming summaries of the whole story (together with Headlines). Bell (1991:176) does not view Headlines as integral parts of printed news discourse, and says (1991:184) that it is inadequate to say that Leads are simply summaries of the story, they are also beginnings of it, because they may contain information that is later not included in the text. Here it is observed that, in radio news, Leads and Main Events merge. That is, Leads are not summaries, not even beginnings of the story, but the principal part of the story itself. There is one exception to this rule: when the news item contains a correspondent's report; in this case it sometimes happens that the information packed into the lead is repeated in the report, probably to allow for lapses of attention on the part of the listeners. Otherwise Leads and Main Events are one and the same in radio news; this is why there was no need to introduce a category of Abstract/Summary, and Story has been transformed to Rest of Story.

Commentary is needed also for the representation of Headlines and Additional Main Event: the parentheses instead of a box indicate that these categories are not simply optional, but rather exceptional. Out of a hundred items, only two have an Additional Main Event. As for the usage of Headlines, a discussion follows in Section 8.6.

The model might seem more complicated than either van Dijk's or Bell's, but this is because the recursiveness of the Attribution category has been overtly represented in the figure. Attribution has proved to be one of the most pervasive categories in the BBC corpus. This is in consonance with Bell's (1991:191) observation that, although broadcast news rarely attributes agencies, the BBC World Service news is distinctive for frequently naming the journalistic sources of its information.

Several other new categories have been introduced, for example, Future Events, in the case where the Main Event was not a future one, and the news item reports on anticipated events. Future Events is not identical with Consequences because, although the latter are subsequent to the Main Event, they might also have taken place in the past (in respect to broadcasting time), while Future Events only take place after broadcast. Further, they are not identical with Expectations either, because Expectations are the brain-children of journalists/correspondents, often directly attributed to them, and their probability is not so strong, while Future Events are dictated by external powers/circumstances, and they are almost sure to happen:

- (9) *The Argentine currency, the peso, has survived its first day of trading on foreign exchange markets since it was devalued, ending just slightly below its opening price against the US dollar. It was the first full day of trading in three weeks since the height of the economic crisis. Some analysts had predicted the peso would collapse. They say the real test may come on Monday. Banks reopen their doors as well having been closed since last month's riots.* (7/12)
- (10) *The White House says President Bush fainted for a few seconds and fell from a couch while watching television in his residence. His doctor said he'd been feeling unwell for a couple of days and he lost consciousness when his heart rate slowed after he failed swallow a pretzel properly. The doctor said a subsequent examination showed that President Bush was fine. He fell to the floor when he fainted slightly cutting his face and bruising his lip.* (7/14)
- (11) *The Indian Interior Minister L. K. Advani is in Washington to seek support for India in its quarrel with Pakistan. India wants the United States to back its corps for Pakistan to take action against cross-border militant groups. Delhi says these groups supported by Pakistan were behind the attack on the Indian parliament last month. But it is not clear how much backing Mr Advani will receive with United States officials on Tuesday welcoming the position taken by the Pakistani president Pervez Musharraf. The State Department spokesman said Pakistan's efforts to curb extremism were the key to reduce in tensions.* (8/9)

The underlined part of example (9) is a scheduled Future Event, subsequent to transmission time. The highlighted section in item (10) is a Consequence, subsequent to the Main Event, but prior to broadcasting time. The underlined sequence in item (11) is an Expectation – or, maybe, lack of expectation –, probably originating from the correspondent.

It has been mentioned that a corpus-based study is able not just to help identify the textual units, but also to indicate just how frequent these categories are. Table 8.4 illustrates this, with the numbers of occurrences rounded and normalised to one hundred news items:

One can easily notice the strikingly great number of Attributions. The results indicate that, in an average news item, there are more than two Attributions. Without this category we would not be able to account for a very typical characteristic of the BBC corpus. The next most common category is Lead; the number shows that a news item contains one and only one Lead. Main Events together make up a hundred, therefore, there is one Main Event in an item, which is not surprising, given that Leads and Main Events are identical in radio news. Almost half of the Main Events are verbal, which, therefore, is the most common subcategory, followed by Actional and Future ones (for a discussion of these categories, see the next section). Previous Events are also

very common: Reasons and Other Previous Events together give 97, which means almost one such category per item. Verbal Reactions, a specific element of the news media, are also very common: more than every other item contains one. This is followed by two of the Story Present units: Context, Details of Main Event, with almost the same frequency. Evaluations and Future Events are present in every fourth item. Next in the row are Consequences (Story Future) and Circumstances (Story Present). The least common categories in radio news are History and Expectations, with History slightly more and Expectations a little less than once in ten items.

Textual category	Number in 100 items	Textual category	Number in 100 items
Verbal Main Event	49	Details of Main Event	35
Actional Main Event	39	Circumstances	18
Future Main Event	13	Context	36
Attributions	220	Future Events	23
Lead	100	Consequences	13
Reasons	26	Verbal Reactions	57
Other previous Events	71	Evaluations	25
History	11	Expectations	7

Table 8.4 The rate of textual categories in the BBC corpus (normalised to a hundred items)

A quantitative observation: the average news item in the BBC corpus contains 4.3 textual units. The maximum number found is 10, which occurs once in a hundred. The minimum is 2, which happens in twelve cases out of a hundred. Besides Main Event, these minimally well-formed radio news items typically contain Previous Events (including Reasons) Context, or Details of Main Event. A deviation from this pattern is experienced when the Main Event is a future one: then, the other structural unit is typically Expectations.

In sum, Leads/Main Events and Attributions are compulsory elements of radio news. These are most typically complemented by mention of near-past previous events and a description of the present situation, possibly accompanied by an evaluation. Less common are reports on the story future, and even less frequent are those elements,

which deal with either the remote past or an uncertain future. Let us now turn to some of the most distinctive textual units of radio news in more detail.

8.4 Lead/Main Event

Main Event has been observed to have three subtypes: Actional, Verbal and Future Main Event. The first is exactly what has been understood in the two models by the (Main) Event of a news story. Verbal Main Event corresponds to what has tangentially been pointed out by van Dijk (1985a:89; 1988a:56) that, although Verbal Reactions is usually ordered toward the end of the article, it may be placed in the front of the article if the information contained in it is sufficiently relevant. But Bell (1991:165) uses the term ‘verbal story’, thus recognising that a whole story can be a Verbal Reaction. It has been found that, in the BBC corpus, quite a lot of news items almost exclusively communicate Verbal Reactions, or their Main Event is verbal:

- (12) *The Indonesian government has announced fuel price increases averaging 22 per cent despite protests in several cities. The energy minister Purnomo Yusgiantoro said subsidies were being cut to reduce the burden on the national budget. He said that the costs of most fuel products would now be pegged to world market prices and that the state energy company would publish a revised price list at the beginning of every month.* (7/17)

Example (12) above is exclusively made up of verbal events. In such cases the Main Event of the news item has been classed as a Verbal Main Event, thus establishing the character of the item, and the rest of the statements have been analysed for the content of the statements, that is, in the above case, the second sentence has been labelled as Reasons, and the third as Future Events.

The third type is Future Main Events. Quite a number of items in the BBC corpus (13%) are news about something that has not happened yet. The suspicion emerged that this is exclusively due to the fact that much of the data have been gathered from morning programmes, which give information on the expectable happenings of the day. However, the mid-day programmes also contained an appreciable proportion (around 9%) of Future Main Event news. This was conducive to the decision that such a category is not far-fetched. An example for such a category is (13):

- (13) *Further moves are expected today towards lifting international sanctions imposed on the former Taleban government of Afghanistan. The United Nations Security Council is expected to vote to end a ban imposed on flight by the Afghan national airline Ariana. And the BBC UN correspondent says that other sanctions are expected to be lifted later in the week.* (6/15)

We have seen that Main Event in radio news merges with Lead. Bell (1991:185) points out that broadcast stories are less likely to pack two events into the lead in the way the press does. The BBC corpus only bears out this statement if by ‘event’ Bell means Main Event. If so, then indeed radio news rarely has two Main Events, as has been mentioned earlier. But it is very common for Leads to contain several events, one of which is the Main Event, others may be Reasons, Other Previous Events, etc. In other words, it is the rule and not the exception to have more than one event in a Lead. Consider the following lead:

- (14) *Israeli aircraft have attacked a Palestinian authority building in the West Bank city of Tulkarm, following the killing of six Israelis by a Palestinian gunman in an attack on a banqueting hall in Northern Israel.* (7/18)

If we analyse lead (14) for propositions, we find that three of them can be formulated: a Palestinian gunman attacked a banqueting hall in Northern Israel; he killed six Israelis; and Israeli aircraft have attacked a Palestinian authority building (in chronological order). Out of these, the third represents the Main Event, and the other two are Previous Events (the conjunction *following* expresses a vague connection suggesting a causal relationship, which, however, is not certain). The second proposition exemplifies a common method of integration (see Section 2.3.3), where an action appears as a gerund. Lead (15) below shows integration by a verb becoming a noun:

- (15) *Another fifteen hundred people in Australia have been evacuated from their homes because of a flare-up in bushfires near Sydney.* (9/7)

On the surface, lead (15) is a single clause, but it disguises two propositions: the bushfires near Sydney, Australia, have flared up (Reason); another 1500 people have been evacuated (Main Event). Leads (16) and (17) present another typical way of condensing information:

- (16) *The American Secretary of State, Colin Powell, who is beginning a visit to Pakistan, says he intends calming the conflict between Pakistan and India and moving it into reverse.* (7/16)
- (17) *The United Nations Secretary General, Kofi Annan, is in Kabul, where he is holding a day of talks with the leaders of the interim Afghan administration.* (7/25)

Both leads contain relative clauses, which are nonrestrictive, but do not describe their antecedents, they rather serve as a way to continue the story as smoothly as possible. Some grammarians (e.g. Thomson and Martinet 1986:88) call these types of relatives ‘connective’, because the sentence could be reformulated with the help of a coordinating conjunction. In our case: the American Secretary of State, Colin Powell is beginning a visit to Pakistan (Actional Main Event); and he says he intends calming down the conflict between Pakistan and India, and he intends to move the conflict into reverse (Verbal Main Event). As we can see, lead (16) is one of the rare examples where there are two Main Events in the story. In the formulation of propositions *he says* has not been analysed as a different proposition, because it classifies as an Attribution for the Verbal Main Event. Or in (17): the UN Secretary General, Kofi Annan is in Kabul (Circumstance); and he is holding a day of talks with the leaders of the interim Afghan administration (Future Main Event). This use of relative clauses is not unique to leads; it is often applied in the rest of the news items as well.

As could be seen, Lead typically comprises more than one event: along the Main Event, one can find Reasons, Other Previous Events, Circumstances etc. present in the Lead. This is distinctive of radio news since, in the press, leads normally contain only Main Event(s).

Another distinctive feature of radio leads is their use of tenses. Logically, Leads expressing Future Main Events use a way of expressing future time, as in the examples below:

- (18) *United States Defence Secretary Donald Rumsfeld is visiting later today the detention camp of suspected al-Qaeda and Taleban prisoners at the American base of Guantanamo Bay in Cuba.* (12:30/27)
- (19) *Parliament in Zimbabwe is due to begin a debate later today on a controversial media bill, which critics say is part of President Mugabe’s drive to silence opposition to his bid for re-election in March.* (6/22)

- (20) *Australia is to resume processing the asylum claims of illegal Afghan immigrants after more than a week of protests at detention centres.* (12:30/24)
- (21) *The European Union's only directly elected body, the European Parliament, will be electing a new president today.* (6/15)

Out of the possible future forms, leads typically give preference to the ones which express a higher degree of certainty, passing on to less probable expressions in the rest of the item:

- (22) *Foreign exchange markets in Argentina reopen shortly, three weeks after violent street demonstrations and looting forced them to close. The national currency, the peso, will be traded for the first time since it was devalued by almost 30% earlier this week, as part of an emergency plan to tackle the country's economic crisis.* (12:30/11)

However, they keep the less sure future form if the outcome of the Future Main Event is doubtful:

- (23) *The Parliament of Kosovo will try again today to elect a president, two months after elections to the assembly.* (7/10)

As for Verbal Main Events, Bell (1991:210) observes that, in broadcast news, present tense is the common form for both the speech verb and the indirect speech, while the equivalent press story would have used past tense throughout. He explains that the use of the present is assumed to help create the impression of immediacy. The BBC corpus verifies these statements: the present tense is common for verbal leads, as in the lead below:

- (24a) *President Musharraf of Pakistan says Osama Bin Laden is probably dead.* (7/18)

However, it is typical for the continuation of the verbal story, both the reporting verb and the reported statement, to switch to past tense. Below is the rest of the above news item:

- (24b) *In an interview with the American television General Musharraf said the Saudi fugitive may have succumbed to a kidney disorder, because he no longer had access to dialysis equipment he needed But the Pakistani leader admitted there*

was a possibility Osama Bin Laden was still hiding in Afghanistan or Pakistan.
(7/18)

Many more examples of such nature could be quoted here. Besides immediacy, the present tense of the lead may be aimed at suggesting that the act of statement is still valid (for example, that it has not been refuted). Variations of the reporting verb may exist in the form of the perfect aspect, but the tense remains present:

(25a) *A former executive with the bankrupt American energy giant, Enron, has said that documents have been destroyed at the firm's headquarters in Texas, long after a court order of forbidding such action.*
(7/22)

No systematic difference is apparent between the function of the present simple or perfect in reporting verbs, which is shown by the fact that the headline belonging to the news item displays a simple structure:

(25b) *A former executive of the collapsed American energy giant Enron says company documents were being shredded at least until two weeks ago despite a court ban.*
(7/22)

The tense of the reported statement has switched to past because of the use of an exact time expression.

As far as Actional Main Events are concerned, Bell (1991:202) again notes that, while the regular expression of time in press news is past tense (optionally plus time adverbial), broadcasting leads with the immediacy of present perfect with no time adverbial. The BBC corpus shows the same tendency. However, much in the same manner as in the case of verbal leads, the tense changes into the past in the rest of the item, as in (25)

(26) *Three Russians with links to the outlawed Japanese sect Aum Shinrikyo have been given jail sentences for plotting to bomb Japanese cities. A court in the Russian city of Vladivostok sentenced the three men to prison terms ranging from four-and-a-half to eight years. Investigators said the men had planned to attack public places in Tokyo and Aomori in a bid to force Japanese authorities to free the leaders of the Aum sect.*
(9/23)

It can be observed that, unlike in Verbal Main Events, the tense of the Verbal Reaction is also past, just like the majority of Rest of Story verbs.

The overwhelming presence of the present perfect in the first parts of radio news is in harmony with the findings listed in the quantitative, grammatical part of the present study (see Section 7.4). Without the qualitative component, however, we would not have found out that this preference for the perfective aspect is concentrated in a particular structural unit, namely, the lead.

To sum up, contrary to Bell's findings, Leads in BBC World Service news typically comprise more than one event: the Main Event and one or more of Reasons, Other Previous Events, Circumstances etc. In conformity with Bell's findings, the tense used in Leads is typically present. This present form refers to the future in Future Main Events, and is accompanied by the perfect aspect in Actional and often Verbal Main Events. With the exception of future stories, however, this present invariably turns into the past in the rest of the story.

8.5 Attributions

Attributions in the BBC corpus most commonly precede indirect statements. Bell (1991:193) observes that, although in the press attribution is often postponed, particularly after a direct quotation, broadcast news follows the pattern of general spoken English, where attribution usually begins a sentence. Attributions may be made to BBC correspondents in various parts of the world, to representatives of certain institutions, acting as information sources or, mostly in the case of Verbal Main Events, to elite news actors whose verbal manifestations are newsworthy in their own right.

Attributions in radio news are not frequently followed by direct quotes, mainly because it is difficult to signal quotation marks in speech. Bell (1991:208) notices an equivalent of quotation marks in speech: 'verbal glosses' such as *what they call*, used as a means for the radio station to distance itself from the overtone of the term that follows. The BBC corpus also contains such verbal glosses, as in the following excerpt:

(27) *The government has implored what it describes as all well-meaning and patriotic Nigerians to go about their normal business.* (7/16)

Interestingly enough, BBC World Service shows detachment from its own correspondents as well, as in (28):

- (28) *The United Nations envoy James LeMoyne was engaged in what the BBC correspondent in the capital, Bogota, calls frantic negotiations to save the peace process.* (8/12)

The disowning is probably due to the stylistic load of the expression used by the correspondent.

Although Bell (loc cit) states that it is not possible to direct-quote an isolated phrase of the newsmaker's speech, the BBC corpus does exemplify some real direct quotes:

- (29) *'Seeing is believing,' says one official here.* (6/16)
 (30) *'This strike in on indefinitely,' say the unions.* (7/16)
 (31) *Sounding relieved, his [Robert Mugabe's] first comment was: 'All's well that ends well,' but his face looked nervous and tired.* (7/15)

That they really are direct quotes is signalled by the postponing of the reporting clause in (29) and (30), and by no backshift after the past tense reporting verb in (31). As we can see, these direct quotes are either sayings – as in (29) and (31) – or short, 'quotable' and well-sounding statements – as in (30).

However, there is a special, 'radiophonic' quote type, namely, soundbites. The direct use of the newsmaker's voice is, according to Bell, a variation of direct quotation conditioned by the technology of the medium (1991:205), and is in principle no different from direct quotation in printed news (1991:149). In the BBC corpus, out of a hundred news items only six contain soundbites. Typically (in four out of six cases) these represent Verbal Reactions, but they may also express Evaluations, Reasons, and Context. And if Verbal Reactions are reserved to elite personalities, then this is even more so in the case of soundbites. Newsactors present in the BBC corpus with their own voices are from the highest circles: President Bush, an Israeli government spokesman, the leader of the terrorist detention camp in Cuba, the US Defence Secretary, the UN spokesman in Afghanistan, the Afghan Finance Minister, the Pakistani leader, the Zimbabwean Foreign Minister, the Colombian president etc. The sources of soundbites correspond to the most extensively covered topics in the BBC corpus (see Section 8.1). Additional soundbite-actors are: a Red Cross, a World Food Programme spokesman and an Oxfam coordinator, and they are all connected to a high priority disaster story, which headed the news for several days on end, the Congolese volcanic eruption.

Despite the BBC World Service's reputation as a faithful indicator of information sources, unnamed or totally unsourced attributions also occur:

- (32) *So far, Israeli officials have reacted to Ahmed Saadat's arrest with caution and with scepticism.* (6/16)
- (33) *The US authorities have said from the start they would be happy to let the Red Cross inspect Camp X-ray.* (6/17)
- (34a) *Hundreds of Chinese farmers are reported to have attacked government officials trying to collect taxes in a township in the central province of Henan.* (9/7)
- (35) *It is alleged that, while in 1993 the state government authorised the clearance of nearly 50,000 hectares of bushland by managed fires, the main method of fire prevention, last year the figure was less than 20,000.* (8/13)
- (36) *There have been various accounts of how the fighting began...* (6/16)

The first two examples display unnamed sources, where their social-political functions, and not their exact identity, are indicated. Examples (34a) to (36) show cases of unsourced attributions. In the BBC corpus, unnamed sources are common, but unsourced attributions rare.

Nevertheless, Attribution remains one of the most pervasive structural units of news in the BBC corpus. The record is represented by the example below – the full item belonging to (34a) above –, where there is at least one attribution in every sentence, giving a total number of 7 attributions in one piece of news:

- (34b) *Hundreds of Chinese farmers are reported to have attacked government officials trying to collect taxes in a township in the central province of Henan. Local officials say the protest on Friday was a small one and deny reports that farmers have been arrested. The officials say many farmers have not paid taxes for more than five years. The farmers say the taxes are far too high and they can't afford them. The BBC correspondent in Beijing says angry protests are becoming an increasingly common occurrence in rural China.* (9/7)

In sum, attributions in the BBC World Service news are most typically followed, and not preceded, by the reported statements. Most of these statements follow the grammatical rules of indirect speech, but some direct quotes are also detectable. A special type of direct quotes is represented by soundbites, which regularly contain the verbal manifestations of the most elite personalities. Reported statements are, as a rule,

accompanied by attributions, but these sometimes do not personalise the source or, occasionally, are left out altogether.

8.6 Headlines

Headlines are discussed in a separate section because radio headlines are so much different from press ones. In printed news, headlines are compulsory, and they give an ultimate abstract of the story, containing its most important macroproposition. According to Bell (1991:150), however, broadcast news has no headlines, except in so far as stories are summarised at the beginning and/or end of a news bulletin. Or, in another formulation (1991:186), headlines are either absent altogether or collected at the beginning and/or end of longer news programmes. This is exactly so in the BBC corpus as well: the five-minute World News bulletins contain a couple of headlines at the beginning of the programme; the two-minute Newshour summaries have no headlines at all. However, it is worth seeing what exactly the conditions of the presence of headlines are and, if they occur, what is their relationship to the news item they belong to.

Headlines are indeed peripheral to broadcast news in the sense that they do not directly precede the item, rather, they form one group at the start of the bulletin. Moreover, while in the press headlines are compulsory, in broadcast news they are not. In other words, not every item is signalled by a headline.

The BBC corpus has been analysed to see which of the items is complemented by a headline. It seems that headlining follows the following rules:

1. If there are no correspondent's reports in the programme, the first three items of the programme are headlined.
2. If there are reports in the programme, the reported news items are headlined.
3. The order of headlines follows the order of the news items in the programme.

These rules are – deliberately or unconsciously – followed in the majority of the news editions (about 60 %). Deviations from the rules are caused by several factors. Some of these are internal, others external. Internal factors are related to the topic of the news items, and external ones to the time of edition or the proximity of other programmes. Let us look at these factors in more detail.

We have seen in Section 8.2 that the order of news items is conditioned by their perceived importance, and topics that get the greatest amount of coverage also tend to be positioned in the first places of bulletins. Rules 1 to 3 above dictate that headlines also follow this order of importance. The headlining conventions related to the reportedness of the item are also connected with importance because, certainly, important topics are reported on by correspondents. However, the degree of the interest that the listeners might take in the particular topics also influences the choice on headlines. It seems that the order of importance is not so strict in headlines as within the news programme itself. So, although the news items do follow the importance order, the headlines may feature less important but more exciting items, which are placed backwards in the bulletin. In the 8 o'clock edition of January 25, for example, instead of the reported item in position 5, we find a headline of the unreported item in position 4, which deals with President Bush's fainting spell caused by a pretzel. In the 7 o'clock edition of the January 9 news, besides the reported items, an additional fourth headline appears, dealing with the US military's expectance of former Taleban ministers to be handed over. Even the last item may surface in the headlines, such as in (8/17), if it is concerned with such an attention-catching topic as the increased possibility of life on Mars.

An interesting case is (7/13), where regulation 2 above overrules law number 3: the headlines follow the order of the reported news (1st, 2nd and 4th), and are followed by an additional (belonging to the item in position 3), which deals with the appealing topic of Prince Harry's alcohol and drug-consumption. This proves that the order of the rules presented above is valid.

An external factor is the time of edition: for example, rules are not so strict if it is weekend. On Saturdays and Sundays there is much more flexibility in headlining: on January 26, for example, items 1, 5, 6 and 7 are headlined. We can see this in other kinds of radio or TV programmes: the most compelling piece is anaphorically signalled at the beginning, but left to the end, to make the listener/viewer attend to the entire programme. The headlining of the weekend editions in the BBC corpus shows affinity with such procedures.

A significant deviation from the rules is existent at the beginning of 6 o'clock news. Consider, for example, the headline below:

- (37) *Today the government and rebels in Colombia give peace talks one more chance to pull back from the brink of confrontation. But what are the prospects?
“We have had a permanent confrontation even during the conversations, and I’m not sure that the guerilla is going to stop this confrontation.”* (6/15)

The headline is followed by a question, a very unusual phenomenon in the news, and by a soundbite, whose source is not even disclosed, only signalled by the reference to the country and the strong accent that it might be President Pastrana of Colombia speaking. The style of the headlines is also different from what we are used to in other editions. Here are two from the same programme:

- (38a) *Today the US Congress is getting its teeth into the Enron case. Nine committees will be trying to get at the truth, but what exactly are they looking for?* (6/24)
- (38b) *President Bush girds America’s loins to the fight against terror by sharply checking up defence spending. Critics line up to say he is bashing the budget.* (6/24)

Both headlines display informal or jocular expressions that decrease the seriousness of the news and make it more like a story than reality itself.

The proximity of other programmes may also influence the way headlines are formulated. For example, the 7 o’clock edition of the January 15 news regularly headlines the reported news (1st, 3rd and 4th), and the 8 o’clock one displays the same discontinuous headlines despite the fact that the reports have been left out of the programme. In this case, an earlier programme has affected the structure of the one in question.

In 6 o’clock news, we often get a headline only partly, or not at all, connected to an item in the following five minutes; in these cases the headline is a forerunner not of the news edition, but of a piece in the programme following the World News. In this case, a subsequent programme has influenced headlining procedures.

The factors identified above do not work independently from each other, for example, Saturday flexibility and the drive to feature interesting news in headlines may result in stronger oscillation from the standard. And certainly, the working habits of the individual news editor also affect the result in ways not analysed in the present study.

Some of these deviations are in close connection with what Fairclough (1995b:9-10) calls the ‘conversationalisation of public language’ and some tensions affecting contemporary media language: between information and entertainment, and between

public and private. Acquiring an easy-going verbal style, with colloquialisms and funny expressions, is a powerful transformation of information into entertainment. Announcing interesting pieces at the beginning of a programme is a market-driven way to 'sell' the programme and the station in general.

A look at the textual structure of the headlines reveals the following: just as has been observed by Bell (1991:187) in connection with press headlines, radio headlines also abstract the story, and are entirely derivable from the story, most commonly from the lead alone. Thus, while Leads in radio news are not abstracts but deliverers of the Main Event information, headlines – when present – do function as abstracts. The textual relationship of Headlines to Leads in the BBC corpus can be placed on a continuum. On one end of the continuum are the cases where Headline deletes information from Lead. In the middle are the cases where Headline contains exactly the same information as Lead. Towards the other end of the scale are the cases where Headline picks out information from Lead, but also displays pieces of knowledge that are not contained by Lead but by later parts of the news item. On the furthest end of the cline are the cases where Headline exclusively contains information taken from Rest of Story. In actuality, the headlines of the BBC corpus group towards the left end of the scale (Figure 8.15):

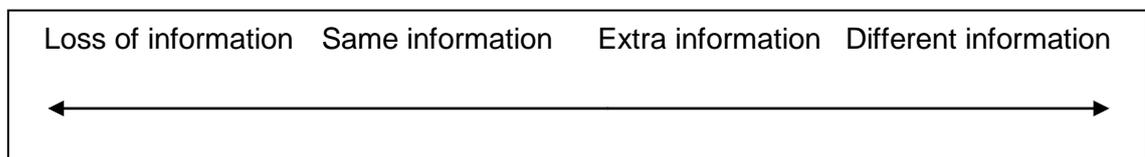


Figure 8.15 The Headline information continuum in the BBC corpus (as compared to Lead)

In structural terms, different types of information can be left out from headlines; for example, Circumstance, Details of Main Event etc., as in the following (in each case, the 'a' part of the example is the headline, 'b' is the lead, in which the underlining highlights the deleted parts):

(39a) *Today the man convicted of masterminding the Lockerbie bombing appeals against his conviction.* (6/23)

(39b) *A Lybian man found guilty of blowing up an airliner over the Scottish town of Lockerbie in 1988 is due to appear before a court in the Netherlands today to appeal against his conviction.* (6/23)

(affiliation of news actor, focus of Previous Event, exact location of Previous Event, time of Previous Event, location of Main Event)

(40a) *President Bush has released information which he says proves that the Palestinian authority was involved in trying to smuggle a boatload of weapons to the West Bank and Gaza.* (7/25)

(40b) *President Bush has provided three Arab countries with what he says is evidence that Yasser Arafat's Palestinian authority was involved in trying to smuggle 50 tons of weapons to the West Bank and Gaza.* (7/25)

(recipient of Main Event action, main actor of Previous Event, amount of patient in Previous Event)

However, headline (40a) also displays information not included in the lead: that the weapons were tried to be smuggled in by sea (*boatload*).

Sometimes the information content of headlines and leads are the same, only the formulation is different, as below:

(41a) *Aid agencies say safe drinking water is a top priority for hundreds of thousands of people left homeless after the volcanic eruption in the Democratic Republic of Congo.* (7/20)

(41b) *Aid agencies trying to help refugees who have fled from the volcanic eruptions in the Democratic Republic of Congo say providing clean drinking water is a main concern.* (7/20)

Despite the different formulation, the two texts contain much the same elements or, if not, the pieces present in one and missing from the other are perfectly guessable from the model they activate: it is just natural that aid agencies try to help, and that people who have fled from their town are left homeless.

Towards the right end of the continuum are the cases where headlines contain more information than leads, as in the following:

(42a) *Kofi Annan has arrived in Afghanistan, the first such visit by a UN Secretary General for more than forty years.* (8/25)

(42b) *The United Nations Secretary General, Kofi Annan, is in Kabul, where he is holding a day of talks with the leaders of the interim Afghan administration.* (8/25)

- (43a) *The chairman of the bankrupt American energy giant, Enron, has resigned, as investigations continue into the company's collapse.* (7/24)
- (43b) *The chairman and chief executive of the American energy giant, Enron, has resigned, a month after the company suffered the biggest corporate collapse in United States history.* (7/24)

In the news item introduced by (42b), the evaluation that this is the first such visit for a while appears in the next sentence; after lead (43b), the fact that the criminal enquiry is still going on is communicated at the very end of the item. Certainly, the headlines also delete information from the lead: that Kofi Annan is there for talks (quite guessable); that Enron's collapse happened a month ago, and that it was the biggest scandal in the history of US economy.

Finally, there are the cases where the information content of headlines and leads are considerably different, as in (44) and (45) below (here, the underlining refers to common information):

- (44a) *Half of the town of Goma in the east of the Democratic Republic of Congo has been engulfed in flames after molten lava from a volcano poured through the streets.* (10/18)
- (44b) *Hundreds of thousands of people have been driven from their homes by volcanic eruptions near the town of Goma in the east of the Democratic Republic of Congo.* (10/18)

(Reason in Headline, Main Event in Lead)

- (45a) *Tens of thousands of people in the Democratic Republic of Congo have spent the night outdoors after streams of volcanic lava destroyed much of the town of Goma.* (7/19)
- (45b) *Lava from the volcano Mount Nyiragongo in the Democratic Republic of Congo is continuing to flow and powerful earth tremors have continued throughout the night.* (7/19)

(Consequences in Headline, Main Event in Lead)

The fact that it is difficult to represent these differences in content by clear cases, and that the same headline–lead pair may display several of the differences outlined above, proves that these relationships are indeed placed on a continuum.

We have seen that the information content of headlines and leads overlap appreciably, but there may be considerable departure as well. This is due to the fact that,

while Lead in radio news is not a summary, but the main part of the story itself, Headline is a summary indeed, and not of Lead alone, but of the whole story. Thus, it may easily pick up information from Rest of Story.

The present study is not interested in the syntactic composition of radio headlines, but a glance at them shows that they are totally different from the telegraphic, formulaic, block-language style of press headlines. They are full sentences which do not show the characteristic rhetorical devices of press headlines either, as mentioned by Bell (1991:189): monosyllabic words, alliteration, punning etc. This must partly be due to the fact that broadcast headlines are intended for the ear, and not for the eye.

To sum up the nature of headlines and their relationship to news items: headlines are peripheral and not compulsory elements of radio news. Their attachment reflects the importance order followed in the news programme itself, and sticks to certain rules connected with whether the news item contains correspondent's reports or not. Deviations from the rules are conditioned by internal factors such as the topic of the item, or external factors, e.g. time of edition, or proximity of other programmes. Headlines are summaries of the news items. The information transmitted by headlines overlaps with that contained by the lead. However, it may well tell less than the lead, more than it, or be almost totally different from it, taking its source from other parts of the story.

8.7 Transformations

This section will examine the linguistic transformation of texts in three cases: (i) when a correspondent's report is later incorporated in a news item; (ii) when a news item is only partially repeated on a later occasion; and (iii) when a long, morning news item becomes a short, afternoon one.

Before starting to analyse the textual transformations in case (i), however, first we need to know what is the textual relationship between a report and its introduction. As we have seen, the lead of a radio news item comprises the Main Event reported on. This makes it superfluous for the report to contain Main Event information. Therefore, a report normally gives Context, Previous Events, Consequences and, often, Verbal Reactions, as in (46) below:

- (46) *The Palestinian authorities announced an investigation into allegations that an arms shipment seized by the Israeli forces in the Red Sea last week had been ordered by Palestinians.*

-----→ Main Event
The ship was boarded by Israeli commanders who detained the crew.

-----→ Previous Event
The captain Omar Akawi told reporters from the jail where he is being held in Israel that he had received his instructions from close aides to the Palestinian leader.

-----→ Verbal Reaction
From Jerusalem here's J. R.:

-----→ Attribution
News of the enquiry came following a meeting held between the Palestinian leader Arafat and the European Union foreign policy chief Javier Solana.

-----→ Context
Palestinian officials say the inquiry will indict and punish everyone involved in the matter. The announcement of this investigation comes after an earlier suggestion to set up a commission with the Israelis and the Americans was not taken up.

-----→ Previous Events
The Israeli government says it holds the Palestinian authority responsible for the weapon-smuggling. The Defence minister Binyamin Ben Eliezer says that the operation was managed and funded by the Palestinian authority in cooperation with Iran and other sources - a claim rejected by the Palestinians.

-----→ Verbal Reactions
 (7/8)

The Previous Events and Verbal Reactions in the introductory part of the item are not the same as the ones in the report. The cohesion between the report and the introduction is achieved through the anaphoric definite article and synonym in the first sentence (*the enquiry* refers back to *an investigation*).

Another typical case is when the lead, besides conveying the Main Event of the story, also displays parts that summarise certain sequences of the report. Consider, for example, (47) below:

- (47) *The leaders of the Turkish and Greek communities on Cyprus are beginning intensive talks today. They are trying to negotiate an end to the 23-year partition of the island.*

-----→ Main Event
A. R. reports from Nicosia:

-----→ Attribution
Ever since 1974, when Turkey invaded and occupied the north of the island, Cyprus has been partitioned. And all attempts to reconcile the Greek Cypriot majority with the Turkish minority have failed.

-----→ History

But the government here hopes to complete its negotiations this year to become a member of the European Union.

-----→ Context
And the prospect of Cyprus joining while its northern third remains under illegal Turkish control has added great urgency to the search for a solution.

-----→ Reason
The talks will focus on finding a new constitutional arrangement, which would give the Turkish Cypriots security and the right to run their own affairs, while ensuring that the country survives as a unified state.

-----→ Future Event
It won't be easy, but there is more will to find a solution now than at any time in the past.

-----→ Evaluation
 (7/16)

Here, *the 23-year partition of the island* underlined in the lead is a summary of the highlighted sentence in the report. (The lead of this item is exceptional in that it is made up of two sentences.)

Items (46) and (47) present the two typical relationships between introductions and reports: the report giving background information to the introduction, and the introduction also containing some summary of the report. Reports often contain cohesive ties to the introduction, exemplified in (46) above, or (48) below, the first sentence of a report, where the reference of the definite article is absolutely unguessable without the previous text:

(48) *The circumstances are still unclear.* (7/9)

Certainly, there are a couple of exceptions from the rules. Consider the following examples:

(49) *Residents of the rebels' safe haven in South-Eastern Colombia say they fear reprisals by right-wing paramilitaries after the collapse of peace-talks between the government and the left-wing FARC movement. About 100 thousand people live in the zone in South-East Colombia, and human rights groups say they fear the paramilitaries will regard them as collaborators once the rebels withdraw and the army moves in.*

C. M. reports from Bogota:

The Colombian army is making its final preparations to move into the demilitarised zone, massing troops along the border. With the deadline for the withdrawal of the revolutionary armed forces at Colombia less than 24 hours away, frantic last-minute negotiations are still taking place. But with the rebels already having pulled back into the jungle, the possibility of a peaceful solution is

remote. United Nations peace envoy James LeMoyne is still trying to keep the talks going. But it seems that he will play a more important role over the coming days in trying to ensure the safety of the local population. There are fears that right-wing paramilitary groups could move in to carry out bloody reprisals against civilians. (7/14)

In (49) the lead features the Main Event that residents of the safe haven fear atrocities, and renders a collapse of peace-talks as Previous Event. However, in the report ‘frantic last-minute negotiations’ are still going on, and we see other Circumstance and Context elements as well. Here, the UN envoy’s ensuring the safety of the population appears as an Expectation, and the fears of the local people as a Reason for the Expectation. In sum, item (49) is an example of a probable disagreement between the correspondent and the editor on what constitutes the Main Event of a story: the editor has picked up what appears as a rather subsidiary story element in the last sentence of the report. (The local cohesive tie is otherwise present here as well, with *the demilitarised zone* in the first sentence of the report referring back to *the rebels’ safe haven* in the lead.)

Now that we know the relationship of reports to their introductions, we may proceed to examine the ways they are incorporated in later items. We have seen in the case of headlines that the transformation of information from leads takes place along a continuum, so if we think of loss of information as being white, and completely different information as being black, we can find greys of different hues along the cline. In the case of reported and unreported news, however, the continuum is not a valid representation of the transformations. If we stay with the same metaphor: here, there are chequered patterns as well, which cannot be placed in the scale. Thus, the transformations show two main tendencies, represented in Figure 8.16:

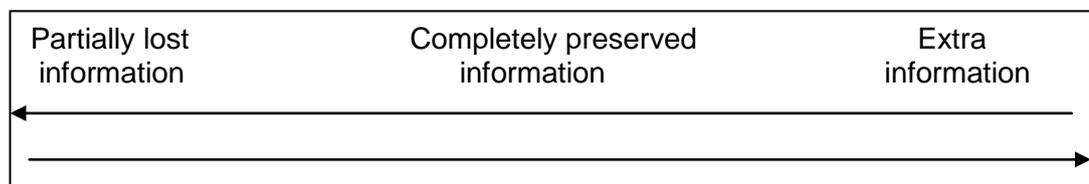


Figure 8.16 The transformation tendencies in the BBC corpus

The left end of the figure is represented by reports whose parts are deleted from the item on next mention. This may take place in the form of a simple cut-off, so that

the voice of the correspondent is kept, but the end of his/her report is left out, as in (50) below:

- (50) *The peace process in Colombia has won a last-minute reprieve after the largest rebel group agreed to attend the negotiating table just hours before government troops were due to move against the rebels' safe haven in the south-east. Foreign diplomats persuaded the Revolutionary Armed Forces of Colombia or FARC to drop their objections to military controls around the enclave.*

C. M. reports from Bogota.

In a televised address to the nation after this dramatic eleventh hour agreement President Andres Pastrana claimed a victory for the peace process. He said that the Revolutionary Armed Forces of Colombia had publicly and undeniably agreed to commit to immediate ceasefire negotiations. He added that the rebels had given in to the strong show of force displayed by the Colombian armed forces in accepting stricter military controls over the safe haven. But the President noted a new deadline: on January 20 the zone could be officially disbanded. The two groups have until then to produce some concrete results. Personally thanking the United Nations envoy James LeMoyne, Pastrana said he would continue his battle for peace. So for the moment the talks are back on, but a final solution to this violent 38-year-old civil war does still seem a long way off. (7/15)

The underlined part of the report (a Verbal Reaction and an Evaluation) has been cut off in another edition on the same day.

Another case of partially lost information is when the next mention does not feature a report anymore, and it incorporates some, but not all, information contained by the report, as in (51a, b):

- (51a) *Israeli troops backed by tanks have blown up the Voice of Palestine radio and television station in the West Bank town of Rumallah. Soldiers laid explosives in the building after clearing its occupants. Israel has frequently accused the Palestinian media of inciting political violence.*

B. P. reports from Jerusalem:

Israeli troops surrounded the hill-top broadcasting complex before dawn on Saturday. Soldiers entered the empty building to lay charges, then blew it up in a series of controlled explosions. A large cloud of smoke and debris billowed into the sky as Palestinian firemen tried to put out the flames. The Voice of Palestine broadcast official positions and Israel has accused it many times of transmitting material that fuels the Mid-East conflict. Palestinians deny the charge. They say the reports reflect the mood of the people and accuse Israel of trying to silence their media. This operation was much larger than in the past. It followed a shooting spree by a Palestinian gunman that killed six Israeli party-goers. (7/19)

- (51b) *Israeli troops backed by tanks have blown up a building housing Palestinian radio and television services in the West Bank town of Rumallah. Soldiers laid charges in the multi-storey building after clearing its occupants. A series of*

controlled explosions sent up a large cloud of smoke and debris leaving Palestinian fire-fighters battling to put out the flames. Israel has frequently accused the Palestinian media of inciting political violence and it is not the first time broadcasting facilities have been targeted. *But the BBC Jerusalem correspondent says this latest operation, part of Israel's retaliation for a Palestinian attack on Thursday, which killed six Israelis, seemed designed to completely destroy the main broadcast facility of Yasser Arafat's administration.* (8/19)

In (51b), the underlining marks parts which are exact repetitions of information carried by the report (though they may not have the same wording). The parts marked by different font types are new information. The preserved information features Details of Main Event and Reasons (though it clearly expresses a causal relationship between the shooting of six Israelis and the blowing up of the media building, a connection which has only been suggested in the report). The new information gives Evaluations (besides a small detail about the size of the building). In (51a), the underlined parts are the ones left out from the next mention: a Previous Event, Context, and a Verbal Reaction, thus overthrowing the balance of the report, which gave the Verbal Reactions of both sides.

In the middle are the cases where almost all the information conveyed by the report is kept in a subsequent item. Items (52a, b) exemplify this process:

(52a) *Following his trip to Pakistan, Mr Blair briefly visited Afghanistan for talks with the head of the interim government, Hamid Karzai, at Bagram airbase. He became the first western leader to visit the country since the fall of the Taliban.*

I. M. reports from Bagram:

In freezing temperatures in a flood-lit airplane hangar in the middle of the night, the British Prime Minister pledged his support for the people of Afghanistan. Britain's commitment, he said, is long-term. Afghanistan was a field-stage and an exporter of terrorism. But under the new interim government, he said, there was a real prospect of progress and stability. Mr Blair also paid tribute to the British troops in Afghanistan. There are now hundreds in Kabul, with more arriving every day. They will form the core of the international security force who will soon be patrolling the streets of the capital.

(7/8)

(52b) *Following his trip to Pakistan, Mr Blair briefly visited Afghanistan for talks with the head of the interim government, Hamid Karzai, at the Bagram airbase. In the first visit by a western leader to the country since the fall of the Taliban, Mr Blair said he had come to reaffirm the long-term commitment of the international community to Afghanistan. He said that, although Afghanistan had been an exporter of terrorism in the past, there was now a real prospect of progress and stability under the new interim government. He also paid tribute to the British*

troops who are there as an advance guard of the international peace-keeping force. (8/8)

In (52a), the underlined parts are those which do not appear in the later version. However, most of the information contained can be inferred from what is repeated: Afghanistan cannot be an exporter of terrorism if it is not a field-stage at the same time; and it is quite natural that troops arrive in hundreds, and – as a peace-keeping force – they patrol the streets. The only really new information that has been left out is that more and more soldiers are arriving every day. Apart from this, and a bit of a change signalled by altered font types, we may experience an entire preservation of information in the above pair.

On the right end are the cases where next mention contains extra information vis-à-vis the earlier report. We have seen such an instance in (51), but here is another example:

(53a) *After years of legal wrangling, Hong Kong's Highest Court has rejected a plea by several thousand Chinese mainlanders who wanted to stay in the former British colony. Some 5000 migrants had asked permission to stay, but lawyers interpreting the complex ruling said only a few hundred had been successful.*

From Shanghai here's D. H.:

Hong Kong's Court of Final Appeal ruled that, out of 25 mainland Chinese applicants, only three had the right to stay in the territory. The 25 were representatives of more than 5000 mainlanders with relatives in Hong Kong. They had been fighting against deportation since an initial court ruling in their favour two and a half years ago was overturned by Chinese parliament. The court ruled that only those in Hong Kong before the territory's return to Chinese sovereignty in 1997 have an automatic right of abode, while those who applied before the first court ruling in 1999 may stay at the discretion of the immigration department. (7/10)

(53b) *After years of legal wrangling Hong Kong's Highest Court has rejected a plea by several thousand Chinese mainlanders to be allowed to stay in the former British colony. Some 5000 migrants had asked permission to stay, but lawyers interpreting the complex ruling said only a few hundred had been successful. The court said only those who arrived before the end of January 1999 could remain in Hong Kong. That was when Beijing made a defensive ruling on residency rights. The immigration officials said they didn't know when the others will be sent back. Some of the claimants have asked for four weeks' grace.*

(8/10)

It is only the exact numbers underlined in (53a) that are missing from (53b). The parts highlighted by different font types contain roughly the same information in

different wording. Much more is the information – underlined in (53b) – that is completely new: a Verbal Reaction and a Consequence.

Most of the above examples demonstrate that clear cases are rare and, for example, deletion and addition of information may take place in the same item. This shows that the tendencies shown in Figure 8.16 are at work at the same time.

There are cases where a reformulation of a report gives rise to serious unclarity, similar to those found by Bell (1998:91) in a *Daily Mirror* story. Consider the following items:

(54a) *South Korea's President Kim Dae-jung has publicly apologised for a series of corruption scandals that have involved high-ranking government officials including members of the presidential office. President Kim expressed deep regret and promised that eradicating corruption would be one of the main priorities of his last year in office.*

C. G. reports from Seoul:

Several times during his New Year press conference President Kim expressed deep regret at a series of corruption scandals involving venture companies in which senior government officials were allegedly implicated. Speaking at his presidential office, President Kim promised to set up an independent agency to investigate the scandals. He also told reporters that he'd accept the resignation of the country's Prosecutor General, Shin Seungnam, who offered to step down on Sunday after the arrest of his younger brother on suspicion of receiving bribes from a businessman now facing trial on charges of embezzlement and stock price manipulation. (7/14)

(54b) *South Korea's President Kim Dae-jung has publicly apologised for a series of corruption scandals that have undermined his administration. Speaking at a news conference President Kim said he was shocked by what had happened. His apology came within hours of the country's Prosecutor General resigning after his brother was arrested on suspicion of receiving bribes. Last week the man nominated by President Kim to lead an anti-corruption drive stepped down after being linked to the jailed businessman at the centre of bribery allegations. Mr Kim also urged North Korea to respond without conditions to offers by the United States to re-start talks, which have been suspended since President Bush came to office.* (8/14)

It is very difficult to figure out that the country's Prosecutor General in (54a) is the same as the man nominated by President Kim to lead an anti-corruption drive in (54b), since the Prosecutor General is said to have resigned a couple of hours ago, and 'the man' last week, though both are said to have had links to an imprisoned businessman. The clue to the mystery is that those hours ago and the previous week are both in fact the day before, given that the day of this broadcast is Monday. Such

unsuccessful reformulations often happen and, while the reader may go back to clarify earlier parts of a press story, if he takes the effort, the listener does not have a chance for such elucidation.

Some additional remarks are needed about the use of Attributions in repeated items. It is to be noticed that the incorporation of reports into unreported news items often happens in the way that only parts of the reports are attributed to the correspondents. Other pieces of information are presented as facts, probably because the BBC regards pieces of knowledge originating from its correspondents as first-hand information. When the reporter is explicitly mentioned as a source, this happens mainly for evaluative comments, as in (51b) above. The attributions that appear within the reports themselves, however, are routinely kept, though the contrary may also take place: for example, on January 9, knowledge of a state of maximum alert in Israeli villages in Gaza is attributed to the Israeli Radio on first mention, but unattributed in a later version.

Turning to instances when unreported long news items are partially repeated, we notice very similar transformations taking place. In one group of items, information is deleted on second mention. See, for example, (55) below:

- (55) *A team from the International Red Cross has arrived at the American base of Guantanamo Bay in Cuba for the first independent inspection of conditions under which Taleban and al-Qaeda prisoners are being held. It's expected to stay several days and have individual access to all of the prisoners starting later today. D. C., a Red Cross spokesman, explains what the group's priorities are: "The most important thing is to have private talks with each prisoner, so that each prisoner can freely talk to us and tell us about the way he feels, the way he feels he's treated, but also anything related to his- his health and access to medical care." After the visit the Red Cross will compile a report on whether the captives are being treated in accordance with the Geneva Convention on prisoners of war. The United States says they are not classified as such and it's describing them as illegal combatants. (7/18)*

In the (8/18) edition of the World News, the whole underlined part of (55), that is, two Verbal Reactions, one in the form of a soundbite and the other as an indirect speech, and Future Events, are left out. Or consider (56):

- (56) *The film drama A Beautiful Mind and the musical Moulin Rouge have picked up most of the main prizes at the Golden Globe Awards in Los Angeles. The Golden Globes is generally regarded as a good indication of which films are likely to*

triumph in the industry's most prestigious award, the Oscars in March. A Beautiful Mind, the story of a mathematician's baffled with schizophrenia, was chosen as best drama, with its star Russell Crowe also winning an award as best actor. The veteran filmmaker, Robert Altman, was named as best director for his murder-mystery satire Gosford Park. And the award for best foreign language film went to No Man's Land from Bosnia. (6/21)

In a subsequent edition, (9/21), all the underlined parts in (56) have been left out, information that may be considered Details of Main Event. The last sentence of the remaining part could have been deleted as well because, without the specification that Crowe rated best actor, the fact that the film came out best drama becomes superfluous, given that its genre is also mentioned in the lead.

In another group of partially repeated items, new information is added to an earlier version. The reverse of (55) is observable in (57):

(57) *The United States military says it wants to move the remaining al-Qaeda and Taleban prisoners it is holding in Afghanistan to its military base at Guantanamo Bay in Cuba as soon as possible. A US military official, Brigadier-General Mike Lehnert, said the present camp was being expanded to hold up to 600 detainees. He said the team from the International Red Cross had been given full access to the prisoners at the base and they had made a number of recommendations: Where it's possible, we are accommodating those recommendations. But I would stress that the detainees have been here about a week now and that the detention facilities are about 13 days old, so this is a work in progress. But I would characterise the ICRC as a very professional group of individuals. The Red Cross is now expected to have a permanent presence at the base. There has been international criticism of the way the prisoners are being treated.* (8/20)

The underlined text in (57) is an addition to the (7/20) version of the item, containing a Verbal Reaction in the form of a soundbite. Another case of new information on next mention is (58):

(58) *Today is the first anniversary of the earthquake which devastated the Indian state of Gujarat, killing about 20 thousand people and destroying the homes of tens of thousands. A year later, as the exact time of the earthquake came round again, many people left their homes in the town of Bhuj, one of the worst affected towns, to stand in the open air. Some villages in the state have been almost totally rebuilt, funded by international aid agencies and Indian charities. A BBC correspondent says that in some areas many families are still living in tents with no running water.* (8/26)

Underlining, again, signals addition, in this case of Previous Events and Context/Consequence of History.

Mixed cases are also possible, as in (59a, b):

(59a) *Thousands of Hindu holy men are meeting in the Indian capital, Delhi, at the end of a week-long march from Ayodhya, where they want to build a Hindu temple on the site of a mosque that was destroyed nine years ago. The Ayodhyan temple issue has dogged the Indian government and threatens to dominate the forthcoming elections in the large Indian state of Uttar Pradesh. Hindu hardliners organised the march to press their case that work should begin immediately on the building of a temple in Ayodhya, which they claim as the birthplace of the Hindu God, Ram.* (7/27)

(59b) *Thousands of Hindu holy men are meeting in the Indian capital, Delhi, at the end of a week-long march from Ayodhya, where they want to build a Hindu temple on the site of a mosque destroyed nine years ago. Hindu hardliners want work to begin immediately on the building at Ayodhya, which they claim as the birthplace of the Hindu God, Ram. Two thousand people died in bitter fighting after the mosque was torn down by Hindu fanatics in 1992. Muslims want the mosque rebuilt.* (8/27)

The highlighted part in (59a) is an Evaluation missing from (59b), which, on its part, is supplemented with new History and Context, underlined in (59b).

As could be seen, these operations may take place in either mid-position or at the end of an item, but never at the beginning. Lead, it seems, is the most fixed part of a radio news item.

As for the transition of a long, World News item to a short, news summary one, it is easier to say what is kept than what is got rid of. In Section 7.5, we have seen that the average number of news items per a two-minute news summary is the same as in a five-minute programme (7), so the difference lies in the length of news items. Indeed, Table 7.2 shows that the average short item is around 2.5 times shorter than a long item. A close look at the transformations reveals that what is invariably kept in these items is the lead plus one more element. Consider the short news items below:

(60) *Pledges of more than two billion dollars for the reconstruction of Afghanistan have been made so far at an International Donor's Conference in Tokyo. The largest offers came from the European Union, Japan, and the United States.* (12:30/21)

- (61) *The American singer Peggy Lee has died at her home in California at the age of 81. Her hits included Fever and The Folks that Live on the Hill. She also wrote songs and appeared in several films.* (12:30/22)

The textual units preserved are Details of Main Event in (60), and some Previous Events in (61).

There are cases in which short news features additional information as well:

- (62) *Thousands of people in Brazil have been paying their last respect to the mayor of a town on the outskirts of Sao Paolo, whose body was found on a dirt road, riddled with bullets. Celso Daniel is the second mayor from the left-wing Workers' Party to be killed in the past four months.* (12:30/21)
- (63) *Armed motorcyclists have attacked the United States Information Service building in the Indian city of Calcutta, killing five policemen. Pakistan has rejected Indian suggestions that the attack has had links with Pakistani Intelligence.* (12:30/22)

The underlined part adds an Evaluation also revealing Previous Event in (62), and a Verbal Reaction in (64), directly attached to the preserved lead of the item.

We can see that, in spite of the fact that it is called 'news summaries', short news is in fact not an abstract. The very rare cases in which it functions as a summary are when, besides the lead, it also contains information from a correspondent's report, as in (64):

- (64a) *Explosive experts have started to blow up the heart of a town in south-west China as part of the controversial Three Gorges dam project. They are the first major demolitions to make way for the dam, the world's largest, which is due to start holding back water next year. 1.2 million people are being moved from their homes for the project, but the relocation has been plagued by corruption. D. H. reports from Shanghai: With the countdown shown live on national television, Fengjie township's four-storey government headquarters collapsed in a cloud of smoke and dust. Next to go was the local power plant. Local residents watched the process from the riverbank far below. State television said the explosions were an alarm bell to remind them that, by the end of this year, the town must be completely evacuated. By the middle of next year waters on this stretch of the Yangtze will rise by dozens of metres as the giant Three Gorges dam begins to hold back water. But for all the official excitement, the project remains highly controversial. Environmentalists warn that the scheme, which is designed to generate electricity and control flooding on the river, could actually cause massive silting and pollution upstream.* (7/20)

(64b) *Chinese demolition crews have begun blowing up buildings in an ancient town in the south-west, in spite of the controversial multi-billion dollar Three Gorges dam project on the Yangtze river. A government building and a power plant in the town of Yichang was the first to be destroyed in an operation shown live on Chinese television.*

(12:30/20)

The highlighted part in (64b) is a summary of the underlined information in (64a), added directly to the lead of the item, supplemented only with the phrase *on the Yangtze river* from the report.

We are now in a position to draw some conclusions about the textual transformations between different subgenres of news: (i) when a correspondent's report is later incorporated in a news item; (ii) when a long news item is partially repeated; and (iii) when a long news item becomes a short one. The transformations indicate two opposite but often simultaneous tendencies at work: one points towards partial loss of information, the other to packing extra knowledge in the repeated item, though there are cases in which the information is exactly retained. In (i), attributions to correspondents are made in the case of evaluative wordings or statements, other sources of information are normally re-mentioned. In (ii), these deletions/additions may be positioned in the middle or at the end of the item, but never at the beginning. In (iii), we find that, although they are designated as 'news summaries', short news items do not summarise long ones, except for some negligible cases of report-abstractions. In all subgenres, Lead is the most stable textual element of radio news, a unit that resists all transformations.

8.8 Summary of findings

At this point, we may reiterate the answers to the qualitative research questions asked in 6.2 in a more focused way.

- Question 5 was concerned with the topics of the BBC World Service news. Twelve topic areas have been identified: Legal issues, Conflicts, Economic issues, Human interest stories, Nature-related, Military, and Immigration issues, Protests, Media-related issues, Elections, Anniversaries and Scientific issues, in decreasing order of coverage amount. In a model offered for the topic structure of the BBC corpus, each topic area branches into topic groups (which may divide into topic subgroups), these

are made up of topics, which, for their part, consist of subtopics. However, not all topics arch over several subtopics, not all topics have topic groups governing them, and topic subgroups are rare. This complex system is a true reflection of the disparate threads of media topics.

- Question 6 sought the answer to whether there is any relationship between the topic and the position of a news item. We started off from the axiom that the amount of coverage of a topic area suggests its perceived importance – the more something is spoken of, the more significant it is felt to be. It has been found that this sense of significance is reflected in positioning as well: more important/covered topics tend to occur in higher positions, and less important/covered ones in lower slots. Thus, the hypothesis that items on more significant topics occur in front positions has been confirmed. However, Position 1 has been found to deviate from this general tendency in that the inner division of topic areas here does not follow the above-mentioned importance order, but a recency principle.
- Question 7 searched for the way in which partly or entirely repeated items change their positions. Roughly speaking, half of the repeated items in the BBC corpus occupy the same position as earlier, one-third of such items slip one slot down, and one-fifth move upwards. Therefore, the assumption that most repeated items slide downwards was wrong.
- Question 8 was interested in the discourse structure of radio news, and how two extant models for the textual make-up of press stories can be applied to radio news. In exploring the structure of the BBC news, a mix of the terminology in the two – otherwise quite similar – models has been applied. In compliance with the hypothesis, it has been found that both models are, to a certain degree, apt for being used in radio news analyses. However, Bell's category of Attributions – absent from Van Dijk's model – has been found indispensable. Some new subcategories, such as Verbal Main Events, Future (Main) Events and Reasons, have been introduced to make up a modified schema of radio news. As for textual structure, it has been found that Leads and Main Events are one and the same in radio news, and that Leads/Main Events and Attributions are compulsory elements of radio news. These are most typically complemented by near-past previous events and a description of the present situation, possibly accompanied by an evaluation. Less common are reports on the story future, and even less frequent are those elements, which deal

with either the remote past or an uncertain future. Some peculiarities of tense usage in radio news have also been identified.

- Question 9 was meant to find out the nature of headlines and their relationship to news items. Thus, headlines are peripheral and optional elements of radio news. Headlining takes place along certain rules, being attached (i) to the first three elements of an unreported news programme; (ii) to the reported elements of a mixed news programme; and (iii) in the same order as that of the news items they belong to. Deviations from these rules are conditioned by internal factors such as the topic of the item; or external factors, e.g. time of edition, or proximity of other programmes. In a structural sense, headlines are summaries of news items, and the information they transmit more or less overlaps with that contained by the lead. The degree of this overlap can be placed along a continuum from loss of information through retaining to extra and different information.
- Question 10 asked about the textual transformations between different subgenres of news: (i) when a correspondent's report is later incorporated in a news item; (ii) when a long news item is partially repeated; and (iii) when a long news item becomes a short one. It is concluded that, in all cases, the transformations indicate two opposite but often simultaneous tendencies at work: one points towards partial loss of information, the other to packing extra knowledge in the repeated item, with cases in the middle in which the information is exactly retained. The supposition that deletion is the most common transformation type, therefore, has not been verified by the data: there is much more to it than simply editing out certain parts. Nor has the hypothesis been confirmed that short news summarises long news. It may well be the case that deletions/insertions and mixtures of these are in fact instances of the same operation of information management, just in opposite directions. To put it simply, the core of a radio news item seems to be the lead, and all other elements are subject to discretionary alterations as dictated by time, availability of extra information, individual working habits, and other possible constraints.

9 A supplementary analytical framework for radio news

The present chapter is meant to be a complementary analytical framework for the one described in Chapter 8. It is suggested that texts in radio news cross the boundaries of individual pieces of news, they arch over several items. The framework presented here is just an outline, needful of more empirical work. However, if it proves to be tenable, it may be applied to texts other than radio news, and reach a more general validity. This chapter, therefore, is in fact a flash on possible future research.

To decide where the boundaries of a news text are, first we must clarify certain basic notions.

9.1 What is a text?

The interpretations of the above term are many. De Beaugrande (1985:48) warns that such definitions as ‘unit above the sentence’ or ‘sequence of sentences’ are at best incomplete and at worst inadequate, because they bypass the status of the text as a communicative event. Therefore, he thinks, the distinction between a text and a non-text cannot be determined by formal definition; it can only be explored as a gradation of human attitudes, actions and reactions. Normally, a non-text occurs only when someone deliberately blocks communication.

First of all, certain distinctions have to be made. For example, there needs to be a clarification of differences between ‘text’ and ‘context’. Schiffrin (1994:363-364) uses the term ‘text’ to differentiate linguistic material (e.g. what is said, assuming a verbal channel) from the environment in which ‘sayings’ (or other linguistic productions) occur (context). In terms of utterances, then, ‘text’ is the linguistic content: the stable semantic meanings of words, expressions, and sentences, but not the inferences available to hearers depending upon the contexts in which words, expressions, and sentences are used.

The other distinction has to be made between the terms ‘text’ and ‘discourse’. The simplest differentiation between the two is made on the basis of the mode they are delivered in. Stubbs (1983:9) thinks that the two expressions, as they are used in the literature, often simply imply slight differences in emphasis. He finds that, first, one often talks of written text versus spoken discourse. Or alternatively, discourse often

implies interactive discourse; whereas text implies non-interactive monologue, whether intended to be spoken aloud or not. For example, one talks of the (written) text of a speech. A second distinction, as Stubbs sees it, is that discourse implies length, whereas a text may be very short. In this usage complete texts include: 'Exit' or 'No smoking'.

Another way of drawing a distinction is, in Stubb's formulation (1983:9-10), 'theoretically loaded': here, Stubbs thinks of van Dijk's (1977:3) usage of the term 'text' to refer to an abstract theoretical construct which is realised in 'discourse'. In other words, text is to discourse as sentence is to utterance. Van Dijk says those utterances which can be assigned textual structure are thus acceptable discourses of the language, that is, they are well-formed and interpretable.

Kress (1985:27) finds that the terms 'text' and 'discourse' have tended to be used without sharp distinction in the literature, both referring to language structure beyond the level of sentence. He thinks the difference in usage is connected to the theoretical framework one works in, so that discussions with a more sociological basis or aim tend to use the term 'discourse', while those with a more linguistic basis or aim tend to use the term 'text'. Instead, he proposes a firmly drawn distinction, with each term having its own quite specific and distinctive area of reference. Thus, discourse is a category that belongs to and derives from the social domain, and text is a category that belongs to and derives from the linguistic domain. The relation between the two is one of realisation: Discourse finds its expression in text. However, this is never a straightforward relation; any one text may be the expression or realisation of a number of sometimes competing and contradictory discourses.

De Beaugrande wrote in (1985:47) that, by then, the consensus had arisen that the text should be defined as 'a natural language occurrence in a communicative setting', and the discourse as 'a set of mutually relevant texts'.

In Togeby's (1994:4579) clear-cut formulation, a text is normally one person's written or spoken utterance, intended as one message to one audience about one coherent topic in one concrete situation. In other words, a text is defined as a communicative event in which the communicator transfers to the audience, by means of language, and with social consequences, some propositional content. In this view, the text is, across time, delimited by a change of one of the factors in the communicative situation - a shift of communicator, of communicator's intention, of topic, of audience, of social relations, of communication channel, or of language establishes the beginning of a new text. Apparently, this is a much narrower sense of text as in de Beaugrande

1985, where a non-text was seen to occur only if someone deliberately blocks communication.

In some interpretations, the terms 'text' and 'discourse' are used interchangeably. In an overview of the literature on discourse, McHoul (1994) identifies three major approaches: the Formal Approach, characterised by a faith in formal linguistic methods of analysis, and represented by Hallidayan systemic-functionalism on the one hand, and the more recent text linguistics of de Beaugrande and the early van Dijk. The second is the Empirical Approach, featuring sociological forms of analysis, and represented by conversation analysis. The third is the Critical Approach, based on Foucauldian traditions, and represented by critical linguistics and critical discourse analysis. If we accept this categorisation, then the present study pursues the first line. Turning back to distinctions, in McHoul's terms (1994:941), the Formal Approach solves the problem by considering discourse identical with text, namely, a relatively discrete subset of a whole language, used for specific social or institutional purposes.

Due to the 'formal' affiliation of the present study, more space will be consecrated here to the conception of text in the first approach mentioned above, represented by Halliday and Hasan's seminal 1976 work on cohesion, and de Beaugrande and Dressler's similarly influential 1981 introduction to text linguistics.

Halliday and Hasan (1976:1) see 'text' as referring to any passage, spoken or written, of whatever length, that does form a unified whole. A text may be spoken or written, prose or verse, dialogue or monologue. It may be anything from a single proverb to a whole play, from a momentary cry for help to an all-day discussion on a committee. However, the linguists make the important statement that the distinction between a text and a collection of unrelated sentences is in the last resort a matter of degree.

Contrary to the now outdated view that a text is a sequence of language longer than the sentence, Halliday and Hasan (1976:2) state that a text is not something that is like a sentence, only bigger; it is something that differs from a sentence in kind. A text is best regarded as a semantic unit: a unit not of form but of meaning. Thus it is related to a clause or sentence not by size but by realisation, the coding of one symbolic system in another. A text does not consist of sentences; it is realised by, or encoded in, sentences.

In a search for what makes the text hang together, Halliday and Hasan (1976:4) propose a concept of cohesion that is a semantic one; it refers to the relations of

meaning that exist within the text, and that define it as a text. Thus, cohesion occurs where the interpretation of some element in the discourse is dependent on that of another. The one presupposes the other, in the sense that it cannot be effectively decoded except by recourse to it. This means that a text, as the researchers view it (1976:6-7), is not a structural unit; and cohesion, in the sense in which they are using the term, is not a structural relation. Since texts cohere, cohesion within a text – texture – depends on something other than structure. There are certain specifically text-forming relations which cannot be accounted for in terms of constituent structure; they are properties of the text as such, and not of any structural unit such as a clause or sentence. Halliday and Hasan's use of the term 'cohesion' refers specifically to these non-structural text-forming relations.

Halliday and Hasan (1976:23) supplement the concept of cohesion with that of register (see Section 1.2.1), since the two together effectively define a text. A text is a passage of discourse which is coherent in these two regards: it is coherent with respect to the context of situation, and therefore consistent in register; and it is coherent with respect to itself, and therefore cohesive. The linguists again state that texture is a matter of degree.

De Beaugrande and Dressler (1981:3) take a more concrete position in saying that a text will be defined as a communicative occurrence which meets seven standards of textuality. If any of these standards is not satisfied, the text will not be communicative. What exactly Halliday and Hasan's 'texture' and de Beaugrande and Dressler's 'textuality' are will be clarified in the next section.

An additional remark about the specific nature of media texts: Garrett and Bell (1998:2) think that, amid the conspicuous lack of agreement on definitions of both 'discourse' and 'text', language in the modern media cannot realistically be seen in terms of the traditional linguistic distinction between these terms, as spoken and written language respectively. To be able to account for media texts as well, a new distinction has emerged: text tends to be used to refer to the outward manifestation of a communicative event, whereas discourse (following Cook) is not concerned with language alone, it also examines the context of communication: who is communicating with whom and why; in what kind of society and situation, through what medium; how different types of communication evolved, and their relationship to each other.

9.2 What makes a text?

Halliday and Hasan (1976:2) view their concept of texture as entirely appropriate to express the property of 'being a text'. A text has texture, and this is what distinguishes it from something that is not a text. It derives this texture from the fact that it functions as a unity with respect to its environment.

In more details (1976:20), the hearer or reader, when he is determining, consciously or unconsciously, the status of a specimen of language, invokes two kinds of evidence, the external as well as the internal: (s)he uses not only linguistic clues but also situational ones. Linguistically, (s)he responds to specific features which bind the passage together, the patterns of connection, independent of structure, that are referred to as cohesion. Situationally, (s)he takes into account all (s)he knows of the environment: what is going on, what part the language is playing, and who are involved."

The reason why Halliday and Hasan (1976:25) may say that, in fact, there are degrees of texture, is because one may at times be uncertain as to whether a particular point marks a continuation of the same text or the beginning of a new one. This is because texture is really a 'more-or-less' affair. A partial shift in the context of situation – say a shift in one situational factor, in the field of discourse or in the mode or tenor – is likely to be reflected in some way in the texture of the discourse, without destroying completely the continuity with what has gone before. This remark will prove extremely useful in the framework outlined in section 9.3.

The linguists (1976:28) point out that it is a characteristic of a text that the sequence of the sentences cannot be disturbed without destroying or radically altering the meaning. A text has meaning as a text, whereas a passage consisting of more than one text has no meaning as a whole; it is simply the sum of its parts. Within a text, the meaning of each sentence depends on its environment, including its cohesive relations with other sentences.

Media language, as signalled by Garrett and Bell above, is problematic. For example, if we translate the above account to radio discourse, this means that, for example, the text of a news programme is not likely to be a single text: it is the sum of the news items building it up. However, if a text derives its texture from the fact that it functions as a unity with respect to its environment, a news programme must be one text.

Textuality, in de Beaugrande and Dressler's (1981:3-10) sense, lies in seven text-forming features. The first two are text-centred notions, namely, cohesion and coherence. Cohesion is the result of the ways in which the components of the surface text, i.e. the actual words we hear or see, are mutually connected within a sequence. Cohesion, therefore, rests upon grammatical dependencies between text elements. Coherence is arrived at by the ways in which the components of the textual world, i.e. the configuration of concepts and relations which underlie the surface text, are mutually accessible and relevant. In van Dijk's (1998:36) sociocognitive approach, coherence is defined relative to models. That is, roughly speaking, a sequence of sentences is coherent if a model can be constructed for it.

The next two criteria of textuality are user-centred ones, the first of which being intentionality, that is, the text producer's attitude that the set of occurrences should constitute a cohesive and coherent text instrumental in fulfilling the producer's intentions, e.g. to distribute knowledge. Then follows acceptability, namely, the text receiver's attitude that the set of occurrences should constitute a cohesive and coherent text having some use or relevance for the receiver, e.g. to acquire knowledge. Further, a text should also display a proper degree of informativity, which is the extent to which the occurrences of the presented text are expected vs unexpected or known vs unknown or uncertain. A text should also meet the requirement of being situational, that is, to possess the factors which make a text relevant to a situation of occurrence. And finally, a text should also be relevant in terms of intertextuality, that is, it should have factors which make the utilisation of one text dependent upon knowledge of one or more previously encountered texts. Intertextuality is, in a general fashion, responsible for the evolution of text types as classes of texts with typical patterns of characteristics. Thus, in de Beaugrande and Dressler's system, a text is really a text if it is cohesive, coherent, intentional on the part of the speaker/writer, acceptable from the point of view of the listener/reader, not less and not more informative than needed, relevant to situation, and if it complies with other texts of the same type.

9.3 The concept of PANTEXT

I propose a new term as an analytical tool for dealing with radio news discourse: PANTEXT²³. The term has been coined from the prefix *pan-* and the lexeme *text*, with recourse to the original meanings of the units combined. The whole of section 9.1 above has been dedicated to defining ‘text’. As for *pan-*, its entry in Merriam-Webster's Collegiate Dictionary (the most all-encompassing of the definitions I have found) goes like this: *1: all: completely 2a: involving all of a (specified) group b: advocating or involving the union of a (specified) group 3: whole: general*. It is sense 2a-b that I take up when interpreting PANTEXT.

In this way, by PANTEXT I mean all the news items broadcast on a radio station (in this case, the BBC) which deal with the same topic. A pantext, therefore, is a text which has been written since the topic was first taken up, and is being written until it is dealt with for the last time. A pantext is concerned, for example, with current affairs. Depending on the ‘hotness’ of the issue, the length of a pantext could range from just two pieces of news to a potentially endless number of items. However, since items are often repeated in the programmes of a radio station – as we have seen in the BBC corpus, too –, the definition has slightly been modified in the following way: a pantext is a collection of all news items on a radio station which deal with, and add new information to, the same topic.

In the above working definition, the concept of ‘topic’ is used in the sense offered by van Dijk (1988a:31): “the summary, gist, upshot or most important information” of a text or, in an earlier formulation (1977:6), as denoting what a discourse or part of it “is about”. Thus, he (loc cit) continues, it may intuitively be said that several sentences in sequence belong to the same topic of discourse. Later, Van Dijk uses two additional labels for levels where textuality, that is, the property of sequential sentences to ‘hang together’ or form a ‘unity’, resides on: local and global coherence. Local coherence is present between propositions of text or talk, being a necessary, but not sufficient, condition for discursive coherence. Another unifying principle is at work, Van Dijk continues, namely that of overall or global coherence, as it is defined by ‘topics’ of paragraphs, large stretches of text or whole discourses. Such

²³ The term would probably make more sense spelled with a hyphen, a possibility which I have abandoned for copyright reasons: a company exists in Germany specialised in translations, named Pan-text.

topics may be formally described as semantic macrostructures that are derived from local microstructures by specific mapping rules (1998:36-38). We can see how the earlier concept of macrostructures is incorporated into the new idea of global coherence.

It is on this level of macrostructure and global coherence that I place the concept of the topic a pantext deals with. However, the topic of a pantext is not just of any kind, it is a special sort that I named PANTOPIC. We have seen in Section 6.5 that the topic structure of the news broadcast on a radio station is quite complicated, made up of topic areas, topic groups, (occasionally) topic subgroups, then topics and subtopics. In this system, represented in Figure 6.1, pantopics are placed on the level of topics in the following way: a topic becomes a pantopic if it has several subtopics. A topic standing by itself is not a pantopic. In other words, a topic becomes a pantopic – constructive of a pantext – if it is dealt with through several news items or days.

We are now in a position to give a final definition of pantexts:

A PANTEXT is the unity of all the news items broadcast on a radio station which deal with, and add new information to, the same PANTOPIC, that is, a topic that is pursued across several items, and has distinct subtopics.

For an illustration of a pantext on a specific pantopic, consider, for example, the following five news items (for the sake of brevity, I have chosen them exclusively from short news):

- (65) *The Zimbabwean Parliament is due shortly to begin reconsidering a controversial election law which was rejected on Tuesday. The opposition complains it will allow the government to rig elections, stifle dissent and restrict the freedom of the Zimbabwean media.* (12:30/9)
- (66) *Parliament in Zimbabwe is to resume debates on controversial new security and public order measures today after a marathon 14-hour session, in which the opposition says it won important concessions.* (12:30/10)
- (67) *Zimbabwe is facing the threat of economic sanctions from the European Union, a day after its parliament approved measures which critics say will enable the government to stifle opposition. The EU could block hundreds of millions of dollars in aid.* (12:30/11)
- (68) *A Southern African regional conference has opened with the host, Malawi, appearing to rule out sanctions against Zimbabwe. The Malawian vice-president instead spoke of the need for support for dialogue inside the country. He said political instability in Zimbabwe was depressing trade and discouraging investment in Southern Africa.* (12:30/12)

- (69) *The Zimbabwean Parliament is due to discuss a media bill today that's been criticised for curbing press freedom in advance of the presidential election in March. The bill seeks to prevent foreign correspondents being based in the country and requires local journalists to apply for a permit.* (12:30/13)

It can easily be noticed, if first on just an intuitive basis, that all five news items deal with the same topic, a pantopic that could be formulated as 'new legal measures in Zimbabwe'. Naturally, every item has a topic of its own, namely, (65): 'reconsidering a controversial election law'; (66): 'debates on controversial new security and public order measures'; (67): 'a threat of economic sanctions from the European Union'; (68): 'a Southern African regional conference to rule out sanctions against Zimbabwe'; and (69): 'discussion of a media bill'. However, if one were to give titles to the above individual items, these five are the ones that could be used as subheadings, and the common heading for all five could be the pantopic suggested above.

The first problem that such a concept may pose is whether a pantext is a text at all. For examining this aspect, I will ponder over the status of pantext on the basis of Beaugrande and Dressler's 1981 and Halliday and Hasan's 1976 criteria of textuality or texture, respectively. For this purpose let us have a look at the Zimbabwe text, at this point *as if it were* a single text, with the news items presented as paragraphs:

The *Zimbabwean* Parliament is due [shortly] to begin reconsidering a controversial **election law** which was rejected on Tuesday. The opposition complains it will allow the *government* to rig **elections**, stifle dissent and restrict the freedom of the *Zimbabwean* media.

Parliament in *Zimbabwe* is to resume debates on controversial new security and public order **measures** [today] after a marathon 14-hour session, in which the opposition says it won important concessions.

Zimbabwe is facing the threat of economic **sanctions** from the European Union, a day after its parliament approved **measures** which critics say will enable the *government* to stifle **opposition**. The EU could block hundreds of millions of dollars in aid.

A Southern African regional conference has opened with the host, Malawi, appearing to rule out **sanctions** against *Zimbabwe*. The Malawian vice-president instead spoke of the need for support for dialogue inside the country. He said political instability in *Zimbabwe* was depressing trade and discouraging investment in Southern Africa.

The *Zimbabwean* Parliament is due to discuss a media **bill** [today] that's been criticised for curbing press freedom in advance of the presidential **election** in March. The **bill** seeks to prevent foreign correspondents being based in the country and requires local journalists to apply for a permit.

The first criterion of textuality suggested by Beaugrande and Dressler is cohesion, that is, the ways in which the components of the surface text i.e. the actual words we hear or see, are mutually connected within a sequence" (1981:3). Cohesion rests upon cohesive ties of the kind Halliday and Hasan present in detail. Out of the types of cohesion listed (1976:279), namely, reference, substitution, ellipsis, conjunction and lexical cohesion the above pantext mainly displays lexical cohesion. There are many cases of reiteration in the text, out of which I will pick only those which permeate through item-boundaries, since it is these which prove the textuality of the pantext. One can notice same-word repetitions of lexical items (printed in similar font types), the shifting of already used elements to different classes (termed as 'partial recurrence' by Beaugrande and Dressler 1981:49) such as *Zimbabwe - Zimbabwean, critics - criticised*; synonyms or near synonyms: *law/ measure/ bill, restrict the freedom of the media/ curb press freedom; begin reconsidering/ resume debates /due to discuss*; and superordinates: *country to Zimbabwe*.

Since the 'paragraphs' of this pantext follow each other with considerable stretches of time in between (speech-mediated discourses evolve in time, not in space), other types of cohesive elements are not characteristic of pantexts. Beaugrande and Dressler point out that, for the successful decoding of reference through e.g. anaphors, or of ellipsis, the previous discourse that the cohesive tie relates to must be in *active storage* (1981:48) in the mind of the listener. This storage is short-term so, in the case of a pantext, there is no way it would be active over often 24-hour lapses of time.

Another type of cohesion, proposed by Beaugrande and Dressler, is represented by surface signals for the relationships among events or situations in a textual world: tense, aspect and junction (1981:49). The fact that a pantext develops over a longer period than other types of speech-mediated discourse makes the sequence usage of tense and aspect impossible. The dominant verb form in the above pantext is the simple present, which is due to the real-time character of each paragraph. This is why, if read one after the other, the word *today* is confusing, since all these actions cannot take place on one and the same day. Present in a pantext is real present, past is real past etc., in accordance with the way the news keeps abreast of current events. In a way, this omnipresence of the present can be interpreted as similar to the 'historical present' used in narratives to reach a dramatic effect of immediacy, or to the kind of present tense used in telling the plots of novels or films.

It may be argued here that what kind of a text is one in which there are no cohesive elements between paragraphs, in the form of linking adverbs such as *However* or *Finally*. Nevertheless, here is what Bell (1998:90) observes about news texts:

“News stories are standardly written as a series of one-sentence paragraphs, and commonly express little linkage between the sentences. With each news sentence usually also its own paragraph, there is no larger unit of text organisation. There is routinely no flow of time sequence from one sentence to the next, and a lack of devices such as adverbs expressing linkages between sentences. It is common for cohesion between sentences to be unclear or nonexistent, and we may be genuinely doubtful what actions within the story belong together, at what point location actually shifted, or what material is attributed to whom.”

The reason for this is given by Bell elsewhere (1991:172): the ideal news story is one which could be cut to end at any paragraph. Whatever the reason, if the general lack of cohesion between paragraphs is true of written news stories, which we, however, never regard as non-texts, we have no reason to expect of a spoken pantext to be more cohesive.

The second criterion of textuality, Beaugrande and Dressler (1981:4) say, is coherence, concerning the ways in which the components of the textual world, i.e. the configuration of concepts and relations which underlie the surface text, are mutually accessible and relevant. The coherence of the Zimbabwe text is obvious. It has a global topic, and the paragraphs present continuity of entities and progression of events. We recognise the temporal succession of events reflected by the order of delivery: first the Zimbabwean Parliament debates the election law in a 14-hour session, then it discusses security and public order measures, and finally the media bill. We realise the causal relationship between the draconian Zimbabwe laws, the EU threats of economic sanctions, and the Southern African reaction to rule out sanctions. (These cause-and-effect relations reveal temporal ones as well.) The pantext is abundant in entities that belong to the same semantic field: *government + opposition + political instability + presidential election; economic sanctions + depressing trade + discouraging investment + dollars in aid* etc.

The next two criteria of textuality is intentionality, that is, whether the producer of the text intends it to be a cohesive and coherent text which distributes knowledge; and acceptability, that is, the extent to which the receiver of a text takes it as a cohesive and coherent text from which (s)he can acquire knowledge. Here, it might seem problematic that these news items have probably not been produced by the same person.

However, it can be argued that news programmes are manifestations of the BBC World Service's institutional voice suggesting, also in openings and closing, something like "This is the BBC World Service speaking to you from London". Similarly, this *you* is also an institutional role of the ideal, regular listener of the BBC World Service. In reality, people rarely listen to all news programmes of a station, making repetition necessary of both entire items or of pieces of information from earlier items. However, when a listener skips a day, (s)he is aware of the fact that some information must have been missed, and accepts it as such, thus acknowledging the information on a pantopic as a flowing unit.

The next text-forming feature is informativity. As we have seen in the Zimbabwe text above, pantexts display much redundancy. However, this does not mean low informativity, because, since a pantext is delivered across several days, listeners might forget certain pieces of knowledge, so a pantext is just as informative as it should be to help recall. Further, as Cameron (2001:33-34) points out, a look at faithful transcriptions of natural conversations reveals that real talk is very repetitive and incoherent, but repetition is useful for the hearer, because it increases the chance that (s)he will be able to take in the information before it disappears. When judging the textuality of a pantext we should not remember that it is intended for the ear, and if we find that it is just as repetitive as natural conversation, we are in a good position in terms of textuality, since we would not think of a conversation as a non-text. Regarding informativity, research suggests that the problem of radio news is often that it is too informative, not that it is not informative enough (see van Dijk 1988a or Bell 1991 for news understanding).

The remaining two criteria of textuality, situationality and intertextuality, are not problematic. Pantexts are absolutely relevant to the situation of a journalist producing a text for the audience and a listener receiving it as such. The listener switches on the set because (s)he wants to find out about the news, and the station complies with this wish. The text type of radio news is thus well-known: the BBC corpus has indicated that the news is uniform, it follows a very well established norm of news-writing, and the audience is also very familiar with this text type.

The concept of intertextuality, however, offers an angle from which the textuality of pantexts is even more acceptable: that of *feuilletons*. Pantexts are similar in a sense to such novels published in instalments. This idea is in consonance with what Bell postulates about news stories (1991:154): that a possible resolution in a news story

is only the latest step in a continuing saga. The news – he continues – is more like a serial than a short story: the criminal was arrested, but the trial is in the future; the accident occurred, but the victims are still in hospital. Thinking of feuilletons, we would undoubtedly consider the instalments parts of one and the same novel. This might give us an inclination to regard separate news items as sequels of the same story written by life.

Some more linguistic evidence for the existence of pantexts is due. Pantexts in the BBC corpus, however, are typically quite long, so for space-saving reasons only relevant parts of a pantext will be presented here, with the full pantext to be found in Appendix 14. Consider, for example, the following item:

- (70) *The first group of about twenty al-Qaeda and Taleban prisoners are on their way from Afghanistan to Cuba, where they will be detained at the American base at Guantanamo. The prisoners are being treated as potentially dangerous, overhooded and shackled, and will remain chained to their seats for the 20-hour journey. Human rights groups have expressed some concern at the treatment of the prisoners. The director of the New York based Human Rights Watch told the BBC he was particularly concerned at reports the prisoners have been drugged or sedated. More prisoner airlifts are expected over the next few days.* (7/11)

The expression *first* at the beginning of the item is cataphoric not only of the last sentence of the item, but also suggests that there will be more groups of prisoners arriving in Cuba, information that the station will surely report on. The next ‘paragraph’ of the pantext goes like this:

- (71) *As 20 Taleban and al-Qaeda prisoners spent their first night in the United States detention camp in Cuba, the American Defence Secretary has said he does not consider them prisoners of war. Mr Rumsfeld said the detainees were illegal combatants, and under that status were not protected by the Geneva Conventions. However, he insisted the prisoners would be treated humanely.* (8/12)

Again, we have an indication of a *first night* in the item, which is anaphoric of the information that they will be detained at the base in the previous item, but also cataphoric of further news about how they spend their time there. The statement of the US Defence Secretary that these detainees cannot be considered prisoners of war, and still, they will be treated humanely, is an answer to the human rights groups’ concern about the way the detainees are being handled.

A subsequent item is a continuation of the dialogue going on between US authorities and people standing up for the prisoners' human rights:

- (72) *The British Foreign Secretary Jack Straw has expressed concern over the detention of at least 50 Taleban and al-Qaeda prisoners at a US military base in Cuba. Mr Straw said the civilised world had a responsibility to uphold their rights. The American Defence Secretary, Donald Rumsfeld, in an interview with the BBC, insisted that conditions at the base were humane, vastly better than those in which the men have been found.* (8/16)

In a later 'paragraph', the story goes on:

- (73) *Another batch of Taleban and al-Qaeda prisoners has arrived at the American military base at Guantanamo Bay in Cuba. It is the third group to go there, bringing the total number of captives to 80. As for the previous prisoners, they were taken blindfolded and chained to outdoor cages where they would be held. The human rights organisation Amnesty International has added its voice to criticism of the conditions in which the prisoners are being held. It said the cages where they were locked up fell below minimum standards for human treatment.* (6/17)

The beginning of the item, *another batch of Taleban and al-Qaeda prisoners*, is an indicator that a similar journey has already taken place. It seems that a radio station would never formulate its news like this if it had not already reported on the event referred back to. Similarly, it would never say *next week's Donor Conference* if it had not signalled earlier that there will be such a conference. If there was no news on this earlier, the formulation would be *a Donor Conference next week*. This is a convincing proof of the existence of pantext. Another cohesive tie to previous items in the above sequence is the formulation that Amnesty International has added its voice to criticism, which refers back to earlier mentions of such critical remarks.

A further cohesive pair is exemplified below:

- (74) *The Red Cross is expected to visit the American military base at Guantanamo Bay in Cuba today to inspect the conditions under which Taleban and al-Qaeda prisoners are being held.* (8/17)
- (75) *He [US military official Brigadier-General Mike Lehnert] said the team from the International Red Cross had been given full access to the prisoners at the base and had interviewed each one privately.* (7/20)

In (74) we get to know that a Red Cross team is going to visit the base. Three days after, in (75) the representatives of the organisation are referred to as *the team from*

the International Red Cross, the definite article clearly showing a second mention. Or, in (76):

- (76) *The United States military says it wants to move the remaining al-Qaeda and Taleban prisoners it is holding in Afghanistan to its military base at Guantanamo Bay in Cuba.* (7/20)

The expression *the remaining al-Qaeda and Taleban prisoners it is holding in Afghanistan* suggests that the great majority of the prisoners have now been moved to Cuba. On the next day comes an answer to human rights groups' concern broadcast eleven days earlier that prisoners are drugged on the way. What is communicated here is related to a more recent airlift, however, it also functions as a justification of the facts the human rights groups got in possession of:

- (77) *Officials at the United States naval base at Guantanamo Bay in Cuba say two suspected Taleban and al-Qaeda prisoners transferred there from Afghanistan on Sunday had to be sedated during the flight because of their disruptive behaviour. A military spokesman says the men were drugged after they began yelling and thrashing around in their aircraft seats.* (7/21)

Many similar examples could be called to help, but the ones above are probably enough to show that the concept of pantext is a tenable one. Just one more remark: in his study of discourse organising elements (DOEs) in radio news, Al-Shabab (1986:117-118) tentatively argues that Opening DOEs (*This is the BBC in London. The news read by XY.*), Medial DOEs (*You're listening to the news from the BBC.*) and Closing DOEs (*That's the end of the news.*) deictically refer to the news as one language event, so they serve as external evidence of the discourse unity of news. However – he continues –, there is the challenge that the news items are on different topics, so they do not form a coherent whole. It is not sure whether the 'discourse unity' Al-Shabab talks about means 'one and the same text', but if it does, this can be applied so much more in the case of pantext-forming news items which really are on the same topic.

To refer back to the theoretical introductory part of this chapter, we find that Togeby's conception of a text, which states that a shift in any of the components of a speech event may signal the end of a text and the beginning of another, is too narrow to comprise pantexts as well. However, de Beaugrande's notion of a text becoming a non-

text only if someone deliberately blocks communication, is perfectly applicable to pantexts, too, where such a blocking would be if a listener decides to give up following the news altogether. Further, if we Halliday and Hasan's position that texture is a matter of degree, without claiming that a pantext is as much of a text as a well-written short story, for example, we have to accept that it is a text. As Halliday and Hasan (1976:11) put it, although in judging whether there is texture or not we certainly have recourse to some feeling about how much the sentences do actually interrelate in meaning, we could not give any very explicit account of the degree of relatedness that is needed or how it is to be measured. Pantexts, therefore, might not be very typical or central members of the category, but may well belong to peripheral sites of the class of texts.

One may wonder: would it not be enough to make a less controversial statement than the postulation that news items occurring days apart form a single text, and say that these are items of the same discourse? However, this would be too wide, since all texts on the radio may be thought of as making up radio discourse. Moreover, this would not be able to account for the clear connection between two news items on, let's say, Prince Harry's bad habits, but the not so apparent relationship between an item on an interview with a paedophile murderer and one on the return of the exiled former Afghan king's son. It is proposed here that, if textuality is really a matter of degree (Halliday and Hasan 1976:1), then a pantext can be considered a real text, albeit not the most typical, central type.

9.4 Future applications

Although the concept of pantext needs further examination, it is suggested here that this notion can be widened to comprise cases other than radio news. Feuilletons have already been mentioned, but soap operas and other kinds of serials can also be thought of as pantexts, this time with visual elements as well. The following is a sketch of possible pantexts, with no claim to completeness:

We have seen that a shared topic is an essential characteristic of pantexts. Therefore, letter, e-mail and SMS-message *exchanges* have been written instead of correspondence in general, to suggest that these are on the same topic. Similarly, out of all medical examinations, only those can be regarded pantexts, which take place between the same doctor and patient, and lead up to the curing of the same ailment. Out of all business meetings, a pantext is made up of discussions between the same business

partners until they reach an agreement on one transaction. Those tutorials are pantexts which take place between the same teacher and student, and lead up to, for example, one BA thesis. Further, those university lectures constitute a pantext which are given by one teacher to one group of students in one semester, and so on.

		PANTEXTS	
		Interactive	Non-interactive
Single-channelled	Written	letter-exchanges e-mail-exchanges SMS message exchanges	feuilletons diaries
	Spoken	same-topic talks repeated medical examinations business meetings tutorials	university lectures campaign speeches radio news voice-diaries
Multi-channelled		film series, serials several-part films	

Table 9.1 A tentative classification of possible pantexts

9.5 Summary

This chapter has outlined a possible framework which may supplement the analytical tools at hand for exploring radio news. The concept of pantext and a corresponding idea of pantopic have been introduced, to capture the unity of a series of news items on the same topic in the output of a radio station. It has been demonstrated through an application of textuality principles that pantexts show many features which may classify them as certain kinds of narrative text types in their own right. Finally, while it has been admitted that further research is necessary to sustain the argument, the concept of pantexts may be generalised to a number of linguistic phenomena other than radio news.

Conclusions

The present study has been a non-critical, product-oriented work aimed at revealing the main grammatical and discourse features of radio news, and its spoken vs written character. The relative novelty of the study stands in its systematic attention to the relatively neglected research area of radio news both on micro- and macrolinguistic levels, and in successfully challenging the assumption that news represents an extreme of written language in the oral medium of radio.

The research questions were formulated so that they can find out whether radio news is any different from press news in its grammatical features and discourse structure, and if subtypes of radio news are dissimilar from each other in grammatical features, size and textual make-up. The study also aimed at discussing whether two frameworks for the textual structure of printed news (van Dijk's and Bell's models) are relevant for radio news as well. Additionally, it was asked whether there is any relationship between the topic, positioning and headlining of radio news items.

After reviewing the literature on media research and orality vs literacy, the study turned to the presentation of the grammatical and textual background against which radio news can be viewed. To obtain a ground for the comparison of radio news with spoken and written text types – both on levels below and above the sentence –, two kinds of reviews were pursued:

For a grammatical background, the microlinguistic features of printed news and conversations were summarised on the basis of a large-scale, corpus-based study (Biber et al. 1999). It was found that, in both press news and conversations, the grammatical features correspond to *what kind* of information has to be communicated on the one hand, and *how* this information is conveyed on the other. News is mainly concerned with current events, newly emerged states, and the acts and words of elite personalities; hence the use of activity and existence verbs, heavily modified nouns and circumstance adverbials. It also packs as much information as possible in a limited space; hence its lexical density, long sentences and phrases, a high frequency of nouns, determiners, prepositions, attributive-affiliative adjectives, agentless passives, and nonfinite clauses. Conversations are preoccupied with on-going activities, and the acts, thoughts and features of conversants or third parties; hence their high frequency of (mostly activity and mental) verbs, adverbs, circumstance adverbials, simple and progressive verb

phrases, size and age adjectives. Colloquies involve personal viewpoints; thus they display a high rate of evaluative adjectives and stance adverbials. Conversations present information in more linear, simple structures; therefore, they display a low lexical density, a low rate of passives, nonfinite clauses and simple sentences, a high frequency of linking adverbials; and adjectives are used both attributively and predicatively. Face-to-face conversations are interpersonal and situational; hence their low noun modification, high rate of pronouns and present tense verbs. Finally, conversations take place in real time; thus they might contain much non-clausal material and non-standard linguistic forms.

For a textual background, two most wide-spread models of the discourse structure of printed news were discussed. It was found that van Dijk's (1985a, 1988a) 'sociocognitive' and Bell's (1991, 1998) 'structural discourse analytical' models are in great consonance. Both models feature Abstract/Summary (Headline plus Lead), Main Events, Consequences, Context, History, Verbal Reactions, Expectations and Evaluations as textual elements of typical press news stories. The main difference between the two frameworks is that Bell has complemented van Dijk's model with the category Attribution, for both the news story as a whole and the individual pieces of information contained. Bell also places great emphasis on the discontinuous time-planes that make up a news story.

The data were gathered so that they provide the best chance of answering the research questions. The BBC corpus collected for the purposes of the present study consists of a 4-hour 30-minute sample of BBC World Service news from five-minute morning and two-minute afternoon programmes for three weeks running. The consecutive-day sampling was important for allowing some of the qualitative questions to be addressed: e.g. textual transformations along news items dealing with the same topic. The news was transcribed in terms of words, which resulted in a body of text made up of approximately 49,000 words (452 news items). The data are fairly representative of the BBC WS news output and, given the prestige and acknowledged exemplary character of the BBC, the findings are prone to be generalised to all Western-type public service radio news production.

The database was quantitatively tagged with the help of QTAG 3.01 part-of-speech tagger, and manually checked for errors and cases in need of clarification. Some shortcomings of the programme were also eliminated. The news was qualitatively

analysed for topics and textual units, and the subgenres were compared in terms of discourse structure.

The study yielded the following quantitative results:

Like press stories, radio news is also concerned with the doings of prominent personalities who have to be easily identified by the listeners. This is reflected by a high ratio of determiners and nouns modified by adjectives, prepositional phrases, participles and relative clauses. The most frequent words in the corpus also reveal the interest of the news to give the circumstances of, and amounts involved in, current events (time and place expressions, quantifiers). The most frequent modals (expressing prediction) are related to the fact that news often deals with expectable events.

Radio news also strives to present as much information as possible in the shortest possible time. Its grammatical features, therefore, display means of integration, such as long sentences, long noun and prepositional phrases, multi-purpose genitives, many gerunds and a high number of nonfinite clauses.

Thus, radio news was found to present some features characteristic of written genres, namely, the length of noun and prepositional phrases, and aspectual marking (a preference for the perfective aspect over the progressive, when not simple). In these features, radio news is much similar to the printed media. In other features, however, radio news differs considerably from literate language forms:

- In lexical density, radio news stands between written news texts and conversations.
- The average noun/verb ratio of radio news (2.3 nouns per verb) places it between news reportage (3 to 4 nouns per verb) and conversations (a 1 to 1 relationship).
- Clause grammar, that is, the rate of finite vs non-finite subclauses, coordination vs subordination and the retention vs omission of *that*, shows a more marked similarity to spoken than written language.
- The extremely common usage of semi-modal auxiliaries makes radio news especially ‘conversation-like’.

This is an impressive list of grammatical features in radio news which approximate it more to spoken than written language. The findings considerably alter

the picture drawn by previous works, which placed this radio genre totally at the writing end of the oral vs literate continuum.

The study identified a distinctive feature of radio news, namely, a remarkably high rate of the perfective aspect in verb phrases. This must be in a direct relationship with the aim of radio news to be of an accentuated present relevance. The strong preference for the perfective proves to be a unique characteristic of radio news.

The subgenres of radio news were found to differ from each other in some ways: reports stand closer to conversations than to the press in their more frequent usage of pronouns, conjunctions and adverbs. Short news displays fewer, but not shorter, sentences than long (reported or unreported) news. However, there is no difference between the number of news items per edition: the average number is seven both in the two-minute and the five-minute programmes.

Therefore, radio news, indeed, shows appreciable similarities to spoken language both overall and, to a different extent, in its various subgenres.

The study reached the following qualitative findings:

Twelve topic areas were identified in the BBC World Service news: Legal issues, Conflicts, Economic issues, Human interest stories, Nature-related, Military, and Immigration issues, Protests, Media-related issues, Elections, Anniversaries and Scientific issues, in decreasing order of coverage amount. In a model offered for the topic structure of the BBC corpus, each topic area branches into topic groups (which may divide into topic subgroups), these are made up of topics, which, for their part, may consist of subtopics. However, not all topics arch over several subtopics, not all topics have topic groups governing them, and topic subgroups are rare. This complex system is a true reflection of the disparate threads of media topics.

As for the relationship between topic and positioning – starting off from the axiom that the amount of coverage of a topic area suggests its perceived importance – it was found that this sense of significance is reflected in positioning as well: more important/covered topics tend to occur in higher positions, and less important/covered ones in lower places. (However, Position 1 deviates from this general tendency in that the inner division of topic areas here does not follow the above-mentioned importance order, but a recency principle.)

If a topic is treated along several items on the same day in the BBC news, half of the repeated items occupy the same position as earlier. This is normal if we take into consideration that the perceived importance of a topic is not likely to change in the

course of a single day. One-third of the repeated news items slip one slot down. This is again understandable: the editor might feel that an item that has lost its 'freshness' by having been heard before is relatively less important. One-fifth of the repeated news items move upwards; this may be due to a previous item having been left out in the next edition, or to a possible negotiation on the part of the editor not being able to decide which topic is more important, so they are presented in order *a-b* in an earlier edition, and *b-a* in a later bulletin.

With regard to how the two models for the textual make-up of press stories can be applied to radio news, it was found that both models are, to a certain degree, apt for being used in radio news analyses. However, Bell's category of Attributions – absent from Van Dijk's model – was found indispensable. Some new subcategories, such as Verbal Main Events, Future (Main) Events and Reasons, were introduced to make up a modified schema of radio news. As for textual structure, it was found that Leads and Main Events merge in radio news, and Leads/Main Events as well as Attributions are compulsory elements. These are most typically complemented by near-past previous events and a description of the present situation, possibly accompanied by an evaluation. Elements which deal with either the remote past or an uncertain future are the least common.

Headlines were found to be peripheral and optional elements of radio news. Headlining takes place according to certain rules, being attached (i) to the first three elements of an unreported news programme; (ii) to the reported elements of a mixed news programme; and (iii) in the same order as the news items they belong to. Deviations from these rules are conditioned by internal factors such as the topic of the item; or external factors, e.g. time of edition, or proximity of other programmes. In a structural sense, headlines are summaries of news items, and the information they transmit more or less overlaps with that contained by the lead. The degree of this overlap can be placed along a continuum from loss of information through retaining knowledge to extra and different information.

The textual transformations between different subgenres of news when (i) a correspondent's report is later incorporated in a news item, (ii) a long news item is partially repeated, and (iii) a long news item becomes a short one, indicate two opposite but often simultaneous tendencies at work. One such direction points towards partial loss of information, the other to packing extra knowledge in the repeated item, with neutral cases in which the information is exactly retained, or mixed cases of the two

tendencies. The core of a radio news item seems to be the lead, and all other elements are subject to discretionary alterations due to several constraints (time, availability of information etc.).

Finally, a supplementary framework was offered for the analysis of radio news. The concept of ‘pantext’ and a corresponding idea of ‘pantopic’ were introduced, to capture the unity of a series of news items on the same topic in the output of a radio station. It was demonstrated through an application of textuality principles that pantexts show many features which may classify them as texts in their own right. The concept of ‘pantexts’ may be generalised to a number of linguistic phenomena other than radio news.

Most of the hypotheses were verified, except for the supposition that short news or reports display shorter sentences than long news; that deletion is the most common transformation type between subgenres of radio news; or that ‘news summaries’ are indeed summaries.

The combination of quantitative and qualitative research methods proved to be useful: the results obtained in one way were confirmed or complemented by the other. For example, the quantitative method of tagging and frequency counts revealed the distinctive grammatical feature of radio news: the usage of the present perfect. The qualitative method of news item analysis, for its part, showed how exactly the present perfect is used in radio news: in the textual unit of Lead, while in later parts of radio news it turns into the simple past. Or another example: the most important topic areas in the BBC corpus were established through the qualitative method of topic-assignment plus the quantitative method of rate-of-coverage counts. The most frequent nouns, verbs, adjectives etc. identified via the tagging process confirmed the results of the textual analysis: e.g. the fact that the most often used verb was *say* is in concordance with Attributions being the most pervasive textual unit of radio news.

It has been indicated in the introduction to this study that a single work cannot, by nature, comprise all grammatical and textual aspects of radio news. Therefore, I suggest some future directions of research:

Surely, there are other distinctive grammatical features of radio news besides the use of the present perfect. The multi-purpose way in which genitives are used is one candidate. Anyway, this aspect requires further research.

The data of this study were drawn from a prestigious public service radio station, and it has been suggested that the results are generalisable for other Western-type public

radio news output. The question arises, however, whether the same is true for news programmes of commercial radio stations. This also asks for scientific analysis.

As pointed out in Chapter 9, the concept of ‘pantext’ lacks experimental support. The tentative list of possible pantexts offered has to be tested on real language samples from the different text types. It is also necessary to examine whether audiences regard these pantexts as single pieces of texts. If not, the concept has to be discarded, however elegant it may seem.

Intonation is a very important aspect of radio news. Cotter (2001:424) refers to her earlier research in which the routinised intonation of radio news is viewed as a way to cue listener expectations in a particular discourse environment. Csölle (1999) points out that intonation distinguishes between two types of textual givenness: relation (that is, grammatically and semantically dependent categories) and reiteration, which, in her view, should be separated from the notion of cohesion and treated as a distinct text characteristic. An analysis of intonation in radio news should support this view. Moreover, an intonation-based givenness analysis could also prove or query the suggested existence of ‘pantexts’.

Ferguson (1983:167) draws attention to routinised ways of reporting the ‘count’ in sports announcer talk as “an example of the general human tendency to routinise recurrent messages in recurrent communicative settings, eliminating both redundancy and stylistic variety”. He proposes that the use of prefabricated modules inserted as needed at appropriate points in communication are also markers of the register itself. Bell (personal communication, 2000) also notices the use of formulaic language in radio news which, because of its brevity and the tightness of its deadlines (often hourly), especially lends itself to doing routine stories in similar phrases – e.g. disasters, elections, etc. This aspect of prefabricated stretches of discourse in radio news certainly deserves more attention.

Cross-linguistic studies are also worth considering both on the concept of ‘pantext’ and on the use of formulaic language. A study in and on Hungarian could reveal whether ‘pantext’ is generalisable to other languages as well, and if Hungarian radio language also displays prefabricated units in news on the same topic area. Moreover, the textual structure models suggested for English news texts – printed or broadcast – and the spoken vs written character of radio news should also be tested on the Hungarian news media.

Appendices

Appendix 1/a

The Birmingham-Lancaster Tagset

This tagset has been jointly agreed upon by Lancaster and Birmingham for a joint project on tagger evaluation. It is a variant of the quite common Brown/Penn-style tagsets.

BE	<i>be</i>	PN	pronoun, indefinite (<i>anyone, nothing</i>)
BEDR	<i>were</i>	POS	possessive particle (' , 's)
BEDZ	<i>was</i>	PP	pronoun, personal (<i>I, he</i>)
BEG	<i>being</i>	PP\$	pronoun, possessive (<i>my, his</i>)
BEM	<i>am</i>	PPX	pronoun, reflexive (<i>myself, himself</i>)
BEN	<i>been</i>	RB	adverb, general (<i>chronically, deep</i>)
BER	<i>are</i>	RBR	adverb, comparative (<i>easier, sooner</i>)
BEZ	<i>is</i>	RBS	adverb, superlative (<i>easiest, soonest</i>)
CC	conjunction, coordinating (<i>and</i>)	RP	adverbial particle (<i>back, up</i>)
CD	number, cardinal (<i>four</i>)	SYM	symbol or formula (<i>US\$500, R300</i>)
CS	conjunction, subordinating (<i>until</i>)	TO	infinitive marker (<i>to</i>)
DO	<i>do</i>	UH	interjection (<i>aah, oh, yes, no</i>)
DOD	<i>did</i>	VB	verb, base (<i>believe</i>)
DOG	<i>doing</i>	VBD	verb, past tense (<i>believed</i>)
DON	<i>done</i>	VBG	verb, -ing (<i>believing</i>)
DOZ	<i>does</i>	VBN	verb, past participle (<i>believed</i>)
DT	determiner, general (<i>a, the, this, that</i>)	VBZ	verb, -s (<i>believes</i>)
EX	existential <i>there</i>	WDT	det, wh- (<i>what, which, whatever, whichever</i>)
FW	foreign word (<i>ante, de</i>)	WP	pronoun, wh- (<i>who, that</i>)
HV	<i>have</i>	WPS	pronoun, possessive wh- (<i>whose</i>)
HVD	<i>had</i> (past tense)	WRB	adv, wh- (<i>how, when, where, why</i>)
HVG	<i>having</i>	XNOT	negative marker (<i>not, n't</i>)
HVN	<i>had</i> (past participle)	!	!
HVZ	<i>has</i>	"	quotation mark
IN	preposition (<i>on, of</i>)	'	apostrophe
JJ	adjective, general (<i>near</i>)	((
JJR	adjective, comparative (<i>nearer</i>)))
JJS	adjective, superlative (<i>nearest</i>)	,	,
MD	modal auxiliary (<i>might, will</i>)	-	-
NN	noun, common singular (<i>action</i>)	.	.
NNS	noun, common plural (<i>actions</i>)
NP	noun, proper singular (<i>Thailand, Thatcher</i>)	:	:
NPS	noun, proper plural (<i>Americas, Atwells</i>)	;	;
OD	number, ordinal (<i>fourth</i>)	?	?
PDT	determiner, pre- (<i>all, both, half</i>)	???	unclassified

Appendix 1/b

The modified Birmingham-Lancaster Tagset used in the present study

Tag	Description
BE	<i>be</i> - passive auxiliary
BEDR	<i>were</i> - passive auxiliary
BEDR G	<i>were</i> - progressive auxiliary
BEDR V	<i>were</i> - main verb
BEDZ	<i>was</i> - passive auxiliary
BEDZ G	<i>was</i> - progressive auxiliary
BEDZ V	<i>was</i> - main verb
BEG	<i>being</i> - passive auxiliary
BEGG G	<i>be</i> - progressive auxiliary
BEGV	<i>being</i> - main verb
BEM	<i>am</i> - passive auxiliary
BEMG	<i>am</i> - progressive auxiliary
BEMV	<i>am</i> - main verb
BEN	<i>been</i> - passive auxiliary
BENG	<i>been</i> - progressive auxiliary
BENV	<i>been</i> - main verb
BER	<i>are</i> - passive auxiliary
BERG	<i>are</i> - progressive auxiliary
BERV	<i>are</i> - main verb
BEV	<i>be</i> - main verb
BEZ	<i>is</i> - passive auxiliary
BEZG	<i>is</i> - progressive auxiliary
BEZV	<i>is</i> - main verb
CC	conjunction, coordinating (<i>and</i>), clause-connective
CCC	sentence initial coordinator (<i>And, Or, But</i>)
CCP	conjunction, coordinating (<i>and</i>), phrase-connective
CD	number, cardinal (<i>four</i>)
CS	conjunction, subordinating (<i>until</i>)

Tag	Description
CST	the complementiser <i>that</i>
DO	<i>do</i> - auxiliary
DOD	<i>did</i> - auxiliary
DODE	<i>did</i> - emphatic auxiliary
DODV	<i>did</i> - main verb
DOE	<i>do</i> - emphatic auxiliary
DOG	<i>doing</i> - auxiliary
DOGV	<i>doing</i> - main verb
DON	<i>done</i> - auxiliary
DONV	<i>done</i> - main verb
DOV	<i>do</i> - main verb
DOZ	<i>does</i> - auxiliary
DOZE	<i>does</i> - emphatic auxiliary
DOZV	<i>does</i> - main verb
DT	determiner, general (<i>a, the, this, that</i>)
EX	existential <i>there</i>
FW	foreign word (<i>ante, de</i>)
HV	<i>have</i> - perfective auxiliary
HVD	<i>had</i> (past tense) - perfective auxiliary
HVDV	<i>had</i> (past tense) - main verb
HVG	<i>having</i> - perfective auxiliary
HVGV	<i>having</i> - main verb
HVN	<i>had</i> (past participle) - perfective auxiliary
HVNV	<i>had</i> (past participle) - main verb
HVV	<i>have</i> - main verb
HVZ	<i>has</i> - perfective auxiliary
HVZV	<i>has</i> - main verb
IN	preposition (<i>on, of</i>)
JJ	adjective, general (<i>near</i>)
JJP	adjective in predicative usage
JJR	adjective, comparative (<i>nearer</i>)
JJRP	comparative adjective in

Tag	Description
	predicative usage
JJS	adjective, superlative (<i>nearest</i>)
JJSP	superlative adjective in predicative usage
MD	modal, present (<i>may, will</i>)
MDD	modal, past (<i>could</i>)
MDS	semi-modal, present (<i>is due to</i>)
MDS D	semi-modal, past (<i>was due to</i>)
MDS N	semi-modal, nonfinite (<i>be due to</i>)
NN	noun, common singular (<i>action</i>)
NNG	noun as genitive (<i>the boy's</i>)
NNGC	noun as classifying genitive (<i>women's</i> [fashion])
NNGT	noun as genitive referring to time (<i>yesterday's</i>)
NNS	noun, common plural (<i>actions</i>)
NP	noun, proper singular (<i>Thailand</i>)
NPG	proper noun as genitive (<i>Afghanistan's</i>)
NPGC	proper noun as classifying genitive (<i>Alzheimer's</i>)
NPGT	proper noun as genitive referring to time (<i>Monday's</i>)
NPS	noun, proper plural (<i>Americas</i>)
OD	number, ordinal (<i>fourth</i>)
PDT	determiner, pre- (<i>all, both, half</i>)
PF	proform (<i>so</i>)
PFN	proform, nominal (<i>one, these</i>)
PN	pronoun, indefinite (<i>anyone, nothing</i>)
POS	possessive particle (<i>' , 's</i>)
PP	pronoun, personal (<i>I, he</i>)
PP\$	pronoun, possessive (<i>mine, his</i>)
PP1	first person pronoun (<i>I</i>)
PP2	second person pronoun (<i>you</i>)
PPX	pronoun, reflexive (<i>myself, himself</i>)

Tag	Description
RB	adverb, general (<i>chronically, deep</i>)
RBR	adverb, comparative (<i>easier, sooner</i>)
RBS	adverb, superlative (<i>easiest, soonest</i>)
RP	adverbial particle (<i>back, up</i>)
SYM	symbol or formula (<i>US\$500, R300</i>)
TO	infinitive marker (<i>to</i>)
UH	interjection (<i>aah, oh, yes, no</i>)
VB	verb, base, nonfinite ([<i>to, may</i>] <i>believe</i>)
VBD	verb, past tense (<i>believed</i>)
VB F	verb, base, finite ([<i>I, you...</i>] <i>believe</i>)
VBG	verb, present participle (<i>believing</i>)
VBGG	verb, gerund (<i>believing</i>)
VB N	verb, past participle (<i>believed</i>)
VBZ	verb, -s (<i>believes</i>)
WDT	det, <i>wh-</i> (<i>what, whichever</i>)
WP	pronoun, <i>wh-</i> (<i>who, that</i>)
WP\$	pronoun, possessive <i>wh-</i> (<i>whose</i>)
WRB	adv, <i>wh-</i> (<i>how, when, where, why</i>)
XNOT	negative marker (<i>not, n't</i>)
'	apostrophe
-	-
!	!
"	quotation mark
((
))
,	,
.	.
...	...
:	:
;	;
?	?
???	unclassified

Appendix 2 to 13

Topic structures in the BBC corpus

Appendix 2/a

Topic structure of the *Legal actions against terrorist suspects* topic group of the *Legal issues* topic area (topics in decreasing order of coverage amount; subtopics in order of broadcast). The numbers refer to the news items dealing with the particular (sub)topic.

Topic area	Topic group	Topics	Subtopics
Legal issues		Fundamentalist prisoners in US military base in Cuba	Twenty prisoners flown (2) First night in camp (2) British concern over conditions (1) Another group travels to camp (2) Red Cross to inspect prison conditions (1) Red Cross talks to prisoners (3) US wishes to move all remaining prisoners to Cuba (2) US releases pictures of camp conditions (1) Two prisoners sedated on way to Cuba (1) US defence secretary to visit camp (1)
		Southern Philippine Muslim faces trial	Leader awaits charges of rebellion (2) Clashes between separatist's supporters and police (3)
	Legal actions against terrorist suspects	Explosives-carrying British plane-passenger faces trial	Briton faces more charges (2) Briton faces preliminary hearing (1)
		American terrorist conspirator to be charged (3)	
		Rights to televise September 11 suspect's trial denied (2)	
		British police to question nine (1)	
		Lockerbie bomber in court (3)	
		Russian supporters of outlawed Japanese sect convicted (2)	
		Former Taleban ministers in Afghanistan surrender (2)	
		Offices of Islamic groups sealed in Pakistan (1)	
Israel sceptical about Palestinian arrest of terrorist leader (3)			

Appendix 2/b

Topic structure of the *Crime* topic group of the *Legal issues* topic area (topic subgroups in decreasing order of coverage amount; topics in order of broadcast). The numbers refer to the news items dealing with the particular topic.

Topic area	Topic group	Topic subgroup	Topics
Legal issues	Crime	Murder	Ugandan opposition activist shot dead (1) Brazilian mayor killed (2) Colombian murderer nun sentenced to 14 years in prison (2) Northern Indonesian separatist leader killed (1) Former Lebanese Christian militia leader assassinated (1) Muslim family killed in Indian-administered Kashmir (2) 16 people killed by rebels in North-Eastern India (2)
		Corruption	South Korean corruption scandals (2) Indonesian parliamentary speaker's corruption involvement (1) Corruption allegations against French president (1) British experts investigate Kenyan corruption cases (1) UN warns Italian government not to interfere in bribery accusations against Prime Minister (3) 17 former government officials arrested in Puerto Rico (2)
			Wanted Peruvian ex-president reappears (1) Peruvian spy chief linked to death squads (2)

Appendix 2/c

Topic structure of the *New legislation in Zimbabwe* topic of the *Legal issues* topic area (subtopics in order of broadcast). The numbers refer to the news items dealing with the particular subtopic.

Topic area	Topic	Subtopic
Legal issues	New legislation in Zimbabwe	New media bill (2) New security bills (3) Rejected election law reconsidered (1) Opposition reactions to new legislation (2) EU sanctions as a result of new laws (2) International election monitors in Zimbabwe (2) President denounces British policy at Southern African summit (1) Southern African summit rules out sanctions (1) Southern African summit calls for free and fair elections (2) Opposition leader's speech at Southern African summit (2) Communiqué at end of Southern African summit (2) New media bill again (2) New media bill re-discussed (3)

Appendix 2/d

Topic structure of the *Court cases* topic group of the *Legal issues* topic area (topics in decreasing order of coverage amount; subtopics in order of broadcast). The numbers refer to the news items dealing with the particular (sub)topic.

Topic area	Topic group	Topic subgroup	Topics	Subtopics
Legal issues	Court cases	Microsoft in court	100 private cases still unsettled (2) AOL sues Microsoft for unfair practices (2)	
			The Milosevic trial (3)	
			Greenpeace activists in court	Serious charges dropped, lesser charges maintained (1) Tresspassers sentenced to two years' probation (1)
			Iranian MP jailed for insult	Speaker of Parliament protests by walk-out (1) Supreme leader pardons MP (1)
			Adulterous Nigerian woman in court (2) French law for the right not to be born (1) German interior minister displeased with postponement of ban on far-right party (1) Indonesian trials of East-Timor killings (1) Portuguese trial against women having abortion (1)	

Appendix 3/a

Topic structure of the *Israeli-Palestinian conflict* topic group of the *Conflicts* topic area (topics in decreasing order of coverage amount; subtopics in order of broadcast). The numbers refer to the news items dealing with the particular (sub)topic.

Topic area	Topic group	Topics	Subtopics
Conflicts	Israeli-Palestinian conflict	Arms shipment caught	Shipment seized in Israel (2) Suspects to be arrested (2) CIA help in locating vessel (1) Bush connects shipment to Palestinian authority (2)
		Israelis in Palestinian-controlled Tulkarm	Incursion (3) Pull-out (3)
		Israeli action in Arafat's headquarter town	Action started (1) Media-building blown up (3)
		Israelis demolish Palestinian houses	Demolishing starts (1) Demolishing stops (1)
		Palestinian-shot Israeli women die of injuries (2) 2 Palestinians killed in Israel (2) Israelis ruin Palestinian airport (2) Israelis kill Hamas member and Palestinian gunmen (2) Israeli missiles fired on Palestinian police dock (2) Israel attacks Palestinian authority building (2) 4 Israelis killed in Palestinian attack (1) Israel and US press Palestinian leader to curb militants (1) Arafat accuses Israelis of exaggeration (1) Palestinians release Islamic militants after protest (1) Israeli-killed Islamic militants buried (1) Israel renews blockade of Palestinian cities (1) Suicide bomb-attack in Central Jerusalem (1)	

Appendix 3/b

Topic structure of the *Conflict resolution attempts* topic group of the *Conflicts* topic area (topics in decreasing order of coverage amount; subtopics in order of broadcast). The numbers refer to the news items dealing with the particular (sub)topic.

Topic area	Topic group	Topic subgroup	Topics	Subtopics
Conflicts	Conflict resolution attempts		US State Secretary's mission tour	Call for peace (3) Attempts to calm down Israeli–Palestinian conflict (2) Talks in India resumed (1) Talks in India finished (1) Talks lead in Nepal (1)
			Blair's mission tour	Talks in Pakistan (2) Talks in Afghanistan (3)
			Pakistani president's speech	Speech prepared while Pakistani police arrest militants (1) Speech denounces terrorism (2) India welcomes Pakistani measures against terrorism (1)
		Religious appeals for reconciliation	International meeting of Christian and Muslim scholars (1) World religions pray for peace in Assisi (1)	
			Greek–Turkish reconciliation talks in Cyprus (2) Pakistani president says Bin Laden is probably dead (2) Archbishop of Wales criticises war in Afghanistan (1) Son of exiled Afghan ex-king to return (1) Ceasefire agreement between Sudanese government and rebels (1)	

Appendix 3/c

Topic structure of the remaining topic groups of the *Conflicts* topic area (topics in decreasing order of coverage amount; subtopics in order of broadcast). The numbers refer to the news items dealing with the particular (sub)topic.

Topic area	Topic group	Topics	Subtopics
Conflicts		Colombian civil unrest	New deadline for rebels (2) UN envoy talks with rebels (1) Troops surround rebels' safe haven (1) Rebels' proposals rejected by president (3) Safe haven residents fear atrocities (2) Army prepares to move into safe haven (1) Last-minute peace talks (3) New deadline for peace-agreement (1) Rebels attack after peace-agreement (1) Foreign diplomats met rebels to salvage peace (2) Last-ditch peace efforts (2) Ceasefire agreement between government and rebels (3)
	Afghan– Pakistani conflict	Afghan military in capital	Ordered to pull out (1) Unable to pull out on time (1)
		Military measures in Pakistan (1)	
	Iraqi– Iranian conflict	Iraqi president denounces US (1)	
		Iraqi prisoners released from Iran (1)	
		Sierra Leonean civil war	Last rebel commanders give up guns (1) Hostilities officially declared over (2)
		Taiwan's independence from China	Separate Taiwanese passports issued (1) Pro-independence Taiwanese party cautious about invitation to China (2)
		Sectarian violence in Northern Ireland (3) Ethnic conflicts in Nigeria (2) Displaced families in rebel-held Sri Lanka (1)	

Appendix 4/a

Topic structure of the *Afghan financial crisis* and the *Enron scandal* topics of the *Economic issues* topic area (subtopics in order of broadcast). The numbers refer to the news items dealing with the particular subtopic.

Topic area	Topics	Subtopics
Economic issues	The Afghan financial crisis	Frozen Taleban funds needed (1) US politician urges Taleban fund release (1) UN backing assured (2) International sanctions lifted (2) US to release Afghan gold reserves (1) End of serious drought in Afghanistan (1) US assures Afghans of golden assets (1) US State Secretary promises support (2) Afghan leader to discuss Saudi support (1)
		International Donors' Conference for raising funds US State Secretary to attend (1) Conference in Tokyo opens (2) More than 2b dollars pledged (1) More than 4.5b dollars pledged (2) Interim leader pleased with outcome (2) Interim leader seeks further support in China (2) UN Secretary General talks with Afghan leader on use of money (2) UN Secretary General leads talks in Iran on Afghan situation (2) Afghan leader talks to Bush on use of money (1)
	The Enron scandal	Criminal investigation started (3) Relations to Bush administration (1) Legal actions against auditors (2) Former executive reveals destruction of documents (3) Chairman resigns (3) Police investigates death of ex-vice-chairman (1)

Appendix 4/b

Topic structure of the remaining topics of the *Economic issues* topic area (topics in decreasing order of coverage amount; subtopics in order of broadcast). The numbers refer to the news items dealing with the particular (sub)topic.

Topic area	Topics	Subtopics
Economic issues	The Argentine financial crisis	The peso devalued (3) President asks for low prices (2) Civil riots against economic measures (2) Foreign exchange restarts (1) Devalued peso first traded (2) President calls for national dialogue (3) IMF patient with overdue debt; Bush supports Argentina (3) Investigation into capital flight begins (1) Hostile reaction to dollar bank accounts turned into pesos (1) Police clashes with demonstrators against economic programme (2)
	Economic slowdown in Germany	Increase in unemployment (1) Lowest growth in economy since early 90's (1)
	Italian prime minister supports EU expansion (1) Debate on state of British railways (1) Poll on British view of developing world (1) Changes in Eastern opium production (1) Indonesian fuel price rises (1) Planes may fly closer together (1) Popular Chilean football club bankrupt (1)	

Appendix 5

Topic structure of the *Human interest stories* topic area (topics in decreasing order of coverage amount; subtopics in order of broadcast). The numbers refer to the news items dealing with the particular (sub)topic.

Topic area	Topic group	Topics	Subtopics
Human interest stories		President Bush has medical problems	Faints of pretzel in White House (1) Further checks after fainting spell (2)
		Saint Laurent retires (3) Prince Harry consumes drug and alcohol (3) Former US State Secretary dies (2) Muhammad Ali is 60 (2) Peggy Lee dies (2) British Chancellor's daughter dies (1) Pavarotti performs after mother dies (1) Mariah Carrey paid to end contract with EMI (1)	
	Significant events	Australian Open tennis championship winners announced	Capriati wins women's finals (2) Johansson wins men's finals (2)
		Golden Globe film festival winners announced (2) Detroit motor-show opens (2)	
		Bereaved Americans grieve with Afghans in mourning (2) US town borrows snow from another (1) New Zealand schoolboy excels in golf (1) Prostitution area in Venice assigned (1) Winning lottery ticket searched in Australian town rubbish tip (1) Thailand students to undergo drug tests (1) Malaysian university population to declare allegiance to authorities (1)	

Appendix 6

Topic structure of the *Military issues* topic area (topics in decreasing order of coverage amount; subtopics in order of broadcast). The numbers refer to the news items dealing with the particular (sub)topic.

Topic area	Topic group	Topics	Subtopics
Military issues	Asian military issues	Indian interior minister seeks US support (2) India tests nuclear-capable missile (2) International agency tests nuclear materials in Iraq (2) Second US–British attack on Iraqi anti-aircraft artillery site (2) Indian army ready for possible war with Pakistan (1)	
	Attacks on diplomatic outposts	Action against USIS building in Indian capital	Building attacked (3) Suspects detained (2)
		US embassy in Singapore prevented from being blown up (1) British High Commission attacked in Pakistani capital (1)	
	American military issues	Bush wishes US to increase defence budget	More spending planned on war against terrorism (1) More money to protect borders (2)
		US military plane crashes in Pakistan (2) US military helicopter crashes in Afghanistan (2)	
	Russian military issues	Russian military base in Cuba shut down (2) Russian Airforce and Pacific Fleet deprived of electricity (1) Russian reaction against US warhead storage (1) Russia to use Azari radar tracking station (1)	
African military issues	East-African initiatives against terrorism (1) Ugandan troops sent to Congo (1)		

Appendix 7

Topic structure of the *Nature-related issues* topic area (topics in decreasing order of coverage amount; subtopics in order of broadcast). The numbers refer to the news items dealing with the particular (sub)topic.

Topic area	Topic groups	Topics	Subtopics
Nature-related issues	Disasters	Volcanic eruptions in Congo	People flee to neighbouring Rwanda (2) Lava continues to flow (1) Displaced people sleep in the open (1) Rwandan appeal for international aid (1) Aid agencies to provide drinking water (2) Refugees reluctant to enter camps (1) Congolese return despite aid agency discouragement (1) More earth-tremors (1) Explosion at Goman petrol station (2) Relief supplies arrive (1)
		Bushfires in Australia	Evacuations continue (2) Fires finally under control (1) New outbreaks of fires (2) Further evacuations needed (1)
		Earth-approaching asteroid (2) Floods in New Zealand (1)	
	Environmental issues	Chinese efforts to reduce pollution (3) Chinese town blown up to make way for dam (3)	

Appendix 8

Topic structure of the *Immigration issues* topic area (topics in decreasing order of coverage amount; subtopics in order of broadcast). The numbers refer to the news items dealing with the particular (sub)topic.

Topic area	Topics	Subtopics
Immigration issues	Australian refugee camp strikes	Protests begin; government stays firm (1) Government continues to defy protests (1) Government threatens to remove children involved (1) Protests spread to a second centre (2) More and more on strike; suicide attempts (2) Government promises to resume processing asylum claims (1) Protesting children taken to hospital (1) Real number of strikers double the official figure (1)
	Chinese mainlanders rejected to stay in Hong Kong (1) Australian anti-immigrationist politician retires (1) Fatal skin disease spreads from Afghan refugees to Pakistani population (1) Protests against tougher Italian immigration laws (1) Romanian refugees help asylum seekers enter Britain (1)	

Appendix 9

Topic structure of the *Protests* topic area (topics in decreasing order of coverage amount; subtopics in order of broadcast). The numbers refer to the news items dealing with the particular (sub)topic.

Topic area	Topics	Subtopics
Protests	General strike in Nigeria	Protests at fuel price rises begins (3) Protests continued despite arrest of union leaders (3) Union leaders re-arrested in continued fuel price protests (1)
	Hindus march to get temple built in Indian town (2) Chinese farmers attack tax collectors (1) Jordanian policeman dies in clash with rioters against boy's death in custody (1)	

Appendix 10

Topic structure of the *Media-related issues* topic area (topics in decreasing order of coverage amount; subtopics in order of broadcast). The numbers refer to the news items dealing with the particular (sub)topic.

Topic area	Topics	Subtopics
Media-related issues	Last independent Russian TV station	Closed down on financial grounds (1) Minister of information to decide who should run it (1) Channel taken off the air (2)
	English news staff protests against dismissals in Malaysia (2) Clandestine interview with paedophile murderer on Belgian TV (2) Bangladeshi journalist released on bail (1)	

Appendix 11

Topic structure of the *Elections* topic area (in decreasing order of coverage amount). The numbers refer to the news items dealing with the particular topic.

Topic area	Topics
Elections	Election preparations in Sierra Leone (3) National elections in Kosovo (2) European Parliament elects new president (2)

Appendix 12

Topic structure of the *Anniversaries* topic area (in decreasing order of coverage amount). The numbers refer to the news items dealing with the particular topic.

Topic area	Topic group	Topics
Anniversaries		1 st of devastating Indian earthquake (2) 10 th of Bosnian Serb Republic (1) 1 st of deposition of Philippine ex-president (1)
	Second World War anniversaries	60 th of Nazi Wannsee Conference (1) 57 th of Auschwitz camp liberation (1)

Appendix 13

Topic structure of the *Scientific achievements* topic area (in decreasing order of coverage amount). The numbers refer to the news items dealing with the particular topic.

Topic area	Topic group	Topics
Scientific achievements	Medical issues	AIDS surpasses plague as worst pandemic (2) Early diagnosis of Alzheimer's (1) New techniques for sterilising men (1)
	Astronomical issues	New astronomical telescope opened in Chile (2) New life forms increases likelihood of life on Mars (1)

Appendix 14

The pantext on the pantopic 'Fundamentalist prisoners in US military base in Cuba' in the BBC corpus

The first group of about twenty al-Qaeda and Taleban prisoners are on their way from Afghanistan to Cuba, where they will be detained at the American base at Guantanamo. The prisoners are being treated as potentially dangerous, overhounded and shackled, and will remain chained to their seats for the 20-hour journey. Human rights groups have expressed some concern at the treatment of the prisoners. The director of the New York based Human Rights Watch told the BBC he was particularly concerned at reports the prisoners have been drugged or sedated. More prisoner airlifts are expected over the next few days. (7/11)

As 20 Taleban and al-Qaeda prisoners spent their first night in the United States detention camp in Cuba, the American Defence Secretary has said he does not consider them prisoners of war. Mr Rumsfeld said the detainees were illegal combatants and, under that status, were not protected by the Geneva Conventions. However, he insisted the prisoners would be treated humanely. Cuba has said it will cooperate with the United States' plan to house the prisoners at Guantanamo Bay, and has offered medical assistance. The statement released to journalists said Cuba agreed with the US on the need to eradicate terrorism, but disagreed on the best way to do it. (8/12)

The British Foreign Secretary, Jack Straw, has expressed concern over the detention of at least 50 Taleban and al-Qaeda prisoners at a US military base in Cuba. Mr Straw said the civilised world had a responsibility to uphold their rights. The American Defence Secretary, Donald Rumsfeld, in an interview with the BBC, insisted that conditions at the base were humane, vastly better than those in which the men had been found. (8/16)

Another batch of Taleban and al-Qaeda prisoners has arrived at the American military base at Guantanamo Bay in Cuba. It is the third group to go there, bringing the total number of captives to 80. As for the previous prisoners, they were taken blindfolded and chained to outdoor cages where they would be held. The human rights organisation Amnesty International has added its voice to criticism of the conditions in which the prisoners are being held. It said the cages where they were locked up fell below minimum standards for human treatment. (6/17)

The Red Cross is expected to visit the American military base at Guantanamo Bay in Cuba today to inspect the conditions under which Taleban and al-Qaeda prisoners are being held. Human rights groups have criticised the size of their cells and the fact that they are open to the elements. Also at issue is the ill-defined status of the captives, who the US authorities say are not prisoners of war, only what they call illegal combatants, therefore not entitled to the protection of the Geneva Convention. The latest batch of prisoners arrived on Wednesday blindfolded and chained, raising the total number to 80. The American commander of the camp insisted that the prisoners are being treated humanely. (8/17)

A team from the International Red Cross will today begin inspecting the conditions under which Taleban and al-Qaeda prisoners are being held at the American base of Guantanamo Bay in Cuba. The team is expected to stay several days and will have individual access to all the prisoners. (7/18)

The United States military says it wants to move the remaining al-Qaeda and Taleban prisoners it is holding in Afghanistan to its military base at Guantanamo Bay in Cuba. A US military official, Brigadier-General Mike Lehnert says the present camp is being expanded to hold up to 600 detainees, who will be transferred as soon as possible. He said the team from the International Red Cross had been given full access to the prisoners at the base, and had interviewed each one privately. The Red Cross is now expected to have a permanent presence at the base, amid international criticism of the way the prisoners are being treated. (7/20)

The US Department of Defence has released photographs of Taleban and al-Qaeda prisoners being held at a US naval base in Guantanamo Bay in Cuba. The pictures, published in newspapers around the world, show the prisoners in conditions of sensory deprivation, forced to wear goggles, ear-muffs and heavy gloves. The prisoners' conditions have been condemned by civil rights groups. (12:30/20)

Officials at the United States naval base at Guantanamo Bay in Cuba say two suspected Taleban and al-Qaeda prisoners transferred there from Afghanistan on Sunday had to be sedated during the flight because of their disruptive behaviour. A military spokesman says the men were drugged after they began yelling and thrashing around in their aircraft seats. They were among a group of 34 prisoners, some of whom are reported to have sustained gunshot wounds in Afghanistan. The American say they are planning to build a field hospital at the base, solely for treating suspected Taleban and al-Qaeda prisoners. (7/21)

United States Defence Secretary Donald Rumsfeld is visiting later today the detention camp of suspected al-Qaeda and Taleban prisoners at the American base of Guantanamo Bay in Cuba. The conditions under which the prisoners are being held have come in for international criticism. (12:30/27)

References

- Al-Shabab, O.S., 1986. *Organizational and Textual Structuring of Radio News Discourse in English and Arabic*. Unpublished PhD dissertation. Birmingham: The University of Aston in Birmingham. [Microfilm format]
- Babbie, E., 1989/1973. *The Practice of Social Research*. Fifth Edition. Belmont CA: Wadsworth Publishing Company
- Baráth E., 1997. 'Are you an Alien? Ideologies at Work in the Hungarian Printed Media.' In Cunningham, J. and J. Horváth (eds.), 1997. *Framing the Issues. British Studies – Media Studies*. Pécs: Janus Pannonius University, 125-148.
- BBC World Service. [Online], URL <http://www.bbc.co.uk.html>
- de Beaugrande, R., 1985. 'Text Linguistics in Discourse Studies.' In T.A. van Dijk (ed.), *Handbook of Discourse Analysis. Vol. 1. Disciplines of Discourse*, London: Academic Press, 41-70.
- de Beaugrande, R. and W.U. Dressler, 1981. *Introduction to Text Linguistics*, London and New York: Longman
- Bell, A., 1984. 'Language style as audience design.' *Language in Society* 13/2: 145-204.
1991. *The Language of News Media*. Oxford UK and Cambridge MA: Blackwell
1994. 'Climate of opinion: Public and media discourse on the global environment.' *Discourse and Society* 5/1: 33-64.
1995. 'Language and the media.' *Annual Review of Applied Linguistics* 15: 23-41.
1998. 'The discourse structure of news stories.' In A. Bell and P. Garrett (eds.), *Approaches to Media Discourse*, London: Blackwell Publishers, 64-105.
- Bell, A. and P. Garrett (eds.), 1998. *Approaches to Media Discourse*, London: Blackwell Publishers
- Bentele, G., 1985. 'Audio-visual analyses and a grammar of presentation forms in news programmes.' In T. A. van Dijk (ed.), *Discourse and Communication. New Approaches to the Analysis of Mass Media Discourse and Communication*. Berlin and New York: Walter de Gruyter, 159-184.
- Biber, D., 1985. 'Investigating macroscopic textual variation through multi-feature/multi-dimensional analyses.' *Linguistics* 23:337-60.

1986. 'Spoken and written textual dimensions in English: resolving the contradictory findings.' *Language* 62:384-414.

1988. *Variation across speech and writing*. Cambridge: Cambridge University Press

Biber, D, Johansson, S, Leech, G, Conrad, S, and Finegan, E., 1999. *Longman Grammar of Spoken and Written English*. London: Longman

Brazil, D., 1985. 'Phonology: intonation in discourse.' In T. A. Van Dijk (ed.), *Handbook of Discourse Analysis. Vol. 2. Dimensions of Discourse*, London: Academic Press, 57-75.

Brookes, H. J., 1996. "'Suit, tie, and a touch of juju" – the ideological construction of Africa: A critical discourse analysis of news on Africa in the British press.' *Discourse and Society* 7/4: 461-494

Cameron, D., 2001. *Working with Spoken Discourse*. London: SAGE Publications

Chafe, W. L., 1982. 'Integration and involvement in speaking, writing, and oral literature.' In D. Tannen (ed.), *Spoken and Written Language: Exploring Orality and Literacy*, Norwood NJ: Ablex, 35-53.

Connell, I., 1980. 'Television news and the Social Contract.' In S. Hall et al. (eds.), *Culture, Media, Language*. London: Routledge and Birmingham: CCCS, 139-156.

Connell, I. and A. Mills, 1985. 'Text, discourse and mass communication.' In T. A. van Dijk (ed.), *Discourse and Communication. New Approaches to the Analysis of Mass Media Discourse and Communication*. Berlin and New York: Walter de Gruyter, 26-43.

Cotter, C., 2001. 'Discourse and Media.' In D. Schiffrin, D. Tannen and H. E. Hamilton (eds.), *The Handbook of Discourse Analysis*. Oxford, UK and Cambridge, MA: Blackwell, 416-436.

Crisell, A., 1994. *Understanding Radio*. Second Edition. London and New York: Routledge

Cruttenden, A., 1981. 'Falls and rises: meanings and universals.' *Journal of Linguistics*. 17/1:77-91.

Csölle, A., 1999. *Discourse cohesion in spoken English narratives*. Unpublished PhD dissertation. Budapest: Eötvös Loránd University

Cunningham, J. and J. Horváth (eds.), 1997. *Framing the Issues. British Studies – Media Studies*. Pécs: Janus Pannonius University

Davis, H. 1985. 'Discourse and media influence'. In T. A. van Dijk (ed.), *Discourse and Communication. New Approaches to the Analysis of Mass Media Discourse and Communication*. Berlin and New York: Walter de Gruyter, 44-59.

Davis, H. and P. Walton, 1983a. 'Editors' Introduction.' In H. Davis and P. Walton (eds.), *Language, Image, Media*, Oxford: Basil Blackwell, 1-7.

Davis, H. and P. Walton, 1983b. 'Death of a premier: consensus and closure in international news.' In H. Davis and P. Walton (eds.), *Language, Image, Media*, Oxford: Basil Blackwell, 8-49.

Davis, H. and P. Walton (eds.), 1983. *Language, Image, Media*, Oxford: Basil Blackwell

DeFelice, W., 1996. 'Review of Allan Bell: *The Language of News Media*. Blackwell, 1991.' *Applied Linguistics* 17/1: 120-122.

van Dijk, T.A., 1977. *Text and Context. Explorations in the Semantics and Pragmatics of Discourse*. London and New York: Longman

1985a. 'Structures of news in the press.' In T. A. van Dijk (ed.), *Discourse and Communication. New Approaches to the Analysis of Mass Media Discourse and Communication*. Berlin and New York: Walter de Gruyter, 69-93.

1985b. 'Preface.' In T. A. van Dijk (ed.), *Discourse and Communication. New Approaches to the Analysis of Mass Media Discourse and Communication*. Berlin and New York: Walter de Gruyter, v-vi.

1985c. 'Introduction: Discourse analysis in (mass) communication research.' In T. A. van Dijk (ed.), *Discourse and Communication. New Approaches to the Analysis of Mass Media Discourse and Communication*. Berlin and New York: Walter de Gruyter, 1-9.

1988a. *News as Discourse*. Hillsdale, NJ: Lawrence Erlbaum Associates

1988b. *News Analysis. Case Studies of International and National News in the Press*. Hillsdale, NJ: Lawrence Erlbaum Associates

1991. *Racism and the Press*. London and New York: Routledge

1993a. 'Editor's foreword to Critical Discourse Analysis.' *Discourse and Society* 4/2: 131-132.

1993b. 'Principles of critical discourse analysis.' *Discourse and Society* 4/2: 249-283.

1994a. 'Editorial: Discourse analysis as social analysis.' *Discourse and Society* 5/2:163-164.

1994b. 'Editorial: Critical discourse analysis.' *Discourse and Society* 5/4: 435-436.

1995. 'Discourse semantics and ideology.' *Discourse and Society* 6/2: 243-289.

1998. 'Opinions and ideologies in the press.' In A. Bell and P. Garrett (eds.), *Approaches to Media Discourse*, London: Blackwell Publishers, 21-63.
2001. 'Critical discourse analysis.' In D. Schiffrin, D. Tannen and H. E. Hamilton (eds.), *The Handbook of Discourse Analysis*. Oxford, UK and Cambridge, MA: Blackwell, 352-371.
- van Dijk, T. A. (ed.), 1985. *Discourse and Communication. New Approaches to the Analysis of Mass Media Discourse and Communication*. Berlin and New York: Walter de Gruyter
- van Dijk, T.A. and W. Kintsch, 1983. *Strategies of Discourse Comprehension*. Orlando, FL: Academic Press
- Downes, W., 1994. 'Register in Literature.' In R.E. Asher (ed.) *The Encyclopedia of Language and Linguistics*. Oxford, New York, Seoul, Tokyo: Pergamon Press, 3509-3511.
- Dreier, P., 1987. 'The corporate complaint against the media.' In D. Lazere (ed.), *American Media and Mass Culture. Left Perspectives*, The University of California Press, 64-80.
- Drew, P., 1994. 'Conversation analysis.' In R.E. Asher (ed.) *The Encyclopedia of Language and Linguistics*. Oxford, New York, Seoul, Tokyo: Pergamon Press, 749-752.
- Elek, J., 1997. 'News: A deeply rooted part of communication culture.' In Cunningham, J. and J. Horváth (eds.), 1997. *Framing the Issues. British Studies – Media Studies*. Pécs: Janus Pannonius University, 7-14.
- Fairclough, N., 1989. *Language and Power*. London and New York: Longman
1993. 'Critical discourse analysis and the marketisation of public discourse: The universities.' *Discourse and Society* 4/2: 133-168
- 1995a. *Critical Discourse Analysis: The Critical Study of Language*. London and New York: Longman
- 1995b. *Media Discourse*. London: Arnold
1996. 'A reply to Henry Widdowson's "Discourse Analysis: A critical view".' *Language and Literature* 5/1: 49-56.
- Fairclough, N. and R. Wodak, 1997. 'Critical discourse analysis.' In T. A. van Dijk (ed.) *Discourse as Social Interaction*, London, Thousand Oaks and New Delhi: SAGE, 258-284.
- Ferguson, C. A. 1983. 'Sports announcer talk: Syntactic aspects of register variation.' *Language in Society* 12: 153-172.
- Fishman, M., 1980. *Manufacturing the News*. Austin: University of Texas Press

Fiske, J., 1990. *Introduction to Communication Studies*. Second edition. London and New York: Routledge

Fiske, J. and J. Hartley, 1978. *Reading Television*. London and New York: Routledge

Fowler, R., 1985. 'Power.' In T. A. van Dijk (ed.), *Handbook of Discourse Analysis. Vol. 4. Discourse Analysis in Society*, London: Academic Press, 61-82.

1991. *Language in the News. Discourse and Ideology in the Press*. London and New York: Routledge

Fowler, R., Hodge, R., Kress, G. and T. Trew, 1979. *Language and Control*. London, Boston and Henley: Routledge and Kegan Paul

Garrett, P. and A. Bell, 1998. 'Media and discourse: A critical overview.' In A. Bell and P. Garrett (eds.), *Approaches to Media Discourse*, London: Blackwell Publishers, 1-20.

Geert, J. 1999. 'Hegemony in the news.' In J. Verschueren (ed.), *Language and Ideology. Selected papers from the 6th International Pragmatics Conference*, Antwerp: International Pragmatics Association, Vol. 1. 254-261.

Gerbner, G. 1985. 'Mass media discourse: Message system analysis as a component of cultural indicators.' In T. A. van Dijk (ed.), *Discourse and Communication. New Approaches to the Analysis of Mass Media Discourse and Communication*. Berlin and New York: Walter de Gruyter, 13-25.

Glasgow University Media Group. 1976. *Bad News*. London: Routledge

1980. *More Bad News*. London: Routledge

Graber, D. A., 1994. 'The infotainment quotient in routine television news: A director's perspective.' *Discourse and Society* 5/4: 483-508.

Greatbatch, D. 1998. 'Conversation analysis: Neutralism in British news interviews.' In A. Bell and P. Garrett (eds.), *Approaches to Media Discourse*, London: Blackwell Publishers, 163-185.

Hall, S., 1980a. 'Introduction to media studies at the Centre.' In S. Hall et al. (eds.), *Culture, Media, Language*. London: Routledge and Birmingham: CCCS, 117-121.

Hall, S., 1980b. 'Encoding/decoding.' In S. Hall et al. (eds.), *Culture, Media, Language*. London: Routledge and Birmingham: CCCS, 128-138.

Hall, S., D. Hobson, A. Lowe and P. Willis (eds.), 1980. *Culture, Media, Language*. London: Routledge and Birmingham: CCCS

Halliday, M.A.K., 1985a. *An Introduction to Functional Grammar*. London: Edward Arnold

1985b. 'Dimensions of Discourse Analysis: Grammar.' In T. A. van Dijk (ed.), *Handbook of Discourse Analysis. Vol. 2. Dimensions of Discourse*. London: Academic Press, 29-56.

Halliday, M.A.K. and R. Hasan, 1976. *Cohesion in English*, London and New York: Longman

Halmari, H. and J. Östman, 2001. 'The soft-spoken, angelic pickax killer: The notion of discourse pattern in controversial news reporting.' *Journal of Pragmatics* 33/6: 805-823.

Hartley, J., 1982. *Understanding News*. London and New York: Routledge

Heck, M. C., 1980. 'The ideological dimension of media messages.' In S. Hall et al. (eds.), *Culture, Media, Language*. London: Routledge and Birmingham: CCCS, 122-127.

Heritage, J. 1985. 'Analysing news interviews: Aspects of the production of talk for an overhearing audience.' In T. A. van Dijk (ed.), *Handbook of Discourse Analysis. Vol. 3. Discourse and Dialogue*. London: Academic Press, 95-117.

1994. 'Conversation analysis: Methodological aspects.' In U. M. Quasthoff (ed.), *Aspects of Oral Communication*, Berlin and New York: Walter de Gruyter, 393-418.

1988. 'Conversation analysis and institutional talk: Analysing distinctive turn-taking systems.' In S. Čmejrková et al., *Dialogue Analysis VI. Proceedings of the 6th Conference, Prague, 1996*. Tübingen: Max Niemeyer Verlag

Heritage, J. and D. Greatbatch, 1991. 'On the institutional character of institutional talk: The case of news interviews.' In D. Boden and D. H. Zimmerman (eds.), *Talk and Social Structure*, Berkeley: University of California Press, 93-137.

Heritage, J. and A. L. Roth, 1995. 'Grammar and institution: Questions and questioning in the broadcast news interview.' *Research on Language and Social Interaction* 28/1: 1-60

Herman, E. S. and N. Chomsky, 1988. *Manufacturing Consent. The Political Economy of the Mass Media*. New York: Pantheon Books

Holly, W., 1995. 'Secondary orality in the electronic media. In U. M. Quasthoff (ed.), *Aspects of Oral Communication*, Berlin and New York: Walter de Gruyter, 340-363.

Hutchby, I., 1996. *Confrontation Talk: Arguments, Asymmetries, and Power on Talk Radio*. Mahwah, NJ: Lawrence Erlbaum Associates

1999a. 'Frame attunement and footing in the organisation of talk radio openings.' *Journal of Sociolinguistics* 3/1: 41-63.

- Hutchby, I., 1999b. 'Beyond agnosticism: Conversation analysis and the sociological agenda.' *Research on Language and Social Interaction* 32/1-2: 85-93.
- Jalbert, P.L., 1983. 'Some constructs for analysing news.' In H. Davis and P. Walton (eds.), *Language, Image, Media*, Oxford: Basil Blackwell, 282-299.
- Kress, G., 1983. 'Linguistic and ideological transformations in news reporting.' In H. Davis and P. Walton (eds.), *Language, Image, Media*, Oxford: Basil Blackwell, 120-138.
- Kress, G., 1985. 'Ideological structures in discourse.' In T. A. van Dijk (ed.), *Handbook of Discourse Analysis, Vol. 4., Discourse Analysis in Society*, 27-42.
1993. 'Against arbitrariness: The social production of the sign as a foundational issue in critical discourse analysis.' *Discourse and Society* 4/2: 169-191.
- Kress, G. and R. Hodge, 1979. *Language as Ideology*. London, Boston and Henley: Routledge and Kegan Paul
- Kress, G. and T. van Leeuwen, 1998. 'Front pages: (The critical) analysis of newspaper layout.' In A. Bell and P. Garrett (eds.), *Approaches to Media Discourse*, London: Blackwell Publishers, 186-219.
- Kronrod, A. and O. Engel, 2001. 'Accessibility theory and referring expressions in newspaper headlines.' *Journal of Pragmatics* 33/5: 683-699
- Ladd, D. R., 1980. *The Structure of Intonational Meaning*. Bloomington and London: Indiana University Press
- Lakoff, G., 1987. *Women, Fire and Dangerous Things. What Categories Reveal about the Mind*. Chicago: The University of Chicago Press
- Lakoff, R. T., 1982. 'Some of my favorite writers are literate: the mingling of oral and literate strategies in written communication.' In D. Tannen (ed.), *Spoken and Written Language: Exploring Orality and Literacy*, Norwood NJ: Ablex, 239-260.
- Lazere, D. (ed.), 1987. *American Media and Mass Culture. Left Perspectives*, The University of California Press
- van Leeuwen, T. 1987. 'Generic strategies in press journalism.' *Annual Review of Applied Linguistics* 10/2:199-220.
1993. 'Genre and field in critical discourse analysis: a synopsis.' *Discourse and Society* 4/2: 193-223.
1995. 'Representing social action.' *Discourse and Society* 6/1: 81-106
- Leitner, G., 1997. 'The sociolinguistics of communication media.' In F. Coulmas (ed.), *The Handbook of Sociolinguistics*, Oxford and Malden: Blackwell, 187-204.

2000. 'Lexical frequencies in a 300 million word corpus of Australian newspapers.' *International Journal of Corpus Linguistics* 5/2:147-178.
- Lerman, C.L., 1983. 'Dominant discourse: the institutional voice and control of topic.' In H. Davis and P. Walton (eds.), *Language, Image, Media*, Oxford: Basil Blackwell, 75-103
- Marosán, L., 1994. 'Adjective or noun?' In L. Varga (ed.) *The EVEN Yearbook. A Collection of Papers by Members of the Department of English Linguistics*. Budapest: Eötvös Loránd University
2000. *The Parts of Speech and Meaning*. Unpublished PhD dissertation. Budapest: Eötvös Loránd University
- Martin, J. R., 1992. *English Text: System and Structure*. Philadelphia: John Benjamins
- Martinet, A., 1994. 'Functional grammar: Martinet's model.' In R.E. Asher (ed.) *The Encyclopedia of Language and Linguistics*. Oxford, New York, Seoul, Tokyo: Pergamon Press, 1323-1327.
- Mason, O., 2001/1994. *QTAG 3.01*. [A part-of-speech tagger].
URL <http://www.english.bham.ac.uk/staff/oliver/software/tagger/>
- McHoul, A. 1994., 'Discourse'. In R. E. Asher (ed.), *The Encyclopedia of Language and Linguistics*, Oxford, New York, Seoul, Tokyo: Pergamon Press, 940-947.
- McLuhan, M., 1964. *Understanding Media. The Extensions of Man*. Cambridge MA and London UK: The MIT Press
- Merriam-Webster's Collegiate Dictionary Online*. 19.03.2002. <http://www.m-w.com/cgi-bin/dictionary>
- Mey, I., 2001. 'The CA/CDA controversy.' *Journal of Pragmatics*. 33/4:609-615.
- Meyers, M., 1994. 'Defining homosexuality: News coverage of the "repeal the ban" controversy.' *Discourse and Society* 5/3: 321-344.
- Miller, J. and R. Cann, 1994. 'Data collection in linguistics.' In R. E. Asher (ed.), *The Encyclopedia of Language and Linguistics*, Oxford, New York, Seoul, Tokyo: Pergamon Press, 815-817.
- Moya, A. J. and J. I. Albentosa, 2001. 'Points of departure in news items and tourist brochures: Choices of theme and topic.' *Text* 2/3: 347-371.
- Murray, D. E., 1988. 'The context of oral and written language: A framework for mode and medium switching.' *Language in Society* 17/3:351-373.
- Nevalainen, T., 1992. 'Intonation and discourse type.' *Text* 12/3:397-427.

- Nir, R. and I. Roeh, 1987. 'Declarative verbs on the radio news.' *Hebrew Computational Linguistics* 25: 19-30.
- Ochs, E., 1979. 'Planned and unplanned discourse.' In T. Givón (ed.), *Syntax and Semantics. Vol. 12. Discourse and Syntax*, New York: Academic Press, 51-80.
- Ong, W. J., 1982. *Orality and Literacy: The Technologizing of the Word*. London: Methuen
- Östman, J., 1999. 'Coherence through understanding through discourse patterns: Focus on news reports.' In W. Bublitz, U. Lenk and E. Ventura (eds.), *Coherence in Spoken and Written Discourse. How to Create It and How to Describe It*. Amsterdam and Philadelphia: John Benjamins, 77-100.
- Pomerantz, A. and B. J. Fehr, 1997. 'Conversation analysis: An approach to the study of social action as sense making practices.' In T. A. van Dijk (ed.) *Discourse as Social Interaction, London, Thousand Oaks and New Delhi: SAGE*, 64-91.
- Postman, N., 1987. 'The teachings of the media curriculum.' In D. Lazere (ed.), *American Media and Mass Culture. Left Perspectives*, The University of California Press, 421-430.
- Psathas, G. 1994. 'Ethnomethodology.' In R.E. Asher (ed.) *The Encyclopedia of Language and Linguistics*. Oxford, New York, Seoul, Tokyo: Pergamon Press, 1160-1164.
1995. *Conversation Analysis. The Study of Talk-in-Interaction*. Thousand Oaks, London and New Delhi: SAGE
- Quirk, R, Greenbaum, S, Leech, G, and J. Svartvik, 1985. *A Comprehensive Grammar of the English Language*. London and New York: Longman
- Richards, J. C. and R. W. Schmidt, 1983. 'Conversational analysis.' In J. C. Richards and R. W. Schmidt (eds.), *Language and Communication*, London and New York: Longman, 117-154.
- Richardson, K., 1998. 'Signs and wonders: Interpreting the economy through television.' In A. Bell and P. Garrett (eds.), *Approaches to Media Discourse*, London: Blackwell Publishers, 220-250.
- Scannell, P. 1998. 'Media – Language – World.' In A. Bell and P. Garrett (eds.), *Approaches to Media Discourse*, London: Blackwell Publishers, 251-267.
- Schiffrin, D., 1987. *Discourse Markers*. Cambridge: Cambridge University Press
- Schiffrin, D., 1994. *Approaches to Discourse*. Oxford UK and Cambridge USA: Blackwell
- Schneider, E. W., 2001. 'Review of Biber et al., *Longman Grammar of Spoken and Written English*, Harlow: Longman, 1999.' *English World-Wide* 22/1:137-143.

Schröder, K. C. 1994. 'Media language and communication.' In R. E. Asher (ed.), *The Encyclopedia of Language and Linguistics*. Oxford, New York, Seoul, Tokyo: Pergamon Press, 2412-2420.

Sebba, M. and S. D. Fligelstone, 1994. 'Corpora.' In R. E. Asher (ed.), *The Encyclopedia of Language and Linguistics*. Oxford, New York, Seoul, Tokyo: Pergamon Press, 769-773.

Stubbs, M., 1983. *Discourse Analysis. The Sociolinguistic Analysis of Natural Language*. Oxford: Basil Blackwell

Tannen, D., 1982. 'The oral/literate continuum in discourse.' In D. Tannen (ed.), *Spoken and Written Language: Exploring Orality and Literacy*, Norwood, NJ: Ablex. 1-16.

Thomson, A. J. and A.V. Martinet, 1986. *A Practical English Grammar*, Fourth edition. Oxford: Oxford University Press

Togoby, D., 1994. 'Text Pragmatics'. In R. E. Asher (ed.), *The Encyclopedia of Language and Linguistics*, Oxford, New York, Seoul, Tokyo: Pergamon Press, 4587-4581.

Toolan, M. 1997. 'What is Critical Discourse Analysis and why are people saying such terrible things about it?' *Language and Literature* 6/2: 83-103.

Trew, T., 1979. "'What the papers say": linguistic variation and ideological difference.' In Fowler et al. (eds.), *Language and Control*, London, Boston and Henley: Routledge and Kegan Paul, 117-156.

Triandafyllidou, A., 1995. 'The Chernobyl accident in the Italian press: a "media storyline".' *Discourse and Society* 6/4: 517-536.

Tuchman, G., 1987. 'Representation and the web of news narrative: The web of facticity.' In D. Lazere (ed.), *American Media and Mass Culture. Left Perspectives*, The University of California Press, 331-344.

Vagle, W., 1991. 'Radio language – spoken or written?' *International Journal of Applied Linguistics* 1/1: 118-131.

Wacha, I., 1988. 'Élő nyelvi (spontán) szövegek megnyilatkozásainak (szintaktikai) vizsgálati szempontjaihoz (a gazdagréti kábeltelevízió élő nyelvi felvételei alapján).' In M. Kontra (ed.), *Beszélt nyelvi tanulmányok*, Budapest: Az MTA Nyelvtudományi Intézete. 103-158.

Weber, J. J. 2002. 'The critical practices of Henry Widdowson.' *Language and Literature* 11/2: 153-160.

Widdowson, H. 1995. 'Discourse analysis: A critical view.' *Language and Literature* 4/3: 157-172.

1996. 'Reply to Fairclough. Discourse and interpretation: Conjectures and refutations.' *Language and Literature* 5/1: 57-69.

2002. 'Verbal art and social practice: A reply to Weber.' *Language and Literature* 11/2:161-167.