A Corpus of Formal British English Speech

The Lancaster/IBM Spoken English Corpus

Edited by
Gerald Knowles,
Briony Williams and Lita Taylor

A Corpus of Formal British English Speech

A Corpus of Formal British English Speech: The Lancaster/IBM Spoken English Corpus

Edited by Gerry Knowles
Briony Williams, L. Taylor

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 1996 by Addison Wesley Longman Limited

Published 2013 by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN
711 Third Avenue, New York, NY 10017, USA

Routledge is an imprint of the Taylor & Francis Group, an informa business

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

ISBN 13: 978-0-582-05639-8 (pbk)

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record has been applied for

Set by 8

Contents

Editor's acknowledgements

The Spoken English Corpus (SEC) project was supported jointly in 1984–5 by the Humanities Research Fund at Lancaster University and by IBM (UK) Ltd, and subsequently by IBM UK Ltd. IBM not only provided financial support, but actively participated in the project.

The editors wish to thank the many people at Lancaster and IBM who contributed to the project, or to the publication of the corpus. The project would not have begun at all but for the continued help and support of Geoffrey Leech at Lancaster and Geoffrey Kaye at IBM. The publication of the corpus in its present form was made possible by Peter Alderson, who later took over as Speech Research Manager at IBM. The text was made ready for publication thanks to the work of Nick Campbell, former research fellow at IBM, Matthew Peters, programmer at IBM, and of students at IBM, in particular Ashley Groombridge.

Without the speakers who produced the texts, there would of course be no spoken corpus at all. The University Orator, Colin Lyas, kindly reproduced two of his Degree Day speeches as M05 and M06. Heather Kempson and Rita Green, at that time MA students, produced the dialogue J06. These three texts were recorded in the Media Services Unit at Lancaster University.

The majority of texts in the corpus were obtained from the BBC, and we wish to thank Norma Jones of BBC Sound Archives for her help in organising contracts, contacting speakers, and providing information during the three years of the project.

Thanks are also due to all those who gave free permission for their work and/or speech to be included in the corpus:

Elizabeth Bell (*Story Time*); Louise Botting (*Money Box*); James Cox (*News*); John Carlin (*From our own Correspondent*); Alina Dadlez and Decca International (extracts from *Betjeman Reads Betjeman*); Isabelle Dean (*Time for Verse*); Paddy Feeny (*Review of the Year*); Dr Robert Fox (*Science and Belief in 18C France*); Susan Hampshire (*Week's Good Cause*); David Henderson (*The Reith Lectures III*); John Hollis (*Listening and Reading*); Martin Jarvis (*Morning Story* and *Time for Verse*); Juliet Johnson (*Money Box*); Catherine Kneafsey and Oxford University Press (extracts from *Streamline English Series*); Doris Lessing (author of *Through the Tunnel*); Colin Lyas (*Nelson Mandela* and *Tom Stephenson speeches*); Christopher Poole (*Review of the Year*); Brian Redhead (*Week's Good Cause*); Diana Ruault and Open University Educational Enterprises Ltd. (OU programmes – *Modern Art, Science and Belief in 18th Century France, Development of Fractions*); Graham Seal (author of *What shall we do if it rains?*); Simon Taylor (*Review of the Year*); Janet Trewin (*News*); The Met Office (*Weather Forecasts*).

Special thanks are due to Molly Price-Owen of BBC World Service Sport for her help with the *Review of the Year* extracts.

The following samples were obtained with the permission of the BBC and are covered by a contract with them covering copyright permission:

A01	In Perspective
A02–A12	From our own Correspondent
B01–B04	News
C01	The Reith Lectures III
E01–E02	Daily Service
F01–F03	Money Box
F04	Review of the Year
G01	Story Time
G02	Listening and Reading
G05	Morning Story
H04–H05	Time for Verse
J01	Review of the Year
K01–K02	Week's Good Cause
M02	Motoring News
M03	Weather Forecast
M04	Programme News
M07	Travel Roundup
M08	Weather Forecast
M09	Programme News

In addition to the copyright agreement with the BBC, permission to use material was obtained from any individuals not employed by the BBC,

The Open University Educational Enterprises Ltd gave permission for the inclusion of samples D01–D03.

Oxford University Press gave permission for the inclusion of *Streamline English* material, texts G03, G04, J02, J03, J04, and J05.

Decca International gave permission for inclusion of the samples of John Betjeman's work, texts H01, H02, H03 and M01.

Publisher's acknowledgements

We are grateful to the following for permission to reproduce copyright material:

The BBC for extracts from their programmes *In Perspective* (Radio 4), *From our own Correspondent* (Radio 4), *News* (Radio 2, 3 and 4), *Daily Service* (Radio 4), *Money Box* (Radio 4), *Review of the Year* (BBC World Service), *Motoring News* (Radio 4), *Weather Forecast* (Radio 4), *Programme News* (Radio 4), *Travel Roundup* (Radio 4); The Commonwealth Society for the Deaf for *This Week's Good Cause* (Radio 4 25.1.87); David Henderson for *The Reith Lectures III* (Radio 4 20.11.85); Heather Kempson and Rita Green for their dialogue J6; Jonathan Clowes Ltd, London on behalf of Doris Lessing/HarperCollins Publishers, Inc for an abridged and complete form (see page 262) of the short story 'Through the Tunnel' from *The Habit of Loving* by Doris Lessing. Copyright 1954 Doris Lessing/Copyright 1955 by Doris Lessing. Originally appeared in The New Yorker. Copyright renewed; Colin Lyas for two of his Degree Day speeches M05 and M06; John Murray (Publishers) Ltd for the poems 'Eunice', 'Felixstowe: *or* The Last of her Order', 'Harrow-on-the-Hill' and an extract from *Summoned by Bells* by John Betjemin; The Open University for 'Modern Art: Berlin Dada', 'Science and Belief' and 'Development of Fractions'; Oxford University Press for extracts from *Streamline English* by Bernard Hartley and Peter Viney (1978); the author's agent on behalf of the author for 'Lion at School' from *The Lion at School and Other Stories* by Philippa Pearce (Viking Kestrel). © 1971 Philippa Pearce; Graham Seal for an extract from his short story 'What shall we do if it rains?'.

We have unfortunately been unable to trace the copyright holder of the item K01 (*This Week's Good Cause*) (Radio 4 18.1.87) and would appreciate any information which would enable us to do so.

Introduction

The Lancaster/IBM Spoken English Corpus began in September 1984 as part of a research project into the automatic assignment of intonation undertaken by members of the Unit for Computer Research on the English Language at Lancaster University in collaboration with the Speech Research Group at IBM UK Scientific Centre. The first task for the project was to collect samples of natural spoken British English which could be used as a database for analysis and for testing the intonation assignment programs. The result is a machine-readable corpus of approximately 52,637 words of contemporary spoken British English.

The original design of the corpus was determined by the need to provide data for research into speech synthesis. As a result, unlike most other corpora currently being used in the computational linguistics field, the SEC exists in various forms. The texts are made available not only in standard orthographic form, but in different versions with a range of special annotations. For instance, in order to study prosodic patterns in the corpus, the texts were transcribed prosodically. A grammatically annotated version was produced using the CLAWS system (Constituent-Likelihood Automatic Word-tagging System) developed at Lancaster, which made possible an analysis of the syntax of spoken texts, and of the connection between syntax and prosody.

However, whatever the original motivation for compiling a corpus, it quickly becomes an object of interest in its own right. New users find it valuable for applications for which it was not designed. For example, the SEC has been used profitably by MA students at Lancaster whose main interest is in English Language Teaching. For most of these users the corpus has been compiled in an appropriate form, e.g. the orthographic texts present the texts in the manner required for concordancing. There are also users who need to make their own independent analysis of the data. They may wish to make a kind of annotation not yet carried out, e.g. phonemic transcription, or check the prosodic transcriptions. For such users copies of the original recordings are available. These were produced mainly by IBM UK Scientific Centre using equipment of sufficient quality to make the tapes suitable for instrumental analysis.

The corpus contains an interesting range of speaking styles. It is impossible in a corpus of 52,000 words to include samples of every style of spoken English; emphasis has been placed on collecting a sizeable sample of the type of spoken English suitable as a model for speech synthesis. This explains the relatively high proportion of prepared or semi-prepared speech produced by trained broadcasters. Small samples of highly stylised speech of the kind used in sermons or poetry reading have been included: these are examples of what might traditionally be regarded as 'good speaking', but are in fact quite inappropriate as models for speech synthesis by computer. These have proved of considerable interest for the study of prosodic style.

An important attribute of a modern corpus is that it is computer-readable: a corpus tends to reside on a hard disk rather than a bookshelf. In presenting the

corpus in this book form, we have taken into account the needs of established corpus linguists, and of those who are not yet familiar with corpora. Anyone who has the corpus on disk can make hard copies of most of the files; but without a special font to print the prosodic symbols, the prosodic texts will be either unprintable or unreadable. For this reason the prosodic version has been chosen for publication. It is also the case that despite a rapidly growing literature on prosody and intonation, it is still difficult to find prosodically transcribed texts of naturally produced speech. It is hoped that this book will be of use to those who are unfamiliar with corpora but need access to prosodically transcribed material.

The original compilation of the SEC was completed by 1987. Since then the corpus has been analysed from several different points of view. Some of the early findings have been published (Knowles et al. 1996). But it is far from finished. Work so far has essentially involved an extension of techniques designed for written corpora (see e.g. Garside et al. 1987). The current aim is to develop a new methodology specially for spoken corpora. This will involve digitising the speech waveforms, conversion of the corpus texts into a speech database, and the development of annotation methods to relate the texts to the phonetic information in the waveforms.

Prosodic characters

The prosodic transcription was based on the system used by O'Connor & Arnold (1973) with some modifications:

- no distinction was made between symbols for onsets and nuclei
- the distinction between high and low variants was extended to all tones
- high and low were defined not with respect to the pitch range, but to the immediately preceding pitch level

These are discussed further below.

The following set of 15 special characters was used:

|	Minor tone-group boundary
||	Major tone-group boundary
ˆ	Caret
ˋ	High fall
ˎ	Low fall
ˊ	High rise
ˏ	Low rise
¯	High level
ˍ	Low level
ˇ	High fall rise
ˬ	Low fall rise
ˆ	Rise fall
·	Stressed but unaccented
↑	Up arrow
↓	Down arrow

Table 1. Tonetic stress marks

Some notes on the interpretation of these characters are necessary:

(1) Stressed and accented syllables:

An accented syllable has an independent pitch movement associated with it, the 'tone'. Tones are marked with iconic symbols representing the pitch movement.

Syllables which are perceived to be stressed but not accented (i.e. they are prominent but have no independent pitch movement) are marked with a dot.

Unstressed syllables are left unmarked.

The pitch of all unaccented syllables is predictable from the tone marks on neighbouring accented syllables.

(2) Pitch direction:

The terms *fall*, *rise*, and *level* describe pitch movements which begin on the vowel of the accented syllable (or in the case of a falling diphthong, on the first element of the diphthong). Any pitch movements before this point are ignored in this terminology.

Most of the pitch movement of a fall is completed on or soon after the accented syllable, leaving a slight drop over the tail of the tone group. A rise might start almost level, with a marked increase in slope towards the end of the tone group.

The level tone is not strictly level except possibly in special styles in which it is intoned. Any rising or falling is insufficient for the tone to be classed as a rise or a fall.

(3) Simple and complex tones:

Simple tones move in only one direction up or down. Complex tones change pitch direction: fall-rise and rise-fall.

The fall of the fall-rise is completed quickly, and the completion of the rise is delayed to the end of the tone group. The rise of the rise-fall is completed quickly on the accented syllable, and the fall is completed as soon as possible thereafter.

A phonetic variant of the fall-rise is the 'shallow fall': instead of falling to low and rising again, the pitch movement is cut off before it reaches low.

(4) High and low:

A distinction is made for all tones between a high and a low variety. A high tone begins higher than the preceding pitch level, and a low tone begins lower than the preceding pitch level.

(5) Up arrow and down arrow:

These are used to indicate significant changes of pitch which are not sufficiently marked by the tone symbols. The up arrow indicates a rise in pitch and the down arrow a drop in pitch.

When used in conjunction with tone marks they indicate a rise or drop in pitch which is significantly greater than that indicated by the high or low position of the tone mark alone.

Used alone, on an unstressed syllable, the arrow marks a pitch pattern which is not predictable from neighbouring tone marks. At the beginning of a tone group, the arrow indicates that the pitch contour begins significantly above or below its expected level towards the bottom of the pitch range.

For further details of the system, see (Knowles et al. 1996) especially chapter 3, 'The formulation of an intonation transcription system for British English'.

The composition of the corpus

In line with the conventions used in the Lancaster-Oslo/Bergen (LOB) corpus project (Johansson et al. 1978, 1986), each text is assigned to a category [A..M] and identified by number within that category. In addition, each text is given an absolute number to indicate its position in the corpus as a whole. For example H03 is the third text in category H, and is also text no. 34.

Figure 1 on page 7 lists the categories, and gives the total number of words in each category, and this figure expressed as a percentage of the 52,637 words in the whole corpus.

Breakdown into categories

Category A - Commentary

A01	In Perspective
A02-A12	From our own Correspondent

Texts in this category include news reports on events happening around the world. They are more informal than those in category B. *In Perspective* covers a religious topic, the reports *From our own Correspondent* all deal with overseas news events:

A02	The PNC meetings in Lebanon
A03	Westmorland and Sharon suing *Time* magazine
A04	The conflict in El Salvador
A05	The economic climate in Rumania
A06	The plight of Turks living in Huttenheim, Germany
A07	The hijack of the TWA passengers and the release of the Shi'ites
A08	Security checks at Athens airport
A09	The new government in Namibia
A10	The plight of the Tamil refugees
A11	Report on Gorbachev and his government
A12	The financial state of banks in Hong Kong

Category		No. of texts	Total words	% Total
A	Commentary	12	9066	17
B	News broadcast	4	5235	10
C	Lecture type I	3	4471	8
D	Lecture type II	1	7451	14
E	Religious broadcast	2	1503	3
F	Magazine-style reporting	4	4710	9
G	Fiction	5	7299	14
H	Poetry	5	1292	2
J	Dialogue	6	6826	13
K	Propaganda	2	1432	3
M	Miscellaneous	9	3352	6
	Grand Total:	53	52637	

Figure 1. Composition of the Spoken English Corpus

Category B - News Broadcasts

B01	Radio 4 News
B02	Radio 4 News
B03	Radio 2 News
B04	Radio 3 News

These are news reports of current and recent events in Great Britain and abroad. B04 consists of one speaker, B01-B03 consist of a main newsreader and additional reporters. The style of the main newsreaders is more formal than that of the reporters.

Category C - Lecture type I

C01	The Reith Lectures - III

This is a lecture addressed to the general public on economics entitled 'Needs, Centralism, and Autarchy'.

Category D - Lecture type II

D01	Open University - Modern Art
D02	Open University - Science and Belief
D03	Open University - Development of Fractions

These lectures are designed to be used as part of an Open University course. D01 covers the Berlin Dada movement in Germany, and so contains some German names and words. D02 is a discussion of theology and science in 18th-century France. D03 describes the development in the notation used in the representation of fractions, and contains some simple mathematical formulae.

Category E - Religious Broadcast

E01	Daily Service
E02	Daily Service

These are religious services with the hymns edited out.

Category F - Magazine-style reporting

F01	Money Box
F02	Money Box
F03	Money Box
F04	Review of the Year

These texts contain magazine-style in-depth reporting of financial news. Topics covered are: the perks of owning shares; the upgrading of state benefits; listeners' trusts; and Building Society rates.

Category G - Fiction (general)

G01	Story Time
G02	Listening and Reading
G03	Streamline English course samples
G04	Streamline English course samples
G05	Morning Story

All these texts are general fiction. G01 is a story aimed at an adult audience, *Through the Tunnel*, by Doris Lessing. G02 is aimed at children aged between 5 and 10, *Lion at School*, by Philippa Pearce. G03 and G04 are stories taken from the ELT course book *Streamline English*. G05 is aimed at adults - *What shall we do if it rains?*, by Graham Seal.

Category H - Poetry

H01	John Betjeman
H02	John Betjeman
H03	John Betjeman
H04	Time for Verse
H05	Time for Verse

This section consists of poetry readings of the work of John Betjeman and Sir Henry Newbolt. H01-H03 are John Betjeman's readings of his own poems: *Eunice*, *The Last of Her Order*, and *Harrow-on-the-Hill*. H04 and H05 are actors reading Sir Henry Newbolt's poems: *The Linnet's Nest*, and *The Nightjar*.

Category J - Dialogue

J01	Review of the Year
J02	course sample from *Streamline English*
J03	course sample from *Streamline English*
J04	course sample from *Streamline English*
J05	course sample from *Streamline English*

J06 Kempson and Green dialogue

These are dialogues of varying degrees of informality. J01 consists of a radio discussion of notable sports events of 1986. J02-J05 are dialogues contrived to illustrate a particular facet of English (although this is not immediately noticeable) for the *Streamline English* ELT course. J06 is an informal dialogue between two MA students about working abroad.

Category K - Propaganda

K01 Week's Good Cause

K02 Week's Good Cause

These two texts are charity appeals.

Category M - Miscellaneous

M01 John Betjeman

M02 Motoring News

M03 Weather Forecast

M04 Programme News

M05 Oratory

M06 Oratory

M07 Travel Roundup

M08 Weather Forecast

M09 Programme News

M01 is a sample of John Betjeman reading a section of prose: *An Unpleasant Nursemaid.* M02 and M07 consist of reports on road conditions. M03 and M08 are weather forecasts. M04 and M09 give details of forthcoming programmes on Radio 4. M05 and M06 are speeches delivered at degree ceremonies before the presentation of honorary degrees to Nelson Mandela and Tom Stephenson.

Speakers

As far as possible, the texts selected for the corpus were spoken in received pronunciation (RP). This was relatively easy where material from the BBC programmes *From our own Correspondent* and the *News* was concerned, as the BBC themselves make similar requirements of their presenters of news or news commentary programmes. Most of the speakers in the corpus have accents reasonably close to RP - even if, in some cases, the concept of 'RP' has been interpreted rather widely - and those with particularly strong regional accents were excluded.

A balance was sought wherever possible between male and female speakers. Of the 53 texts in the corpus, 17 were produced wholly or in part by female speakers. This represents 30 per cent of the corpus. In the highly stylised texts - poetry, religious broadcast, propaganda, and dialogue, there is a reasonably good balance. The higher percentage of male speakers in the News and Commentary categories reflects the tendency of the BBC to use mainly male speakers in these types of programme.

The following lists give the speakers in the corpus by category. In some cases it has not been possible to identify speakers by name, and so they are simply listed as 'male speaker' or 'female speaker'.

A01 Rosemary Hartill
A02 Gerald Butt
A03 Jon Silverman
A04 John Carlin
A05 James Morgan
A06 David Smeeton
A07 Laurie Margolis
A08 Keith Graves
A09 Graham Leach
A10 Alan MacDonald
A11 Peter Ruff
A12 Jim Biddulph

B01 Brian Perkins
Mike Wooldridge
Laurie Margolis
Janet Trewin

B02 Brian Perkins
James Cox
Clive Small
Kevin Bocquet
Ann Cadwallader
Peter Burden

B03 David Geary
Peter Smith
Colin Blane
David Davis

B04 Peter Bragg

C01 David Henderson

D01 Dawn Adies
D02 Dr Robert Fox
D03 Graham Flegg

E01 Frances Gumley

E02 Rev Stephen Oliver

F01 Louise Botting
Peter Day

F02 Louise Botting
Vincent Duggleby

F03 Louise Botting
Juliet Johnson
Frances McDonald

F04 Kevin Geary
Chris Florence
Harry Peart
Simon Taylor
Linda Spurr
Chris Poole
Mike Costello

G01 Elizabeth Bell
G02 John Hollis
G03 Male speaker

G04 Female speaker
G05 Martin Jarvis

H01 John Betjeman
H02 John Betjeman
H03 John Betjeman

H04 Isabelle Dean
H05 Martin Jarvis

J01 Kevin Geary
Martin Fookes
Paddy Feeny

J02 Male speakers
J03 Male speakers

J04 Male speaker
Female speaker
J05 Male speaker
Female speaker
J06 Heather Kempson
Rita Green

K01 Brian Redhead

K02 Susan Hampshire

M01 John Betjeman
M02 Male speaker
M03 Male speaker
M04 Male speaker
M05 Colin Lyas

M06 Colin Lyas
M07 Male speaker
M08 Male speaker
M09 Male speaker

Dates of composition and recording

For material obtained from BBC radio programmes the date of broadcast was as follows:

A01	In Perspective	24th November, 1984
A02-A06	From our own Correspondent	24th November, 1984
A07-A12	From our own Correspondent	22nd June, 1985
B01	News (R4)	24th November, 1984
B02	News (R4)	22nd June, 1985
B03	News (R2)	5th December, 1985
B04	News (R3)	14th January, 1986
C01	The Reith Lectures III	20th November, 1985
E01	Daily Service	26th November, 1985
E02	Daily Service	27th November, 1985
F01	Money Box	24th November, 1984
F02-F03	Money Box	22nd June, 1985
F04	Review of the Year	December, 1986
G01	Story Time	25th June, 1985
G02	Listening and Reading	28th January, 1987
G05	Morning Story	26th November, 1986
H04-H05	Time for Verse	26th November, 1986
J01	Review of the Year	December, 1986
K01	Week's Good Cause	18th January, 1987
K02	Week's Good Cause	25th January, 1987
M02	Motoring News	18th January, 1987
M03	Weather Forecast	18th January, 1987
M04	Programme News	18th January, 1987
M07	Travel Roundup	25th January, 1987
M08	Weather Forecast	25th January, 1987
M09	Programme News	25th January, 1987

For material prepared at the Media Services Unit, Lancaster University, the actual dates of recording were:

J06	Kempson and Green dialogue	11th March, 1987

M05	Nelson Mandela speech	29th January, 1987
M06	Tom Stephenson speech	29th January, 1987

The samples from *Betjeman Reads Betjeman* are all dated 'since 1954':

H01 Eunice
H02 The Last of Her Order
H03 Harrow-on-the-Hill

M01 An Unpleasant Nursemaid

The *Streamline English* texts were first published in 1982:

G03 SE Unit 10 - A funny thing happened...
G04 SE Unit 19 - Night flight

J02 SE Unit 16 - Inside Story
J03 SE Unit 25 - Murder at Gurney Manor
J04 SE Unit 72 - Getting things done
J05 SE Unit 75 - Messages

The material obtained from the Open University unfortunately contained no information on date of composition or publication. The texts concerned are:

D01 OU Modern Art
D02 OU Science and Belief in 18th-Century France
D03 OU Development of Fractions

The dates of composition of the Newbolt poems were:

H04	The Linnet's Nest	May, 1924
H05	The Nightjar	May, 1925

The duration of text extracts

The total duration of the corpus is 339 minutes 18 seconds. The average length of sample is 6 minutes, but individual texts vary greatly from this. Texts were not kept to a fixed duration as it was considered important to be able to study complete sections of speech (e.g. to observe the ways in which monologues begin and end). A predetermined cut-off point based on a number of words or minutes would have resulted in an unnatural-sounding endpoint to speech samples. The following table gives details of extract lengths in the corpus.

Cat	m:s		Cat	m:s	
A01	15:00		G01	20:00	
A02	4:28		G02	8:56	
A03	4:01		G03	2:39	
A04	5:41		G04	5:30	[Tot G]
A05	4:48		G05	9:20	[46:25]
A06	4:32				
A07	3:54		H01	1:41	
A08	4:08		H02	2:03	
A09	5:12		H03	1:00	
A10	4:26		H04	2:59	[Tot H]
All	4:15	[Tot A]	H05	1:17	[9:00]
A12	4:05	[64:30]			
			J01	7:58	
B01	9:32		J02	1:31	
B02	9:40		J03	2:04	
B03	5:00	[Tot B]	J04	0:27	
B04	5:00	[29:12]	J05	1:28	[Tot J]
			J06	24:00	[37:28]
C01	30:00	[30:00]			
			K01	4:32	[Tot K]
D01	19:00		K02	4:09	[8:41]
D02	19:00	[Tot D]			
D03	19:00	[57:00]	M01	0:41	
			M02	1:10	
E01	6:48	[Tot E]	M03	0:48	
E02	4:30	[11:18]	M04	1:40	
			M05	4:33	
F01	3:48		M06	7:05	
F02	3:32		M07	1:06	
F03	4:54	[Tot F]	M08	0:47	[Tot M]
F04	13:16	[25:30]	M09	2:24	[20:14]

SEC text details

The following table gives for each text:

- the category letter and text number in the category (e.g. B02)
- the text number in the corpus as a whole (e.g. 14)
- the programme title
- the name of the speaker(s)
- the number of words

Category A - Commentary: 9066 words

A01	01	In Perspective	Rosemary Hartill	793
A02	02	From our own Correspondent	Gerald Butt	734
A03	03	From our own Correspondent	Jon Silverman	620
A04	04	From our own Correspondent	John Carlin	977
A05	05	From our own Correspondent	James Morgan	804
A06	06	From our own Correspondent	David Smeeton	828
A07	07	From our own Correspondent	Laurie Margolis	716
A08	08	From our own Correspondent	Keith Graves	618
A09	09	From our own Correspondent	Graham Leach	787
A10	10	From our own Correspondent	Alan MacDonald	800
A11	11	From our own Correspondent	Peter Ruff	785
A12	12	From our own Correspondent	Jim Biddulph	604

Category B - News Broadcasts: 5235 words

B01	13	Radio 4 News	Brian Perkins Mike Wooldridge Laurie Margolis Janet Trewin	1722
B02	14	Radio 4 News	Brian Perkins James Cox Clive Small Kevin Bocquet Ann Cadwallader Peter Burden	1720
B03	15	Radio 2 News	David Geary Peter Smith Colin Blane David Davis	940

B04	16	Radio 3 News	Peter Bragg	853

Category C - Lecture Type I: 4471 words

C01	17	The Reith Lectures - III	David Henderson	4471

Category D - Lecture Type II: 7451 words

D01	18	OU Modern Art	Dawn Adies	2410
D02	19	OU Science and Belief	Dr Robert Fox	2434
D03	20	OU Development of Fractions	Graham Flegg	2607

Category E - Religious Broadcast: 1503 words

E01	21	Daily Service	Frances Gumley	915
E02	22	Daily Service	Rev Stephen Oliver	588

Category F - Magazine-style reporting: 4710 words

F01	23	Money Box	Louise Botting Peter Day	671
F02	24	Money Box	Louise Botting Vincent Duggleby	667
F03	25	Money Box	Louise Botting Juliet Johnson Frances MacDonald	850
F03	26	Review of the Year	Kevin Geary Chris Florence Harry Peart Simon Taylor Linda Spurr Chris Poole Mike Costello	2522

Category G - Fiction: 7299 words

G01	27	Story Time	Elizabeth Bell	3163
G02	28	Listening and Reading	John Hollis	1221
G03	29	A funny thing happened...	Male	442
G04	30	Night Flight	Female	810
G05	31	Morning Story	Martin Jarvis	1663

Category H - Poetry: 1292 words

H01	32	Eunice	John Betjeman	248
H02	33	The Last of Her Order	John Betjeman	286
H03	34	Harrow-on-the-Hill	John Betjeman	157
H04	35	The Linnet's Nest	Isabelle Dean	405
H05	36	The Nightjar	Martin Jarvis	196

Category J - Dialogue: 6826 words

J01	37	Review of the Year	Kevin Geary Martin Fookes Paddy Feeny	1674
J02	38	Inside Story	Males	279
J03	39	Murder at Gurney Manor	Males	375
J04	40	Getting things done	Male & Female	74
J05	41	Messages	Male & Female	277
J06	42	Kempson and Green dialogue	Rita Green Heather Kempson	4147

Category K - Propaganda: 1432 words

K01	43	Week's Good Cause	Brian Redhead	798
K02	44	Week's Good Cause	Susan Hampshire	634

Category M - Miscellaneous: 3352 words

M01	45	An Unpleasant Nursemaid	John Betjeman	93
M02	46	Motoring News	Male	200
M03	47	Weather Forecast	Male	140
M04	48	Programme News	Male	298
M05	49	Nelson Mandela speech	Colin Lyas	738
M06	50	Tom Stephenson speech	Colin Lyas	1112

M07	51	Travel Roundup	Male	187
M08	52	Weather Forecast	Male	143
M09	53	Programme News	Male	441