1. Introduction

This manual contains a description of the design principles underlying the British National Corpus (BNC), and detailed information about the way in which it is encoded, in particular, a definition of the SGML document type declaration (DTD) used. A list giving brief bibliographic details for each text making up the corpus is also included.

This edition of the manual is a revised version of the document released with version 1.0 of the corpus, as distributed in May 1995. It describes the BNC World Edition, released in October 2000. Further information about the BNC is also available from its World Wide Web server at http://www.hcu.ox.ac.uk/BNC

The material presented in this manual derives from a number of BNC Project internal documents, with original contributions from all the participants in the project. Factual errors, chiefly relating to the composition of the corpus, have been corrected and the description of the encoding scheme has been modified in line with the changes introduced in this version. In other respects, this version of the documentation has been unchanged from the first release of the corpus. A brief list of the revisions made to the corpus encoding is given in section 9 (Compatibility issues).

1.1. Acknowledgments

The BNC was created by an academic-industrial consortium whose original members were:

- Oxford University Press
- Longman Group Ltd
- Chambers Harrap
- Oxford University Computing Services
- Unit for Computer Research on the English Language (Lancaster University)
- British Library Research and Development Department

Creation of the corpus was funded by the UK Department of Trade and Industry and the Science and Engineering Research Council under grant number IED4/1/2184 (1991-1994), within the DTI/SERC Joint Framework for Information Technology. Additional funding was provided by the British Library and the British Academy.

Management of the project was co-ordinated by an executive committee whose members were as follows:

**OUP** Tim Benbow; Simon Murison-Bowie

**Longman** Della Summers; Rob Francis
1.1. Acknowledgments

**Chambers Harrap** John Clement
**OUCS** Lou Burnard
**UCREL** Geoffrey Leech
**British Library** Terry Cannon
**DTI observers** Gerry Gavigan; Donald Bell

An Advisory Council supervised the running of the project 1991-1994. Members of this Council were:

- Dr Michael Brady
- Christopher Butler
- Professor David Crystal
- Sir Antony Kenny (chair)
- Dr Nicholas Ostler
- Professor Sir Randolph Quirk
- Tim Rix
- Dr Henry Thompson

Many people within each member organization made major contributions to the success of the project. It is a pleasure to acknowledge their hard work and dedication here.

**OUP** Lyndsay Brown; Jeremy Clear (project manager 1991-2); Caroline Davis; Ginny Frewer; Frank Keenan; Tom McLean; Anita Sabin; Ray Woodall (project manager 1992-4)
**Longman** Steve Crowdy (project manager); Denise Denney; Duncan Pettigrew
**Chambers Harrap** Robert Allen; Ilona Morison
**OUCS** Glynis Baguley; Gavin Burnage; Tony Dodd; Dominic Dunlop (project manager 1992-4)
**UCREL** Tom Barney; Michael Bryant (project manager 1991-3); Elizabeth Eyes; Jean Forrest; Roger Garside; Mary Hodges; Mary Kinane; Nicholas Smith; Xungfeng Xu.

The project also benefited greatly from the advice and support of many external consultants. Listing all those who have influenced our thinking and to whom we are indebted would be very difficult, but chief amongst them we would like to thank:

- Sue Atkins
- Clive Bradley
- Ann Brumfitt
- Charles Clark
- James Clark
- Bruce Heywood
- Mark Lefanu
- Michael Rundle
- Richard Sharman
- Michael Sperberg-McQueen
- Anna-Brita Stenström
- Russell Sweeney
2.2. General definitions

After the completion of the first edition of the BNC, a phase of tagging improvement was undertaken at Lancaster University with funding from the Engineering and Physical Sciences Research Council (Research Grant No. GR/F 99847). This tagging enhancement project was led by Geoffrey Leech, Roger Garside and Tony McEnery. The main objective was to correct as many tagging errors as possible, using an enhanced version of Claws4. In addition, a new tool was developed (the Template Tagger) for ‘patching’ the corpus in such a way as to eliminate further sets of errors by rule. This tool was developed by Michael Pacey, building on a prototype written by Steven Fligelstone. The research team working on tagging improvement was Nicholas Smith (lead researcher), Martin Wynne and Paul Baker.

Correction and validation of the bibliographic and contextual information in all the BNC Headers was carried out at OUCS by Lou Burnard, with assistance at various stages from Andrew Hardie and Paul Groves, who helped check demographic details for all spoken texts, and in particular from David Lee, who checked bibliographic and classification information for the bulk of the written texts. Thanks are also due to the many users of the original version of the BNC who took the time to notify us of errors they found.

Thanks are also due to Sebastian Rahtz for his help in the production of this manual.

2. Design of the corpus

This section discusses some of the basic design issues underlying the creation of the BNC. It summarizes the kinds of uses for which the corpus is intended, and the principles upon which it was created. Some summary information about the composition of the corpus is also included.

2.1. Purpose

The uses originally envisaged for the British National Corpus were set out in a working document called Planned Uses of the British National Corpus BNCW02 (11 April 91). This document identified the following as likely application areas for the corpus:

- reference book publishing
- academic linguistic research
- language teaching
- artificial intelligence
- natural language processing
- speech processing
- information retrieval

The same document identified the following categories of linguistic information derivable from the corpus:

- lexical
- semantic/pragmatic
- syntactic
- morphological
- graphological/written form/orthographical

2.2. General definitions

The British National Corpus is:

- a **sample** corpus: composed of text samples generally no longer than 45,000 words.
- a **synchronic** corpus: the corpus includes imaginative texts from 1960, informative texts from 1975.
• a **general** corpus: not specifically restricted to any particular subject field, register or genre.
• a **monolingual** British English corpus: it comprises text samples which are substantially the product of speakers of British English.
• a **mixed** corpus: it contains examples of both spoken and written language.

### 2.3. Composition

There is a broad consensus among the participants in the project and among corpus linguists that a general-purpose corpus of the English language would ideally contain a high proportion of spoken language in relation to written texts. However, it is significantly more expensive to record and transcribe natural speech than to acquire written text in computer-readable form. Consequently the spoken component of the BNC constitutes approximately 10 per cent (10 million words) of the total and the written component 90 per cent (90 million words). These were agreed to be realistic targets, given the constraints of time and budget, yet large enough to yield valuable empirical statistical data about spoken English. In the BNC sampler, a two per cent sample taken from the whole of the BNC, spoken and written language are present in approximately equal proportions, but other criteria are not equally balanced.

From the start, a decision was taken to select material for inclusion in the corpus according to an overt methodology, with specific target quantities of clearly defined types of language. This approach makes it possible for other researchers and corpus compilers to review, emulate or adapt concrete design goals. This section outlines these design considerations, and reports on the final make-up of the BNC.

This and the other tables in this section show the actual make-up of the second version of the British National Corpus (the BNC World Edition) in terms of

- texts: number of distinct samples not exceeding 45,000 words
- S-units: number of `<s>` elements identified by the CLAWS system (more or less equivalent to sentences)
- W-units: number of `<w>` elements identified by the CLAWS system (more or less equivalent to words)

For further explanation of `<s>` and `<w>` elements, see section 5.4 (Segments and words).

The BNC World Edition contains 4054 texts and occupies (including SGML markup) 1,508,392 Kbytes, or about 1.5 Gb. In total, it comprises just over 100 million orthographic words (specifically, 100,467,090), but the number of w-units (POS-tagged items) is slightly less: 97,619,934. The total number of s-units identified by CLAWS is just over 6 million (6,053,093). Counts for these and all the other elements tagged in the corpus are provided below in 10.1 (Elements defined by the BNC DTD)

In the following tables both an absolute count and a percentage are given for all the counts. The percentage is calculated with reference to the relevant portion of the corpus, for example, in the table for “written text domain”, with reference to the total number of written texts. These reference totals are given in the first table below.

<table>
<thead>
<tr>
<th>Text type</th>
<th>Texts</th>
<th>Kbytes</th>
<th>W-units</th>
<th>S-units</th>
<th>percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spoken demographic</td>
<td>153</td>
<td>4206058</td>
<td>4.30</td>
<td>610563</td>
<td>10.08</td>
</tr>
<tr>
<td>Spoken context-governed</td>
<td>757</td>
<td>6135671</td>
<td>6.28</td>
<td>428558</td>
<td>7.07</td>
</tr>
<tr>
<td>All Spoken</td>
<td>910</td>
<td>10341729</td>
<td>10.58</td>
<td>1039121</td>
<td>17.78</td>
</tr>
<tr>
<td>Written books and periodicals</td>
<td>2688</td>
<td>78580018</td>
<td>80.49</td>
<td>4403803</td>
<td>72.75</td>
</tr>
</tbody>
</table>
3.1. Sampling basis: production and reception

All texts are also classified according to their date of production. For spoken texts, the date was that of the recording. For written texts, the date used for classification was the date of production of the material actually transcribed, for the most part; in the case of imaginative works, however, the date of first publication was used. Informative texts were selected only from 1975 onwards, imaginative ones from 1960, reflecting their longer “shelf-life”, though most (75 per cent) of the latter were published no earlier than 1975.

Table 2. Date of production

<table>
<thead>
<tr>
<th>Creation date</th>
<th>texts</th>
<th>w-units</th>
<th>%</th>
<th>s-units</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>162</td>
<td>1814051</td>
<td>1.85</td>
<td>127132</td>
<td>2.10</td>
</tr>
<tr>
<td>Before 1974</td>
<td>47</td>
<td>1741624</td>
<td>1.78</td>
<td>121323</td>
<td>2.00</td>
</tr>
<tr>
<td>1974 to 1983</td>
<td>156</td>
<td>4621950</td>
<td>4.73</td>
<td>255057</td>
<td>4.21</td>
</tr>
<tr>
<td>1984 to 1994</td>
<td>3689</td>
<td>89442309</td>
<td>91.62</td>
<td>5549581</td>
<td>91.68</td>
</tr>
</tbody>
</table>

Spoken and written components of the corpus are discussed separately in the next two sections.

3. Design of the written component

3.1. Sampling basis: production and reception

While it is sometimes useful to distinguish in theory between language which is received (read and heard) and that which is produced (written and spoken), it was agreed that the selection of samples for a general-purpose corpus must take account of both perspectives.

Text that is published in the form of books, magazines, etc., is not representative of the totality of written language that is produced, as writing for publication is a comparatively specialized activity in which few people engage. However, it is much more representative of written language that is received, and is also easier to obtain in useful quantities, and thus forms the greater part of the written component of the corpus.

There was no single source of information about published material that could provide a satisfactory basis for a sampling frame, but a combination of various sources furnished useful information about the totality of written text produced and, particularly, received, some sources being more significant than others. They are principally statistics about books and periodicals that are published, bought or borrowed.

Catalogues of books published per annum tell us something about production but little about reception as many books are published but hardly read.

A list of books in print provides somewhat more information about reception as time will weed out the books that nobody bought (or read): such a list will contain a higher proportion of books that have continued to find a readership.

The books that have the widest reception are presumably those that figure in bestseller lists, particularly prize winners of competitions such as the Booker or Whitbread. Such works were certainly candidates for inclusion in the corpus, but the statistics of book-buying are such that very few texts achieve high sales while a vast number sell only a few or in modest numbers. If texts had been selected in strict arithmetical proportion to their sales, their range would have
been severely limited. However, where a text from one particular subject domain was required, it was appropriate to prefer a book which had achieved high sales to one which had not.

Library lending statistics, where these are available, also indicate which books enjoy a wide reception and, like lists of books in print, show which books continue to be read.

Similar observations hold for magazines and periodicals. Lists of current magazines and periodicals are similar to catalogues of published books, but perhaps more informative about language reception, as it may be that periodicals are bought and read by a wider cross-section of the community than books. Also, a periodical that fails to find a readership will not continue to be published for long.

Periodical circulation figures have to be treated with the same caution as bestseller lists, as a few titles dominate the market with a very high circulation. To concentrate too exclusively on these would reduce the range of text types in the corpus and make contrastive analysis difficult.

Published written texts were selected partly at random from Whitaker’s Books in Print for 1992 and partly systematically, according to the selection features outlined in section 3.2 (Selection features) below.

Available sources are concerned almost exclusively with published books and periodicals. It is much more difficult to obtain data concerning the production or reception of unpublished writing. Intuitive estimates were therefore made in order to establish some guidelines for text sampling in the latter area.

3.2. Selection features

Texts were chosen for inclusion according to three selection features: domain (subject field), time (within certain dates) and medium (book, periodical, etc.).

The purpose of these selection features was to ensure that the corpus contained a broad range of different language styles, for two reasons. The first was so that the corpus could be regarded as a microcosm of current British English in its entirety, not just of particular types. The second was so that different types of text could be compared and contrasted with each other.

3.2.1. Selection Procedure

Each selection feature was divided into classes (e.g. “Medium” into books, periodicals, unpublished etc.; “Domain” into imaginative, informative, etc.) and target percentages were set for each class. These percentages are quite independent of each other: there was no attempt, for example, to make 25 per cent of the selected periodicals imaginative.

Seventy-five per cent of the samples were to be drawn from informative texts, and the remaining 25 per cent from imaginative texts.

Titles were to be taken from a variety of media, in the following proportions: 60 per cent from books, 30 per cent from periodicals, 10 per cent from miscellaneous sources (published, unpublished, and written to be spoken).

Half of the books in the “Books and Periodicals” class were selected at random from Whitaker’s Books in Print 1992. This was to provide a control group to validate the categories used in the other method of selection: the random selection disregarded Domain and Time, but texts selected by this method were classified according to these other features after selection.

3.2.2. Sample size and method

For books, a target sample size of 40,000 words was chosen. No extract included in the corpus exceeds 45,000 words. For the most part, texts which in their entirety were shorter than 40,000 words were further reduced by ten per cent for copyright reasons; a few texts longer than the target size were however included in their entirety. Text samples normally consist of a continuous stretch of discourse from within the whole. A convenient breakpoint (e.g. the end of a section or chapter) was chosen as far as possible to begin and end the sample so that
high-level discourse units were not fragmented. Only one sample was taken from any one text. Samples were taken randomly from the beginning, middle or end of longer texts. (In a few cases, where a publication included essays or articles by a variety of authors of different nationalities, the work of non-UK authors was omitted.)

Some types of written material are composite in structure: that is, the physical object in written form is composed of more than one text unit. Important examples are issues of a newspaper or magazine which, though editorially shaped as a document, contain discrete texts, each with its specific authorship, stylistic characteristics, register and domain. The BNC attempts to separate these discrete texts where appropriate and to classify them individually according to the selection and classification features. As far as possible, the individual stories in one issue of a newspaper were grouped according to domain, for example as “Business” articles, “Leisure” articles, etc.

The following subsections discuss each selection criterion, and indicate the actual numbers of words in each category included.

### 3.2.3. Domain

Classification according to subject field seems hardly appropriate to texts which are fictional or which are generally perceived to be literary or creative. Consequently, these texts are all labelled imaginative and are not assigned to particular subject areas. All other texts are treated as informative and are assigned to one of the eight domains listed below.

<table>
<thead>
<tr>
<th>Domain</th>
<th>texts</th>
<th>w-units</th>
<th>%</th>
<th>s-units</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applied science</td>
<td>370</td>
<td>7104635</td>
<td>8.14</td>
<td>357067</td>
<td>7.12</td>
</tr>
<tr>
<td>Arts</td>
<td>261</td>
<td>6520634</td>
<td>7.47</td>
<td>321442</td>
<td>6.41</td>
</tr>
<tr>
<td>Belief and thought</td>
<td>146</td>
<td>3007244</td>
<td>3.44</td>
<td>151418</td>
<td>3.01</td>
</tr>
<tr>
<td>Commerce and finance</td>
<td>295</td>
<td>7257542</td>
<td>8.31</td>
<td>382717</td>
<td>7.63</td>
</tr>
<tr>
<td>Imaginative</td>
<td>477</td>
<td>1637726</td>
<td>18.76</td>
<td>1356458</td>
<td>27.05</td>
</tr>
<tr>
<td>Leisure</td>
<td>438</td>
<td>12187946</td>
<td>13.96</td>
<td>760722</td>
<td>15.17</td>
</tr>
<tr>
<td>Natural and pure science</td>
<td>146</td>
<td>3784273</td>
<td>4.33</td>
<td>183466</td>
<td>3.65</td>
</tr>
<tr>
<td>Social science</td>
<td>527</td>
<td>13906182</td>
<td>15.93</td>
<td>700122</td>
<td>13.96</td>
</tr>
<tr>
<td>World affairs</td>
<td>484</td>
<td>17132023</td>
<td>19.62</td>
<td>800560</td>
<td>15.96</td>
</tr>
</tbody>
</table>

The evidence from catalogues of books and periodicals suggests that imaginative texts account for significantly less than 25 per cent of published output, and unpublished reports, correspondence, reference works and so on would seem to add further to the bulk of informative text which is produced and consumed. However, the overall distribution between informative and imaginative text samples is set to reflect the influential cultural role of literature and creative writing. The target percentages for the eight informative domains were arrived at by consensus within the project, based loosely upon the pattern of book publishing in the UK during the past 20 years or so, as reflected in the categorized figures for new publications that appear annually in Whitaker’s Book list.

### 3.2.4. Medium

This categorisation is broad, since a detailed taxonomy or feature classification of text medium could have led to such a proliferation of subcategories as to make it impossible for the BNC adequately to represent all of them. The labels used here are intended to be comprehensive in the sense that any text can be assigned with reasonable confidence to these macro categories. The labels we have adopted represent the highest levels of a fuller taxonomy of text medium.
3.3.1. Author information

Table 4. Written medium

<table>
<thead>
<tr>
<th>Medium</th>
<th>texts</th>
<th>w-units</th>
<th>%</th>
<th>s-units</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Book</td>
<td>1414</td>
<td>49891770</td>
<td>57.16</td>
<td>2895652</td>
<td>57.75</td>
</tr>
<tr>
<td>Periodical</td>
<td>1208</td>
<td>28356005</td>
<td>32.48</td>
<td>1487725</td>
<td>29.67</td>
</tr>
<tr>
<td>Published miscellanea</td>
<td>238</td>
<td>4197450</td>
<td>4.80</td>
<td>288004</td>
<td>5.74</td>
</tr>
<tr>
<td>Unpublished miscellanea</td>
<td>249</td>
<td>3508500</td>
<td>4.01</td>
<td>222438</td>
<td>4.43</td>
</tr>
<tr>
<td>To-be-spoken</td>
<td>35</td>
<td>1324480</td>
<td>1.51</td>
<td>120153</td>
<td>2.39</td>
</tr>
</tbody>
</table>

The ‘Miscellaneous published’ category includes brochures, leaflets, manuals, advertisements. The ‘Miscellaneous unpublished’ category includes letters, memos, reports, minutes, essays. The ‘written-to-be-spoken’ category includes scripted television material, play scripts etc.

3.3. Descriptive features

Written texts may be further classified according to sets of descriptive features. These features describe the sample texts; they did not determine their selection. This information is recorded to allow more delicate contrastive analysis of particular sets of texts. As a simple example, the gross division into two time periods in the selection features can, of course, be refined and subcorpora defined over the BNC for more specific dates. However, the relative sizes of such subcorpora are undefined by the BNC design specification.

These descriptive features were monitored during the course of the data gathering, and text selection, in cases where a free choice of texts was available, took account of the relative balance of these features. Thus although no relative proportions were defined for different target age groups (for example), we ensured that the corpus does contain texts intended for children as well as for adults.

The following tables summarize the results for the first release of the corpus. Note that many texts remain unclassified.

3.3.1. Author information

Information about authors of written texts was included only where it was readily available, for example from the dust-wrapper of a book. Consequently, the coverage of such information is very patchy. The authorship of a written text was characterized as ‘corporate’ where it was produced by an organization and no specific author was given, and as ‘multiple’ in cases where several authors were named. Author sex was classified as “mixed” where more than one author of either sex was specified, and “unknown” where it could not reliably be determined from the author’s name. Note that ‘author age’ means the author’s age at the time of creation of the work concerned.

Table 5. Type of author

<table>
<thead>
<tr>
<th>Author type</th>
<th>texts</th>
<th>w-units</th>
<th>%</th>
<th>s-units</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>211</td>
<td>3750668</td>
<td>4.29</td>
<td>175027</td>
<td>3.49</td>
</tr>
<tr>
<td>Corporate</td>
<td>347</td>
<td>6497415</td>
<td>7.44</td>
<td>471152</td>
<td>9.39</td>
</tr>
<tr>
<td>Multiple</td>
<td>1323</td>
<td>34284025</td>
<td>39.28</td>
<td>1813636</td>
<td>36.17</td>
</tr>
<tr>
<td>Sole</td>
<td>1263</td>
<td>42746097</td>
<td>48.97</td>
<td>2554157</td>
<td>50.94</td>
</tr>
</tbody>
</table>

Table 6. Author sex
3.3.2. Target audience

Some attempt was made to characterize the kind of audience for which written texts were produced in terms of age, sex and ‘level’ (a subjective assessment of the text’s technicality or difficulty). The last of these proved very difficult to assess and was very frequently confused with circulation size or audience size; for that reason, no figures for it are included here.

### Table 7. Author age group

<table>
<thead>
<tr>
<th>Author age</th>
<th>texts</th>
<th>w-units</th>
<th>%</th>
<th>s-units</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>2519</td>
<td>65,457,159</td>
<td>74.99</td>
<td>37,076,000</td>
<td>73.94</td>
</tr>
<tr>
<td>0-14</td>
<td>3</td>
<td>5,907</td>
<td>0.06</td>
<td>3,447</td>
<td>0.06</td>
</tr>
<tr>
<td>15-24</td>
<td>19</td>
<td>53,725</td>
<td>0.61</td>
<td>29,862</td>
<td>0.59</td>
</tr>
<tr>
<td>25-34</td>
<td>67</td>
<td>228,693</td>
<td>2.62</td>
<td>163,079</td>
<td>3.25</td>
</tr>
<tr>
<td>35-44</td>
<td>191</td>
<td>666,060</td>
<td>7.63</td>
<td>410,324</td>
<td>8.18</td>
</tr>
<tr>
<td>45-59</td>
<td>205</td>
<td>715,798</td>
<td>8.20</td>
<td>410,717</td>
<td>8.19</td>
</tr>
<tr>
<td>60+</td>
<td>140</td>
<td>511,197</td>
<td>5.86</td>
<td>288,943</td>
<td>5.76</td>
</tr>
</tbody>
</table>

### Table 8. Author domicile

<table>
<thead>
<tr>
<th>Author domicile</th>
<th>texts</th>
<th>w-units</th>
<th>%</th>
<th>s-units</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>2273</td>
<td>56,750,777</td>
<td>65.02</td>
<td>314,457,800</td>
<td>62.71</td>
</tr>
<tr>
<td>UK and Ireland</td>
<td>843</td>
<td>295,700,976</td>
<td>33.88</td>
<td>181,255,000</td>
<td>36.14</td>
</tr>
<tr>
<td>Commonwealth</td>
<td>12</td>
<td>407,076</td>
<td>0.46</td>
<td>257,627</td>
<td>0.51</td>
</tr>
<tr>
<td>Continental Europe</td>
<td>6</td>
<td>232,275</td>
<td>0.26</td>
<td>124,691</td>
<td>0.24</td>
</tr>
<tr>
<td>USA</td>
<td>8</td>
<td>243,177</td>
<td>0.27</td>
<td>156,770</td>
<td>0.31</td>
</tr>
<tr>
<td>Elsewhere</td>
<td>2</td>
<td>74,803</td>
<td>0.08</td>
<td>29,361</td>
<td>0.05</td>
</tr>
</tbody>
</table>

3.3.2. Target audience

Some attempt was made to characterize the kind of audience for which written texts were produced in terms of age, sex and ‘level’ (a subjective assessment of the text’s technicality or difficulty). The last of these proved very difficult to assess and was very frequently confused with circulation size or audience size; for that reason, no figures for it are included here.

### Table 9. Target age group

<table>
<thead>
<tr>
<th>age group</th>
<th>texts</th>
<th>w-units</th>
<th>%</th>
<th>s-units</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child</td>
<td>42</td>
<td>89,5413</td>
<td>1.02</td>
<td>81,085</td>
<td>1.61</td>
</tr>
<tr>
<td>Teenager</td>
<td>77</td>
<td>176,9940</td>
<td>2.02</td>
<td>135,583</td>
<td>2.70</td>
</tr>
<tr>
<td>Adult</td>
<td>2915</td>
<td>813,45838</td>
<td>93.20</td>
<td>462,5633</td>
<td>92.25</td>
</tr>
<tr>
<td>Any</td>
<td>110</td>
<td>326,7014</td>
<td>3.74</td>
<td>171,671</td>
<td>3.42</td>
</tr>
</tbody>
</table>

### Table 10. Target sex

<table>
<thead>
<tr>
<th>sex</th>
<th>texts</th>
<th>w-units</th>
<th>%</th>
<th>s-units</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>707</td>
<td>20,113,523</td>
<td>23.04</td>
<td>113,5038</td>
<td>22.63</td>
</tr>
</tbody>
</table>
3.4.1. Bestsellers

<table>
<thead>
<tr>
<th>Gender</th>
<th>Texts</th>
<th>W-units</th>
<th>%</th>
<th>S-units</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>61</td>
<td>2366396</td>
<td>2.71</td>
<td>136564</td>
<td>2.72</td>
</tr>
<tr>
<td>Female</td>
<td>176</td>
<td>6882659</td>
<td>7.88</td>
<td>507713</td>
<td>10.12</td>
</tr>
<tr>
<td>Mixed</td>
<td>2200</td>
<td>57915627</td>
<td>66.35</td>
<td>3234657</td>
<td>64.51</td>
</tr>
</tbody>
</table>

3.3.3. Miscellaneous classification information

Written texts were also characterized according to their place of publication and the type of sampling used.

<table>
<thead>
<tr>
<th>Region</th>
<th>Texts</th>
<th>W-units</th>
<th>%</th>
<th>S-units</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>690</td>
<td>14583761</td>
<td>16.70</td>
<td>790465</td>
<td>15.76</td>
</tr>
<tr>
<td>UK (unspecific)</td>
<td>264</td>
<td>7124424</td>
<td>8.16</td>
<td>383046</td>
<td>7.63</td>
</tr>
<tr>
<td>Ireland</td>
<td>37</td>
<td>567046</td>
<td>0.64</td>
<td>31825</td>
<td>0.63</td>
</tr>
<tr>
<td>UK (North)</td>
<td>192</td>
<td>3778114</td>
<td>4.32</td>
<td>230008</td>
<td>4.58</td>
</tr>
<tr>
<td>UK (Midlands)</td>
<td>93</td>
<td>2622554</td>
<td>3.00</td>
<td>192379</td>
<td>3.83</td>
</tr>
<tr>
<td>UK (South)</td>
<td>1854</td>
<td>58066891</td>
<td>66.53</td>
<td>3365045</td>
<td>67.11</td>
</tr>
<tr>
<td>United States</td>
<td>14</td>
<td>535415</td>
<td>0.61</td>
<td>21204</td>
<td>0.42</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample type</th>
<th>Texts</th>
<th>W-units</th>
<th>%</th>
<th>S-units</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>1583</td>
<td>35240809</td>
<td>40.37</td>
<td>1994357</td>
<td>39.77</td>
</tr>
<tr>
<td>Whole text</td>
<td>270</td>
<td>6463415</td>
<td>7.40</td>
<td>433833</td>
<td>8.65</td>
</tr>
<tr>
<td>Beginning sample</td>
<td>585</td>
<td>20890666</td>
<td>23.93</td>
<td>1121658</td>
<td>22.37</td>
</tr>
<tr>
<td>Middle sample</td>
<td>512</td>
<td>18344188</td>
<td>21.01</td>
<td>1055383</td>
<td>21.04</td>
</tr>
<tr>
<td>End sample</td>
<td>119</td>
<td>4271138</td>
<td>4.89</td>
<td>253413</td>
<td>5.05</td>
</tr>
<tr>
<td>Composite</td>
<td>75</td>
<td>2067989</td>
<td>2.36</td>
<td>155328</td>
<td>3.09</td>
</tr>
</tbody>
</table>

In addition to the above, standard bibliographic details such as author, title, publication details, extent, topic keywords etc. were recorded for the majority of texts, as further described below (see 8 (The header)).

3.4. Selection procedures employed

3.4.1. Books

Roughly half the titles were randomly selected from available candidates identified in Whitaker’s *Books in Print* (BIP), 1992, by students of Library and Information Studies at Leeds City University. Each text randomly chosen was accepted only if it fulfilled certain criteria: it had to be published by a British publisher, contain sufficient pages of text to make its incorporation worthwhile, consist mainly of written text, fall within the designated time limits, and cost less than a set price. The students noted the ISBN, author, title and price of each book thus selected; the final selection weeded out texts by non-UK authors.

Half of the books having been selected by this method, the remaining half were selected systematically to make up the target percentages in each category. The selection proceeded as follows.

3.4.1.1. Bestsellers

Because of their wide reception, bestsellers were obvious candidates for selection. The lists used were those that appeared in the *Bookseller* at the end of the years 1987 to 1993 inclusive.
Some of the books in the lists were rejected, for a variety of reasons. Obviously books that had already been selected by the random method were excluded, as were those by non-UK authors. In addition, a limit of 120,000 words from any one author was imposed, and books belonging to a domain or category whose quota had already been reached were not selected. Other bestseller lists were obtained from *The Guardian*, the British Council, and from Blackwells Paperback Shop.

The titles yielded by this search were mostly in the Imaginative category.

3.4.1.2. Literary prizes

The criteria for inclusion were the same as for bestsellers. The prize winners, together with runners-up and shortlisted titles, were taken from several sources, principally Anne Strachan, *Prizewinning literature: UK literary award winners*, London, 1989. For 1990 onwards the sources used were: the last issue of the *Bookseller* for each year; *The Guardian Index, 1989-*; entries under the term “Literature”; and *The Times Index, 1989-*; entries under the term “Literature — Awards”.

Literary prizes are in the main awarded to works that fall into the Imaginative category, but there are some Informative ones also.

3.4.1.3. Library loans

The source of statistics in this category was the record of loans under Public Lending Right, kindly provided by Dr J. Parker, the Registrar. The information comprised lists of the hundred most issued books and the hundred most issued children’s books, in both cases for the years 1987 to 1993.

The lists consist almost exclusively of imaginative literature, and many titles found there also appear in the lists of bestsellers and prize winners.

3.4.1.4. Additional texts

As collection proceeded, monitoring disclosed potential shortfalls in certain domains. A further selection was therefore made, based on the “Short Loan” collections of seven University libraries. (Short Loan collections typically contain books required for academic courses, which are consequently in heavy demand.)

3.4.2. Periodicals and magazines

Periodicals, magazines and newspapers account for 30 per cent of the total text in the corpus. Of these, about 250 titles were issues of newspapers. These were selected to cover as wide a spectrum of interests and language as possible. Newspapers were selected to represent as wide a geographic spread as possible: *The Scotsman* and the *Belfast Telegraph* are both represented, for example.

3.4.3. Other media

In addition to samples from books, periodicals, and magazines, the written part of the corpus contains about seven million words classified as “Miscellaneous Published”, “Miscellaneous Unpublished”, or as “Written to be spoken”. The distinction between “published” and “unpublished” is not an easy one; the former category largely contains publicity leaflets, brochures, fact sheets, and similar items, while the latter has a substantial proportion of school and university essays, unpublished creative writing or letters, and internal company memoranda. The “written to be spoken” material includes scripted material, intended to be read aloud such as television news broadcasts; transcripts of more informal broadcast materials such as discussions or phone-ins are included in the spoken part of the corpus.

3.4.4. Copyright permissions

Before a selected text could be included, permissions had to be obtained from the copyright owner (publisher, agent, or author). A standard Permissions Request was drafted with considerable care, but some requests were refused, or simply not answered even after prompting, so that the texts concerned had to be excluded or replaced.
4. Design of the spoken component

Lexicographers and linguists have long hoped for corpus evidence about spoken language, but the practical difficulties of transcribing sufficiently large quantities of text have prevented the construction of a spoken corpus of over one million words. The British National Corpus project undertook to produce five to ten million words of orthographically transcribed speech, covering a wide range of speech variation. A large proportion of the spoken part of the corpus — over four million words — comprises spontaneous conversational English. The importance of conversational dialogue to linguistic study is unquestionable: it is the dominant component of general language both in terms of language reception and language production.

As with the written part of the corpus, the most important considerations in constructing the spoken part were sampling and representativeness. The method of transcription was also an important issue.

The issues of corpus sampling and representativeness have been discussed at great length by many corpus linguists. With spoken language there are no obvious objective measures that can be used to define the target population or construct a sampling frame. A comprehensive list of text types can be drawn up but there is no accurate way of estimating the relative proportions of each text type other than by a priori linguistically motivated analysis. An alternative approach, one well known to sociological researchers, is demographic sampling, and this was broadly the approach adopted for approximately half of the spoken part of the corpus. The sampling frame was defined in terms of the language production of the population of British English speakers in the United Kingdom. Representativeness was achieved by sampling a spread of language producers in terms of age, gender, social group, and region, and recording their language output over a set period of time.

We recognised, however, that many types of spoken text are produced only rarely in comparison with the total output of all “speech producers”: for example, broadcast interviews, lectures, legal proceedings, and other texts produced in situations where — broadly speaking — there are few producers and many receivers. A corpus constituted solely on the demographic model would thus omit important spoken text types. Consequently, the demographic component of the corpus was complemented with a separate text typology intended to cover the full range of linguistic variation found in spoken language; this is termed the context-governed part of the corpus.

4.1. The demographically sampled part of the corpus

The approach adopted uses demographic parameters to sample the population of British English speakers in the United Kingdom. Established random location sampling procedures were used to select individual members of the population by personal interview from across the country taking into account age, gender, and social group. Selected individuals used a portable tape recorder to record their own speech and the speech of people they conversed with over a period of up to a week. In this way a unique record of the language people use in everyday conversation was constructed.

4.1.1. Sampling procedure

124 adults (aged 15+) were recruited from across the United Kingdom. Recruits were of both sexes and from all age groups and social classes. The intention was, as far as possible, to recruit equal numbers of men and women, equal numbers from each of the six age groups, and equal numbers from each of four social classes.

Additional recordings were gathered for the BNC as part of the University of Bergen COLT Teenager Language Project. This project used the same recording methods and transcription scheme as the BNC, but selected only respondents aged 16 or below.
The tables below give figures for the amount of transcribed material collected by each respondent, classified by their age, class, and sex.

**Table 13. Age group of demographic respondent**

<table>
<thead>
<tr>
<th>Age group</th>
<th>texts</th>
<th>w-units</th>
<th>%</th>
<th>s-units</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-14</td>
<td>26</td>
<td>265382</td>
<td>6.30</td>
<td>41035</td>
<td>6.72</td>
</tr>
<tr>
<td>15-24</td>
<td>36</td>
<td>660847</td>
<td>15.71</td>
<td>97994</td>
<td>16.04</td>
</tr>
<tr>
<td>25-34</td>
<td>29</td>
<td>848162</td>
<td>20.16</td>
<td>121750</td>
<td>19.94</td>
</tr>
<tr>
<td>35-44</td>
<td>22</td>
<td>839622</td>
<td>19.96</td>
<td>126688</td>
<td>20.74</td>
</tr>
<tr>
<td>45-59</td>
<td>20</td>
<td>957382</td>
<td>22.76</td>
<td>136540</td>
<td>22.36</td>
</tr>
<tr>
<td>60+</td>
<td>20</td>
<td>634663</td>
<td>15.08</td>
<td>86556</td>
<td>14.17</td>
</tr>
</tbody>
</table>

**Table 14. Social class of demographic respondent**

<table>
<thead>
<tr>
<th>Social class</th>
<th>texts</th>
<th>w-units</th>
<th>%</th>
<th>s-units</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>7</td>
<td>37363</td>
<td>0.88</td>
<td>5339</td>
<td>0.87</td>
</tr>
<tr>
<td>AB</td>
<td>59</td>
<td>1363571</td>
<td>32.41</td>
<td>197804</td>
<td>32.39</td>
</tr>
<tr>
<td>C1</td>
<td>36</td>
<td>1097023</td>
<td>26.08</td>
<td>169384</td>
<td>27.74</td>
</tr>
<tr>
<td>C2</td>
<td>31</td>
<td>1080654</td>
<td>25.69</td>
<td>144877</td>
<td>23.72</td>
</tr>
<tr>
<td>DE</td>
<td>20</td>
<td>627447</td>
<td>14.91</td>
<td>93159</td>
<td>15.25</td>
</tr>
</tbody>
</table>

**Table 15. Sex of demographic respondent**

<table>
<thead>
<tr>
<th>Sex</th>
<th>texts</th>
<th>w-units</th>
<th>%</th>
<th>s-units</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>5</td>
<td>16151</td>
<td>0.38</td>
<td>2407</td>
<td>0.39</td>
</tr>
<tr>
<td>Male</td>
<td>73</td>
<td>1730592</td>
<td>41.14</td>
<td>248247</td>
<td>40.65</td>
</tr>
<tr>
<td>Female</td>
<td>75</td>
<td>2459315</td>
<td>58.47</td>
<td>359909</td>
<td>58.94</td>
</tr>
</tbody>
</table>

Recruits who agreed to take part in the project were asked to record all of their conversations over a two to seven day period. The number of days varied depending on how many conversations each recruit was involved in and was prepared to record. Results indicated that most people recorded nearly all of their conversations, and that the limiting factor was usually the number of conversations a person had per day. The placement day was varied, and recruits were asked to record on the day after placement and on any other day or days of the week. In this way a broad spread of days of the week including weekdays and weekends was achieved. A conversation log allowed recruits to enter details of every conversation recorded, and included date, time and setting, and brief details of other participants.

**4.1.2. Recording procedure**

All conversations were recorded as unobtrusively as possible, so that the material gathered approximated closely to natural, spontaneous speech. In many cases the only person aware that the conversation was being taped was the person carrying the recorder. Although an initial unnaturalness on the part of the recruit was not uncommon this soon seemed to disappear. Similarly, where non-intrusive recording was not possible, for example at a family gathering where everyone is aware they are being recorded, the same initial period of unease sometimes occurred, but in our experience again vanished quickly. The guarantee of confidentiality and complete anonymity (all references to full names and addresses have been removed from the
corpus and the log), and the fact that there was an intermediary between those being recorded and those listening to the recordings certainly helped.

For each conversational exchange the person carrying the recorder told all participants they had been recorded and explained why. Whenever possible this happened after the conversation had taken place. If any participant was unhappy about being recorded the recording was erased. During the project around 700 hours of recordings were gathered.

4.1.3. Sample size

The number of people recruited may seem small in comparison to some demographic studies of the population of the United Kingdom. As with any sampling method, some compromise between what was theoretically desirable and what was feasible within the constraints of the BNC project had to be made. There is no doubt that recruiting 1000 people would have given greater statistical validity but the practical difficulties and cost implications of recruiting 1000 people and transcribing 50–100 million words of speech made this impossible. given that we were not attempting to represent the complete range of age and social groups within each region we considered a sample size between 100 and 130 would be adequate. It is also important to stress that the total number of participants in all conversations was well in excess of a thousand.

4.1.4. Piloting the demographic sampling approach

Because this approach to spoken corpus sampling had to our knowledge never previously been attempted a detailed piloting project was carried out to investigate:

- the likelihood that enough material would be obtained from a sample of around 100 people
- any problems that might be encountered during the recruitment and collection stages
- any problems or difficulties experienced by recruits during taping or with logging details of conversations and participants
- any areas where the documentation designed for the project could be improved
- whether the recording quality under a wide range of conditions would be good enough for accurate transcription
- whether the predicted throughput rates for tape editing, transcription and checking were accurate.

The results of the pilot generally confirmed predictions and allowed some procedures to be refined for the full project.

4.2. The context-governed part of the corpus

As mentioned above, the spoken texts in the demographic part of the corpus consists mainly of conversational English. A complementary approach was developed to create what is termed the context-governed part of the corpus. As in other spoken corpora, the range of text types was selected according to a priori linguistically motivated categories. At the top layer of the typology is a division into four equal-sized contextually based categories: educational, business, public/institutional, and leisure. Each is divided into the subcategories monologue (40 per cent) and dialogue (60 per cent). Each monologue subcategory therefore totals 10 per cent of the context-governed part of the corpus, and each dialogue subcategory 15 per cent.

Within each subcategory a range of text types was defined. This range was not fixed, and the design was flexible enough to allow the inclusion of additional text types. The sampling methodology was different for each text type but the overall aim was to achieve a balanced selection within each, taking into account such features as region, level, gender of speakers, and topic. Other features, such as purpose, were applied on the basis of post hoc judgements.
4.2.1. Sampling procedure

For the most part, a variety of text types were sampled within three geographic regions. However, some text types, such as parliamentary proceedings, and most broadcast categories, apply to the country as a whole and were not regionally sampled. Different sampling strategies were required for each text type, and these are outlined below.

4.2.1.1. Educational and informative:

**Lectures, talks, educational demonstrations**  Within each sampling area a university (or college of further education) and a school were selected. A range of lectures and talks was recorded, varying the topic, level, and speaker gender.

**News commentaries**  Regional sampling was not applied, but both national and regional broadcasting companies were sampled. The topic, level, and gender of commentator was varied.

**Classroom interaction**  Schools were regionally sampled and the level (generally based on student age) and topic were varied. Home tutorials were also included.

4.2.1.2. Business:

**Company talks and interviews**  Sampling took into account company size, areas of activity, and gender of speakers.

**Trade union talks**  Talks to union members, branch meetings and annual conferences were all sampled.

**Sales demonstrations**  A range of topics was included.

**Business meetings**  Companies were selected according to size, area of activity, and purpose of meeting.

**Consultations**  These included medical, legal, business and professional consultations.

All categories under this heading were regionally sampled.

4.2.1.3. Public/ or institutional:

**Political speeches**  Regional sampling of local politics, plus speeches in both the House of Commons and the House of Lords.

**Sermons**  Different denominations were sampled.

**Public/government talks**  Regional sampling of local inquiries and meetings, plus national issues at different levels.

**Council meetings**  Regionally sampled, covering parish, town, district, and county councils.

**Religious meetings**  Includes church meetings, group discussions, and so on.

**Parliamentary proceedings**  Sampling of main sessions and committees, House of Commons and House of Lords.

**Legal proceedings**  Royal Courts of Justice, and local Magistrates and similar courts were sampled.

4.2.1.4. Leisure:

**Speeches**  Regionally sampled, covering a variety of occasions and speakers.

**Sports commentaries**  Exclusively broadcast, sampling a variety of sports, commentators, and TV/radio channels.

**Talks to clubs**  Regionally sampled, covering a range of topics and speakers.

**Broadcast chat shows and phone-ins**  Only those that include a significant amount of unscripted speech were selected from both television and radio.

**Club meetings**  Regionally sampled, covering a wide range of clubs.
4.2.2. Sample size

Each monologue text type contains up to 200,000 words of text, and each dialogue text type up to 300,000 words. The length of text units within each text type vary — for example, news commentaries may be only a few minutes long (several hundred words), lectures are typically up to one hour (10,000 words), and some business meetings and parliamentary proceedings may last for several hours (20,000 words+). For the context-governed part of the corpus an upper limit of 10,000 words per text unit was generally imposed, although a few texts are slightly above this.

4.2.3. Composition of the spoken component

A total of 757 texts (6,153,671 words) make up the context-governed part of the corpus. The following contexts are distinguished:

| Table 16. Context in which spoken text was captured |
|--------------------------------------|--------|--------|--------|--------|
| Context                           | texts  | w-units | %     | s-units | %     |
| Educational/Informative           | 169    | 1633303 | 26.61 | 119252 | 27.82 |
| Business                          | 131    | 1285938 | 20.95 | 108101 | 25.22 |
| Public/Institutional              | 262    | 1655263 | 26.97 | 96504  | 22.51 |
| Leisure                           | 195    | 1561167 | 25.44 | 104701 | 24.43 |

In addition, the following classifications are applicable to both demographic and context-governed spoken texts:

| Table 17. Region where spoken text captured |
|---------------------------------------------|--------|--------|--------|--------|
| Region                                      | texts  | w-units | %     | s-units | %     |
| Unknown                                     | 35     | 446584  | 4.31  | 27706  | 2.66  |
| South                                       | 312    | 4658232 | 45.04 | 458253 | 44.10 |
| Midlands                                    | 213    | 2471184 | 23.89 | 240320 | 23.12 |
| North                                       | 350    | 2765729 | 26.74 | 312842 | 30.10 |

| Table 18. Interaction type for spoken text |
|--------------------------------------------|--------|--------|--------|--------|
| Interaction type                           | texts  | w-units | %     | s-units | %     |
| Monologue                                  | 212    | 1578614 | 15.26 | 94272  | 9.07  |
| Dialogue                                   | 698    | 8763115 | 84.73 | 944849 | 90.92 |

5. Basic structure

The mark-up scheme chosen for the British National Corpus is an application of ISO 8879, the Standard Generalized Mark-Up Language. This international standard provides, amongst other things, a method of specifying an application-independent document grammar, in terms of the elements which may appear in a document, their attributes, and the ways in which they may legally be combined. It is also a superset of the language XML, the extensible markup language currently proposed by the World Wide Web Consortium for general use on the World Wide Web. A brief summary of the encoding format used in the BNC to represent SGML constructs is given in section 5.1 (Markup conventions) below; more detailed information about SGML and XML is readily available in many places.
The original BNC encoding format was strongly influenced by the proposals of the Text Encoding Initiative (TEI). This international research project resulted in the development of a set of comprehensive guidelines for the encoding and interchange of a wide range of electronic texts amongst researchers. An initial report appeared in 1991, and a substantially revised and expanded version in early 1994. A conscious attempt was made to conform to TEI recommendations, where these had already been formulated, but in the first version of the BNC there were a number of differences in tag names, and models. In the present edition of the BNC, the tagging scheme has been changed to conform as far as possible with the published Recommendations of the TEI. Unless otherwise stated, elements used here have the same meaning as those of the published TEI scheme. More information about the relationship between the BNC’s markup and both its original CDIF format and the TEI standard are given in section 9 (Compatibility issues).

Section 5 (Basic structure) describes the basic structure of the British National Corpus, in terms of the SGML elements distinguished and the tags used to mark them up. Section 6 (Written texts) describes the elements which are peculiar to written texts, and section 7 (Spoken texts) those peculiar to spoken texts. In each case, a distinction is made between those elements which are marked up in all texts and those which (for technical or financial reasons) are not always so distinguished, and hence appear in some texts only.

Section 8 (The header) describes the structure of the <teiHeader> element attached to each component of the corpus, and also to the whole corpus itself. Sections 6 (Written texts) and 7 (Spoken texts) informally describe the elements specific to written and to spoken texts respectively. It should be noted that by no means all of the features described here will be present in every text of the corpus, nor, if present, will they necessarily be tagged. A list of elements actually used in the whole corpus is given below in 10.1 (Elements defined by the BNC DTD).

5.1. Markup conventions

The BNC texts use the “reference concrete syntax” of SGML, in which all elements are delimited by the use of tags. There are two forms of tag, a start-tag, marking the beginning of an element, and an end-tag marking its end. Tags are delimited by the characters < and >, and contain the name of the element (its gi, for generic identifier), preceded by a solidus (/) in the case of an end-tag.

For example, a heading or title in a written text will be preceded by a tag of the form <head> and followed by a tag in the form </head>. Everything between these two tags is regarded as the content of an element of type <head>.

Attributes applicable to element instances, if present, are also indicated within the start-tag, and take the form of an attribute name, an equal sign and the attribute value, which may be a number, a string literal or a quoted literal. Attribute values are used for a variety of purposes, notably to represent the part of speech codes allocated to particular words by the CLAWS tagging scheme.

For example, the <head> element may take an attribute type which categorizes it in some way. A main heading will thus appear with a start tag <head type="main">, and a subheading with a start tag <head type="sub">.

In XML (but not always in SGML), case is significant in all tag or attribute names. A consistent style has been adopted throughout the corpus. This style uses lower-case letters for identifiers, unless they are derived from more than one word, in which case the first letter of the second and any subsequent word is capitalized.
SGML (but not XML) permits various kinds of minimization, or abbreviatory conventions. Only two such are used: end-tag omission and attribute-name omission. These conventions apply only to the elements <s>, <w> and <c> (i.e., for sentences, words, and punctuation).

For all other non-empty elements, every occurrence in the distributed form of the corpus has both a start-tag and an end-tag, and any attributes specified are supplied in the form attribute name=value (in the body of the texts), or attribute name="value" (in the headers). For the elements <s>, <w> and <c>, and all empty elements, end-tags are routinely omitted. For these three elements only, attribute values are given without any associated attribute name. See section 5.4 (Segments and words) for some examples.

In the present release of the corpus, the headers are marked up using XML; this means that empty-tags take a slightly different form and that attribute values are always quoted.

Only a restricted range of characters is used in element content: specifically, the upper- and lower-case alphabetics, digits, and a subset of the common punctuation marks. All other characters are represented by SGML entity references, which take the form of an ampersand (&) followed by a mnemonic for the character, and terminated by a semicolon (;) where this is necessary to resolve ambiguity.

For example, the pound sign is represented by the string &pound;, the character é by the string &eacute; and so forth. The French word “été” (summer), if it appeared in the corpus, would be represented as

&eacute;te;

The mnemonics used are taken from standard entity sets, and are listed in section 10.2 (Character entities defined by the BNC DTD).

Finally, although this is not mandated by either XML or SGML, in the present form of the corpus, tags are never broken across linebreaks. Additionally, an attempt has been made to avoid linebreaks within the content of a single <s> element, so as to simplify processing of the text.

5.2. Global attributes

Three global attributes are defined, each of which may potentially be specified for any element. In practice their use is limited to certain specific functions, which are discussed at the appropriate place below, but for convenience their use is also summarized here:

- **id** system-generated identifier of an item, unique within the corpus
- **n** any name or identifier for an element, not necessarily unique within the corpus
- **rend** the rendition or appearance of an element

5.3. Corpus and text elements

The British National Corpus contains a large number of text samples, some spoken and some written. Each such sample has some associated descriptive or bibliographic information particular to it, and there is also a large body of descriptive information which applies to the whole corpus.

In SGML terms, the British National Corpus consists of a single SGML element, tagged <bnc>. This element contains a single <teiHeader> element, followed by a sequence of <bncDoc> elements. Each such <bncDoc> element contains its own <teiHeader>, followed by either a <text> element (for written texts) or an <stext> element (for spoken texts). The last named element is an extension of the TEI scheme, but the others are all standard TEI elements, possibly renamed as permitted by the TEI scheme.

The components of the header are fully documented in section 8 (The header). Further discussion of SGML concepts and practices is provided in section 11 (Software for the BNC).
Note that different elements are used for spoken and written texts because each has a different substructure; this represents a departure from TEI recommended practice.

Both `<text>` and `<stext>` elements take the following attributes in addition to the attributes globally available:

- **org** specifies how the content of the text is organised. Legal values are:
  - `composite` composite content: i.e. no claim is made about the sequence in which elements inferior to this one are to be processed, or their inter-relationships
  - `seq` sequential content: i.e. the elements contained by this one form a logical unit, to be processed in the sequence given

- **decls** supplies the identifiers of any specific encoding or editorial conventions defined in the corpus header and applicable to this specific text

The `org` attribute is used to characterize the internal organization of written texts. All demographically collected spoken texts have the same internal organization: each `<stext>` element collects together all the conversations for a given respondent, each distinct conversation being represented by a `<div>` element (see further 7.1 (Basic structure: spoken texts)). Since the order of these `<div>` elements is not significant, the `org` attribute always has the value “composite”.

### 5.4. Segments and words

At the lowest level, the corpus consists of `<w>` (word) and `<c>` (punctuation) elements, grouped into `<s>` (segment) elements:

- `<s>` a segment of spoken or written text as identified by the CLAWS segmentation scheme. The global `n` attribute is always supplied for `<s>` elements.
- `<w>` represents a grammatical (not necessarily orthographic) word. Note that the CLAWS definition of a ‘word’ does not correspond with the conventional orthographic definition. Attributes include:
  - **type** specifies the word class assigned to this form by the CLAWS system.
- `<c>` represents a punctuation character. Attributes include:
  - **type** specifies the class assigned to this character by the CLAWS system.

For this edition of the BNC, the word class tagging system has been extensively revised. A detailed description of the tagging procedures and their application is provided by the *Manual to accompany The British National Corpus (Version 2) with Improved Word-class Tagging* by Geoffrey Leech and Nicholas Smith, which is distributed with the corpus in electronic form. A short list of the POS codes used for the `type` attribute on `<w>` and `<c>` is also provided in section 10.9 (Word class codes) below. As noted above, the representation of this attribute used by the current version of the corpus is minimized, so, for example, the word *difficulty* tagged as a singular noun, appears as

```html
<w>NN1</w>Difficulty
```

rather than as the equivalent XML encoding:

```xml
<w type="NN1">Difficulty</w>
```

The `<s>` element is the basic organizational principle for the whole corpus: every text, spoken or written, may be regarded as an end-to-end sequence of `<s>` elements, possibly grouped into higher-level constructs, such as paragraphs or utterances.
Here is a simple example:

```xml
<s n=11>
  <w NN1>Difficulty</w> <w VBZ>is</w> <w VBG>being</w>
  <w VVN>expressed</w> <w PRP>with</w> <w AT0>the</w>
  <w NN1>method</w> <w TO0>to</w> <w VBI>be</w> <w VVB>used</w>
  <w TO0>to</w> <w VVI>launch</w> <w AT0>the</w> <w NN1>scheme.</w>
</s>
```

The `n` attribute is specified for each `<s>` element and gives its sequence number within the text from which it comes. The code within each `<w>` or `<c>` tag is the word class code assigned by the CLAWS tagging system. These codes are listed below, in section 10.9 (Word class codes).

In most cases, `<s>` elements will correspond with regular orthographic sentences, and `<w>` elements with regular orthographic words. However, it should be noted that several common phrases are treated as single `<w>` elements, typically prepositional phrases such as “in spite of”, while some single orthographic forms such as “can’t” and possessive forms such as “man’s” are decomposed into two `<w>` elements. Further discussion of these non-orthographic word forms is given in the accompanying Manual to accompany The British National Corpus (Version 2) with Improved Word-class Tagging by Geoffrey Leech and Nicholas Smith.

Fragmentary sentences such as headlines or labels in lists are sometimes encoded as `<s>` elements, as in the following example:

```xml
<div1 org=seq>
  <head>
    <s n=1>
      <w NPO>THEOBALD</w> <w POS>’S</w> <w NN1>ROAD</w>
    </head>
    <p>
      <s n=2>
        <w PNP>He</w> <w VVD>walked</w> <w PRP>through</w> <w AT0>the</w>
        <w AJ0>white</w> <w NN2>corridors</w>, <w PRP>past</w>
        <w AT0>the</w> <w NN1>notice</w> <w NN2>boards</w>.
      </s>
    </p>
  </div1>
```

Partly for this reason, the white space (if any) following each orthographic word has been retained in the encoded text. Simply removing the tags will in general produce a correctly punctuated text. (Note however that some punctuation marks are represented as entity references:

```xml
<s n=00024>
  <w PHP>It</w> <w VBD>was</w> <w AT0>the</w> <w NN1>sort</w>
  <w PRP>of</w> <w NN1>sight</w> &mdash; <w NN1>VVB>the</w>
  <w AJO>poor</w>, <w AT0>the</w> <w AJO>strange</w> &mdash;
  <w AHO>which</w> <w AV0>usually</w> <w VVD>alarmed</w>
  <w NPO>Graham</w>.
</s>
```

Dashes used to separate numbers are represented in a similar way, using the `ndash` entity.

Quotation marks are also represented by entity references The reference name used will depend on whether or not the usage of quotation marks in the text has been normalized. Information in the header should describe the course taken for a particular text, as described in section 8.2.1 (Documentary components of the encoding description).

Where the quoted text is a true quotation (that is, a phrase or sequence attributed to someone other than the current narrator or writer) the `<quote>` element discussed in section 6.2.2 (Quotations) may optionally be used. This does not apply to dialogue in fictional works, which is not marked, except by the presence of the quotation mark entities, as in the following example:

```xml
<p>
  <s n=0022>
    ...
  </s>
</p>
```
5.5. Editorial indications

Editorial changes made to the texts during transcription are recorded using the following elements:

- `<gap>` marks the spot where some part of the original source text has been omitted for some reason. Attributes include:
  - `desc` brief description of the material omitted e.g. "name and address".
  - `extent` extent of omitted material e.g. "six words".
  - `reason` brief explanation e.g. "anonymization", "inaudible".
  - `resp` code identifying the agency responsible for marking up the omission.

The `<gap>` element is typically used to indicate where words identifying persons or places have been removed during transcription, where labels etc. have been suppressed for ease of processing, or where material has simply not been transcribed because it is inaudible, illegible or not transcribable (e.g. figures, graphs).

- `<corr>` any editorial correction or regularization, e.g. of material obviously mistranscribed or misspelled, or of variant spellings. Attributes include:
  - `sic` supplies the original form of the word or phrase marked.
  - `resp` code identifying the agency responsible for making the correction.
  - `<sic>` a word or phrase which has not been corrected, but which is in doubt; for example, a spoken word which the transcribers cannot recognise, or a dubious spelling. Attributes include:
    - `reg` supplies a corrected form for the word or phrase marked.
    - `resp` code identifying the agency responsible for noting the need for correction.

In general, the `<corr>` element is used wherever a word appears to be misspelled in the source, and the `<sic>` element where the transcriber is unable to propose a correction, but believes the original to be erroneous. The `<sic>` element is also used to mark words which are intentionally misspelled, for example to indicate non-standard pronunciation; in this case, the `corr` attribute is used to supply a standard spelling.

Slightly different transcription policies have been followed by different transcribers, and consequently these elements may not appear in all texts. The `<editorialDecl>` element of the header described in section 8.2.1 (Documentary components of the encoding description) gives further details of the editorial principles applied across the corpus. The value of the `decls` attribute for an individual text will indicate which principle or set of principles applies to it. The `<tagsDecl>` element in each text’s header may also be consulted for an indication of the usage of these and other elements within it (see further section 8.2 (The encoding description)).
Users are cautioned that the corpus contains a significant number of errors, both in transcription and encoding. Every attempt has been made to reduce the incidence of such errors to an acceptable level, using a number of automatic and semi-automatic validation and correction procedures, but exhaustive proof-reading of a corpus of this size was not economically feasible. The corrections indicated by the tags discussed above are included only where errors have been detected, and no claim should be inferred that no other errors remain.

5.5.1. Some examples

In the following example, the start of a chapter has been deleted at OUCS for sampling reasons:

```
<div1 complete=n n=7 org=seq>
  <gap reason="sampling strategy" desc="beginning of chapter" resp="OUCS">%
  <p>
    <s n=00001>
    <w DPS>Her <w AJ0>thin <w NN1>voice
    <w VVD>trailed <w AVP>off <w PRP>into
    <w AJ0>thin <w NN1>air<c PUN>,
```

In the following example, a list of proper names has been deleted:

```
<head>
  <s n=00081>
  <hi r=ul>
  <w CRD>27.6.90 </hi>
  <w NN2>Minutes <w PRF>of <w AT0>a <w NN1>meeting
  <w PRF>of <w AT0>the <w NP0>Juniper <w NP0>Green
  <w NP0>Village <w NN0>Association <w VVD-VVN>held
  <w PRP>in <w AT0>the <w NP0>Village <w NP0>Hall
  <w PRP>on <w NP0>Wednesday<c PUN>, <w ORD>27th
  <w NP0>June <w PRP>at <w CRD>7.30 <w AV0>pm<c PUN>.
  </head>
  <gap desc="committee members present and absentees" resp=oup>
```

In the following example, a telephone number has been omitted:

```
<s n="1541">
  <w VVB-NN1>Ring <gap desc='telephone number' reason=anonymization><c PUN>.
```

In the following examples, typographic errors in the original are corrected:

```
...<br>
<s n="237"> <w PNP>It ....
  <w VVZ>tries <w AV0>soulfully <w XX0>not <w TO0>to</w TO0>to</w TO0>
  <corr sic=loose> <w VVI>lose </corr> <w DPS>its <w NN1>integrity<c PUN>.
  ...</s n="490"> <w PNP>I <w VBD>was
  <corr sic=fourteen> <w CRD>fourteen </corr> <w AV0>then<c PUN>,
```

In the following example, typographic variation in the original has been regularized:

```
<s n=00029>
  <w AT0>The <w NN1>sum <w PRF>of <w NN0>&pound;60
  <w VHD>had <w VBD>been <w VBN>raised</w VBN>for <w PRP>the <w AT0>The <w NN1>Telethon <w NN1>Appeal</w NN1> Appeal</w NN1>This <w AT0>will <w VBD>be <w VNN>further <corr resp=oucs sic="week end"></w VNN>in <w AT0>the <w NN1>weekend</corr>
  <w VBN>on <w VBD>the <w AT0>the <w NN1>programme</w NN1>programme<c PUN>.
```

In the following example, the transcriber has expressed a doubt as to the validity of the word “memorandising”, but no correction has been made, as it has for the misspelling “bedeviled” which follows it:
6.1. Divisions of written texts

23

6.1. Divisions of written texts

6.1. Divisions of written texts

Parts of a text are normally transcribed in the same order as they appear in the source text. In certain circumstances, however, parts of a text have been moved from the position in which they appear in the source to simplify linguistic processing. There are two common situations where this is necessary:

- where a caption or note appears in the middle of a syntactic unit
- where speakers overlap

Where re-ordering of the first type has occurred, the moved element is generally re-located to the end of the paragraph or similar element in which it appears. Its original position is recorded using a pointer element (<ptr>), an empty tag whose target attribute supplies the identifier of the relocated element. In the following example, the note which originally appeared between the words “roughie-toughie” and “types” has been relocated to the end of the paragraph. The note itself is given an automatically-generated identifier C87NT000 which is then supplied as the value of the target attribute. For example,

This mechanism is also used to represent captions, notes, etc which interrupt the normal reading sequence. By far the commonest use of the <ptr> element, however, is to represent alignment of synchronous speech; see further section 7.4 (Alignment of overlapping speech).

6. Written texts

6.1. Divisions of written texts

Written texts exhibit a bewildering variety and richness of different structural forms. Some have very little organization at levels higher than the paragraphs; others may have a complex hierarchy of parts, sections, chapters etc. Novels are divided into chapters, newspapers into sections, reference works into articles and so forth. The following elements are used to represent all such textual divisions:

<div1> major subdivision of a written text, e.g. chapter.
6.2. Paragraph-level elements and chunks

Written texts may be organized into structural units containing more than one `\langle s\rangle` element and smaller than any of the divisions discussed in section 6.1 (Divisions of written texts) above. The most commonly found such element is the `\langle p\rangle` (paragraph), but there are several others. Their common identifying feature is that they may appear directly within divisions (that is, directly within `\langle div1\rangle`, `\langle div2\rangle` etc., or within `\langle text\rangle` elements, not nested within some other element such as a paragraph).

An alphabetically ordered list of these elements follows:

- `\langle bibl\rangle` a loosely structured bibliographic citation appearing within a corpus text (see 6.2.6 (Notes and citations)).
6.2.1. Headings and captions

Headings and captions serve a variety of functions in written texts. The BNC scheme currently distinguishes between <head> elements, which can appear only at the start of a text division and are logically associated with it (for example, chapter titles, newspaper headlines etc.) and <caption> elements which are logically independent of the position they may have within a textual division (for example, captions attached to pictures or figures, “pull-quotes” embedded within the text, “by-lines” identifying authorship and provenance of a newspaper or periodical article).

One or more <head> elements may appear in sequence at the start of any <div1>, <div2>, <div3> or <div4> element, or at the start of a <list> or <poem>, as in the following example:

```xml
<div1 type="u" n=1>
  <head type=MAIN>
```
6.2.2. Quotations

A quotation is an extract from some other work than the text itself which is embedded within it, for example as an epigraph or illustration. It is marked up using the `<quote>` element. This may contain any combination of other chunks (for example paragraphs, poems, lists) but may not directly contain phrase-level elements. Any reference for the citation should also be contained within it.

For example:

```xml
<quote><p>This <w>NNI</w>way <w>PRP</w>for <w>ATO</w>the <w>AJO</w>sorrowful <w>NNI</w>city.<w>PUN</w>.</p></quote>
```
6.2.4. Poems

Poems or fragments of verse or song may appear both within and between paragraphs. The `<l>` (line) element is used to mark each metrical line, and any titles or headings present are marked with `<head>` elements. Each such group of lines is marked as a `<poem>` element, with no indication of its completeness.

6.2.3. Spoken paragraphs

As noted above, the `<sp>` element is used to mark parts of a written text which were or are intended to be spoken, for example the speeches in a dramatic text or a published interview. Such parts are generally readily identifiable by the use of such conventions as speaker prefixes (the label supplying the name of the speaker) and stage directions, for which the following specific tags are defined:

- `<spkr>` contains the speech prefix used in the original source to identify the speaker of a passage written to be spoken.
- `<stage>` contains any kind of stage direction within a dramatic text.

The `<sp>` element is used only for speaker turns identified as such in a written text, by contrast with the element `<u>` discussed in section 7.2 (Utterances), which is used only for speaker turns identified in a spoken text, i.e. one which has been transcribed from audio tape.

If present, a `<spkr>` element should appear at the start of the `<sp>` element, followed by one or more `<p>` elements containing the actual speech. Any `<stage>` element present will usually be relocated to the end of the paragraph in which it occurs and replaced by a `<ptr>` element, as discussed in section 5.6 (Pointers).

For example:

```xml
<sp>
  <spkr>
    <s n=00156>
      M
    </spkr>
    <p>
      <s n=00157>
        Give her medicine. I'll kill myself if she dies.
      </p>
    </p>
  </sp>
  <stage id=HHWST01C type=u>
    <s n=00159>
      (Sinking to knees and banging head)
    </stage>
  </sp>
```

6.2.4. Poems

Poems or fragments of verse or song may appear both within and between paragraphs. The `<l>` (line) element is used to mark each metrical line, and any titles or headings present are marked with `<head>` elements. Each such group of lines is marked as a `<poem>` element, with no indication of its completeness.
28

6.2.5. Lists

No provision is made for marking units of verse such as stanzas, verse paragraphs etc. A part attribute is defined for the <l> which allows incomplete lines to be indicated, but in the current version of the corpus this always takes the value “u” (for unknown).

For example:

```xml
<poem>
  <l part=u>
    <s n=0900>
      <w PHP>1 <w VVB>send <w DPS>my <w NN1>soul <w PRP>through 
      <w NN1>time <w CJC>and <w NN1>space <w TO0>to <w VVI>greet 
      <w PHP>youn</w PHP>
    </l>
    <l part=u>
      <s n=0901>
        <w PHP>1 <w VBD>were <w AT0>a <w NN1>poet</w PHP>
      </l>
    </s n=0902>
    <w PHP>1 <w VM0>will <w VVI>understand</w PHP>
  </poem>
```

Note that the <l> element is not used to mark typographic lineation; on the few occasions where this has been recorded, it is marked with the <lb> tag discussed in section 6.3.2 (Miscellaneous phrase-level elements) below.

6.2.5. Lists

A list is a collection of distinct items flagged as such by special layout in written texts, often functioning as a single syntactic unit. Lists may appear within or between paragraphs. Where marked, lists are tagged with the <list> element.

A <list> element consists of an optional <head> element, followed by one or more <item> elements, each of which may optionally be preceded by a <label> element, used to hold the identifier or tag sometimes attached to a list item, for example “(a)”. It may also contain a word or phrase used for a similar purpose.

The <item> element may appear only inside lists. It contains the same mixture of elements as a paragraph, and may thus contain one or more nested lists. It may also contains a series of paragraphs, each marked with a <p> element.

Here is an example of a simple list:

```xml
<list>
  <item>
    <s n=0087>
      <w VBZ>Is <w DPS>your <w NN1>nylon <hi r=it> <w NN1>nightie </hi>
      <w AJ0>fireproof</w AJ0>?
    </item>
    <item>
      <s n=0088>
        <w ATO>The <w NN1>hurricane <w VBD>was <hi r=it> <w AJ0>mighty </hi>
        <w AJ0>fierce</w AJ0>.<pb n=78>
    </item>
    <item>
      <s n=0089>
        <w VMO>Will <w PHP>you <hi r=it> <w VVI>mow </hi>
        <w ATO>the <w NN1>lawn</w NN1>?
    </item>
    <item>
      <s n=0090>
        <w VDD>Did <w PHP>you <hi r=it> <w VVI>know </hi>
        <w ATO>the <w NN1>time</w NN1>?
    </item>
</list>
```

Here is an example of a labelled list:

```xml
<list>
  <item>
    <s n=0087>
      <w VBZ>Is <w DPS>your <w NN1>nylon <hi r=it> <w NN1>nightie </hi>
      <w AJ0>fireproof</w AJ0>?
    </item>
    <item>
      <s n=0088>
        <w ATO>The <w NN1>hurricane <w VBD>was <hi r=it> <w AJ0>AVO</w AJ0>mighty </hi>
        <w AJ0>fierce</w AJ0>.<pb n=78>
    </item>
    <item>
      <s n=0089>
        <w VMO>Will <w PHP>you <hi r=it> <w VVI>mow </hi>
        <w ATO>the <w NN1>lawn</w NN1>?
    </item>
    <item>
      <s n=0090>
        <w VDD>Did <w PHP>you <hi r=it> <w VVI>know </hi>
        <w ATO>the <w NN1>time</w NN1>?
    </item>
</list>
```
6.2.6. Notes and citations

Annotations occurring in written texts, and bibliographic citations or references, have been marked up in some texts, using the `<note>` element. This element has the following additional attributes:

- **type** identifies the provenance of the note, i.e. editorial or authorial. Legal values are:
  - `ed` note supplied by BNC transcriber or encoder
  - `orig` note present in the original source text

- **ed** code for the person or organization responsible for BNC-supplied note. Legal values are:
  - `lancs` Note supplied by UCREL grammarians
  - `longm` Note supplied by Longman transcribers
  - `oucs` Note supplied by OUCS staff
  - `oup` Note supplied by OUP transcribers
  - `undef` Provenance of note unknown or unspecified

- **place** specifies the location of an original note in the source text. Legal values are:
  - `foot` foot of page
  - `end` end of current division or text
  - `side` left or right margin
  - `unspec` unknown or unspecified.

Notes within headers are tagged using a distinct `<bibNote>` element, which is a departure from TEI-recommended practice, as is the use of the `<note>` element for both original and supplied annotation. The two usages are distinguished by the type attribute.
Here for example is a typical transcriber’s note:

```xml
<note type=ed>
  <s n=0001>
    <w NN1-NP0>Page <w NN2>numbers <w XX0>not <w AJ0>available
  </note>
```

Original notes may contain any mixture of other chunks, and may also contain paragraphs:
they may appear in written texts only. They will normally be relocated to the end of the section
in which they appear, and their original position marked by a `<ptr>` element, as discussed in
section 5.6 (Pointers).

For example:

```xml
<s n=053>
  <w CJS-PRP>As <w AT0>the <w NP0>UK<w POS>'s <w AJ0>main <w AJ0>independent
  <w NN1>AIDS <w AV-P>care <w NN1-VVB>provider c PUN>,
  <w PHP>we <w VVD>cared <w AVP-PRP>for <w PRP>around <w NNO>25% <w PRF>of <w DT0>all <w DT0>those <w PHN>who <w VVD>had <w NN0>AIDS <w ORD>last <w NN1>year <ptr target=A02NT001><c PUN>.</s>
```

Note the use of the `n` attribute to carry the original footnote number in the above example.

Bibliographic citations or references within running texts may also be marked, using the
`<bibl>` element; this is done in some texts only in the present version of the corpus.

For example:

```xml
<bibl>
  <s n=1379>
    <w NPO>Mordechai <w NP0>Chaim <w NP0>Rumkowski<c PUN>,
    <w AJS>Eldest <w PRF>of <w AT0>the <w NN2>Jews <w PRP>in
    <w AT0>the <w NN1-NP0>Lodz <w NN1>ghetto<c PUN>,
    <w VVG>speaking <w PRP>in <w CRD>1942</bibl>
```

6.3. Phrase-level elements

Phrase-level elements are elements which cannot appear directly within a textual division, but
must be contained by some other element. In practice, this means they will be contained within
an `<s>` element.

6.3.1. Highlighted phrases

Typographic highlighting in the original may not be marked in the transcript at all. Alternatively,
highlighted phrases, and the kind of highlighting used, may be recorded in one of two ways:
6.3.2. Miscellaneous phrase-level elements

- using the global \textit{rend} (rendition) attribute
- using the \texttt{<hi>} (highlighted) element

The former is used where the function of the highlighting is clear, for example to mark a heading, and where the boundaries of the highlighted phrase therefore coincide with the boundaries of some other cdif element. The latter is used where the function is not clear, where the DTD does not provide a tag to identify the feature concerned or where the highlighted phrase is not coterminous with some other element.

When the \texttt{<hi>} element is used, its \textit{rend} attribute must be supplied. On all other cdif elements, the \textit{rend} attribute is optional. Its value indicates the nature of the highlighting used, e.g. italic font, quoted, small caps etc. A list of the values used for this attribute is given in section 10.4 (Rendition codes) below.

It should be noted that the purpose of the \textit{rend} attribute is not to provide information adequate to the needs of a typesetter, but simply to record some qualitative information about the original. In particular, the present version of the corpus includes no indication of size of type or style of writing.

Like all other phrase-level elements, each \texttt{<hi>} element must be entirely contained by an \texttt{<s>} element. This implies that where, for example, a bolded passage contains more than one sentence, or an italicised phrase begins in one verse line and ends in another, the \texttt{<hi>} element must be closed at the end of the enclosing element, and then re-opened within the next.

For example, in the following four lines of verse, the first three are rendered in italics, and the \textit{rend} attribute is therefore specified for each \texttt{<l>} element. In the fourth line, only the first few words are in italics: a \texttt{<hi>} element must be used within the \texttt{<l>} to carry this information.

\begin{verbatim}
<l part=u rend=it>
<s n=394><w PNP>It <w VBD>was <w CRD>one <w PRF>of <w AT0>a <w NNO>pair<\textit{c PUN}>.
<s n=395><w DPS>Its <w AJ0>precious <w NN1>twin</l>
<l part=u rend=it>
<s n=396><w VBD>was <w VVN>stolen <w PRP>by <w AT0>the <w NH2>soldiers<\textit{c PUN}>.
<s n=397><w DTO>All <w AT0>the <w NN1>time</l>
<l part=u r=it>
<s n=398><w DPS>her <w NH1>uncle <w VVD>stood <w AVO>there
</l>
<l part=u>
<s n=399><w hi rend=it> <w DPS>his <w AJ0>big <w NN1>fist </hi> <c PUN>&mdash <w AV0>so</c PUN>!
<s n=400><w PNP>She <w VDZ>does <w DT0>a little <w NH1>mime<\textit{c PUN}>.
</l>
\end{verbatim}

6.3.2. Miscellaneous phrase-level elements

The following miscellaneous phrase-level elements also appear within \texttt{<s>} elements in written texts:

- \texttt{<pb>} marks the start of a new page in the original source; used to indicate where e.g. articles in periodicals are split across several pages.
- \texttt{<lb>} marks the start of a new (printed) line in the original source.
- \texttt{name} proper name of a person, place or institution.
- \texttt{salute} a formulaic greeting or form of address appearing at the start or the end of a spoken or a written text.

In this example, the presence of a page break between two verse lines is indicated by the \texttt{<pb>} element:
7. Spoken texts

7.1. Basic structure: spoken texts

Spoken texts are organized quite differently from written texts. In particular, a complex hierarchy of divisions and subdivisions is inappropriate. The following structural elements are used to represent the organization of spoken texts:

- `<stext>` an individual spoken text.
- `<div>` any subdivision or grouping of the utterances (etc.) making up a spoken text.

In demographically sampled spoken texts, each distinct conversation recorded by a given respondent is treated as a distinct `<div>` element. All the conversations from a single respondent are grouped together to form a single `<stext>` element. Each `<div>` element within a demographically sampled spoken text consists of a sequence of `<u>` elements (see section 7.2 (Utterances)), interspersed with a variety of empty elements used to indicate paralinguistic phenomena noticed by the transcribers (see section 7.3 (Paralinguistic phenomena)).

Context-governed spoken texts do not use the `<div>` element; each `<stext>` element containing a context-governed spoken text consists of a sequence of `<u>` elements again interspersed with a variety of empty elements used to indicate para-linguistic phenomena noticed by the transcribers.
7.3. Paralinguistic phenomena

To handle overlapping utterances, TEI recommends the use of a device known as an alignment map, discussed in section 7.4 (Alignment of overlapping speech) below. A single alignment map, represented by the \texttt{<align>} element, may be defined for a whole spoken text, or for each division of it: if overlap is present, the alignment map is given at the start of the division or text concerned.

Each utterance is further subdivided into \texttt{<s>} elements, and then into \texttt{<w>} and \texttt{<c>} elements, in the same way as for written texts.

The methods and principles applied in transcription and normalisation of speech are discussed in TGCW21 \textit{Spoken Corpus Transcription Guide} and summarised in the appropriate part of the corpus header. The editorial tags discussed in section 5.5 (Editorial indications) above are also used to represent normalisation practice when dealing with transcribed speech.

7.2. Utterances

An utterance is a discrete sequence of speech produced by one participant, or group of participants, in a conversation; it is represented by the \texttt{<u>} element, which has the following additional attribute:

\texttt{who} identifies the person or group responsible for the utterance.

The who attribute is mandatory: its function is to identify the person or group of people making the utterance, using the unique code defined for that person in the appropriate section of the header (see section 8.3.2 (The \texttt{<langUsage>} element)). A simple example follows:

\begin{verbatim}
<u who=w0001>
<s n=00010>
<w ITJ>Mm <w ITJ>mm <c PUN>.
</s>
</u>
\end{verbatim}

The code \texttt{W0001} used here will be specified as the value for the \texttt{id} attribute of some \texttt{<person>} element within the header of the text from which this example is taken. The code \texttt{PS000} is used where the speaker cannot be identified and the code \texttt{PS001} is used for a group of unidentified speakers. Where there are several distinct, but unidentified, speakers within a text, distinct identifiers are used. For example, if text \texttt{xyz} contains two different but unidentified speakers, one of them will be given the identifier \texttt{XYZSP001}, and the other \texttt{XYZSP002}.

7.3. Paralinguistic phenomena

In transcribing spoken language, it is necessary to select from the possibly very large set of distinct paralinguistic phenomena which might be of interest. In the texts transcribed for the BNC, encoders were instructed to mark the following such phenomena:

- **voice quality** for example, whispering, laughing, etc., both as discrete events and as changes in voice quality affecting passages within an utterance.
- **non-verbal but vocalised sounds** for example, coughs, humming noises etc.
- **non-verbal and non-vocal events** for example passing lorries, animal noises, and other matters considered worthy of note.
- **significant pauses** silence, within or between utterances, longer than was judged normal for the speaker or speakers.
- **unclear passages** whole utterances or passages within them which were inaudible or incomprehensible for a variety of reasons.
- **speech management phenomena** for example truncation, false starts, and correction.
- **overlap** points at which more than one speaker was active.
Other aspects of spoken texts are not explicitly recorded in the encoding, although their headers contain considerable amounts of situational and participant information.

The elements used to mark these phenomena are listed below in alphabetical order:

- `<event>` any non-verbal and non-vocal event (such as a door slamming) occurring during a conversation and regarded as worthy of note. Attributes include:
  - `desc` description of the event.
  - `dur` duration of the event in seconds.

- `<pause>` a marked pause during or between utterances. Attributes include:
  - `dur` duration of the pause in seconds.

- `<shift>` a marked change in voice quality for any one speaker. Attributes include:
  - `new` description of the voice quality after the shift.

- `<trunc>` a word or phrase which has been truncated during speech.

- `<unclear>` a point in a spoken text at which it is unclear what is happening, e.g. who is speaking or what is being said. Attributes include:
  - `dur` the duration of the passage in seconds.
  - `who` the person or group responsible for the unclear piece of speech.

- `<vocal>` a non-linguistic but communicative sound made by one of the participants in a spoken text. Attributes include:
  - `desc` the kind of sound made
  - `dur` duration of the sound in seconds.

The value of the `dur` attribute is normally specified only if it is greater than 5 seconds, and its accuracy is only approximate.

With the exception of the `<trunc>` element, which is a special case of the editorial tags discussed in section 5.5 (Editorial indications) above, all of these elements are empty, and may appear anywhere within a transcription.

The following example shows an event, several pauses and a patch of unclear speech:

```xml
<u who=d00011>
<s n=00011>
  <event desc="radio on">You <w VVD>got<w TO0>ta <unclear><w NN1>Radio <w CRD>Two <w PRP>with <w DT0>that <c PUN>.<
  <s n=00012>
  <pause dur=6><w AJ0>Bloody <w NN1>pirate
  <w NN1>station <w VMO>would<w XX0>n't
  <w PHP>you <c PUN>?
</u>
```

Where the whole of an utterance is unclear, that is, where no speech has actually been transcribed, the `<unclear>` element is used on its own, with an optional `who` attribute to indicate who is speaking, if this is identifiable. For example:

```xml
<u who=xx><s>....</u>
<brунк who=yy>
<u who=xx><s>....</u>
```

Here YY’s remarks, whatever they are, are too unclear to be transcribed, and so no `<u>` element is provided.
The values used for the `desc` attribute of the `<event>` element are not constrained in the current version of the corpus. Some common examples follow:

```xml
<event desc="a lot of people talking">
<event desc="door closes">
<event desc="tuning in radio">
<event desc="radio advertisements playing">
```

As noted above, a distinction is made between discrete vocal events, such as laughter, and changes in voice quality, such as words which are spoken in a laughing tone. The former are encoded using the `<vocal>` element, as in the following example:

```xml
<vocal desc="laugh">you'll have to take that off there yeah
<vocal desc="laugh">you can take that off there yeah
</vocal>
```

The `<shift>` element is used instead where the laughter indicates a change in voice quality, as in the following example:

```xml
<shift new="laughing">
<shift>
</shift>
```

Here the passage between the tags `<shift new="laughing">` and `<shift>` is spoken with a laughing intonation.

A list of values currently used for the new attribute is given below in section 10.5 (Voice quality codes).

### 7.4. Alignment of overlapping speech

By default it is assumed that the events represented in a transcription are non-overlapping and that they are transcribed in temporal sequence. That is, unless otherwise specified, it is implied that the end of one utterance precedes the start of the next following it in the text, perhaps with an interposed `<pause>` element. Where this is not the case, the following elements are used:

- `<align>` defines an alignment map used to synchronise points within a spoken text.
- `<loc>` a synchronisation point within an alignment map to which other elements may refer.
- `<ptr>` an empty tag pointing from one part of a text to some other element. Attributes include:

  ```
  target supplies the identifier of some other element in a text; for alignment, specifically, a <loc> element within an alignment.
  ```

For each point of synchrony, i.e. at each place where the number of simultaneous utterances, events, vocals etc. increases or decreases, a `<loc>` element is defined within an `<align>` element, which appears at the start of the enclosing `<div>`, if any. At each place to be
synchronised within the text, a <ptr> element is inserted. The target (target) attributes of these <ptr> elements are then used to specify the identifier of the <loc> with which each is to be synchronised.

The following example demonstrates how this mechanism is used to indicate that one speaker’s attempt to take the floor has been unsuccessful:

```
<u who=w0014>
  <w AJ0>Poor <w AJ0>old <w NPO>Luxembourg'<w VBZ>s <w AJ0-VVN>beaten</w VN>.
</u>
```

This encoding is the CDIF equivalent of what might be presented in a conventional playscript as follows:

```
W0001: Poor old Luxembourg’s beaten. You, you’ve, you’ve absolutely just
gone straight over it --
W0014: (interrupting) I haven’t.
W0001: (at the same time) and forgotten the poor little country.
```

8. The header

The header of a TEI-conformant text generally provides a highly structured description of its contents, analogous to the title page and front matter provided for conventional printed books. Such information is all too often missing in electronic texts; or if supplied, provided only in the form of external documentation such as this manual. The component elements of a TEI header are intended to provide in machine-processable form all the information needed to make sensible use of the Corpus.

Every separate text in the British National Corpus (i.e. each <bncDoc> element) has its own header, referred to below as a text header. In addition, the corpus itself has a header, referred to below as the corpus header, containing information which is applicable to the whole corpus, possibly with some local over-riding, as described in section. Both corpus and text headers are represented by <teiHeader> elements. This element carries the following attributes:

- **type** specifies the kind of document to which the header is attached. Legal values are:
  - corpus the header is attached to the corpus.
  - text the header is attached to a single text.

- **creator** specifies the agency responsible for creating the header.
- **status** specifies the revision status of the associated document. Legal values are:
  - new for BNC release 1.0.
  - update for all subsequent releases.

**update** specifies the date on which the header content was last changed or created.
In the remainder of this section, we describe the components of the `<teiHeader>` element, as used within the BNC. The headers used in this release of the corpus are fully TEI compatible, but do not use all features of the TEI as documented in P3. A TEI header contains a file description (section 8.1 (The file description)), an encoding description (section 8.2 (The encoding description)), a profile description (section 8.3 (The profile description)) and a revision description (section 8.4 (The revision description)), represented by the following four elements:

- `<fileDesc>` contains a full bibliographic description of the corpus itself or of a text within it.
- `<encodingDesc>` documents the relationship between an electronic text and the source or sources from which it was derived.
- `<profileDesc>` provides further information about various aspects of a text, specifically the language used, the situation and date of its production, the participants and their setting, and a descriptive classification for it.
- `<revisionDesc>` summarizes the revision history of a file.

8.1. The file description

The file description (`<fileDesc>`) is the first of the four main constituents of the header. It documents the electronic file itself, i.e. (in the case of a corpus header) the whole corpus, or (in the case of a text header) any characteristics peculiar to an individual file within it. In each case, it contains the following five subdivisions:

- `<titleStmt>` contains title information, identifying the corpus, or a text within it.
- `<editionStmt>` contains additional information relating to a particular version of the corpus (not used with individual corpus texts).
- `<extent>` describes the approximate size of the electronic file as stored on some carrier medium.
- `<publicationStmt>` formally describes the publication or distribution of the corpus and its constituent texts.
- `<sourceDesc>` supplies a bibliographic description for the copy text(s) from which a particular corpus text was derived or generated.

Further detail for each of these is given in the following subsections.

8.1.1. The title statement

As used in the BNC, the title statement (`<titleStmt>`) element has a simpler structure than the equivalent TEI element: it contains one or more `<title>` elements, followed by zero or more `<author>`, `<editor>`, or `<respStmt>` elements. These sub-elements are used throughout the header, wherever the title of a work or a statement of responsibility are required.

- `<title>` the title or chief name of a work, including any alternative titles or subtitles.
- `<respStmt>` supplies information about any person or institution responsible for the intellectual content of a text, edition, or electronic transcription, using the following two sub-elements:
  - `<resp>` contains a phrase describing the nature of a person’s or institution’s intellectual responsibility.
  - `<name>` proper name of a person, place or institution.

For title elements in this version of the BNC, a standardized form of words has been used which includes the following components:
8.1.2. The edition statement

- for written texts, a (possibly shortened) version of the original source title, or, if there is none, a descriptive phrase enclosed in square brackets
- an indication of the size and type of the document
- a note indicating the domain or subject matter of the document

Here are some typical examples:

```xml
<title>British intelligence services in action. Sample containing about 39645 words from a book (domain: world affairs)</title>
<title>Harlow Women's Institute committee meeting. Sample containing about 246 words speech recorded in public context</title>
<title>The Scotsman: Arts section. Sample containing about 48246 words from a periodical (domain: arts)</title>
<title>32 conversations recorded by 'Frank' (PS09E) between 21 and 28 February 1992 with 9 interlocutors, totalling 3193 s-units, 20607 words, and 3 hours 22 minutes 23 seconds of recordings.</title>
<title>[Leaflets advertising goods and products]. Sample containing about 23409 words of miscellanea (domain: commerce)</title>
```

A `<respStmt>` element is used to indicate each agency responsible for any significant effort in the creation of the text. Since responsibilities for data encoding and storage, and for enrichment, are the same for all texts, they are not repeated in each text header. The responsibility for original data capture and transcription varies text by text, and is therefore stated in each text header, in the following manner:

```xml
<respStmt>
<resp>
Data capture and transcription
</resp>
<name>
Longman ELT
</name>
</respStmt>
```

Author and editor information for the source from which a text is derived (e.g. the author of a book) is not included in the `<filedesc>` element but in the `<sourceDesc>` element discussed below (8.1.5 (The source description)).

8.1.2. The edition statement

The standard TEI `<editionStmt>` element is used to specify an edition for each file making up the corpus. For the corpus header this takes the following form:

```xml
<editionStmt n="2.0">
</editionStmt>
```

Note the use of the n attribute to specify that the World Edition is the second major release of the corpus.

The form of words used for each text in the current release of the corpus is as follows:

```xml
<editionStmt>
<para>BNC World Edition:
Header automatically generated by mkhdr 0.30
```
8.1.4. The publication statement

Note that here, as elsewhere, the element `<para>` is used to structure the textual note within a header element. This element is defined as a renaming of the standard TEI `<p>` element, since the `<p>` element itself has been redefined with a more restrictive content model (see further section 9 (Compatibility issues).)

8.1.3. The extent statement

The standard TEI `<extent>` element is used to specify the size of the whole corpus, in the corpus header, or of an individual text, in each text header as in the following example:

```
<extent>Approximately 88 Kbytes running text, containing about 5890 orthographically-defined words; for encoding details see &lt;tagUsage&gt; element.</extent>
```

The specified size does not include the size of the header itself. As the text specifies, the size in Kbytes is only approximate (and may vary on different operating systems). The number of words is calculated according to a simple algorithm which defines words as blank-delimited strings. It is not identical to the number of `<w>` elements actually tagged in the text. For example, the sequence “she’s” would count as one word for the purposes of calculating the extent since it does not contain a blank, but it would be tagged as two distinct `<w>` elements, whereas the sequence “in spite of” counts as three orthographic words, although this sequence is treated as a single `<w>` element.

Counts are provided for each element actually tagged in a text, as further discussed below (8.2.2 (The tagging declaration)).

8.1.4. The publication statement

The standard TEI `<publicationStmt>` element is used to specify publication and availability information for an electronic text. It contains three parts:

**name and address of distributor** These are tagged using the standard `<distributor>` and `<address>` elements respectively; for BNC World, in both corpus and text headers, the name and address given is as follows:

```
<distributor>
Oxford University Computing Services
</distributor>
<address>
<addrLine>13 Banbury Road, Oxford OX2 6NN U.K.</addrLine>
<addrLine>Telephone: +44 1865 273221</addrLine>
<addrLine>Facsimile: +44 1865 273275</addrLine>
<addrLine>Internet mail: natcorp@oucs.ox.ac.uk</addrLine>
</address>
```

**identification numbers for the published text** These are tagged using the standard `<idno>` element. For the corpus header only one such element is specified, as follows:

```
{idno type="BNC">BNC-W</idno>
```

For individual text headers, two identification numbers are supplied, distinguished by the value for the type attribute.

```
{idno type="bnc">A0A</idno>
{idno type="old">CAMfct</idno>
```
The second identifier (of type old) is the old-style mnemonic or numeric code attached to BNC texts in early releases of the corpus, and used to label original printed source materials in the BNC Archive. The first three character code (of type bnc) is the standard BNC identifier. It is used both for the filename in which the text is stored and as the value supplied for the id attribute on the <bncDoc> element containing the whole text.

**availability information** For contractual reasons, the corpus header includes a brief rehearsal of the terms and conditions under which the BNC is made available; this is reproduced in section 11.2 (The BNC corpus header) below. A similar brief notice is also provided in the same place for each individual text, as in the following example:

```xml
<availability status="restricted"><para>
Available worldwide
THIS TEXT IS AVAILABLE THROUGHOUT THE WORLD only as part of the British National Corpus at nominal charge FOR ACADEMIC RESEARCH PURPOSES SUBJECT TO A SIGNED END USER LICENCE HAVING BEEN RECEIVED BY OXFORD UNIVERSITY COMPUTING SERVICES, from whom forms and supporting materials are available.
THIS TEXT IS NOT AVAILABLE FOR COMMERCIAL RESEARCH AND EXPLOITATION unless terms have first been agreed with the BNC Consortium Exploitation Committee. Apply in the first instance to Oxford University Computing Services.
It is your responsibility, as a user, to ensure that an End User Licence is in place. For your information, the Terms of the End User Licence are set out in the corpus header, which is likely to have a file name similar to "corphdr" or "CORPHDR".
Distribution of any part of the corpus under the terms of the Licence must include a copy of the corpus header.
Distribution of this corpus text under the terms of the Licence must include this header embodying this notice.
Permissions granter for World: CAMRA (imprint) of St Albans
</para></availability>
```

Note the inclusion at the end of the notice of the name and address of the agency owning rights in the text concerned, which has granted permission for its inclusion in the corpus. If no such agency is named, permission for rights additional to those explicitly given by the licencing arrangements in place should be sought from the BNC Consortium in the first instance. Note that the BNC world edition includes only texts for which world rights have been cleared by the BNC Consortium.

**date of publication** The BNC was first officially published on 24 November 1994. The present edition has 31st October 2000 as its official publication date.

8.1.5. The source description

The standard TEI <sourceDesc> element is used to supply bibliographic details for the original source material from which an electronic text derives. In the case of a BNC text, this might be a book, pamphlet, newspaper etc., or a recording. One of the following two elements available within the <sourceDesc> will be used, as appropriate:

- `<recordingStmt>` describes a set of recordings used in transcription of a spoken text, either as a series of paragraphs or as a formally structured recording element (8.1.5.1 (The recording statement)).
8.1.5.1. The recording statement

The recording statement (<recordingStmt>) contains a structured bibliographic citation, in which only bibliographic sub-elements appear and in a specified order. (8.1.5.2 (Structured bibliographic record))

These elements are not used within the corpus header, which simply contains a note about the sources from which the corpus was derived, tagged as a <para> (paragraph). The headers of individual texts each contain one of the above elements to specify their source.

Context-governed spoken texts derived from broadcast or similar ‘published’ material may have either a recording statement or a bibliographic record as their source.

8.1.5.1. The recording statement

The recording statement (<recordingStmt>) element contains one or more <recording> elements, defined as follows:

- <recording> details of a particular audio recording used as the source of a spoken text, either directly or from a public broadcast. Attributes include:
  - date specifies the date of the recording
  - dur specifies the duration of the recording, in seconds
  - time specifies the time of day when the recording was made.
  - type characterizes the recording in terms of the equipment used to make it. Legal values include:
    - DAT recording made on Digital Audio tape
    - Unknown recording equipment or quality unknown
    - Walkman recording made on Walkman

The standard TEI global attribute n is used (for this element only) to provide the number of the audio tape holding the original recording, as deposited with the National Sound Archive in London.

The BNC version of this TEI element has two additional attributes (date and time) and it may contain only character data, rather than the more complex substructure permitted by the TEI equivalent. In the following simple example, typical of most of the ‘context-governed’ parts of the BNC, the <recording> element has no content at all:

```
<recordingStmt>
  <recording n="121101" date="1994-02-09" time="11:00" type="DAT"></recording>
</recordingStmt>
```

When, as is often the case for the spoken demographic parts of the BNC, a text has been made up by transcribing several different recordings made by a single respondent over a period of time, each such recording will have its own <recording> element, as in the following example:

```
<recordingStmt>
  <recording n="018201" dur="322" date="1991-11-28" time="18:15+" type="Walkman" id="KB7RE000"></recording>
  <recording n="018202" dur="253" date="1991-11-28" time="18:15+" type="Walkman" id="KB7RE001"></recording>
  <!-- ... -->
  <recording n="018301" dur="75" date="1991-11-29" time="12:15+" type="Walkman" id="KB7RE007">
    <!-- ... -->
  </recording>
</recordingStmt>
```
Note the presence of an id attribute on each of the above recordings. The value given here is used to indicate the recording from which a given part of the text was transcribed. Each recording is transcribed as a distinct <div> (division) element within an <stext>, with its identifier supplied as the value of a decls attribute. Thus, in the body of the text from which the above example was taken, there will be a <div> element starting as follows:

```xml
<div decls="KB7RE0077">
```

which will contain the part of text transcribed from that recording. As noted above the identifier supplied on the n attribute is quite distinct, and specifies the original tape on which the recording was made.

8.1.5.2. Structured bibliographic record

The standard TEI <biblStruct> element is used to record bibliographic information for each non-spoken component of the BNC. As defined in the TEI, this element has a complex structure designed to support a wide range of standard bibliographic practices. In the BNC, its structure is restricted as further described below.

At the highest level, all BNC <biblStruct> element will contain a <monogr> element holding other elements that describe the item in question. In a few cases, this may be preceded by an <analytic> element, as further described at the end of this section.

At least one <monogr> element must be present in a <biblStruct> element. It may contain the following elements:

- <title> the title or chief name of a work, including any alternative titles or subtitles; this must be given first. In several cases, a generated title or descriptive paraphrase is used, generally enclosed within square brackets. In the current version of the corpus, subtitles, alternative or series titles are not distinguished from the main title, other than by the use of conventional punctuation.
- <author> the name of an author (personal or corporate) of a work; names are generally given in canonical form, with surnames preceding forenames. Unlike the TEI equivalent element of the same name, the BNC version has two additional attributes:
  - **domicile** specifies the author’s domicile, as established for the purposes of the BNC “Britishness” test.
  - **born** specifies the author’s year of birth, where available.
- <editor> the name of the editor (personal or corporate) for a work.
- <imprint> groups information relating to the publication or distribution of a bibliographic item.
- <biblScope> defines the scope of a bibliographic reference, for example as a list of page numbers, or a named subdivision of a larger work. Attributes include:
  - **type** identifies the type of information conveyed by the element. In the present edition of the corpus the only value used is pp, i.e. page numbers.

A <title> element must be present and is always given first. None of the other components is mandatory, but if any of them are supplied, they must be in the following order, following the title:

- one or more statements of intellectual responsibility (i.e. <author> or <editor> elements)
8.1.5.2. Structured bibliographic record

- one or more <imprint> elements

The `n` attribute is used with both <author> and <imprint> elements to supply a six-letter code identifying the author or imprint concerned. The values used should be unique across the corpus, but this is not validated by the current release of the DTD.

For published texts at least one <imprint> element is supplied, containing the following elements in the order given:

- `<publisher>` name of a publisher.
- `<pubPlace>` place of publication.
- `<date>` date of publication of the edition transcribed, usually given in normalized format. Note that this may not be the same as the date specified by the <creation> element. Attributes include:
  - `value` specifies standard value for this date in ISO 8601 format

The following example demonstrates how these elements are used to record bibliographic details for a typical book:

```
<biblStruct>
  <monogr>
    <title>It might have been Jerusalem. </title>
    <author n="HealyT1" domicile="Scotland">Healy, Thomas</author>
    <imprint n="POLYGO1">
      <publisher>Polygon Books</publisher>
      <pubPlace>Edinburgh</pubPlace>
      <date value="1991">1991</date>
    </imprint>
  </monogr>
  <biblScope type="pp">1-81</biblScope>
</biblStruct>
```

The following example is typical of the case where a collection of leaflets or newsletters has been treated as a single text:

```
<biblStruct>
  <monogr>
    <title>[Potato Marketing Board leaflets]</title>
    <imprint n="POTATO1">
      <publisher>Potato Marketing Board</publisher>
      <pubPlace>London</pubPlace>
      <date value="1991">1991</date>
    </imprint>
  </monogr>
</biblStruct>
```

As noted above, a <monogr> element may be preceded by an <analytic> element. The <analytic> element is used on a few occasions where the item encoded as a single text in the BNC is in fact only a part of a larger bibliographic item. In the following example, the source for the text is a single article which was published in a collection with a different title and author:

```
<biblStruct>
  <analytic>
    <title>Damages on death</title>
    <author n="SauntT1">Saunt, Thomas</author>
  </analytic>
</biblStruct>
```
8.2.1. Documentary components of the encoding description

Where 'series' information is available for a given title, this is not normally tagged distinctly. Instead the series title is given as part of the monographic title, usually preceded by a colon.

This level of bibliographic description has not been carried out with complete consistency across the current release of the corpus.

8.2. The encoding description

The second major component of the TEI header is the encoding description (<encodingDesc>). This contains information about the relationship between an encoded text and its original source and describes the editorial and other principles employed throughout the corpus. It also contains reference information used throughout the corpus.

The standard TEI <encodingDesc> element has the following six components:

- <projectDesc> describes in detail the purpose for which an electronic file was encoded, together with any other relevant information concerning the process by which it was assembled or collected.
- <samplingDecl> contains a prose description of the rationale and methods used in sampling texts in the creation of the corpus.
- <editorialDecl> provides details of editorial principles and practices applied during the encoding of a text.
- <tagsDecl> provides detailed information about the tagging applied to a corpus text.
- <refsDecl> specifies how canonical references are constructed for a text.
- <classDecl> contains a series of <category> elements, defining the classification codes used for texts within the corpus.

In the BNC, one of each of these elements appears in the corpus header, with the exception of the <tagsDecl> element which is also given in the individual text headers.

8.2.1. Documentary components of the encoding description

The <projectDesc> element for the corpus gives a brief description of the goals, organization and results of the BNC project. It is reproduced in section 11.2 (The BNC corpus header) below.

The <samplingDecl> element for the corpus reads as follows:

```xml
<samplingDecl>
<para>Different parts of the BNC were constructed using different sampling policies, as further described in the BNC Design Documentation. The policies are summarized below. Note that information about which policy resulted in the selection of a particular text is not available.
</para>
</samplingDecl>
```
The declaration contains an introductory paragraph (tagged with the standard TEI \textit{<p> —
renamed as \textit{<para> — element}) followed by a list, each item of which defines a different
sampling strategy used in the corpus. The values given for the \textit{id} attribute on each \textit{<item>}
will be specified in the list of identifier values supplied as the value for the target attribute of the
\textit{<catRef>} element prefixed to a particular text's header. For example, the header of a spoken
demographic text will include a \textit{<catRef> element like the following:
\texttt{<catRef target='... SD003 ...'/>}
where the dots indicate other declarations applicable to this text.

An exactly equivalent method is used to indicate the various editorial practices applicable
in different portion of the corpus. The list of practices is given in an \textit{<editorialDecl> element, reproduced in section 11.2 (The BNC corpus header) below, each item of which has an
identifying code which is subsequently referenced via the \textit{decls} attribute on the \textit{<div> or
\textit{<text> element to which that editorial practice applies.

The following table summarizes the codes used in the present version of the corpus:

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN000</td>
<td>Errors tagged with \textit{&lt;sic&gt;} when seen; no normalization</td>
</tr>
<tr>
<td>CN001</td>
<td>Errors tagged with \textit{&lt;sic&gt;} if seen; normalisation with \textit{&lt;corr&gt;}</td>
</tr>
<tr>
<td>CN002</td>
<td>Normalized to standard British English or control list member</td>
</tr>
<tr>
<td>CN004</td>
<td>Corrections and normalizations applied silently</td>
</tr>
<tr>
<td>HN000</td>
<td>Smart elision of line-end hyphens; &amp;rehy used for remainder</td>
</tr>
<tr>
<td>HN001</td>
<td>Dumb elision of line-end hyphens; true hyphens hand-reinstated</td>
</tr>
<tr>
<td>HN002</td>
<td>Line-end hyphens removed by hand where appropriate</td>
</tr>
<tr>
<td>HN003</td>
<td>Source material contains no line-end hyphens</td>
</tr>
<tr>
<td>QN000</td>
<td>Open and close quote normalized to &amp;bquo, &amp;equo</td>
</tr>
<tr>
<td>QN001</td>
<td>Open and close quote normalized to &amp;quo</td>
</tr>
<tr>
<td>QN002</td>
<td>Quotation may be represented using \textit{&lt;shift&gt;}</td>
</tr>
<tr>
<td>SN000</td>
<td>Segmentation carried out by CLAWS5.</td>
</tr>
</tbody>
</table>

\subsection*{8.2.2. The tagging declaration}

The standard tagging declaration (\textit{<tagsDecl> element is used slightly differently in corpus
and in text headers. In the corpus header, it is used to list every element name actually used
within the corpus, together with a brief description of its function. In text headers, it is used to
specify the number of elements actually tagged within each text. In either case it consists of a
number of \textit{<tagUsage> elements, defined as follows:

\textit{<tagUsage>} supplies information about the usage of a specific element within a
\textit{<text>}. Attributes include:

- \texttt{gi} the name (generic identifier) of the element indicated by the tag.
- \texttt{occurs} the number of occurrences of this element within the text.
In the corpus header, each `<tagUsage>` element contains a brief description of the element specified by its `<gi>` element; the occurs attribute is not supplied, as in the following extract:

```
<tagUsage gi="div4">
Written text division, level 4
</tagUsage>
<tagUsage gi="event">
Non-verbal event in spoken text
</tagUsage>
<tagUsage gi="gap">
Point where source material has omitted
</tagUsage>
<tagUsage gi="head">
Header or headline in written text
</tagUsage>
```

In text headers, the `<tagUsage>` elements are empty, but the occurs attribute is always supplied, and indicates the number of such elements which appear within the text, as in the following example, taken from a typical written text:

```
<tagsdecl>
<tagsDecl>
<tagUsage gi="body" occurs="1"></tagUsage>
<tagUsage gi="c" occurs="4649"></tagUsage>
<tagUsage gi="caption" occurs="77"></tagUsage>
<tagUsage gi="div1" occurs="2"></tagUsage>
<tagUsage gi="div2" occurs="10"></tagUsage>
<tagUsage gi="div3" occurs="49"></tagUsage>
<tagUsage gi="gap" occurs="1"></tagUsage>
<tagUsage gi="head" occurs="55"></tagUsage>
<tagUsage gi="hi" occurs="50"></tagUsage>
<tagUsage gi="item" occurs="26"></tagUsage>
<tagUsage gi="list" occurs="4"></tagUsage>
<tagUsage gi="note" occurs="3"></tagUsage>
<tagUsage gi="p" occurs="378"></tagUsage>
<tagUsage gi="pb" occurs="65"></tagUsage>
<tagUsage gi="ptr" occurs="79"></tagUsage>
<tagUsage gi="s" occurs="1713"></tagUsage>
<tagUsage gi="sic" occurs="1"></tagUsage>
<tagUsage gi="text" occurs="1"></tagUsage>
<tagUsage gi="w" occurs="40394"></tagUsage>
</tagsDecl>
```

8.2.3. The reference and classification declarations

The `<refsDecl>` element for the corpus header defines the approved format for references to the corpus. It takes the following form

```
<refsDecl>
<para>Canonical references in the British National Corpus are to text segment (&lt;s&gt;) elements, and are constructed by taking the value of the n attribute of the &lt;cdif&gt; element containing the target text, and concatenating a dot separator, followed by the value of the n attribute of the target &lt;s&gt; element.</para>
</refsDecl>
```

The standard TEI `<classDecl>` element is used in the BNC Corpus Header to formally define the text classification scheme applied to the corpus, and the particular codes within it. It consists of a set of `<category>` elements, each representing a particular textual classification feature and a value for that feature.

```
<classDecl>
<category> contains an individual descriptive category or feature-value pair.
```
8.3. The profile description

<catDesc> describes some category within a taxonomy or text typology, in the form of a brief prose description.

For example, the following <category> elements appear within the BNC <classDecl> element in the header:

```
<category id="wridom">
  <catDesc>Domain for written corpus texts</catDesc>
</category>
<category id="wridom1">
  <catDesc>Imaginative</catDesc>
</category>
<category id="wridom2">
  <catDesc>Informative: natural & pure science</catDesc>
</category>
<category id="wridom3">
  <catDesc>Informative: applied science</catDesc>
</category>
<category id="wridom4">
  <catDesc>Informative: social science</catDesc>
</category>
<category id="wridom5">
  <catDesc>Informative: world affairs</catDesc>
</category>
<category id="wridom6">
  <catDesc>Informative: commerce & finance</catDesc>
</category>
<category id="wridom7">
  <catDesc>Informative: arts</catDesc>
</category>
<category id="wridom8">
  <catDesc>Informative: belief & thought</catDesc>
</category>
<category id="wridom9">
  <catDesc>Informative: leisure</catDesc>
</category>
```

The <catDesc> element contained by the outer <category> element here (that with identifier wridom) is understood to apply also to each <catDesc> contained by each of its constituent (daughter) <category> elements. That is, the full description for category wridom3 is “Domain for written corpus texts: informative: natural science”.

The category descriptions applicable to a given text are specified by the <catRef> element within its header, as described above. Its target lists the identifiers of all <category> elements applicable to that text. Thus, the header of a written text assigned to the social science domain which has a corporate author will include a <catRef> element like the following:

```
<catref target='... wriaty1 wridom4...'>
```

The dots above represent the identifiers of all other category codes applicable to this text.

A full list of all category codes, and the numbers of texts so classified in the current release of the corpus is provided in section 10.8 (Text and genre classification codes).

Information about the classification and categorization of an individual text is held within the <textClass> element discussed below (8.3.5 (Text classification))

8.3. The profile description

The third component of the TEI header is the profile description (<profileDesc>) element, which has the following components:

<creation> contains information about the creation of a text.
8.3.3. The participant description

The participant description (particDesc) element is used to provide information about speakers of texts transcribed for the BNC. In its basic structure it is close to the element defined by the TEI but it has been modified to include some more specific elements provided for the BNC. It appears both within the corpus header, to define the generic “unknown participant”, and also within individual spoken text headers to define the participants specific to those texts.
8.3.3.1. The person element

It contains a series of <person> elements describing the participants whose speech is transcribed in this text, followed by an optional <particLinks> element describing any relationships or links amongst them.

8.3.3.1. The person element

Each <person> element describes a single participant in a language interaction and may take one or more of the following attributes:

- **id** (mandatory) supplies a unique code used to identify this speaker and their utterances in the transcription.
- **role** specifies the role of this participant with respect to the respondent, as specified by the respondent.
- **sex** specifies the sex of the participant. Possible values are:
  - m male
  - f female
  - u unknown or inapplicable
- **age** specifies the age group to which the participant belongs. Possible values are:
  - 0 Under 15 years
  - 1 15 to 24 years
  - 2 25 to 34 years
  - 3 35 to 44 years
  - 4 45 to 59 years
  - 5 Over 59 years
  - X Unknown
- **flang** specifies the first language or mother tongue of the participant. Possible values are listed in section 10.6 (Regional codes).
- **dialect** specifies any dialect spoken by the participant, as specified by the respondent. Possible values are listed in section 10.6 (Regional codes).
- **soc** specifies the social class of the participant. Legal values are:
  - AB AB (top or middle management, administrative or professional)
  - CI C1 (junior management, supervisory or clerical)
  - C2 C2 (skilled manual)
  - DE DE (semi-skilled or unskilled)
  - UU Class unknown
- **educ** specifies the age at which the participant ceased full-time education. Possible values are:
  - 0 Still in education
  - 1 Left school aged 14 or under
  - 2 Left school aged 15 or 16
  - 3 Left school aged 17 or 18
  - 4 Education continued until age 19 or over
  - X Information not available
- **resp** (for spoken demographic participants only) specifies the identifier of the respondent in whose data this participant’s interactions are recorded.

The global id attribute is required for each participant whose speech is included in a text, and its value is unique within the corpus. Although a given individual will always have the same identifier within a single text, there is no way of identifying the same individual appearing
in different texts. For this reason, all demographically sampled conversations collected by a single respondent are treated together as a single text.

The value for the flang attribute consists of a two-letter language code taken from ISO 639 (normally EN for English), optionally suffixed by a three-letter country code taken from ISO 3166. Thus “EN-GBR” is English as spoken in the United Kingdom; “EN-CAN” is English as spoken in Canada, and “FR-FRA” is French as spoken in France.

The value for the dialect attribute is also a three-letter code taken from a local extension to ISO 3166. A full list of codes used and their meanings is given in section 10.6 (Regional codes).

Where available, any additional information about a participant will be provided as text within the <person> element, enclosed within a <para> element. In the BNC DTD, the following extra elements are also provided for this purpose:

- **age** specified more exactly than by the age attribute, which groups respondents into age bands.
- **name** a proper name used for the person.
- **occupation** characterization of the person’s occupation.
- **dialect** characterization of the person’s dialect.

In each case, the information provided is that given by the respondent and is taken from the log books issued to all participants in the demographic part of the corpus. It has not been normalized.

Here is a typical example from the demographic part of the corpus:

```xml
<person age="0" dialect="XLO" id="PS5A1" role="self" sex="m" soc="C2">
  <name>Terry</name>
  <age>14</age>
  <occupation>student</occupation>
  <dialect>London</dialect>
</person>
```

Here is a typical example from the context-defined part of the corpus:

```xml
<person age="X" educ="X" id="PS2AY" n="W0003" sex="m" soc="UU">
  <name>Frank Harasikva</name>
  <occupation>politician</occupation>
  <para>Euro candidate presenting self for selection</para>
</person>
```

### 8.3.3.2. The particLinks element

Relationships between participants, where specified, are represented using the standard TEI <relation> element which has the following description and attributes:

- **<relation>** describes any kind of relationship or linkage amongst a specified group of participants. Attributes include:
  - **active** identifies the “active” participants in a directed relationship, or all the participants in a mutual one.
  - **desc** supplies a name for the relationship, seen from the point of view of the active participant in a directed relationship.
  - **mutual** indicates whether the relationship holds equally amongst all participants. Legal values are:
    - **Y** the relationship is mutual
8.3.4. The setting description

N the relationship is directed
passive identifies the “passive” participants in a directed relationship.

A list of the different types of relationship identified amongst participants is given in section 10.7 (Relationship codes).

Following the TEI Guidelines, we distinguish between mutual relationships, in which all participants are on an equal footing, and directed relationships, in which the roles of the participants are typically described differently. The roles applicable to a directed relationship are arbitrarily classed here as either active or passive. For example, the relationships “colleague” or “spouse” would be classed as mutual, while “employee” or “wife” would be classed as directed. A relationship such as “sister” may or may not be directed, depending on whether it obtains between two women or between a man and a woman.

For a mutual relationship, only the active attribute will be supplied; for a directed one, both active and passive attributes will be supplied. In either case, these attributes take as value a list of the identifiers of the <person> elements understood to be involved in the relationship concerned.

In the current edition of the corpus, relation information is provided in this form only for the context-governed participants. The relationships between participants in the demographically-sampled part of the corpus are indicated by the role attribute as discussed above.

8.3.4. The setting description

The TEI <settDesc> element is used to document the context within which a spoken text takes place. It appears once in the header of each spoken text, and contains one or more <setting> elements for each distinct recording.

<setting> describes one particular setting in which a language interaction takes place. The following attributes are used in the BNC:

- **who** supplies the identifiers of the participants at this setting.
- **id** supplies a unique identifier for this setting.
- **n** supplies the identifier used for the <recording> element corresponding with this setting.

The content of each <setting> element supplies additional details about the place, time of day, and other activities going on, using the following additional elements:

- **<name>** contains a place name, usually prefixed by the name of the English county in which it is located.
- **<locale>** contains a brief informal description of the nature of a place, for example a room, a restaurant, a park bench etc.
- **<activity>** contains a brief informal description of what a participant in a language interaction is doing other than speaking, if anything. Bears an additional attribute:

  - **spont** indicates the degree of spontaneity associated with the activity as either H (high) M (medium) or L (low)

Some typical examples follow:

```
<setting n="020901" who='PS000 DCJPS000 DCJPS001'>
  <name>Essex: Harlow </name>
  <locale> Harlow College</locale>
  <activity spont="M"> A’level lecture </activity>
</setting>
```
8.3.5. Text classification

The TEI provides a number of ways in which classification or text-type information may be specified for a text, grouped together within a `<textClass>` element, which appears once in the header of each text. Classifications may be represented using references to internally defined classifications provided in the `<classCode>` element (such as the BNC classification scheme described in section 8.2.3 (The reference and classification declarations)), by reference to some other predefined classification system, or by an open set of keywords. All three methods are used in the BNC, using the following elements:

- `<catRef>` specifies one or more defined categories within some taxonomy or text typology. Attributes include:
  - `target` identifies all the categories concerned.
- `<classCode>` contains the code used for this text in an externally-defined classification system: in this release of the BNC, the genre codes defined by David Lee are used.
- `<keywords>` contains a list of keywords or phrases identifying the topic or nature of a text, each of which is tagged as a `<term>` element.

A `<catRef>` element is provided in the header of each text. Its target attribute contains values for each of the classification codes listed in the following table and defined in the corpus header. In each case, the classification code consists of an alphabetic prefix (e.g. `alltim`) identifying the category (e.g. “date”), followed by a single digit indicating a value for that category. Thus the code `alltim1` indicates “dated 1960-1974”. The value 0 is always used to indicate missing or unknown values. A list of the values used is given in section 10.8 (Text and genre classification codes) below.

This taxonomy is that originally defined for selection and description of texts during the design of the corpus, as further discussed elsewhere. It is of course possible to classify the texts in many other ways, and no claim is made that this method is universally applicable or even generally useful, though it does serve to identify broadly distinct sub-parts of the corpus for investigation. The reader is also cautioned that, although an attempt has been made in the current edition of the corpus to correct the more egregious classification errors noted in the first edition, unquestionably many errors and inconsistencies remain. In particular, the categories `wrilev` (perceived level of difficulty) and `wrista` (estimated circulation size) were incorrectly differentiated during the preparation of the corpus and cannot be relied on.

A `<classCode>` element is also provided for every text in the corpus. It contains the code assigned to this text in David Lee’s genre-based analysis carried out at Lancaster University since publication of the first edition of the BNC.

In the first release of the BNC, most texts were assigned a set of descriptive keywords, tagged within the `<keywords>` element. These terms were not taken from any particular descriptive thesaurus or closed vocabulary; the words or phrases used are those which seemed useful to the data preparation agency concerned, and are thus often inconsistent or even misleading. They have been retained unchanged in the present version of the BNC, pending a more thorough revision.

In this edition of the BNC, a second set of keywords has been supplied for the majority of written texts. These keywords are also tagged using a `<keywords>` element, but with
9. Compatibility issues

a value for the source attribute of COPAC, indicating that the terms so tagged are derived from a different source. The source used is a major online library catalogue service (see http://www.copac.ac.uk), from which we have taken the subject keywords provided for each title identifiable as forming part of the BNC. Like other public access catalogue systems, COPAC uses a well-defined controlled list of keywords for its subject indexing, details of which are not further given here.

Here is an example showing how one text (BND) is classified in each of these three ways:

```xml
<textClass>
  <catRef target="alltim3 allava2 alltyp3 wriag0 wriad0 wriase3 wriasy2 wriaud3 wrilev2 wrimed1 wripp5 wriam1 wrista2 wrias3">
    <classCode scheme="DLee">W_religion</classCode>
    <keywords scheme="COPAC">
      <term>Marriage - Religious aspects - Christianity</term>
      <term>Marriage - Christian viewpoints</term>
    </keywords>
    <keywords>
      <term>Christian guide to marriage</term>
    </keywords>
  </textClass>
```

8.4. The revision description

The revision description (<revisionDesc>) element is the fourth and final element in the standard TEI header. In the BNC, it consists of a series of <change> elements, each containing a <date>, a <respStmt>, and a <para> element. A new <change> element was added at the start of the list for each major change made in the text or header during preparation of the current edition of the corpus.

Here is the start of a typical example:

```xml
<change><date>2000-09-01</date>
  <respStmt><resp>ed</resp><name>OUCS</name></respStmt>
  <para>Check all tagcounts</para></change>
<change><date>2000-06-23</date>
  <respStmt><resp>ed</resp><name>OUCS</name></respStmt>
  <para>Resequenced s-units and added headers</para></change>
<change><date>2000-01-29</date>
  <respStmt><resp>ed</resp><name>OUCS</name></respStmt>
  <para>Revised participant details</para></change>
<change><date>1994-01-10</date>
  <respStmt><resp>DD</resp></respStmt></change>
```

When any significant change is made to any component of the corpus, the following steps are taken:
- a <change> element is added to the <revisionDesc> of the text affected
- the update attribute of the text header is changed to the date of the change
- the value of the status attribute of the text header is set to “update”

9. Compatibility issues

The first version of the BNC was released slightly in advance of the publication of the TEI’s definitive Recommendations, and over a year before publication of the Corpus Encoding Standard. Although all three standards have much in common (in particular, CDIF - Corpus
9.1. Differences between the BNC DTD and TEI

The present version of the BNC document type declaration (DTD) can be expressed as a set of extensions against the standard TEI dtd, using the extension mechanism recommended by that standard. Full details of the procedure are given in chapter 3 of the TEI Guidelines. Essentially, the procedure requires the definition of two ‘extension’ files, called here bncMods.ent and bncMods.dtd, the former containing definitions of parameter entities needed for this set of extensions, and the latter containing the actual SGML element and attribute definitions which make up the required modifications. Copies of these files are included in the present release, along with the DTD derived from them. The present section describes their content informally.

The DTD described elsewhere in this document makes use of several elements already defined in other TEI tagsets, in particular the base tag sets for prose and for spoken texts, and the additional tagsets for language corpora and analysis. To combine all of these with the extension files mentioned above, a TEI conformant document will begin as follows:

```xml
<!DOCTYPE TEI.2 SYSTEM "tei2.dtd" [  
  <!ENTITY % TEI.prose  "INCLUDE">  
  <!ENTITY % TEI.spoken  "INCLUDE">  
  <!ENTITY % TEI.general  "INCLUDE">  
  <!ENTITY % TEI.analysis  "INCLUDE">  
  <!ENTITY % TEI.corpora  "INCLUDE">  
  <!ENTITY % TEI.extensions.ent SYSTEM "bncMods.ent">  
  <!ENTITY % TEI.extensions.dtd SYSTEM "bncMods.dtd">  
] ]> 
```

This file can be compiled to form a one file version of the DTD, in which all parameter references have been resolved, and any redundant declarations removed, using software such as the TEI PizzaChef.

The file bncMods.ent consists of a number of SGML parameter entity definitions, which over-ride the definitions provided in the TEI DTD itself. These declarations have the following effects:

- to exclude from the DTD a large number of standard TEI elements which are not actually used in the BNC DTD;
- to provide alternative names for some standard TEI elements;
- to exclude from the TEI DTD some elements which are redefined, either with a stricter content model, or with differing attribute lists, in the BNC DTD;
- to specify the location within the TEI class system of some elements not defined in the TEI DTD.

Taking these in turn, some 114 standard TEI elements are excluded from the DTD by means of parameter entity declarations like the following:
9.1. Differences between the BNC DTD and TEI

The following is a complete list of standard TEI elements excluded in this way: `<ab>`, `<abbr>`, `<add>`, `<affiliation>`, `<alt>`, `<altGrp>`, `<anchor>`, `<argument>`, `<authority>`, `<back>`, `<biblFull>`, `<birth>`, `<broadcast>`, `<byline>`, `<cb>`, `<channel>`, `<cit>`, `<cl>`, `<constitution>`, `<correction>`, `<dateline>`, `<dateRange>`, `<del>`, `<derivation>`, `<distinct>`, `<div0>`, `<div5>`, `<div6>`, `<div7>`, `<divGen>`, `<docAuthor>`, `<docDate>`, `<docEdition>`, `<docImprint>`, `<docTitle>`, `<domain>`, `<education>`, `<emph>`, `<epigraph>`, `<equipment>`, `<expan>`, `<factuality>`, `<firstLang>`, `<foreign>`, `<front>`, `<fsdDecl>`, `<funder>`, `<gloss>`, `<group>`, `<headLabel>`, `<headItem>`, `<hyphenation>`, `<interp>`, `<interpGrp>`, `<interpretation>`, `<join>`, `<joinGrp>`, `<kinesic>`, `<link>`, `<linkGrp>`, `<ln>`, `<measure>`, `<meeting>`, `<metDecl>`, `<mentioned>`, `<milestone>`, `<normalization>`, `<notesStmt>`, `<num>`, `<open>, `<orig>`, `<personGrp>`, `<phr>`, `<postBox>`, `<postCode>`, `<preparedness>`, `<principal>`, `<purpose>`, `<q>`, `<quote>`, `<rendition>`, `<residence>`, `<rs>`, `<reg>`, `<scriptStmt>`, `<seg>`, `<segmentation>`, `<series>`, `<seriesStmt>`, `<signed>`, `<soCalled>`, `<soccStatus>`, `<span>`, `<spanGrp>`, `<sponsor>`, `<state>`, `<stdVals>`, `<step>`, `<street>`, `<symbol>`, `<textDesc>`, `<time>`, `<timeRange>`, `<titlePage>`, `<titlePart>`, `<trailer>`, `<variantEncoding>`, `<when>`, `<writing>`, `<xptr>`, `<xref>.

Four elements in the TEI DTD are given different names in the BNC DTD. For example, the TEI element `<speaker>` is renamed `<spkr>`. The declarations below effect this and the other renamings required, by changing the value of the relevant parameter entity:

```
<!ENTITY % n.teiCorpus.2 "bnc">
<!ENTITY % n.TEI.2 "bncDoc">
<!ENTITY % n.p "para">
<!ENTITY % n.speaker "spkr">
```

The next part of the bncExtns.ent file contains IGNORE declarations like those above, which have the effect of removing the existing definitions for 22 TEI elements which are to be redefined. The redefinitions are provided in the second of the two BNC extension files, bncExtn.dtd, along with definitions for some new elements not otherwise available. Their effects are summarized in the following table.

### Table 19. Summary of differences between TEI and BNC

<table>
<thead>
<tr>
<th>TEI Element</th>
<th>Difference in BNC dtd</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>&lt;TEI.2&gt;</code></td>
<td>Changed content model to allow either text or stext; renamed as bncDoc</td>
</tr>
<tr>
<td><code>&lt;activity&gt;</code></td>
<td>Simplified content model; added attribute</td>
</tr>
<tr>
<td><code>&lt;age&gt;</code></td>
<td>New element</td>
</tr>
<tr>
<td><code>&lt;align&gt;</code></td>
<td>New element</td>
</tr>
<tr>
<td><code>&lt;author&gt;</code></td>
<td>Simplified content model; added attributes</td>
</tr>
<tr>
<td><code>&lt;body&gt;</code></td>
<td>Simplified content model</td>
</tr>
<tr>
<td><code>&lt;c&gt;</code></td>
<td>redefined to use endtag and shortref minimization</td>
</tr>
<tr>
<td><code>&lt;caption&gt;</code></td>
<td>New element</td>
</tr>
<tr>
<td><code>&lt;change&gt;</code></td>
<td>Changed content model</td>
</tr>
<tr>
<td><code>&lt;dialect&gt;</code></td>
<td>New element</td>
</tr>
</tbody>
</table>
<div>
  Changed content model, specific to speech
</div>
<div1>
  Simplified content model
</div1>
<div2>
  Simplified content model
</div2>
<div3>
  Simplified content model
</div3>
<div4>
  Simplified content model
</div4>
<item>
  Changed to disallow mixed content
</item>
<loc>
  New element
</loc>
<p>
  New simplified element (TEI p is renamed para)
</p>
<person>
  Simplified content model; added attributes
</person>
<poem>
  New element
</poem>
<quote>
  Changed content model to disallow mixed content
</quote>
<recording>
  Simplified content model; added attributes
</recording>
<s>
  Simplified content model
</s>
<shift>
  Mandatory attribute made optional
</shift>
<sp>
  Changed content model
</sp>
<stext>
  New element
</stext>
<text>
  Simplified content model
</text>
<trunc>
  New element
</trunc>
<unclear>
  Simplified content model; added attributes
</unclear>
<w>
  redefined to use endtag and shortref minimization
</w>

Finally, as mentioned above, there are six elements defined in the BNC DTD which do not appear in the TEI DTD. These must be added to the appropriate element class in the TEI content model. The following declarations in the bncMods.ent file have that effect:

```xml
<!ENTITY % x.chunk "p!">  
<!ENTITY % x.common "caption|poem|unclear|">  
<!ENTITY % x.divtop "align|">  
<!ENTITY % x.seg "trunc|">  
```

Detailed discussion of the extension mechanism and general conformance issues relating to the use of the TEI is given in chapters 28 and 29 of the TEI Guidelines and is not further discussed here. For an explanation of the mechanisms used above, the detailed presentation of the general organization of the TEI DTD provided in chapter 3 of the Guidelines may also be helpful.

### 9.2. Differences between the BNC DTD and CDIF

This section lists significant differences between the current BNC DTD and CDIF 1.0. It lists elements whose names have been changed, elements whose attributes have changed, and elements whose content has been changed in such a way that CDIF-conformant files will not parse against the new DTD.

The following CDIF elements have been given different names:

- `<avail>` is now `<availability>`
- `<biblskop>` is now `<biblScope>`
- `<biblstr>` is now `<bib1Struct>`
- `<bibnote>` is now `<note>`
- `<clasdecl>` is now `<classDecl>`
- `<corr>` is now an `<item>` within `<encodingDesc>`
- `<editdecl>` is now `<editorialDecl>`
9.2. Differences between the BNC DTD and CDIF

• `<ednstmt>` is now `<editionStmt>`
• `<encdesc>` is now `<encodingDesc>`
• `<header>` is now `<teiHeader>`
• `<hyph>` is now an `<item>` within `<encodingDesc>`
• `<partics>` is now `<particDesc>`
• `<profdesc>` is now `<profileDesc>`
• `<projdesc>` is now `<projectDesc>`
• `<pubstmt>` is now `<publicationStmt>`
• `<quot>` is now an `<item>` within `<encodingDesc>`
• `<rec>` is now `<recording>`
• `<recstmt>` is now `<recordingStmt>`
• `<reg>` is now `<corr>`
• `<relation>` is now `<particLinks>`
• `<revdesc>` is now `<revisionDesc>`
• `<segm>` is now an `<item>` within `<encodingDesc>`
• `<settdesc>` is now `<settingDesc>`
• `<srcdesc>` is now `<sourceDesc>`
• `<titstmt>` is now `<titleStmt>`
• `<txtclass>` is now `<textClass>`

The following elements have significantly different attributes

• `<activity>` has acquired the attribute `spont`, formerly present on its parent `<setting>`
• `<sic>`, `<gap>`, and `<corr>` have all acquired attributes `resp` (rather than `ed`)
• `<corr>` and `<sic>` no longer have a `cause` attribute
• `<gap>` has acquired the attribute `reason` (in place of `cause`)
• the complete attribute has been removed from `<text>` and `<stext>`
• `<w>` has a different set of values for its type attribute

The content models for elements in the BNC DTD are generally less restrictive than those of CDIF. In the following list, we specify only those elements where an element conforming to the CDIF model would not also conform to the BNC model.

<table>
<thead>
<tr>
<th>element</th>
<th>CDIF model</th>
<th>BNC model</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>&lt;address&gt;</code></td>
<td>#PCDATA</td>
<td>(addrLine+</td>
</tr>
<tr>
<td><code>&lt;avail&gt;</code></td>
<td>#PCDATA</td>
<td>(para)*</td>
</tr>
<tr>
<td><code>&lt;change&gt;</code></td>
<td>(date, respStmt+)</td>
<td>(date, respStmt+.para)</td>
</tr>
<tr>
<td><code>&lt;clasdecl&gt;</code></td>
<td>(category+)</td>
<td>(taxonomy+)</td>
</tr>
<tr>
<td><code>&lt;editdecl&gt;</code></td>
<td>(corr</td>
<td>quot</td>
</tr>
<tr>
<td><code>&lt;ednstmt&gt;</code></td>
<td>#PCDATA</td>
<td>((edition, respStmt*)</td>
</tr>
<tr>
<td><code>&lt;encDesc&gt;</code></td>
<td>(projDesc, sampDecl</td>
<td>editDecl)*, refsDecl+, tagsDecl?, clasDecl?)</td>
</tr>
<tr>
<td><code>&lt;imprint&gt;</code></td>
<td>(pubPlace</td>
<td>name</td>
</tr>
<tr>
<td><code>&lt;langusg&gt;</code></td>
<td>#PCDATA</td>
<td>(para</td>
</tr>
</tbody>
</table>
This section consists of a series of tables identifying a number of codes used in various aspects of the corpus and its encoding.

The following code tables are provided:

- **10.1 (Elements defined by the BNC DTD)** lists all SGML elements used in the corpus, with a brief description of each
- **10.2 (Character entities defined by the BNC DTD)** lists all SGML entities used in the corpus, with a brief description of each
- **10.3 (Division types)** lists all values actually used in the corpus for the type attribute on division elements (<div1>, <div2> etc.)
- **10.4 (Rendition codes)** lists all values used in the corpus for the r (rendition) attribute, chiefly on <hi> elements, to indicate typographic rendering of the source
- **10.5 (Voice quality codes)** lists all values used in the corpus for the new attribute on the <shift> element, to indicate changes in voice quality for spoken texts
- **10.6 (Regional codes)** lists the codes used to identify regional origins of participants, as specified in the <person> element in the header
- **10.7 (Relationship codes)** lists the codes used to identify relationships documented between participants, as specified in the <relation> element in the header
- **10.9 (Word class codes)** lists all part of speech codes in the C5 tagset, used to specify the linguistic category for all <w> and <c> elements

In addition, a list of ‘non-orthographic words’ recognized by the CLAWS system (i.e. multiword items and clitics) which was included in this section in the first edition of this document is now available in the accompanying Manual to accompany The British National Corpus (Version 2) with Improved Word-class Tagging by Geoffrey Leech and Nicholas Smith.

The list of text classification codes present in the first edition of this document is now included only as part of the presentation of the corpus header in section 11.2 (The BNC corpus header).
10.1. Elements defined by the BNC DTD

The following list gives a brief description of each element defined in the BNC document type definition (DTD). Elements are listed in alphabetical order. Descriptions prefixed by “(H)” are for elements which appear only in the text headers. Counts are given for elements occurring within texts.

- `<activity>` (H) participants’ activity during recording
- `<address>` (H) postal or other address
- `<align>` alignment map for synchronizing overlap points (3461)
- `<analytic>` (H) analytic bibliographic entry
- `<author>` (H) author in bibliographic entry
- `<avail>` (H) availability code for file
- `<bibl>` loosely structured bibliographic reference (1037)
- `<biblScope>` (H) page range within bibliographic entry
- `<biblStruct>` (H) structured bibliographic entry
- `<bnc>` the BNC itself
- `<bncDoc>` an individual text in the BNC
- `<body>` the body of a written text in the BNC (3136)
- `<c>` a punctuation mark (13620069)
- `<caption>` a floating heading or caption (89935)
- `<catDesc>` (H) description of a category
- `<category>` (H) a category-value pair
- `<catRef>` (H) category codes applicable to a text
- `<change>` (H) change note
- `<classDecl>` (H) description of classification scheme
- `<corr>` editorial correction (8323)
- `<creation>` (H) information about creation of a text
- `<date>` (H) a date
- `<classCode>` (H) externally-defined classification code for a text
- `<div>` any subdivision of a spoken text (3779)
- `<div1>` first-level subdivision of a written text (84777)
- `<div2>` second-level subdivision of a written text (72697)
- `<div3>` third-level subdivision of a written text (38122)
- `<div4>` fourth-level subdivision of a written text (12506)
- `<editorialDecl>` (H) descriptions of editorial policies
- `<edition>` (H) edition in a bibliographic entry
- `<editionStmt>` (H) information about a particular edition
- `<encodingDesc>` (H) encoding description
- `<event>` non-verbal event within a spoken text (6565)
- `<extent>` (H) size of a corpus text
- `<fileDesc>` (H) documentation of an electronic text
- `<gap>` a spot where part of source text has been omitted (95959)
- `<head>` any form of heading or title (222876)
- `<hi>` typographically highlighted phrase (210927)
- `<idno>` (H) identifying number for a text
- `<imprint>` (H) imprint within a bibliographic entry
- `<item>` item within a list (117207)
- `<keyWords>` (H) descriptive keywords for topics of a text
- `<l>` line of verse (51559)
<label> label of a list item (65664)
<langUsage> (H) description of languages used in a text
<b> line break in printed source (169)
<lg> group of verse lines (1)
<list> list of items (19758)
<loc> synchronisation point within an alignment map (244975)
<locale> (H) description of a place where speech recorded
<monogr> (H) monographic bibliographic entry
<name> (H) name of place where speech recorded
<note> note or comment of any kind (17206)
<p> paragraph in written text (1515002)
<particDesc> (H) description of spoken text participants
<pause> noticeable pause in spoken text (217916)
<pb> page break in written text (153642)
<person> (H) information about a speaker
<poem> group of verse lines in a written text (3048)
<profileDesc> (H) additional information about a text
<projectDesc> (H) background information about BNC project
<ptr> link to a displaced element or to synchronisation point (578248)
<pubPlace> (H) place of publication within bibliographic entry
<quote> quotation from some other work (15221)
<recording> (H) information about a single recording
<recordingStmt> (H) information about the recordings from which a transcript was made
<refsDecl> (H) description of reference system used
<reg> an editorial regularization (8363)
<relation> (H) relationship between participants in a spoken text
<resp> (H) nature of responsibility
<respStmt> (H) statement of responsibility in a bibliographic entry
<revisionDesc> (H) revision description
<s> sentence-like linguistic segment (6053093)
<salute> salutation or greeting (444)
<samplingDecl> (H) description of sampling policy
<settingDesc> (H) description of the settings in which speech occurs
<setting> (H) an individual setting in which speech occurs
<shift> change in voice quality (36216)
<sic> apparently erroneous transcription (7797)
<sp> speech in a written text (29858)
<spkr> speaker of a speech in a written text (23708)
<sourceDesc> (H) description of the source for a text
<stage> stage direction in a written text (508)
<stext> an individual spoken text (918)
<tagsDecl> (H) list of tags used in a particular text
<tagUsage> (H) count for a particular tag in a text
<teiHeader> meta-information describing a corpus text
<term> (H) individual term in a list of keywords
<text> an individual written text (3136)
10.2. Character entities defined by the BNC DTD

The following list gives a brief description of each character entity used within the text of the BNC. Declarations for these entities may be found either in standard entity sets or from the entity definitions supplied as part of the BNC document type definition, in the file BNCents.dtd. In either case, system specific values should be supplied for the characters described below. The number in parentheses indicates the number of times this entity reference appears in the current version of the corpus.

Table 21. Character entities used in the BNC

<table>
<thead>
<tr>
<th>entity</th>
<th>description</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aacute</td>
<td>capital A, acute accent</td>
<td>51</td>
</tr>
<tr>
<td>aacute</td>
<td>small a, acute accent</td>
<td>2145</td>
</tr>
<tr>
<td>abreve</td>
<td>small A, breve</td>
<td>4</td>
</tr>
<tr>
<td>Acirc</td>
<td>capital A, circumflex accent</td>
<td>8</td>
</tr>
<tr>
<td>acirc</td>
<td>small a, circumflex accent</td>
<td>772</td>
</tr>
<tr>
<td>acute</td>
<td>acute accent</td>
<td>3</td>
</tr>
<tr>
<td>AElig</td>
<td>capital AE diphthong (ligature)</td>
<td>324</td>
</tr>
<tr>
<td>aelig</td>
<td>small ae diphthong (ligature)</td>
<td>232</td>
</tr>
<tr>
<td>agr</td>
<td>small alpha, Greek</td>
<td>1237</td>
</tr>
<tr>
<td>Agrave</td>
<td>capital A, grave accent</td>
<td>43</td>
</tr>
<tr>
<td>agrave</td>
<td>small a, grave accent</td>
<td>878</td>
</tr>
<tr>
<td>Amacr</td>
<td>capital A, macron</td>
<td>20</td>
</tr>
<tr>
<td>amacr</td>
<td>small a, macron</td>
<td>581</td>
</tr>
<tr>
<td>amp</td>
<td>ampersand</td>
<td>18931</td>
</tr>
<tr>
<td>aogon</td>
<td>small a, ogonek</td>
<td>2</td>
</tr>
<tr>
<td>ape</td>
<td>approximate, equals</td>
<td>5</td>
</tr>
<tr>
<td>Aring</td>
<td>capital A, ring</td>
<td>141</td>
</tr>
<tr>
<td>aring</td>
<td>small a, ring</td>
<td>65</td>
</tr>
<tr>
<td>ast</td>
<td>asterisk</td>
<td>919</td>
</tr>
<tr>
<td>atilde</td>
<td>small a, tilde</td>
<td>244</td>
</tr>
<tr>
<td>Auml</td>
<td>capital A, dieresis or umlaut mark</td>
<td>6</td>
</tr>
<tr>
<td>auml</td>
<td>small a, dieresis or umlaut mark</td>
<td>967</td>
</tr>
<tr>
<td>Bgr</td>
<td>capital Beta, Greek</td>
<td>186</td>
</tr>
<tr>
<td>bgr</td>
<td>small beta, Greek</td>
<td>1122</td>
</tr>
<tr>
<td>bquo</td>
<td>normalized begin quote mark</td>
<td>771009</td>
</tr>
<tr>
<td>bsol</td>
<td>reverse solidus</td>
<td>226</td>
</tr>
<tr>
<td>bull</td>
<td>round bullet, filled</td>
<td>2150</td>
</tr>
<tr>
<td>cacute</td>
<td>small c, acute accent</td>
<td>186</td>
</tr>
<tr>
<td>Character entity</td>
<td>Description</td>
<td>Hex Code</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------</td>
<td>----------</td>
</tr>
<tr>
<td>Ccaron</td>
<td>capital C, caron</td>
<td>31</td>
</tr>
<tr>
<td>ccaron</td>
<td>small c, caron</td>
<td>143</td>
</tr>
<tr>
<td>Ccedil</td>
<td>capital C, cedilla</td>
<td>17</td>
</tr>
<tr>
<td>ccedil</td>
<td>small c, cedilla</td>
<td>1320</td>
</tr>
<tr>
<td>ccirc</td>
<td>small c, circumflex accent</td>
<td>2</td>
</tr>
<tr>
<td>cent</td>
<td>cent sign</td>
<td>3</td>
</tr>
<tr>
<td>check</td>
<td>tick, check mark</td>
<td>13</td>
</tr>
<tr>
<td>cir</td>
<td>circle, open</td>
<td>21</td>
</tr>
<tr>
<td>circ</td>
<td>circumflex accent</td>
<td>100</td>
</tr>
<tr>
<td>commat</td>
<td>commercial at</td>
<td>189</td>
</tr>
<tr>
<td>copy</td>
<td>copyright sign</td>
<td>65</td>
</tr>
<tr>
<td>darr</td>
<td>downward arrow</td>
<td>17</td>
</tr>
<tr>
<td>dash</td>
<td>hyphen (true graphic)</td>
<td>1</td>
</tr>
<tr>
<td>dcaron</td>
<td>small d, caron</td>
<td>1</td>
</tr>
<tr>
<td>deg</td>
<td>degree sign</td>
<td>4067</td>
</tr>
<tr>
<td>Dgr</td>
<td>capital Delta, Greek</td>
<td>251</td>
</tr>
<tr>
<td>dgr</td>
<td>small delta, Greek</td>
<td>151</td>
</tr>
<tr>
<td>die</td>
<td>dieresis</td>
<td>13</td>
</tr>
<tr>
<td>divide</td>
<td>divide sign</td>
<td>58</td>
</tr>
<tr>
<td>dollar</td>
<td>dollar sign</td>
<td>24552</td>
</tr>
<tr>
<td>dstrok</td>
<td>small d, stroke</td>
<td>18</td>
</tr>
<tr>
<td>dtrif</td>
<td>dn tri, filled</td>
<td>2</td>
</tr>
<tr>
<td>Eacute</td>
<td>capital E, acute accent</td>
<td>269</td>
</tr>
<tr>
<td>eacute</td>
<td>small e, acute accent</td>
<td>16078</td>
</tr>
<tr>
<td>Ecaron</td>
<td>capital E, caron</td>
<td>2</td>
</tr>
<tr>
<td>ecaron</td>
<td>small e, caron</td>
<td>67</td>
</tr>
<tr>
<td>Ecirc</td>
<td>capital E, circumflex accent</td>
<td>2</td>
</tr>
<tr>
<td>ecirc</td>
<td>small e, circumflex accent</td>
<td>718</td>
</tr>
<tr>
<td>eegr</td>
<td>small eta, Greek</td>
<td>48</td>
</tr>
<tr>
<td>Egr</td>
<td>capital Epsilon, Greek</td>
<td>170</td>
</tr>
<tr>
<td>egr</td>
<td>small epsilon, Greek</td>
<td>207</td>
</tr>
<tr>
<td>Egrave</td>
<td>capital E, grave accent</td>
<td>14</td>
</tr>
<tr>
<td>egrave</td>
<td>small e, grave accent</td>
<td>2577</td>
</tr>
<tr>
<td>emacr</td>
<td>small e, macron</td>
<td>4</td>
</tr>
<tr>
<td>eogon</td>
<td>small e, ogonek</td>
<td>6</td>
</tr>
<tr>
<td>equals</td>
<td>equals sign</td>
<td>3</td>
</tr>
<tr>
<td>equo</td>
<td>normalized end quote mark</td>
<td>752621</td>
</tr>
<tr>
<td>eth</td>
<td>small eth, Icelandic</td>
<td>4</td>
</tr>
<tr>
<td>Euml</td>
<td>capital E, dieresis or umlaut mark</td>
<td>15</td>
</tr>
<tr>
<td>euml</td>
<td>small e, dieresis or umlaut mark</td>
<td>482</td>
</tr>
<tr>
<td>flat</td>
<td>musical flat</td>
<td>154</td>
</tr>
<tr>
<td>formula</td>
<td>mathematical formula</td>
<td>6466</td>
</tr>
<tr>
<td>frac12</td>
<td>fraction one-half</td>
<td>2795</td>
</tr>
<tr>
<td>frac13</td>
<td>fraction one-third</td>
<td>68</td>
</tr>
<tr>
<td>frac14</td>
<td>fraction one-quarter</td>
<td>575</td>
</tr>
<tr>
<td>frac15</td>
<td>fraction one-fifth</td>
<td>20</td>
</tr>
<tr>
<td>frac16</td>
<td>fraction one-sixth</td>
<td>6</td>
</tr>
<tr>
<td>frac17</td>
<td>fraction one-seventh</td>
<td>2</td>
</tr>
<tr>
<td>Character Entity</td>
<td>Description</td>
<td>Value</td>
</tr>
<tr>
<td>------------------</td>
<td>------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>frac18</td>
<td>Fraction one-eighth</td>
<td>60</td>
</tr>
<tr>
<td>frac19</td>
<td>Fraction one-ninth</td>
<td>1</td>
</tr>
<tr>
<td>frac23</td>
<td>Fraction two-thirds</td>
<td>50</td>
</tr>
<tr>
<td>frac25</td>
<td>Fraction two-fifths</td>
<td>9</td>
</tr>
<tr>
<td>frac34</td>
<td>Fraction three-quarters</td>
<td>325</td>
</tr>
<tr>
<td>frac35</td>
<td>Fraction three-fifths</td>
<td>5</td>
</tr>
<tr>
<td>frac38</td>
<td>Fraction three-eighths</td>
<td>52</td>
</tr>
<tr>
<td>frac45</td>
<td>Fraction four-fifths</td>
<td>5</td>
</tr>
<tr>
<td>frac47</td>
<td>Fraction four-sevenths</td>
<td>1</td>
</tr>
<tr>
<td>frac56</td>
<td>Fraction five-sixths</td>
<td>1</td>
</tr>
<tr>
<td>frac58</td>
<td>Fraction five-eighths</td>
<td>33</td>
</tr>
<tr>
<td>frac78</td>
<td>Fraction seven-eighths</td>
<td>7</td>
</tr>
<tr>
<td>ft</td>
<td>Feet indicator</td>
<td>630</td>
</tr>
<tr>
<td>ge</td>
<td>Greater-than-or-equal</td>
<td>18</td>
</tr>
<tr>
<td>Ggr</td>
<td>Capital gamma, Greek</td>
<td>33</td>
</tr>
<tr>
<td>ggr</td>
<td>Small gamma, Greek</td>
<td>499</td>
</tr>
<tr>
<td>grave</td>
<td>Grave accent</td>
<td>2</td>
</tr>
<tr>
<td>Gt</td>
<td>Dbl greater-than sign</td>
<td>8</td>
</tr>
<tr>
<td>gt</td>
<td>Greater-than sign</td>
<td>1102</td>
</tr>
<tr>
<td>half</td>
<td>Fraction one-half</td>
<td>74</td>
</tr>
<tr>
<td>hearts</td>
<td>Heart suit symbol</td>
<td>1</td>
</tr>
<tr>
<td>hellip</td>
<td>Ellipsis (horizontal)</td>
<td>77286</td>
</tr>
<tr>
<td>horbar</td>
<td>Horizontal bar</td>
<td>2</td>
</tr>
<tr>
<td>hstrok</td>
<td>Small h, stroke</td>
<td>2</td>
</tr>
<tr>
<td>Iacute</td>
<td>Capital I, acute accent</td>
<td>3</td>
</tr>
<tr>
<td>iacute</td>
<td>Small i, acute accent</td>
<td>1278</td>
</tr>
<tr>
<td>Icirc</td>
<td>Capital I, circumflex accent</td>
<td>21</td>
</tr>
<tr>
<td>icirc</td>
<td>Small i, circumflex accent</td>
<td>225</td>
</tr>
<tr>
<td>iexcl</td>
<td>Inverted exclamation mark</td>
<td>21</td>
</tr>
<tr>
<td>igr</td>
<td>Small iota, Greek</td>
<td>2</td>
</tr>
<tr>
<td>igrave</td>
<td>Small i, grave accent</td>
<td>39</td>
</tr>
<tr>
<td>imacr</td>
<td>Small i, macron</td>
<td>10</td>
</tr>
<tr>
<td>infin</td>
<td>Infinity</td>
<td>15</td>
</tr>
<tr>
<td>ins</td>
<td>Inches indicator</td>
<td>2306</td>
</tr>
<tr>
<td>iquest</td>
<td>Inverted question mark</td>
<td>11</td>
</tr>
<tr>
<td>Iuml</td>
<td>Capital I, dieresis or umlaut mark</td>
<td>1</td>
</tr>
<tr>
<td>iuml</td>
<td>Small i, dieresis or umlaut mark</td>
<td>507</td>
</tr>
<tr>
<td>kgr</td>
<td>Small kappa, Greek</td>
<td>29</td>
</tr>
<tr>
<td>khgr</td>
<td>Small chi, Greek</td>
<td>300</td>
</tr>
<tr>
<td>Lacute</td>
<td>Capital L, acute accent</td>
<td>2</td>
</tr>
<tr>
<td>lacute</td>
<td>Small l, acute accent</td>
<td>2</td>
</tr>
<tr>
<td>larr</td>
<td>Leftward arrow</td>
<td>1</td>
</tr>
<tr>
<td>lcub</td>
<td>Left curly bracket</td>
<td>345</td>
</tr>
<tr>
<td>le</td>
<td>Less-than-or-equal</td>
<td>23</td>
</tr>
<tr>
<td>lgr</td>
<td>Small lambda, Greek</td>
<td>104</td>
</tr>
<tr>
<td>lowbar</td>
<td>Low line</td>
<td>134</td>
</tr>
<tr>
<td>lsqb</td>
<td>Left square bracket</td>
<td>34752</td>
</tr>
<tr>
<td>Lstrok</td>
<td>Capital L, stroke</td>
<td>3</td>
</tr>
<tr>
<td>Character</td>
<td>Description</td>
<td>Frequency</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>lstrok</td>
<td>small l, stroke</td>
<td>25</td>
</tr>
<tr>
<td>Lt</td>
<td>double less-than sign</td>
<td>3</td>
</tr>
<tr>
<td>It</td>
<td>less-than sign</td>
<td>2295</td>
</tr>
<tr>
<td>mdash</td>
<td>em dash</td>
<td>275695</td>
</tr>
<tr>
<td>Mgr</td>
<td>capital Mu, Greek</td>
<td>1</td>
</tr>
<tr>
<td>mgr</td>
<td>small mu, Greek</td>
<td>376</td>
</tr>
<tr>
<td>micro</td>
<td>micro sign</td>
<td>1487</td>
</tr>
<tr>
<td>middot</td>
<td>middle dot</td>
<td>253</td>
</tr>
<tr>
<td>nacute</td>
<td>small n, acute accent</td>
<td>21</td>
</tr>
<tr>
<td>natur</td>
<td>music natural</td>
<td>6</td>
</tr>
<tr>
<td>ncaron</td>
<td>small n, caron</td>
<td>27</td>
</tr>
<tr>
<td>ncedil</td>
<td>small n, cedilla</td>
<td>2</td>
</tr>
<tr>
<td>ndash</td>
<td>en dash</td>
<td>43489</td>
</tr>
<tr>
<td>ngr</td>
<td>small nu, Greek</td>
<td>88</td>
</tr>
<tr>
<td>Ntilde</td>
<td>capital N, tilde</td>
<td>4</td>
</tr>
<tr>
<td>ntilde</td>
<td>small n, tilde</td>
<td>771</td>
</tr>
<tr>
<td>num</td>
<td>number sign</td>
<td>138</td>
</tr>
<tr>
<td>Oacute</td>
<td>capital O, acute accent</td>
<td>17</td>
</tr>
<tr>
<td>oacute</td>
<td>small o, acute accent</td>
<td>1328</td>
</tr>
<tr>
<td>Ocirc</td>
<td>capital O, circumflex accent</td>
<td>7</td>
</tr>
<tr>
<td>ocirc</td>
<td>small o, circumflex accent</td>
<td>754</td>
</tr>
<tr>
<td>OElig</td>
<td>capital OE ligature</td>
<td>2</td>
</tr>
<tr>
<td>oelig</td>
<td>small oe ligature</td>
<td>55</td>
</tr>
<tr>
<td>Ogr</td>
<td>capital Omicron, Greek</td>
<td>11</td>
</tr>
<tr>
<td>ogr</td>
<td>small omicron, Greek</td>
<td>52</td>
</tr>
<tr>
<td>ograve</td>
<td>small o, grave accent</td>
<td>73</td>
</tr>
<tr>
<td>OHgr</td>
<td>capital Omega, Greek</td>
<td>1</td>
</tr>
<tr>
<td>ohgr</td>
<td>small omega, Greek</td>
<td>23</td>
</tr>
<tr>
<td>ohm</td>
<td>ohm sign</td>
<td>15</td>
</tr>
<tr>
<td>omacr</td>
<td>small o, macron</td>
<td>4</td>
</tr>
<tr>
<td>Oslash</td>
<td>capital O, slash</td>
<td>16</td>
</tr>
<tr>
<td>oslash</td>
<td>small o, slash</td>
<td>304</td>
</tr>
<tr>
<td>Otilde</td>
<td>capital O, tilde</td>
<td>1</td>
</tr>
<tr>
<td>otilde</td>
<td>small o, tilde</td>
<td>4</td>
</tr>
<tr>
<td>Ouml</td>
<td>capital O, dieresis or umlaut mark</td>
<td>344</td>
</tr>
<tr>
<td>ouml</td>
<td>small o, dieresis or umlaut mark</td>
<td>1284</td>
</tr>
<tr>
<td>percent</td>
<td>percent sign</td>
<td>144</td>
</tr>
<tr>
<td>Pgr</td>
<td>capital Pi, Greek</td>
<td>37</td>
</tr>
<tr>
<td>pgr</td>
<td>small pi, Greek</td>
<td>95</td>
</tr>
<tr>
<td>PHgr</td>
<td>capital Phi, Greek</td>
<td>9</td>
</tr>
<tr>
<td>phgr</td>
<td>small phi, Greek</td>
<td>107</td>
</tr>
<tr>
<td>plus</td>
<td>plus sign</td>
<td>198</td>
</tr>
<tr>
<td>plussmn</td>
<td>plus-or-minus sign</td>
<td>122</td>
</tr>
<tr>
<td>pound</td>
<td>pound sign</td>
<td>71698</td>
</tr>
<tr>
<td>Prime</td>
<td>double prime or second</td>
<td>59</td>
</tr>
<tr>
<td>prime</td>
<td>prime or minute</td>
<td>128</td>
</tr>
<tr>
<td>PSgr</td>
<td>capital Psi, Greek</td>
<td>15</td>
</tr>
<tr>
<td>psgr</td>
<td>small psi, Greek</td>
<td>16</td>
</tr>
</tbody>
</table>
## 10.2. Character entities defined by the BNC DTD

<table>
<thead>
<tr>
<th>Entity</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>quot</td>
<td>quotation mark</td>
<td>142126</td>
</tr>
<tr>
<td>racute</td>
<td>small r, acute accent</td>
<td>3</td>
</tr>
<tr>
<td>radic</td>
<td>surd = radical (square root)</td>
<td>9</td>
</tr>
<tr>
<td>rarr</td>
<td>rightward arrow</td>
<td>182</td>
</tr>
<tr>
<td>Rcaron</td>
<td>capital R, caron</td>
<td>1</td>
</tr>
<tr>
<td>rcaron</td>
<td>small r, caron</td>
<td>103</td>
</tr>
<tr>
<td>rcub</td>
<td>right curly bracket</td>
<td>343</td>
</tr>
<tr>
<td>reg</td>
<td>registered sign</td>
<td>11</td>
</tr>
<tr>
<td>rehy</td>
<td>maps to soft hyphen</td>
<td>3905</td>
</tr>
<tr>
<td>rgr</td>
<td>small rho, Greek</td>
<td>81</td>
</tr>
<tr>
<td>rsqb</td>
<td>right square bracket</td>
<td>34807</td>
</tr>
<tr>
<td>Sacute</td>
<td>capital S, acute accent</td>
<td>13</td>
</tr>
<tr>
<td>sacute</td>
<td>small s, acute accent</td>
<td>24</td>
</tr>
<tr>
<td>Scaron</td>
<td>capital S, caron</td>
<td>85</td>
</tr>
<tr>
<td>scaron</td>
<td>small s, caron</td>
<td>257</td>
</tr>
<tr>
<td>Scedil</td>
<td>capital S, cedilla</td>
<td>7</td>
</tr>
<tr>
<td>scedil</td>
<td>small s, cedilla</td>
<td>465</td>
</tr>
<tr>
<td>scirc</td>
<td>small s, circumflex accent</td>
<td>14</td>
</tr>
<tr>
<td>sect</td>
<td>section sign</td>
<td>52</td>
</tr>
<tr>
<td>Sgr</td>
<td>capital Sigma, Greek</td>
<td>13</td>
</tr>
<tr>
<td>sgr</td>
<td>small sigma, Greek</td>
<td>150</td>
</tr>
<tr>
<td>sharp</td>
<td>musical sharp</td>
<td>93</td>
</tr>
<tr>
<td>shilling</td>
<td>British shilling</td>
<td>228</td>
</tr>
<tr>
<td>sim</td>
<td>similar</td>
<td>68</td>
</tr>
<tr>
<td>sol</td>
<td>solidus</td>
<td>355</td>
</tr>
<tr>
<td>sup1</td>
<td>superscript one</td>
<td>2</td>
</tr>
<tr>
<td>sup2</td>
<td>superscript two</td>
<td>45</td>
</tr>
<tr>
<td>sup3</td>
<td>superscript three</td>
<td>10</td>
</tr>
<tr>
<td>szlig</td>
<td>small sharp s, German (sz ligature)</td>
<td>19</td>
</tr>
<tr>
<td>tcaron</td>
<td>small t, caron</td>
<td>1</td>
</tr>
<tr>
<td>tcedil</td>
<td>small t, cedilla</td>
<td>26</td>
</tr>
<tr>
<td>tgr</td>
<td>small tau, Greek</td>
<td>68</td>
</tr>
<tr>
<td>THgr</td>
<td>capital Theta, Greek</td>
<td>13</td>
</tr>
<tr>
<td>thgr</td>
<td>small theta, Greek</td>
<td>193</td>
</tr>
<tr>
<td>THORN</td>
<td>capital THORN, Icelandic</td>
<td>15</td>
</tr>
<tr>
<td>thorn</td>
<td>small thorn, Icelandic</td>
<td>13</td>
</tr>
<tr>
<td>times</td>
<td>multiply sign</td>
<td>2300</td>
</tr>
<tr>
<td>trade</td>
<td>trade mark sign</td>
<td>12</td>
</tr>
<tr>
<td>Uacute</td>
<td>capital U, acute accent</td>
<td>7</td>
</tr>
<tr>
<td>uacute</td>
<td>small u, acute accent</td>
<td>326</td>
</tr>
<tr>
<td>Ucirc</td>
<td>capital U, circumflex accent</td>
<td>1</td>
</tr>
<tr>
<td>ucirc</td>
<td>small u, circumflex accent</td>
<td>107</td>
</tr>
<tr>
<td>Ugr</td>
<td>capital Upsilon, Greek</td>
<td>1</td>
</tr>
<tr>
<td>ugr</td>
<td>small upsilon, Greek</td>
<td>3</td>
</tr>
<tr>
<td>ugrave</td>
<td>small u, grave accent</td>
<td>38</td>
</tr>
<tr>
<td>umacr</td>
<td>small u, macron</td>
<td>3</td>
</tr>
<tr>
<td>uml</td>
<td>umlaut mark</td>
<td>3</td>
</tr>
<tr>
<td>uring</td>
<td>small u, ring</td>
<td>9</td>
</tr>
</tbody>
</table>
10.4. Rendition codes

The following codes are used to indicate the kind of typographic rendition associated with an element which is typographically distinct in some way. These codes are mostly used as values for the `rend` attribute of the `<hi>` element, but may be used on any element bearing this attribute.

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uuml</td>
<td>capital U, dieresis or umlaut mark</td>
<td>25</td>
</tr>
<tr>
<td>uuml</td>
<td>small u, dieresis or umlaut mark</td>
<td>2246</td>
</tr>
<tr>
<td>verbar</td>
<td>vertical bar</td>
<td>269</td>
</tr>
<tr>
<td>wcirc</td>
<td>small w, circumflex accent</td>
<td>2</td>
</tr>
<tr>
<td>xgr</td>
<td>small xi, Greek</td>
<td>8</td>
</tr>
<tr>
<td>yacute</td>
<td>small y, acute accent</td>
<td>138</td>
</tr>
<tr>
<td>Ycirc</td>
<td>capital Y, circumflex accent</td>
<td>1</td>
</tr>
<tr>
<td>ycirc</td>
<td>small y, circumflex accent</td>
<td>6</td>
</tr>
<tr>
<td>yen</td>
<td>yen sign</td>
<td>120</td>
</tr>
<tr>
<td>Yuml</td>
<td>capital Y, dieresis or umlaut mark</td>
<td>2</td>
</tr>
<tr>
<td>yuml</td>
<td>small y, dieresis or umlaut mark</td>
<td>47</td>
</tr>
<tr>
<td>zacute</td>
<td>small z, acute accent</td>
<td>3</td>
</tr>
<tr>
<td>Zcaron</td>
<td>capital Z, caron</td>
<td>9</td>
</tr>
<tr>
<td>zcaron</td>
<td>small z, caron</td>
<td>67</td>
</tr>
<tr>
<td>zdot</td>
<td>small z, dot above</td>
<td>1</td>
</tr>
<tr>
<td>Zgr</td>
<td>capital Zeta, Greek</td>
<td>3</td>
</tr>
<tr>
<td>zgr</td>
<td>small zeta, Greek</td>
<td>33</td>
</tr>
</tbody>
</table>

10.3. Division types

The type attribute on each `<div1>`, `<div2>` (etc) element of a written text may be used to supply a value which characterizes the function of the corresponding subdivision in some way. The following table lists the values used and their frequencies in the current version of the Corpus:

<table>
<thead>
<tr>
<th>Type</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>advertisement</td>
<td>6</td>
</tr>
<tr>
<td>blurb</td>
<td>1</td>
</tr>
<tr>
<td>cartoon</td>
<td>2</td>
</tr>
<tr>
<td>commentary</td>
<td>1</td>
</tr>
<tr>
<td>contents</td>
<td>2</td>
</tr>
<tr>
<td>front</td>
<td>115</td>
</tr>
<tr>
<td>introduction</td>
<td>1</td>
</tr>
<tr>
<td>magazine</td>
<td>1</td>
</tr>
<tr>
<td>part</td>
<td>85</td>
</tr>
<tr>
<td>section</td>
<td>59</td>
</tr>
<tr>
<td>story</td>
<td>178</td>
</tr>
<tr>
<td>subsection</td>
<td>109</td>
</tr>
<tr>
<td>toc</td>
<td>1</td>
</tr>
<tr>
<td>advertisement</td>
<td>6</td>
</tr>
<tr>
<td>appendix</td>
<td>1</td>
</tr>
<tr>
<td>body text</td>
<td>3</td>
</tr>
<tr>
<td>chapter</td>
<td>1018</td>
</tr>
<tr>
<td>competition</td>
<td>1</td>
</tr>
<tr>
<td>element</td>
<td>1</td>
</tr>
<tr>
<td>headlines</td>
<td>2</td>
</tr>
<tr>
<td>item</td>
<td>12</td>
</tr>
<tr>
<td>paper</td>
<td>11</td>
</tr>
<tr>
<td>poster</td>
<td>1</td>
</tr>
<tr>
<td>short story</td>
<td>8</td>
</tr>
<tr>
<td>strip cartoon</td>
<td>1</td>
</tr>
<tr>
<td>title page</td>
<td>1</td>
</tr>
<tr>
<td>sub</td>
<td>2</td>
</tr>
<tr>
<td>title page obverse</td>
<td>1</td>
</tr>
</tbody>
</table>
10.5. Voice quality codes

Changes in voice quality in spoken texts are indicated by values for the <new> attribute on a <shift> element, at the point where the speaker’s voice change. The following list summarises the values used in the present version of the corpus:

- cheering
- crying
- eating
- giggling
- humming
- hummin the stripper’s song
- imitates woman’s voice
- imitating a monkey
- imitating a sexy woman’s voice
- imitating Chinese voice
- imitating drunken voice
- imitating Italian accent
- imitating man’s voice
- imitating posh voice
- imitating woman’s voice
- in a boyish voice
- in the distance
- laughing
- laughing+reading
- laughing+shouting
- mimicking
- mimicking American accent
- mimicking American accent from Wayne’s World
- mimicking an upper class person
- mimicking baby voice
- mimicking Birmingham accent
- mimicking Chinese speaking
- mimicking Cilla Black’s accent
- mimicking crying
- mimicking deep voice
- mimicking Donald Duck
- mimicking finance lady
- mimicking Geordie accent
- mimicking German accent
- mimicking girllie voice
- mimicking Henry Cooper
- mimicking Jamaican accent
- mimicking Manchester accent
- mimicking mentally handicapped
- mimicking northern accent
- mimicking Pakistani accent
- mimicking refined accent
- mimicking Scottish accent
- mimicking stupid man’s voice
- mimicking Swedish accent
- mimicking telephone voice
- mimicking the German accent
- mimicking whining
- mimicking witch
- mimicking Yorkshire accent
- mimicking+shrieking
- moaning
- mumbling
- muttering
- on telephone
- praying
- quoting
• raising voice
• rapping
• reading
• reading+laughing
• reading+shouting
• reading+whispering
• screaming
• shouting
• shouting+laughing
• shouting+spelling
• sighing
• singing
• singing+laughing
• singing+mimicking
• singing+shouting
• singing+whispering
• singing+yawning
• speaking as if mentally handicapped
• speaking dramatically
• speaking with mouth full
• spelling
• talking with mouth full
• whingeing
• whining
• whispering
• whispering+laughing
• yawning
• yawning+reading

10.6. Regional codes

The codes used to mark places of origin, regions, and dialects in the TEI Header are all derived from the same set of ISO 3-letter codes. The codes used are listed here:

CAN Canada
CHN China
DEU Germany
FRA France
GBR United Kingdom
IND India
IRL Ireland
USA United States
XXX Unknown
ZZG Europe
XDE accent: German
XEA accent: East Anglia
XFR accent: French
XHC accent: Home Counties
XHM accent: Humberside
XIR accent: Irish
XIS accent: Indian subcontinent
XLC accent: Lancashire
XLO accent: London
XMC accent: central Midlands
XMD accent: Merseyside
XME accent: north-east Midlands
XMI accent: Midlands
XMS accent: south Midlands
XMW accent: north-west Midlands
XNC accent: central northern England
XNE accent: north-east England
XNO accent: northern England
XOT accent: other or unidentifiable
XSD accent: Scottish
XSL accent: lower south-west England
XSS accent: central south-west England
XSU accent: upper south-west England
XUR accent: European
XUS accent: U.S.A.
XWA accent: Welsh
XWE accent: West Indian
10.7. Relationship codes
Where relationships between individual participants in spoken texts can be identified, they will be specified by means of the <relation> element within the text header (as discussed in section 8.3.3 (The participant description)). The type attribute of this element may take any of the values listed below. The number in parentheses indicates the number of times this value appears in the current version of the corpus.

- acquaint acquaintance (6)
- audience (4)
- aunt (8)
- aunt-i-l aunt-in-law (1)
- b-friend boyfriend (5)
- b-i-l brother-in-law (13)
- b-sitter baby sitter (2)
- brother (53)
- chairman (8)
- child (2)
- church-m church member (1)
- cl-m-i-l common law mother-in-law (1)
- client (1)
- collegu colleague (123)
- cous-i-l cousin-in-law (1)
- cousin (7)
- customer (3)
- d-i-l daughter-in-law (11)
- daughter (84)
- doctor (77)
- employee (4)
- employer (9)
- f-i-l father-in-law (16)
- father (73)
- fiance (1)
- fiancee (2)
- friend (123)
- g-aunt great-aunt (1)
- g-daught granddaughter (15)
- g-fath grand-father (11)
- g-friend girlfriend (5)
- g-moth grandmother (21)
- g-niece great-niece (1)
- g-son grandson (17)
- gg-daugh great-granddaughter (1)
- gg-moth great-grandmother (1)
- hairdres hairdresser (1)
- host (1)
- housekee housekeeper (1)
- husband (103)
- intervee interviewee (42)
- lecturer (4)
- m-i-l mother-in-law (21)
- mother (117)
- neighbour neighbour (13)
- neph-i-l nephew-in-law (1)
- nephew (7)
- niece (9)
- parent (5)
- patient (76)
- s-daught step-daughter (1)
- s-father step-father (1)
- secretar secretary (3)
- server (2)
- sib-i-l sibling-in-law (1)
- sibling (1)
- sis-i-l sister-in-law (12)
- sister (48)
- son (71)
- son-i-l son-in-law (18)
- speaker (9)
- stranger (13)
- student (31)
- teacher (26)
- trainee (1)
- trainer (2)
- tutor (4)
- uncle (6)
- visitor (2)
- wife (104)

10.8. Text and genre classification codes
Texts are classified in several different ways in the BNC, as described in section 8.3.5 (Text classification). Each text carries a number of text classification codes, specified a string of
values on the target attribute of its `<catRefs>` element. Possible values for these codes and their significance are listed in the corpus header (see 11.2 (The BNC corpus header)). These values are also used in the BNC indexing files described in section 11.3 (Creating a subcorpus) and distribution tables showing the number of texts, words, and sentences classified under most of them are given above in section 2 (Design of the corpus).

One of the codes listed below is also supplied for each text as the content of a `<classCode>` element in its text header, as an alternative way of characterising each text. Full details of the analysis scheme used and its rationale are provided in an article by David Lee (Genres, registers, text types and styles: clarifying the concepts and navigating a path through the BNC Jungle, to be published in Language Learning and Technology, vol 5 no 3, September 2001) who has also generously agreed to make some of the results of this work available with the current release of the BNC.

Table 23. Genre classification for spoken texts

<table>
<thead>
<tr>
<th>code</th>
<th>texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>S_brdcast_discussn</td>
<td>54</td>
</tr>
<tr>
<td>S_brdcast_documentary</td>
<td>10</td>
</tr>
<tr>
<td>S_brdcast_news</td>
<td>12</td>
</tr>
<tr>
<td>S_classroom</td>
<td>59</td>
</tr>
<tr>
<td>S_consult</td>
<td>128</td>
</tr>
<tr>
<td>S_conv</td>
<td>153</td>
</tr>
<tr>
<td>S_courtroom</td>
<td>13</td>
</tr>
<tr>
<td>S_demonstratn</td>
<td>6</td>
</tr>
<tr>
<td>S_interview</td>
<td>13</td>
</tr>
<tr>
<td>S_interview_oral_history</td>
<td>119</td>
</tr>
<tr>
<td>S_lect_commerce</td>
<td>3</td>
</tr>
<tr>
<td>S_lect_humanities_arts</td>
<td>4</td>
</tr>
<tr>
<td>S_lect_nat_science</td>
<td>4</td>
</tr>
<tr>
<td>S_lect_polit_law_edu</td>
<td>7</td>
</tr>
<tr>
<td>S_lect_soc_science</td>
<td>13</td>
</tr>
<tr>
<td>S_meeting</td>
<td>132</td>
</tr>
<tr>
<td>S_parliament</td>
<td>6</td>
</tr>
<tr>
<td>S_pub_debate</td>
<td>16</td>
</tr>
<tr>
<td>S_sermon</td>
<td>16</td>
</tr>
<tr>
<td>S_speech_scripted</td>
<td>26</td>
</tr>
<tr>
<td>S_speech_unscripted</td>
<td>51</td>
</tr>
<tr>
<td>S_sportslive</td>
<td>4</td>
</tr>
<tr>
<td>S_tutorial</td>
<td>18</td>
</tr>
<tr>
<td>S_unclassified</td>
<td>44</td>
</tr>
</tbody>
</table>

Table 24. Genre classification for written texts

<table>
<thead>
<tr>
<th>code</th>
<th>texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>W_ac_humanities_arts</td>
<td>87</td>
</tr>
<tr>
<td>W_ac_medicine</td>
<td>24</td>
</tr>
<tr>
<td>W_ac_nat_science</td>
<td>43</td>
</tr>
<tr>
<td>W_ac_polit_law_edu</td>
<td>187</td>
</tr>
<tr>
<td>W_ac_soc_science</td>
<td>142</td>
</tr>
</tbody>
</table>
### 10.9. Word class codes

A full discussion of the principles and practice underlying the CLAWS word class annotation scheme used in the BNC is provided by the document *Manual to accompany The British
**National Corpus (Version 2) with Improved Word-class Tagging**, which is distributed with the BNC World Edition in HTML format.

For convenience, a brief list of the codes used by this scheme extracted from that manual is also provided here.

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AJ0</td>
<td>Adjective (general or positive) (e.g. good, old, beautiful)</td>
</tr>
<tr>
<td>AJC</td>
<td>Comparative adjective (e.g. better, older)</td>
</tr>
<tr>
<td>AJS</td>
<td>Superlative adjective (e.g. best, oldest)</td>
</tr>
<tr>
<td>AT0</td>
<td>Article (e.g. the, a, an, no)</td>
</tr>
<tr>
<td>AV0</td>
<td>General adverb: an adverb not subclassified as AVP or AVQ (see below) (e.g. often, well, longer (adv.), furthest)</td>
</tr>
<tr>
<td>AVP</td>
<td>Adverb particle (e.g. up, off, out)</td>
</tr>
<tr>
<td>AVQ</td>
<td>Wh-adverb (e.g. when, where, how, why, wherever)</td>
</tr>
<tr>
<td>CJC</td>
<td>Coordinating conjunction (e.g. and, or, but)</td>
</tr>
<tr>
<td>CJS</td>
<td>Subordinating conjunction (e.g. although, when)</td>
</tr>
<tr>
<td>CJT</td>
<td>The subordinating conjunction that</td>
</tr>
<tr>
<td>CRD</td>
<td>Cardinal number (e.g. one, 3, fifty-five, 3609)</td>
</tr>
<tr>
<td>CPS</td>
<td>Possessive determiner-pronoun (e.g. your, their, his)</td>
</tr>
<tr>
<td>DT0</td>
<td>General determin-pronoun: i.e. a determiner-pronoun which is not a DTQ or an AT0.</td>
</tr>
<tr>
<td>DTQ</td>
<td>Wh-determiner-pronoun (e.g. which, what, whose, whichever)</td>
</tr>
<tr>
<td>EX0</td>
<td>Existential there, i.e. there occurring in the there is ... or there are ... construction</td>
</tr>
<tr>
<td>ITJ</td>
<td>Interjection or other isolate (e.g. oh, yes, mhm, wow)</td>
</tr>
<tr>
<td>NN0</td>
<td>Common noun, neutral for number (e.g. aircraft, data, committee)</td>
</tr>
<tr>
<td>NN1</td>
<td>Singular common noun (e.g. pencil, goose, time, revelation)</td>
</tr>
<tr>
<td>NN2</td>
<td>Plural common noun (e.g. pencils, geese, times, revelations)</td>
</tr>
<tr>
<td>NP0</td>
<td>Proper noun (e.g. London, Michael, Mars, IBM)</td>
</tr>
<tr>
<td>ORD</td>
<td>Ordinal numeral (e.g. first, sixth, 77th, last)</td>
</tr>
<tr>
<td>PNP</td>
<td>Personal pronoun (e.g. I, you, them, ours)</td>
</tr>
<tr>
<td>PNQ</td>
<td>Wh-pronoun (e.g. who, whoever, whom)</td>
</tr>
<tr>
<td>PNX</td>
<td>Reflexive pronoun (e.g. myself, yourself, itself, ourselves)</td>
</tr>
<tr>
<td>POS</td>
<td>The possessive or genitive marker ‘s or ’</td>
</tr>
<tr>
<td>PRF</td>
<td>The preposition of</td>
</tr>
<tr>
<td>PRP</td>
<td>Preposition (except for of) (e.g. about, at, in, on, on behalf of, with)</td>
</tr>
<tr>
<td>PUL</td>
<td>Punctuation: left bracket - i.e. ( or [</td>
</tr>
<tr>
<td>PUN</td>
<td>Punctuation: general separating mark - i.e. . , ! , : ; - or ?</td>
</tr>
<tr>
<td>PUQ</td>
<td>Punctuation: quotation mark - i.e. ’ or “</td>
</tr>
<tr>
<td>PUR</td>
<td>Punctuation: right bracket - i.e. ) or ]</td>
</tr>
<tr>
<td>TO0</td>
<td>Infinitive marker to</td>
</tr>
<tr>
<td>UNC</td>
<td>Unclassified items which are not appropriately considered as items of the English lexicon.</td>
</tr>
<tr>
<td>VBB</td>
<td>The present tense forms of the verb BE, except for is, ‘s: i.e. am, are, ’m, ’re and be [subjunctive or imperative]</td>
</tr>
<tr>
<td>VBD</td>
<td>The past tense forms of the verb BE: was and were</td>
</tr>
</tbody>
</table>
10.9. Word class codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>VBG</td>
<td>The -ing form of the verb BE: being</td>
</tr>
<tr>
<td>VBI</td>
<td>The infinitive form of the verb BE: be</td>
</tr>
<tr>
<td>VBN</td>
<td>The past participle form of the verb BE: been</td>
</tr>
<tr>
<td>VBZ</td>
<td>The -s form of the verb BE: is, 's</td>
</tr>
<tr>
<td>VDB</td>
<td>The finite base form of the verb BE: do</td>
</tr>
<tr>
<td>VDD</td>
<td>The past tense form of the verb DO: did</td>
</tr>
<tr>
<td>VDG</td>
<td>The -ing form of the verb DO: doing</td>
</tr>
<tr>
<td>VD1</td>
<td>The infinitive form of the verb DO: do</td>
</tr>
<tr>
<td>VDN</td>
<td>The past participle form of the verb DO: done</td>
</tr>
<tr>
<td>VDZ</td>
<td>The -s form of the verb DO: does, 's</td>
</tr>
<tr>
<td>VHB</td>
<td>The finite base form of the verb HAVE: have, 've</td>
</tr>
<tr>
<td>VHD</td>
<td>The past tense form of the verb HAVE: had, 'd</td>
</tr>
<tr>
<td>VHG</td>
<td>The -ing form of the verb HAVE: having</td>
</tr>
<tr>
<td>VHI</td>
<td>The infinitive form of the verb HAVE: have</td>
</tr>
<tr>
<td>VHN</td>
<td>The past participle form of the verb HAVE: had</td>
</tr>
<tr>
<td>VHZ</td>
<td>The -s form of the verb HAVE: has, 's</td>
</tr>
<tr>
<td>VM0</td>
<td>Modal auxiliary verb (e.g. will, would, can, could, 'll, 'd)</td>
</tr>
<tr>
<td>VVB</td>
<td>The finite base form of lexical verbs (e.g. forget, send, live, return)</td>
</tr>
<tr>
<td>VVD</td>
<td>The past tense form of lexical verbs (e.g. forgot, sent, lived, returned)</td>
</tr>
<tr>
<td>VVG</td>
<td>The -ing form of lexical verbs (e.g. forgetting, sending, living, returning)</td>
</tr>
<tr>
<td>VVI</td>
<td>The infinitive form of lexical verbs (e.g. forget, send, live, return)</td>
</tr>
<tr>
<td>VVN</td>
<td>The past participle form of lexical verbs (e.g. forgotten, sent, lived,</td>
</tr>
<tr>
<td></td>
<td>returned)</td>
</tr>
<tr>
<td>VVZ</td>
<td>The -s form of lexical verbs (e.g. forgets, sends, lives, returns)</td>
</tr>
<tr>
<td>XX0</td>
<td>The negative particle not or n’t</td>
</tr>
<tr>
<td>ZZ0</td>
<td>Alphabetical symbols (e.g. A, a, B, b, c, d)</td>
</tr>
</tbody>
</table>

In addition to the basic 57 codes tabulated above, the BNC World Edition uses thirty ‘portmanteau’ or ‘ambiguity’ tags. These are applied wherever the probabilities assigned by the CLAWS automatic tagger to its first and second choice tags were considered too low for reliable disambiguation. So, for example, the ambiguity tag AJ0–AV0 indicates that the choice between adjective (AJ0) and adverb (AV0) is left open, although the tagger has a preference for an adjective reading. The mirror tag, AV0–AJ0, again shows adjective-adverb ambiguity, but this time the more likely reading is the adverb.

The following table lists the ambiguity codes used in BNC World:

<table>
<thead>
<tr>
<th>Ambiguity tag</th>
<th>Ambiguous between</th>
<th>More probable tag</th>
</tr>
</thead>
<tbody>
<tr>
<td>AJ0–NN1</td>
<td>AJ0 or NN1</td>
<td>AJ0</td>
</tr>
<tr>
<td>AJ0–VVD</td>
<td>AJ0 or VVD</td>
<td>AJ0</td>
</tr>
<tr>
<td>AJ0–VVG</td>
<td>AJ0 or VVG</td>
<td>AJ0</td>
</tr>
<tr>
<td>AJ0–VVN</td>
<td>AJ0 or VVN</td>
<td>AJ0</td>
</tr>
<tr>
<td>AV0–AJ0</td>
<td>AV0 or AJ0</td>
<td>AV0</td>
</tr>
<tr>
<td>AVP–PRP</td>
<td>AVP or PRP</td>
<td>AVP</td>
</tr>
<tr>
<td>AVQ–CJS</td>
<td>AVQ or CJS</td>
<td>AVQ</td>
</tr>
<tr>
<td>CJS–AVQ</td>
<td>CJS or AVQ</td>
<td>CJS</td>
</tr>
</tbody>
</table>
11. Software for the BNC

A design goal of the original BNC project was that it should not be delivered in a format which was proprietary or which required the use of any particular piece of software. This, together with the desire to conform to emerging international standards, was a key factor in determining the choice of SGML as the vehicle for the corpus interchange format. Six years after this decision, SGML is still a widely used international standard format for which many public domain and commercial utilities exist. Indeed, in the shape of XML, which is a simplified version of the original standard, SGML now dominates development of the world wide web, and hence of most sectors of the information processing community. New XML software appears almost every week, and it has been adopted by current ‘major players’ from Sun and IBM to Microsoft.

That said, it must be recognised that the requirements of corpus linguists and others wishing to make use of the BNC are often rather specialist, and therefore unlikely to be supported by mainstream commercially produced software. For this and other reasons, the research user of the BNC should expect to have to do some programming. This is another reason behind the choice of SGML or XML as a vehicle for the system: because of the wide take up of these formalisms, there exist many utility libraries and generic programming interfaces which greatly simplify such processes as extracting the tags from a file, selecting portions of the text according to its logical structure, picking out files with certain attributes by searching their headers, and so on.

The BNC uses SGML in a simple and straightforward way described in the rest of this manual: simple programs can be readily written using standard UNIX utilities such as grep or perl to access the corpus just as plain text files. More reliably, programs can be written to application programming interfaces (APIs) such as the W3C’s Document Object Model (DOM) or the Simple API for XML (SAX), using application libraries developed for almost
every modern programming language (C, Perl, Python, tcl etc.). Information about such resources is not provided here, but is readily found on the World Wide Web: currently, one good place to start looking is www.xml.com.

When the BNC was first published, the top of the range personal computer might have as much as 50 or even 100 megabytes of disk storage and 8 Mb of RAM. At the time of writing, 20 or 30 gigabyte hard disks and 128 Mb of RAM are commonplace on entry level machines. It is thus quite likely that software capable of efficiently handling the 1.5 gigabytes of text which make up the BNC will also soon become commonplace. For the moment, however, it has to be recognized that general purpose tools for SGML and XML do not always cope very well with the large size of the whole corpus, although they can still be very useful for processing subsets extracted from it. To handle the whole of the corpus, special purpose indexing software will usually be necessary. Although such systems exist, they are often expensive or difficult to implement. For that reason, the BNC Project also developed its own low-cost alternative, the SARA package, which is documented separately. It should be emphasized however that use of the BNC is not synonymous with use of SARA. Most generic tools developed for corpus linguistics and NLP can be used with the BNC, although the tools may be vary in the extent to which they can make use of the markup in the corpus.

Whatever software is used, the programmer must have a clear understanding of the various elements tagged in the corpus, the contexts in which they may appear, and their intended semantics. The syntax of an SGML document is defined by a document type definition and by an SGML declaration. For TEI conformant texts, the TEI Header provides additional meta-information. The semantics of SGML elements are provided by documentation such as that provided elsewhere in this manual.

11.1. The BNC delivery format

An SGML document like the BNC must have the following components:

- an SGML declaration defining various SGML-specific limits;
- a document type declaration defining the elements entities and attributes which are legal in the document;
- an SGML document instance, that is, the document text itself.

These three components are all included as part of the standard release of the corpus.

11.1.1. Text files

The BNC is delivered in compressed format, using the GNU tar utility. When expanded, it comprises 4054 distinct files, ranging in size from 1 to 45 Kbytes, and totalling about 1.5 Gbytes. Each file contains a single BNC document, i.e. a TEI header and its associated spoken or written text, and has the same name as the value of the id attribute on its <bncDoc> element. Files are grouped according to their names into a three-level hierarchy. For example, all files with names beginning AA are in a subdirectory AA, which is within a subdirectory A (along with all other subdirectories beginning with the letter A). Not all possible three-letter filenames are actually used.

Each single-letter subdirectory (A to K, excluding I) is delivered as a separate compressed archive file. The whole corpus should be unpacked into a single hierarchy, which, as delivered, is called BNC/Texts. The full name for the corpus text with identifier ABC is thus BNC/Texts/A/AB/ABC.

Note that the three-character identifiers used (and hence the directory structure) are entirely arbitrary and do not convey any information about the type of text contained. Each text contains a TEI Header which specifies all such meta information, either directly, or by reference to the corpus header, as described in section 8 (The header). For convenience,
11.1.2. SGML components

All ancillary files relating to the SGML structure and processing of the corpus are included in the standard release within a subdirectory called SGML. This contains the following files:

**bnc.dec** The BNC SGML declaration provided has been modified to allow for the use of either an XML or an SGML format: this is necessary since the TEI Headers provided with the current release of the corpus are distributed in XML format. The corpus texts themselves are however still distributed in SGML format.

**bnc.dtd** The BNC document type declaration, as a single SGML file. This file is automatically generated from the standard TEI DTD using a pair of ‘extension files’ as further discussed in section 9 (Compatibility issues).

**bncMods.dtd** and **bncMods.ent** The two extension files used to parameterize the TEI for BNC usage.

**bncChars.dtd** SGML declarations for all character entities used in the BNC.

**bncDocs.dtd** SGML declarations for all documents making up the BNC.

**driver1.sgm** and **driver2.sgm** Example SGML driver files for processing the BNC.

**bncfinder.dat** and **bncIndex.xml** Ancillary data files which may be used to facilitate access to the corpus (see 11.3 (Creating a subcorpus) below.)

The remainder of this section discusses how these files may be used together as an SGML document. This is by no means the only way of processing the corpus, of course, and is intended solely to demonstrate the function of the various files listed above. Some basic understanding of the components of an SGML system is assumed.

To process a single text from the corpus (say, text ABC), a driver file like the following could be used:

```xml
<!DOCTYPE bnc SYSTEM
 "http://www.hcu.ox.ac.uk/TEI/Guidelines/DTD/tei2.dtd" [
 <ENTITY % TEI.prose "INCLUDE">
 <ENTITY % TEI.spoken "INCLUDE">
 <ENTITY % TEI.general "INCLUDE">
 <ENTITY % TEI.analysis "INCLUDE">
 <ENTITY % TEI.corpus "INCLUDE">
 <ENTITY % TEI.extensions.ent SYSTEM "/home/BNC/SGML/bncMods.ent">
 <ENTITY % TEI.extensions.dtd SYSTEM "/home/BNC/SGML/bncMods.dtd">
 <ENTITY % BNCchars SYSTEM "/home/BNC/SGML/BNCchars.ent">
 %BNCchars;
 <ENTITY corphdr SYSTEM "/home/BNC/Texts/corphdr">
 <ENTITY text SYSTEM "/home/BNC/Texts/A/AB/ABC">
 ]>
 <bnc>
 &corphdr;
 &text;
 </bnc>
```

This driver assumes that the standard TEI DTD is available from the URL given (which was true as of the date of this manual), and that the files from the BNC World distribution have been installed under /home/BNC. Alternatively, if the driver file is to be used offline, using the ‘compiled’ version of the BNC dtd, it might look like the following:
11.1.2. SGML components

To process more than one file from the corpus, a set of declarations like the one given above for the entity text would be necessary, one for each text concerned. For convenience, a file containing such declarations for every text in the corpus is also provided: this file, bncdocs.dtd, consists of declarations like the following:

```xml
<!DOCTYPE bnc SYSTEM "BNC/Texts/A/AB/ABC">
<!DOCTYPE bnc SYSTEM "BNC/Texts/A/AB/ABD">
```

With these declarations in force, it becomes possible to refer to the corpus file ABC simply by means of the entity reference &ABC;, as in the following example:

```xml
<!DOCTYPE bnc SYSTEM "BNC/Texts/A/AB/ABC">
```

The first line declares that what follows is an SGML document and that the dtd describing it is located in the file with the SYSTEM identifier given (/home/BNC/SGML/bnc.dtd). The next few lines (the portion within square brackets) comprise the DTD subset declaration: declarations here are to be processed before the content of the DTD. It comprises three entity declarations.

The first, for BNCdocs, associates that name with the external entity containing declarations for all the documents making up the BNC itself (i.e. the file bncDocs.ent), and then immediately references that entity. The percent sign is a syntactic convention of SGML which need not concern us here: the effect is that each file in the corpus can now be referenced using a name such as &ABC;. The second, for BNCchars, does almost exactly the same thing, but for the character references used within the corpus (see further 10.2 (Character entities defined by the BNC DTD)). The third associates the name BNChdr with the file containing the corpus header.

Following this, the driver file contains the SGML document itself, beginning with the <bnc> start-tag, and ending with the </bnc> end-tag. Between these tags are entity references, one for the corpus header, followed by one for each file to be included in this view of the corpus.
11.2. The BNC corpus header

As discussed in section 5 (Basic structure) above, the BNC consists of an overall corpus header, and a large number of distinct BNC documents, each with its own header. The corpus header must be present for an SGML processor to work with any part of the Corpus, because the corpus header contains declarations of elements (such as the classification records) referred to by almost every part of the corpus.

The various elements making up the header and their functions are discussed in section 8 (The header). The corpus header itself is included in the file corphdr. Its contents are reproduced below, reformatted for legibility.

<teiHeader type="corpus" creator="dominic" status="update" date.updated="2000-10-17" id="BNC-W">

<File Description>

<titleStmt>

The British National Corpus: World Edition

Lead partner in consortium: Oxford University Press

Text selection for miscellaneous and unpublished written materials: WR Chambers
Text selection, data capture and transcription for spoken texts and for 14% of published written texts: Longman ELT
Text selection for 86% published written texts: Oxford University Press
Data capture and transcription for all miscellaneous and unpublished written texts and for 86% of published written texts: Oxford University Press

Encoding, storage and distribution: Oxford University Computing Services

Text enrichment: Unit for Computer Research into the English Language, University of Lancaster
</titleStmt>

<editionStmt n="2.0">

First World Edition
</editionStmt>

<extent>

Approximately 100 million words
</extent>

<publicationStmt>

Publication

<distributor>

Oxford University Computing Services
</distributor>
11.2. The BNC corpus header

</distributor>

<address>
    13 Banbury Road, Oxford OX2 6NN U.K.
    Telephone: +44 1865 273221
    Facsimile: +44 1865 273275
    Internet mail: natcorp@oucs.ox.ac.uk
</address>

$idno type="BNC"> BNC-W $$idno>

<availability>
The British National Corpus is distributed worldwide by Oxford University Computing Services on a not-for-profit basis, and under the terms of a standard license agreement. Each copy of the corpus must include a copy of this corpus header and any redistribution or republishing of the corpus texts (the "BNC Processed Material") is strictly forbidden.

For information, the conditions of the Standard License Agreement are as follows:

• The BNC Consortium grants according to the terms and conditions set out herein and in consideration of the payments specified herein a non-exclusive, non-transferable Licence to the Licensee to use the BNC Processed Material for the purposes of linguistic research and/or the development of language products.

• Distribution of the BNC Processed Material is restricted to the Licensee or in the event of the Licensee being an organisation, to the Licensee’s research group. This group is defined as consisting only of those Licensee’s employees whom the Licensee authorises to perform the work using the BNC Processed Material for the purposes described in paragraph (a).

• Members of the said research group must not, except as herein provided, copy, publish or otherwise give to any third party access to the whole or any part of the BNC Processed Material. It is the responsibility of the Licensee to ensure that the members of the said research group understand and abide by this restriction, and to supervise their activities with respect to the BNC Processed Material. Neither the Licensee nor members of the Licensee’s said research group may assign, transfer, lease, sell, rent, charge or otherwise encumber the BNC Processed Material.

• The BNC Processed Material may be installed at the place or places of work of the said research group. The place of work is defined as the computing systems that the members of a research group normally use to conduct their research activities. It can include both work and home computers, and is not restricted to a particular machine or building.

• Copies of the BNC Processed Material may be made for backup purposes, or for the purposes of making data available to members of the research group but the Licensee shall ensure that the BNC Consortium’s copyright notice is reproduced on all copies or parts thereof of the BNC Processed Material. Any such copies will be deemed to be part of the BNC Processed Material.

• There is no restriction on the use of the Licensee’s Results except that the Licensee may not publish in print or electronic form or exploit commercially in any form whatsoever any extracts from the BNC Processed Material other than as permitted under the provisions of the relevant copyright laws.
• The BNC Consortium does not grant to the Licensee any rights whatsoever to reproduce the BNC Texts or use all or any part of the BNC Texts in commercial products or services in any way other than would be permitted under the fair dealings provision of copyright law.

</availability>
<date value="2000-11-1">
31 October 2000
</date>

</publicationStmt>
<sourceDesc>
The British National Corpus has no single source document. For details of the source or sources used in the creation of each electronic text, see the individual text headers. The principles and practices underlying selection and design of the corpus is documented in the BNC User Reference Guide, a copy of which should be supplied with it.
</sourceDesc>
</fileDesc>
<encodingDesc>

**Encoding**

<projectDesc>

**Goals**  The British National Corpus (BNC) Consortium was formed in 1990, and started work in 1991 on the three-year task of producing a hundred-million word corpus of modern British English for use in commercial and academic research. The first edition was published in 1994.

This second edition was produced during between 1998 and 2000. It contains a thorough revision of the part of speech tagging, several corrections to the headers, and some minor revision of the SGML tagging used.

**The Consortium Participants**  The BNC is the result of of a unique collaboration between three major U.K. dictionary publishers, two universities, and the British Library. The dictionary publishers are Chambers Harrap, Longman, and Oxford University Press; the universities are The Unit for Computer Research into the English Language (UCREL) at Lancaster University; and Oxford University Computing Services (OUCS).

**Funding**  The development of the BNC was funded by the commercial partners in the consortium with assistance from the the U.K. government’s Department of Trade and Industry (DTI) and Science and Engineering Research Council (SERC) under the Joint Framework for Information Technology (JFIT).

**Design**  The British National Corpus is

- A large corpus: its hundred million words are made up of ninety million from written and ten million from spoken sources.
- A **sample** corpus: it is composed of text samples, generally of no more than 40,000 words, rather than of complete works.
11.2. The BNC corpus header

- A **synchronic** corpus: it includes imaginative texts dating from the 1960s to 1994; informative texts dating from 1975 to 1994; and spoken texts gathered primarily between 1990 and 1994.
- A **general** corpus: it is not specifically restricted to any particular subject field, register or genre. It includes language from all age and social groups and a broad spread of U.K. regions.
- A **monolingual** British English Corpus: text samples are substantially the product of British English speakers. A small proportion of the words in the corpus are in a foreign language or non-British English.
- A **TEI-conformant** Corpus: texts in the corpus are uniformly marked up according to the recommendations of the Text Encoding Initiative (TEI), an international consortium concerned with the mark-up of texts for use in academic research. These recommendations are an application of Standardized General Markup Language (SGML), defined by International Standard IS 8879:1986.

**Uses**: 

- Lexicography: The corpus provides a body of new data on word meaning, grammar and usage. It yields empirical data on word frequencies, word classes and spelling preferences, among other things. It also reveals hitherto undocumented evidence about the spoken language, with consequences that go far beyond the immediate impact on dictionary-writing.
- Linguistic research: The corpus provides a standard basis for investigating phenomena and testing competing linguistic theories.
- Language technology: Statistical techniques, requiring very large samples of text, are increasingly used in machine translation, speech recognition, speech synthesizers, spelling and grammar checkers for word-processing and desk-top publishing, hand-held electronic books and other developments in information technology.
- Teaching: The corpus provides a rich source of examples of current usage for English Language Teaching, allowing more frequent patterns of use to be distinguished from less frequent. In addition, the corpus provides a valuable didactic resource for use in many areas of higher education.
- As a model: Future TEI-conformant corpora, in English and other languages, may base their designs on the experience gained in the production of the BNC.

</projectDesc>

<samplingDecl>
Different parts of the BNC were constructed using different sampling policies, as further described in the BNC Design Documentation. The policies are summarized below. Note that information about which policy resulted in the selection of a particular text is not available.

**SD000** Published: chosen selectively from candidate population
**SD001** Published: chosen at random from candidate population
**SD002** Unpublished: chosen according to relevant design criteria
**SD003** Spoken: obtained from demographic sample of UK population
**SD004** Spoken: obtained in context determined by design criteria
The following editorial policies were applied in creating the BNC. The DECLS attribute indicates which policies apply to a given <text> or <div> element; but not all policies are necessarily marked. Policies are identified by ID codes as follows:

**Correction policies**:
- CN000: Errors tagged with <sic> when seen; no normalization
- CN001: Errors tagged with <sic> if seen; normalisation with <corr>
- CN002: Normalized to standard British English or control list member
- CN004: Corrections and normalizations applied silently

**Hyphenation policies**:
- HN000: Smart elision of line-end hyphens; &rehy used for remainder
- HN001: Dumb elision of line-end hyphens; true hyphens hand-reinstated
- HN002: Line-end hyphens removed by hand where appropriate
- HN003: Source material contains no line-end hyphens

**Quotation policies**:
- QN000: Open, close quote normalized to &bquo, &equo
- QN001: Open and close quote normalized to &quo
- QN002: Quotation may be represented using <shift>

**Segmentation**
In this version of the Corpus, all segmentation and word-class marking was carried out in the same way, using CLAWS5.

**Transcription methods**:
- TN000: Copy-typed from hard-copy into OUP format; transduced to CDIF
- TN001: Copy-typed from hard-copy into Longman format; transduced to CDIF
- TN002: Scanned from hard-copy into OUP format; transduced to CDIF
- TN003: Scanned from hard-copy into Longman format; transduced to CDIF
- TN004: Transduced from M-R into OUP format; transduced to CDIF
- TN005: Transduced from M-R into Longman format; transduced to CDIF
- TN006: Recording transcribed into Longman format; transduced to CDIF
11.2. The BNC corpus header

<tagsDecl>

<refDecl>

Canonical references in the British National Corpus are to text segment (<s>) elements, and are constructed by taking the value of the n attribute of the <bncDoc> element containing the target text, and concatenating a dot separator, followed by the value of the n attribute of the target <s> element. Segments are numbered sequentially within each text or stext, starting at 1. There may be gaps in the numeric sequence, as a consequence of post-segmentation corrections.

</refDecl>

<classDecl>
11.2. The BNC corpus header

DLee: David Lee’s Register classification as documented at http://members.xoom.com/davidlee00/genre_register.zip

COPAC: Keyword classifications as supplied by the UK’s COPAC service at http://copac.ac.uk

allava Text availability
- allava0 Ownership has not been claimed
- allava1 Worldwide rights cleared
- allava2 Worldwide rights cleared
- allava3 Not available in North America
- allava4 Not available in U.S.A.
- allava5 Not available outside the European Union
- allava6 Not available in U.S.A. & Philippines
- allava7 Not available in N America & Philippines

alltyp Text type
- alltyp1 Spoken demographic
- alltyp2 Spoken context-governed
- alltyp3 Written books and periodicals
- alltyp4 Written-to-be-spoken
- alltyp5 Written miscellaneous

alltim Publication date
- alltim1 1960-1974
- alltim2 1975-1984
- alltim3 1985-1993
- alltim0 Unknown

scgdom Domain for context-governed spoken material
- scgdom1 Educational/Informative
- scgdom2 Business
- scgdom3 Public/Institutional
- scgdom4 Leisure

sdeage Age band for demographic respondent
- sdeage1 0-14
- sdeage2 15-24
- sdeage3 25-34
- sdeage4 35-44
- sdeage5 45-59
- sdeage6 60+

sdecla Social class for demographic respondent
- sdecla0 Unknown
- sdecla1 AB
- sdecla2 C1
- sdecla3 C2
- sdecla4 DE

sdesex Sex of demographic respondent
- sdesex0 Unknown
- sdesex1 Male
- sdesex2 Female

spolog Interaction type for spoken text
- spolog1 Monologue
- spolog2 Dialogue
11.2. The BNC corpus header

sporeg Region where spoken text captured
  sporeg0 Unknown
  sporeg1 South
  sporeg2 Midlands
  sporeg3 North

wriaag Author age band for written material
  wriaag0 Unknown
  wriaag1 0-14
  wriaag2 15-24
  wriaag3 25-34
  wriaag4 35-44
  wriaag5 45-59
  wriaag6 60+

wriad Author domicile
  wriad1 UK and Ireland
  wriad2 Commonwealth
  wriad3 Continental Europe
  wriad4 USA
  wriad5 Elsewhere
  wriad0 Unknown

wriase Written: author sex
  wriase0 Unknown
  wriase1 Male
  wriase2 Female
  wriase3 Mixed
  wriase4 Unknown

wriaty Written: type of author
  wriaty1 Corporate
  wriaty0 Unknown
  wriaty2 Multiple
  wriaty3 Sole
  wriaty4 Unknown

wriaud Written: audience age
  wriaud0 Unknown
  wriaud1 Child
  wriaud2 Teenager
  wriaud3 Adult
  wriaud4 Any

wridom Domain for written corpus texts
  wridom1 Imaginative
  wridom2 Informative: natural & pure science
  wridom3 Informative: applied science
  wridom4 Informative: social science
  wridom5 Informative: world affairs
  wridom6 Informative: commerce & finance
  wridom7 Informative: arts
  wridom8 Informative: belief & thought
  wridom9 Informative: leisure
11.2. The BNC corpus header

wrilev Written: perceived level of difficulty:
   wrilev0 Unknown
   wrilev1 Low
   wrilev2 Medium
   wrilev3 High

wrimed Medium for written corpus texts
   wrimed1 Book
   wrimed2 Periodical
   wrimed3 Miscellaneous: published
   wrimed4 Miscellaneous: unpublished
   wrimed5 To-be-spoken

wripp Place of publication
   wripp0 Unknown
   wripp1 UK (unspecific)
   wripp2 Ireland
   wripp6 United States
   wripp3 UK: North (north of Mersey-Humber line)
   wripp4 UK: Midlands (north of Bristol Channel-Wash line)
   wripp5 UK: South (south of Bristol Channel-Wash line)

wrisam Written text sample type
   wrisam0 Unknown
   wrisam1 Whole text
   wrisam2 Beginning sample
   wrisam3 Middle sample
   wrisam4 End sample
   wrisam5 Composite

wrista Written: estimated circulation size
   wrista0 Unknown
   wrista1 Low
   wrista2 Medium
   wrista3 High

writas Written: target audience sex
   writas0 Unknown
   writas1 Male
   writas2 Female
   writas3 Mixed
   writas4 Unknown

</classDecl>
</encodingDesc>

<profileDesc>

Profile

<creation>

This version of the corpus contains only texts accessioned on or before 1994-11-04.

</creation>
11.2. The BNC corpus header

<langUsage>
The language of the British National Corpus is modern British English. Words, fragments, and passages from many other languages, both ancient and modern, occur within the corpus where these may be represented using a Latin alphabet. Long passages in these languages, and material in other languages, are generally silently deleted. In no case is the lang attribute used to indicate the language of a word, phrase or passage, nor are alternate writing system definitions used.
</langUsage>

<particDesc>
<person age="X" id="PS000" n="W0000" role="other" sex="u" soc="UU">
PS000: Unknown speaker
</person>
<person age="X" id="PS001" n="W000M" role="other" sex="u" soc="UU">
PS001: Group of unknown speakers
</person>
</particDesc>

<revisionDesc>
Revision History

2000-10-11 2000-10-11
ed
Final manual corrections for BNC-W

2000-10-18 2000-10-18
ed
Further manual corrections for BNC-W

2000-01-08 2000-01-08
ed
Manually changed catdescriptions etc. for BNC-W

Edit
Corpus Revision

</revisionDesc>
</teiHeader>
11.3. Creating a subcorpus

Two files are provided with this version of the corpus to assist in the selection of files according to their classification codes: bncfinder.dat and bncIndex.xml.

11.3.1. Using bncfinder.dat

The file bncfinder.dat is a straightforward ASCII format data file, containing one record for each file in the corpus. Within each record, the following blank-delimited fields are present:

- three-character identifier
- size of text in Kbytes
- number of <p> or <u> elements
- number of <s> elements
- number of <w> elements
- number of orthographic words
- all classification codes assigned to this text

Here are two typical records from this file (wrapped to fit on the page: in the original, each is a single line):

A00 107 112 423 6673 6894 alltim3 allava2 alltyp6 wriaag0 wriad0 wriase0 wriady2 wriaud3 wridom4 wrilev2 wrimed3 wripp5 wriam6 wriast2 writas3
KSP 41 259 306 1543 1427 alltim3 allava2 alltyp1 sdeage1 sdecla1 sdesex1 spolog2 sporeg1

This file can be rapidly searched with simple Unix utilities such as grep to identify subcorpora having particular characteristics: for example, the following command line will select records for all spoken demographic texts collected by female respondents:

$grep sdesex2 bncfinder.dat

The same information could of course be obtained by searching through the corpus texts themselves; however the above is likely to be much quicker.

The classification codes used in the bncfinder.dat file are listed in section 10.8 (Text and genre classification codes).

The lines selected by such a procedure can be processed in many ways. Here for example is a program written in the perl language which just creates the references needed to embed texts in a driver file like those above:

```
open(out,">bncrefs.sgm") || die "Cannot create bncrefs.sgm: $!");
while (<>) {
    ($id,$k,$p,$s,$o) = split;
    $ntexts ++; $kbsz += $k; $nsents += $s; $nwords += $o;
    print OUT "id
";
}
print "You have selected $ntexts texts, totalling $kbsz Kb, $nsents s-units and $nwords orthographic words\n";
```

Assuming that this program is stored in the file subcorp.prl, a command line like the following might be used to create a bncrefs.sgm file defining a subcorpus comprising all the spoken demographic texts collected by female respondents:

$grep sdesex2 bncfinder.dat | perl subcorp.prl
11.3.2. Using bncIndex.xml

The file bncIndex.xml contains similar information to that provided in bncfinder.dat, but formatted for processing as an XML document. It contains a single <bncIndex> element which encloses a series of <doc> elements, one for each text in the BNC. Each <doc> element contains the following subelements:

- <idno> The three character identifier of the text
- <title> Either the short title of the text, taken from the <title> in the <fileDesc>, or the phrase [Unscripted conversation]
- <class> A classification code applied to the text, as given by the target attribute of the <catRef> element. One <class> is given for each classification and the following attributes are used to specify it:
  - type the identifying code specified by some <category> element in the corpus header
  - value a numeric code which, when appended to the type value, gives the classification code for this text.
- <genre> The genre code assigned to this text in David Lee’s classification scheme, as recorded in the <classCode> element in its header
- <counts> Size of this text measured in various ways, as specified by the following attributes:
  - kb size in Kbytes
  - ow size in orthographic words
  - w number of <w> elements
  - s number of <s> elements

Here are the XML elements for the two texts mentioned above (reformatted slightly to fit on the page):

```xml
<doc>
  <idno>A00</idno>
  <title>[ACET factsheets & newsletters]</title>
  <class type="alltim" value="3"/><class type="allava" value="2"/>
  <class type="alltyp" value="5"/><class type="uriaag" value="0"/>
  <class type="uriad" value="0"/><class type="uriase" value="0"/>
  <class type="wriaty" value="2"/><class type="uriawd" value="3"/>
  <class type="wridom" value="4"/><class type="wrilev" value="2"/>
  <class type="wrimed" value="3"/><class type="wripp" value="5"/>
  <class type="wrisam" value="5"/><class type="urista" value="2"/>
  <genre>W_non_ac_medicine</genre>
  <counts kb="107" ow="6894" w="6673" s="423"/>
</doc>

<!-- ... -->

<doc>
  <idno>KSP</idno>
  <title>[Spontaneous conversation]</title>
  <class type="alltim" value="3"/><class type="allava" value="2"/>
  <class type="alltyp" value="1"/><class type="sdeage" value="1"/>
  <class type="sdecla" value="1"/><class type="sdesex" value="1"/>
  <class type="spolog" value="2"/><class type="sporeg" value="1"/>
  <genre>S_conv</genre>
  <counts kb="41" ow="1427" w="1543" s="306"/>
</doc>
```
Although XML files do not require a DTD, one is also provided in the directory for convenience. It is called bncIndex.dtd and listed below:

```
<!ELEMENT bncIndex (doc+) >
<!ELEMENT doc (idno, title, class+, genre, count)>
<!ELEMENT idno (#PCDATA) >
<!ELEMENT title (#PCDATA) >
<!ELEMENT class EMPTY >
<!ATTLIST class type CDATA #REQUIRED
    value CDATA #REQUIRED>
<!ELEMENT genre (#PCDATA) >
<!ELEMENT counts EMPTY >
<!ATTLIST counts kb CDATA #REQUIRED
    ow CDATA #REQUIRED
    w CDATA #REQUIRED
    s CDATA #REQUIRED>
```

XML files may be processed in many different ways, but one of the most convenient is to use an XSLT stylesheet to transform it for display or search it. XSLT is a very high-level programming language defined by the W3C, which offers the ability to transform and process XML documents in a variety of ways. It is (at the time of writing) the language of choice for manipulating XML on the web, where a large number of free tools and tutorials may also be found.

To give a flavour of the language, the following XSLT stylesheet will process the bncIndex.xml selecting only texts with classification "wridom4", and displaying their titles, identifiers, and size as an HTML format table:

```
<xsl:stylesheet
    xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
    version="1.0"
    >
    <xsl:output method="html"/>

    <xsl:template match="bncIndex">
        <html>
            <center>
                <h2>There are
                    <xsl:value-of select="count(doc/class[@type='wridom' and @value='4'])"/>
                    BNC World Files with wridom classified as 4</h2>
            </center>
            <table>
                <xsl:apply-templates select="doc"/>
            </table>
        </html>
        </xsl:template>

    <xsl:template match="doc/class[@type='wridom' and @value='4']"
                    >
        <tr>
            <td><xsl:value-of select="../idno"/></td>
            <td><xsl:value-of select="../title"/></td>
            <td><xsl:value-of select="../genre"/></td>
            <td><xsl:value-of select="../counts/@kb"/></td>
            <td><xsl:value-of select="../counts/@ow"/></td>
            <td><xsl:value-of select="../counts/@w"/></td>
            <td><xsl:value-of select="../counts/@s"/></td>
        </tr>
    </xsl:template>

</xsl:stylesheet>
```
12. List of works excerpted

A00 [ACET factsheets & newsletters]. Aids Care Education & Training , London (1991-09). Sample containing about 6894 words of miscellanea (domain: social science)


A05 Authors. Miller, Karl, Oxford University Press, Oxford (1989), 60-163. Sample containing about 42316 words from a book (domain: arts)


A0A [CAMRA fact sheets]. CAMRA, St Albans. Sample containing about 5890 words of miscellanea (domain: leisure)


A0C Caterer & Hotelkeeper. Reed Publishing Group, Sutton, Surrey (1991-19-19). Sample containing about 31254 words from a periodical (domain: social science)


A0J [Health promotion and education leaflets]. u.p.. Sample containing about 24788 words of miscellanea (domain: natural sciences)


12. List of works excerpted

A0N **King Cameron.** Craig, David, Carcanet Press, Manchester (1991), 15-113. Sample containing about 41893 words from a book (domain: imaginative)


A0R **Nudists may be encountered.** Scott, Mary, Serpent’s Tail, London (1991), 7-115. Sample containing about 43573 words from a book (domain: arts)

A0S [**Official leaflets.**] u.p.. Sample containing about 1525 words of miscellanea (domain: commerce)


A0U **So very English.** Rowe, Martha (ed.), Serpent’s Tail, London (1990). Sample containing about 36950 words from a book (domain: imaginative)

A0V **Tennis World.** Presswatch, Sussex (1991-08). Sample containing about 20581 words from a periodical (domain: leisure)


A0X **Woodworker.** Argus Specialist Publications, Hemel Hempstead (1991-08). Sample containing about 32599 words from a periodical (domain: leisure)

A0Y [**Age Concern fact sheets.**] Age Concern England, Mitcham. Sample containing about 19008 words of miscellanea (domain: social science)

A10 [**Age Concern newsletter and information.**] Age Concern England, Mitcham. Sample containing about 23795 words of miscellanea (domain: social science)


A13 [**CAMRA information leaflets.**] CAMRA, St Albans. Sample containing about 4561 words of miscellanea (domain: leisure)

A14 **What’s brewing.** CAMRA, St Albans (1991-10). Sample containing about 22899 words of miscellanea (domain: leisure)

A15 **Climber and Hill Walker.** George Outram & Company Ltd, Glasgow (1991-08). Sample containing about 30722 words from a periodical (domain: leisure)

A16 **Do It Yourself Magazine.** Link House Magazines Ltd, Croydon (1991-09). Sample containing about 33154 words from a periodical (domain: leisure)


A19 **Electronics and Wireless World.** Reed Business Publishing Group Ltd, Sutton, Surrey (1991-09). Sample containing about 41084 words from a periodical (domain: applied science)


A1B **Studies in Ezra Pound.** Davie, Donald,
12. List of works excerpted


(1989-19-19). Sample containing about 23055 words from a periodical (domain: world affairs)


12. List of works excerpted


12. List of works excerpted


thought)


Sample containing about 12270 words from a periodical (domain: leisure)


12. List of works excerpted


A64 One step backwards, two steps forward. Pethybridge, Roger, Clarendon Press, Oxford (1990), 22-144. Sample containing about 42467 words from a book (domain: world affairs)

12. List of works excerpted


A67 [RAFA journal and miscellaneous info]. Royal Air Force Association. Sample containing about 33454 words of miscellanea (domain: leisure)


A6E *I was a teenage sex pistol.* Silverton, Pete and Matlock, Glen, Omnibus Press, London (1990). Sample containing about 25648 words from a book (domain: arts)


A6R *Angler’s Mail.* mixed, u.p.. Sample containing about 28963 words from a periodical (domain: leisure)


A6W *Autocar and Motor.* Haymarket
12. List of works excerpted

Magazines Ltd, Teddington, Middlesex (1990-10-10; 1990-03-1). Sample containing about 31551 words from a periodical (domain: leisure)


A7F Caterer & Hotelkeeper. Reed Publishing Group, Sutton, Surrey (1991-01-9725). Sample containing about 27546 words from a periodical (domain: social science)

A7G [Charity leaflets and letters]. u.p.. Sample containing about 19719 words of miscellanea (domain: social science)


A7K [Church magazines and leaflets]. u.p.. Sample containing about 24874 words of unpublished miscellanea (domain: belief and thought)


A8A The Guardian, electronic edition of
12. List of works excerpted


12. List of works excerpted


12. List of works excerpted

69 words from a periodical (domain: applied science)


12. List of works excerpted


12. List of works excerpted


AAY Adkin. Harnel Heath Ltd (1992). Mov-
12. List of works excerpted


ABU Ministers decide. Fowler, Nor-


from a periodical (domain: leisure)


**ACR** Farmers' Weekly. Reed Business Publishing Group, Haywards Heath, Sussex (1991-01-9726). Sample containing about 63472 words from a periodical (domain: applied science)


**ACW** Frankie. Highsmith, Domini, Bantam (Corgi), London (1990). Sample containing about 32090 words from a book (domain: imaginative)


**ADK** Making an impact. Thomas, Harvey
12. List of works excerpted


AHC Daily Telegraph, electronic edition of


12. List of works excerpted


<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
</table>
12. List of works excerpted

19). Sample containing about 15255 words from a periodical (domain: commerce)


**AL7** Leaflets advertising services]. u.p.. Sample containing about 5894 words of miscellanea (domain: commerce)


**ALC** Effective training for family and part-time farmers. Birkbeck, David, The Arkleton Trust, Langholm, Dumfriesshire (1985), 9-96. Sample containing about 28812 words of miscellanea (domain: applied science)


**ALF** Disadvantaged rural Europe. Moore, David, The Arkleton Trust, Langholm, Dumfriesshire (1979), 1-34. Sample containing about 11362 words of miscellanea (domain: applied science)


**ALL** Bell in the tree. Chisnall, Edward,


AM0 Club 18-30 summer holiday brochure 1990. ILG Travel Ltd (1990-19-19). Sample containing about 34489 words of miscellaneous (domain: leisure)


AM5 Do It Yourself. Link House Magazines Ltd, Milton Keynes (1992-02). Sample containing about 30227 words from a periodical (domain: leisure)


AM7 Education management for the 1990s. Osborne, Allan, Davies, Brent, West-Burnham, John and Ellison, Linda, Longman Group UK Ltd, Harlow (1990), 1-91. Sample containing about 29230 words from a book (domain: commerce)

AM8 The best future for Britain. Conservative Central Office, London (1992), 1-50. Sample containing about 30316 words from a periodical (domain: world affairs)

AM9 Changing Britain for good. Liberal Democrat Publications, Dorchester (1992), 2-51. Sample containing about 18348 words from a periodical (domain: world affairs)

AMA It's time to get Britain working again.


AMH A guide to exporting. MW Publications Ltd, Belfast, 1-53. Sample containing about 15618 words from a periodical (domain: commerce)


AMX McDonald’s food: the facts. u.p.. Sample containing about 2639 words of miscellanea (domain: natural sciences)

AMY Henley Golf Club – the first 80 years. Henley Golf Club, Henley-on-Thames, Oxon (1986), 1-80. Sample containing about 34287 words of miscellanea (domain: world affairs)

AN0 The intelligence game. Rusbridger, James, I B Tauris & Company Ltd, London (1991), 11-84. Sample containing about 36678 words from a book (domain: world affairs)


AN2 Know your Land Rover. Ivins, Robert, LRO Books Ltd, Diss, Norfolk (1991), 5-88. Sample containing about 35724 words from a book (domain: leisure)

AN3 Latin American society. Cubitt, Tessa,

AN5 The legal context of teaching. Johnstone, Susan, Pearce, Penelope and Harris, Neville, Longman Group UK Ltd, Harlow (1992), 60-154. Sample containing about 38172 words from a book (domain: social science)

AN7 Maggie Jordan. Blair, Emma, Bantam (Corgi), London (1990), 7-94. Sample containing about 40491 words from a book (domain: imaginative)


ANK Murder in old Berkshire. Long, Roger, Barracuda Books Ltd, Buckingham (1990), 11-122. Sample containing about 35956 words from a book (domain: social science)


ANM NCT Birmingham Central Branch Newsletter. u.p.. Sample containing about 27664 words of unpublished miscellanea (domain: social science)

ANP The National Trust Magazine. The National Trust, London (1990-19-19). Sample containing about 21626 words from a periodical (domain: arts)


ANS National curriculum English (ages 5-16). Department of Ed. & Science, UK (1989-07). Sample containing about 28115 words of miscellanea (domain: social science)


AP1  *[Collection of letters from OUP]*. u.p.. Sample containing about 24303 words of unpublished miscellanea (domain: commerce)

AP5  *[Official leaflets]*. u.p.. Sample containing about 17281 words of miscellanea (domain: commerce)

AP6  *Monster Raving Loony Party’s draft manifesto for General Election 1992*. u.p.. Sample containing about 1919 words of miscellanea (domain: commerce)

AP7  *I don’t feel old*. Abendstern, Michele, Thompson, Paul and Itzin, Cathy, Oxford University Press, Oxford (1990), 1-106. Sample containing about 38087 words from a book (domain: social science)

AP8  *The old galleries of Cumbria and the early wool trade*. Allen, Phoebe, Abbot Hall Art Gallery, Kendal (1984), 1-56. Sample containing about 11802 words of miscellanea (domain: world affairs)


APJ  *Report on visit to Peto Institute*. Eccleshall, J and Davis, J, u.p.. Sample containing about 3549 words of unpublished miscellanea (domain: social science)

APK  *Photography*. Argus Specialist Publications, Hemel Hempstead (1992-02). Sample containing about 11489 words from a periodical (domain: leisure)


APS  *Postmodernism and contemporary fic-
12. List of works excerpted


**APX Royal Institute of British Architects Journal. RIBA Magazines Ltd, London (1990-10). Sample containing about 29542 words from a periodical (domain: arts)**


**ARA Stopping crime starts with you. Northern Ireland Office, Belfast, 1-42. Sample containing about 9272 words of miscellanea (domain: social science)**


**ARE Scuba world. Freestyle Publications Ltd, Poole (1991-10). Sample containing about 18455 words from a periodical (domain: leisure)**

ARG *In search of happiness*. Houston, James, Lion Publishing plc, Oxford (1990), 165-275. Sample containing about 35217 words from a book (domain: belief and thought)


ARM *Skateboard!*. Lancastrian Holdings Ltd, Dorset (1991-03). Sample containing about 9984 words from a periodical (domain: leisure)

ARP *Skinhead*. Knight, Nick, Omnibus Press, London (1982), 5-52. Sample containing about 13762 words from a book (domain: social science)


ARW *Spare Rib*. u.p.. Sample containing about 32315 words from a periodical (domain: social science)


ARY *The fundamentals of stability testing*. IFSCC, Weymouth (1992), 5-23. Sample containing about 7733 words of miscellanea (domain: natural sciences)


AS7 *Tales of the loch*. Sandison, Bruce, Mainstream Publishing Company Ltd, Edinburgh (1990), 1-102. Sample containing about 38218 words from a book (domain: imaginative)

Sample containing about 37344 words from a book (domain: leisure)


ASJ  Travel. Seddon, Sue, Alan Sutton/Thomas Cook, Stroud/Peterborough (1991), 11-144. Sample containing about 27522 words from a book (domain: leisure)


AT1  The wedding present. Hodkinson, Mark, Omnibus Press, London (1990), 5-91. Sample containing about 39181 words from a book (domain: arts)


AT4  Who, sir? Me, sir?. Peyton, K M, Oxford University Press, Oxford (1988), 5-
12. List of works excerpted

127

138. Sample containing about 47798 words from a book (domain: imaginative)


AT7 The wingless bird. Cookson, Catherine, Bantam (Corgi), London (1990), 11-128. Sample containing about 40027 words from a book (domain: imaginative)


AT9 Women priests: which way will you vote?. SPCK, London (1990). Sample containing about 6845 words of miscellanea (domain: belief and thought)


AYM [Events and entertainments leaflets]. u.p.. Sample containing about 20733 words of miscellanea (domain: leisure)

AYP [Financial leaflets]. u.p.. Sample containing about 53481 words of miscellanea (domain: commerce)

AYR [Financial letters]. u.p.. Sample containing about 9374 words of unpublished miscellanea (domain: commerce)

AYX [Leaflets advertising services]. u.p.. Sample containing about 5820 words of miscellanea (domain: commerce)


B02 Agriculture and nature conservation in conflict. Smith, Malcolm, The Arkleton Trust, Langholm, Dumfriesshire (1985), 1-94. Sample containing about 31239 words of miscellanea (domain: applied science)

B03 The Alton Herald. Farnham Castle Newspapers Ltd, Farnham, Surrey (1992-04-17), 1-32. Sample containing about 59362 words from a periodical (domain: world affairs)


B05 Approaches to spiritual direction. Long, Anne, Grove Books Ltd, Nottingham (1988), 3-25. Sample containing about 10323 words of miscellanea (domain: belief and thought)


<table>
<thead>
<tr>
<th>12. List of works excerpted</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B0G</strong> <em>Cornish times reflected.</em> Lister, Martin, Tamara Publications, Saltash, Cornwall (1988). Sample containing about 23702 words from a book (domain: belief and thought)</td>
</tr>
<tr>
<td><strong>B0H</strong> <em>Cabinet.</em> Hennessy, Peter, Basil Blackwell Ltd, Oxford (1990), 34-122. Sample containing about 38718 words from a book (domain: world affairs)</td>
</tr>
<tr>
<td><strong>B0R</strong> <em>Coleridge and Wordsworth in the West Country.</em> Mayberry, Tom, Alan Sutton Publishing Ltd, Gloucester (1992), 4-131. Sample containing about 34186 words from a book (domain: arts)</td>
</tr>
<tr>
<td><strong>B0S</strong> <em>The councillor.</em> Prophet, John, Shaw &amp; Sons Ltd, London (1987), 1-76. Sample containing about 15164 words from a book (domain: world affairs)</td>
</tr>
<tr>
<td><strong>B0U</strong> <em>A crowd is not company.</em> Kee, Robert, Sphere Books Ltd, London (1991), 22-121. Sample containing about 41954 words from a book (domain: world affairs)</td>
</tr>
<tr>
<td><strong>B0W</strong> <em>Dementia and home care.</em> Askham, Janet and Thompson, Catherine, Age Concern England, Mitcham (1990), 9-118. Sample containing about 32143 words from a book (domain: social science)</td>
</tr>
<tr>
<td><strong>B0Y</strong> <em>The Dickens index.</em> Burgis, Nina, Slater, Michael and Bentley, Nicolas, Oxford University Press, Oxford (1990), 182-239. Sample containing about 43356 words from a book (domain: arts)</td>
</tr>
<tr>
<td><strong>B10</strong> <em>Discipline: a positive guide for parents.</em> Herbert, Martin, Basil Blackwell Ltd, Oxford (1989), 1-130. Sample contain-</td>
</tr>
</tbody>
</table>
12. List of works excerpted


- **B13** Edenderry Memorial Methodist Church. u.p. (n.d.), 1-61. Sample containing about 20720 words of miscellanea (domain: world affairs)


- **B1J** In good faith. Lamont, Stewart, St Andrew Press, Edinburgh (1989), 3-89. Sample containing about 40137 words from a book (domain: belief and thought)


- **B1M** [Health promotion and education leaflets]. u.p.. Sample containing about 12096 words of miscellanea (domain: natural sciences)
<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Author</th>
<th>Publisher</th>
<th>Year</th>
<th>Pages</th>
<th>Sample size (words)</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1R</td>
<td>How to use homeopathy.</td>
<td>Hammond, Christopher</td>
<td>Element Books Ltd, Shaftesbury, Dorset (1991)</td>
<td>1-134</td>
<td></td>
<td>35817</td>
<td>applied science</td>
</tr>
<tr>
<td>B21</td>
<td>Learn to live.</td>
<td>Spindler, Stephanie</td>
<td>Element Books Ltd, Shaftesbury, Dorset (1991)</td>
<td>1-93</td>
<td></td>
<td>23668</td>
<td>applied science</td>
</tr>
<tr>
<td>B25</td>
<td>Methods of social investigation.</td>
<td>Mann, Peter H</td>
<td>Basil Blackwell Ltd, Oxford (1985)</td>
<td></td>
<td></td>
<td>39333</td>
<td>social science</td>
</tr>
<tr>
<td>B27</td>
<td>[Midland bank leaflets].</td>
<td>u.p.</td>
<td>Sample containing about 7306 words of miscellanea</td>
<td></td>
<td></td>
<td>7306</td>
<td>commerce</td>
</tr>
<tr>
<td>B29</td>
<td>[Northern Ireland tourist information].</td>
<td>u.p.</td>
<td>Sample containing about 29239 words of miscellanea</td>
<td></td>
<td></td>
<td>29239</td>
<td>leisure</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
12. List of works excerpted

University Press, Oxford (1990), 321-452. Sample containing about 19271 words from a periodical (domain: applied science)


**B31** The birds of Sussex. Shrub, Michael, Phillimore & Company Ltd, Chichester (1979), 1-152. Sample containing about 38639 words from a book (domain: leisure)


**B35** Tour de France 92. Channel 4 (1992). Sample containing about 8355 words of miscellanea (domain: leisure)


**B3B** An introduction to postgraduate studies. University of Ulster, Coleraine (1991). Sample containing about 13940 words of miscellanea (domain: social science)


**B3K** [Tourist information: York]. u.p.. Sample containing about 38231 words of miscellanea (domain: leisure)


**B72** New Scientist. IPC Magazines Ltd, London (1991-02-16). Sample containing about 44553 words from a periodi-
12. List of works excerpted

- **B74** New Scientist. IPC Magazines Ltd, London (1991-02-16). Sample containing about 38360 words from a periodical (domain: applied science)
- **B75** New Scientist. IPC Magazines Ltd, London (1991-02-16). Sample containing about 45342 words from a periodical (domain: applied science)
- **B77** New Scientist. IPC Magazines Ltd, London (1991-02-16). Sample containing about 45363 words from a periodical (domain: applied science)
- **B7C** New Scientist. IPC Magazines Ltd, London (1991-02-16). Sample containing about 47149 words from a periodical (domain: applied science)
- **B7D** New Scientist. IPC Magazines Ltd, London (1991-02-16). Sample containing about 38693 words from a periodical (domain: applied science)
- **B7G** New Scientist. IPC Magazines Ltd, London (1991-02-16). Sample containing about 53321 words from a periodical (domain: applied science)
- **B7N** New Scientist. IPC Magazines Ltd, London (1991-02-16). Sample containing about 42952 words from a periodical (domain: applied science)
- **BLW** Forty plus. Batchelor, Mary, Lion Publishing plc, Tring, Herts (1988), 48-166. Sample containing about 40414 words from a book (domain: belief and thought)
- **BLY** Curriculum and reality in African primary schools. Hawes, Colette, Aarons, Audrey, Hawes, Hugh and Croft, John, Longman Group UK Ltd, Harlow (1979), 10-121. Sample containing...
134 12. List of works excerpted

taining about 39163 words from a book (domain: social science)


BM5 Autocar and Motor. Haymarket Magazines Ltd, Teddington, Middlesex (1990-10-10; 1990-03-1). Sample containing about 36341 words from a periodical (domain: commerce)


BM9 Britain’s deaf heritage. Jackson, Peter, British Deaf Association, Haddington, E. Lothian (1990), 67-265. Sample containing about 33408 words from a book (domain: social science)


BMC CD Review. CD Review Ltd, Berkhamstead (1992-05), 4-97. Sample containing about 83844 words from a periodical (domain: arts)


BMM Linford Christie: an autobiography. Christie, Linford and Ward, Tony, Ar-
<table>
<thead>
<tr>
<th>ISBN</th>
<th>Title</th>
<th>Year</th>
<th>Publisher</th>
<th>Location</th>
<th>Pages</th>
<th>Sample Length</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMP</td>
<td><em>Way of life: dominant ideologies and local communities.</em> Jackson, A (ed.)</td>
<td>1980</td>
<td>Social Science Research Council</td>
<td>London</td>
<td>1-76</td>
<td>23321 words</td>
<td>world affairs</td>
</tr>
<tr>
<td>BMR</td>
<td><em>Dirty tricks.</em> Dibdin, Michael</td>
<td>1991</td>
<td>Faber &amp; Faber Ltd</td>
<td>London</td>
<td>5-142</td>
<td>39105 words</td>
<td>imaginative</td>
</tr>
<tr>
<td>BMS</td>
<td><em>Gate-crashing the dream party.</em> Leonard, Alison</td>
<td>1990</td>
<td>Walker Books Ltd</td>
<td>London</td>
<td>7-170</td>
<td>41548 words</td>
<td>imaginative</td>
</tr>
<tr>
<td>BMT</td>
<td><em>The elements of earth mysteries.</em> Heselton, Philip</td>
<td>1991</td>
<td>Element Books Ltd</td>
<td>Shaftesbury, Dorset</td>
<td>1-100</td>
<td>36727 words</td>
<td>belief and thought</td>
</tr>
<tr>
<td>BMU</td>
<td><em>The distance enchanted.</em> Gervaise, Mary</td>
<td>1983</td>
<td>Goodchild Publishers, Wendover</td>
<td>Bucks</td>
<td>1-109</td>
<td>40592 words</td>
<td>imaginative</td>
</tr>
<tr>
<td>BMV</td>
<td><em>Europe in the central middle ages.</em> Brooke, Christopher</td>
<td>1987</td>
<td>Longman Group UK Ltd</td>
<td>Harlow</td>
<td>1-100</td>
<td>39773 words</td>
<td>world affairs</td>
</tr>
<tr>
<td>BWY</td>
<td><em>Natural creation &amp; the formative mind.</em> Davidson, John</td>
<td>1991</td>
<td>Element Books Ltd</td>
<td>Shaftesbury, Dorset</td>
<td>44-141</td>
<td>40903 words</td>
<td>belief and thought</td>
</tr>
<tr>
<td>BN1</td>
<td><em>And thus will I freely sing.</em> Davidson, Tony (ed.)</td>
<td>1989</td>
<td>Polygon Books</td>
<td>Edinburgh</td>
<td>41-186</td>
<td>36954 words</td>
<td>imaginative</td>
</tr>
<tr>
<td>BN4</td>
<td><em>Here's health: the green guide.</em> Wheater, Caroline and Smyth</td>
<td>1990</td>
<td>Argus Books</td>
<td>Hemel Hempstead, Herts</td>
<td>21-126</td>
<td>37899 words</td>
<td>applied science</td>
</tr>
<tr>
<td>BN5</td>
<td><em>Rosemary Conley's hip and thigh diet.</em> Conley, Rosemary</td>
<td>1989</td>
<td>Arrow Books</td>
<td>London</td>
<td>18-96</td>
<td>15954 words</td>
<td>leisure</td>
</tr>
<tr>
<td>BN7</td>
<td><em>Miscellaneous leaflets on health.</em> u.p.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25608 words</td>
<td>social science</td>
</tr>
<tr>
<td>BN8</td>
<td><em>Hollow promises?.</em> Rogers, A, Keith M (eds.)</td>
<td>1991</td>
<td>Mansell Publishing Ltd</td>
<td>London</td>
<td>1-83</td>
<td>28709 words</td>
<td>world affairs</td>
</tr>
<tr>
<td>BN9</td>
<td><em>The Cricketer.</em> Sporting Magazines &amp; Pubs</td>
<td>1992</td>
<td>Tunbridge Wells</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
12. List of works excerpted

ple containing about 42009 words from a periodical (domain: leisure)

BNA *How to interview and be interviewed.* Brandreth, Gyles and Brown, Michele, Sheldon Press, London (1992), 9-123. Sample containing about 35787 words from a book (domain: belief and thought)

BNC *It might have been Jerusalem.* Healy, Thomas, Polygon Books, Edinburgh (1991), 1-81. Sample containing about 30831 words from a book (domain: imaginative)


BNJ *Northamptonshire Rose of the Shires.* u.p.. Sample containing about 22612 words of miscellanea (domain: leisure)

BNK *Nonesuch: the University of Bristol magazine.* University of Bristol, Bristol (1991/1992). Sample containing about 25350 words from a periodical (domain: leisure)

BNL *One’s company.* Underwood, Lynn, Ashford, Southampton (1989), 11-141. Sample containing about 39981 words from a book (domain: belief and thought)


BNX [Leaflets advertising various products]. u.p.. Sample containing about 12631 words of miscellanea (domain: commerce)


BP1 *The rich pass by*. Pope, Pamela, Century Hutchinson, London (1990), 17-117. Sample containing about 36953 words from a book (domain: imaginative)


BP3 *Sainsbury’s wine*. Sainsbury’s (1992). Sample containing about 6307 words of miscellanea (domain: leisure)


BP6 *Welcome to Somerset*. u.p., Sample containing about 6741 words of miscellanea (domain: leisure)


BPC *UK Tourist Brochures*. u.p., Sample containing about 24000 words of miscellanea (domain: leisure)


BPE *UniView*: The National Student Newspaper. YY4ME (1992-08). Sample containing about 9287 words from a periodical (domain: leisure)


BPJ *Yorkshire Life*. Town & County Magazines, Preston (1992-05). Sample containing about 18402 words from a periodical (domain: leisure)

BPK *Fairs, feasts and frolics*. Smith, Ju-
12. List of works excerpted

lia, Smith Settle Ltd, Otley, West Yorks (1989), 22-129. Sample containing about 39983 words from a book (domain: social science)


C87 Zzap 64!. Newspaper Publications, Ludlow, Shropshire (1992-11). Sample containing about 32964 words from a periodical (domain: leisure)


C89 The Artist’s and Illustrator’s Magazine. Artist’s & Illustrator’s Mag, London (1991-10). Sample containing about 19656 words from a periodical (domain: arts)


C8F Calming traffic in residential areas. Tolley, Rodney, Brefi Press, Dyfed (1990), 5-122. Sample containing about 41345 words from a book (domain: world affairs)


C8N The chocolate teapot. Lawrence, David, Scripture Union, London (1992), 17-96. Sample containing about 18226 words from a book (domain: belief and thought)


C8S *Destiny.* Beauman, Sally, Bantam (Corgi), London (1987), 201-306. Sample containing about 35631 words from a book (domain: imaginative)

C8T *Devices and desires.* James, P D, Faber & Faber Ltd, London (1989), 17-120. Sample containing about 41117 words from a book (domain: imaginative)


C8V *Doubt.* Guinness, Os, Lion Publishing plc, Tring, Herts (1976), 7-115. Sample containing about 37912 words from a book (domain: belief and thought)

C8X *In the echoey tunnel.* Reid, Christopher, Faber & Faber Ltd, London (1991), 1-36. Sample containing about 3545 words from a book (domain: imaginative)

C8Y *Caring for elderly parents.* Deeping, Eleanor, Constable & Company Ltd, London (1979), 7-125. Sample containing about 40567 words from a book (domain: social science)


C9E *Golf Monthly.* Specialist & Leisure Magazines Group, London (1991-02,
1992-11). Sample containing about 30402 words from a periodical (domain: leisure)


CA0 Polo. Cooper, Jilly, Bantam (Corgi), London (1991), 162-267. Sample containing about 40478 words from a book (domain: imaginative)


12. List of works excerpted

(1991), 45-160. Sample containing about 37737 words from a book (domain: belief and thought)


CA7 Madeira: the complete guide. Farrow, Susan and Farrow, John, Robert Hale Ltd, London (1990), 22-121. Sample containing about 41184 words from a book (domain: leisure)


CAA New Millennium summer holidays. u.p.. Sample containing about 6383 words of miscellanea (domain: leisure)


CAM The negotiator. Forsyth, Frederick, Bantam (Corgi), London (1989), 186-278. Sample containing about 35796 words from a book (domain: imaginative)


CAR Orkney: a place of safety?. Black, Robert, Canongate Press plc, Edinburgh (1992), 9-103. Sample containing about 38076 words from a book (domain: so-
12. List of works excerpted


CB7 A sensible life. Wesley, Mary, Bantam (Corgi), London (1990), 94-226. Sample containing about 39515 words from a book (domain: imaginative)


CBE Today. News Group Newspapers Ltd,
London (1992-12). Sample containing about 46974 words from a periodical (domain: world affairs)


**CBH** *Some psychic and political dimensions of work experiences.* Gordon Lawrence, W., Tavistock Institute of Human Relations (1982), 1-35. Sample containing about 20036 words of miscellanea (domain: social science)

**CBJ** *The Tiftord family 1547-1947.* Tiftord, John, Phillimore & Company Ltd, Chichester (1989), 84-165. Sample containing about 37052 words from a book (domain: world affairs)


**CBM** *Ulster Newsletter.* Century Newspapers Ltd, Belfast (1991-11-05). Sample containing about 16888 words from a periodical (domain: world affairs)


**CBP** *The complete video course.* Brookes, Keith, Boxtree, London (1989), 7-119. Sample containing about 39083 words from a book (domain: leisure)


**CC0** *The Artist’s and Illustrator’s Magazine.* Artist’s & Illustrator’s Mag, London (1992-19-19). Sample containing about 20271 words from a periodical (domain: arts)

**CC1** *Queen’s Park Baptist church magazines.* u.p.. Sample containing about 4126 words of unpublished miscellanea (domain: belief and thought)

**CC2** *Queen’s Park Baptist church magazines.* u.p.. Sample containing about 3058 words of unpublished miscellanea
12. List of works excerpted

CC3 [Queen’s Park Baptist church magazines]. u.p.. Sample containing about 4728 words of unpublished miscellanea (domain: belief and thought)

CC4 [Queen’s Park Baptist church magazines]. u.p.. Sample containing about 2160 words of unpublished miscellanea (domain: belief and thought)

CC5 [Queen’s Park Baptist church magazines]. u.p.. Sample containing about 3794 words of unpublished miscellanea (domain: belief and thought)

CC6 [Queen’s Park Baptist church magazines]. u.p.. Sample containing about 4172 words of unpublished miscellanea (domain: belief and thought)

CC7 [Queen’s Park Baptist church magazines]. u.p.. Sample containing about 6839 words of unpublished miscellanea (domain: belief and thought)

CC8 [Queen’s Park Baptist church magazines]. u.p.. Sample containing about 3640 words of unpublished miscellanea (domain: belief and thought)

CC9 [Queen’s Park Baptist church magazines]. u.p.. Sample containing about 7851 words of unpublished miscellanea (domain: belief and thought)

CCA A bad spell for the worst witch. Murphy, Jill, Puffin, Harmondsworth (1988), 7-128. Sample containing about 12295 words from a book (domain: imaginative)


CCC British intelligence services in action. Lindsay, Kennedy, Dunrod Press, Dun- dalk, Ireland (1980), 74-176. Sample containing about 39645 words from a book (domain: world affairs)


CCS  *Commandoes and rangers of World War II.* Ladd, James, Macdonald & Jane Publishers, London (1979), 12-82. Sample containing about 35663 words from a book (domain: world affairs)


CCU  *[Collection of publishing contracts].* u.p.. Sample containing about 35883 words of unpublished miscellanea (domain: commerce)


CCX  *Do It Yourself.* Link House Magazines Ltd, Croydon (1992-19-19), 4-62. Sample containing about 31460 words from a periodical (domain: leisure)

CCY  *Do It Yourself Magazine.* Link House Magazines Ltd, Croydon (1992-19-19), 4-70. Sample containing about 26220 words from a periodical (domain: leisure)


CDA  *Foxbat.* Cave, Peter, Severn House, Wallington, Surrey (1979), 11-146. Sample containing about 38923 words from a book (domain: imaginative)


CDC  *Unfinished: George Appleton remembers and reflects.* Appleton, George, Collins, UK (1990), 5-117. Sample containing about 39928 words from a book (domain: belief and thought)


CDE  *Guilty parties.* Street, Pamela, Robert


12. List of works excerpted

1-146. Sample containing about 41761 words from a book (domain: social science)


CED Appropriate technology in the developing world. Woodall, N, u.p.. Sample containing about 9754 words of unpublished miscellanea (domain: applied science)


CEP Today: sports pages. News Group
Newspapers Ltd, London (1992-12). Sample containing about 202283 words from a periodical (domain: leisure)


**CET** The great unsolved mysteries of science. Grant, John, Quintet Publishing Ltd, London (1990), 8-123. Sample containing about 40486 words from a book (domain: natural sciences)


**CF7** East Anglian Daily Times. East Anglian Daily Times Company, Ipswich (1993-03). Sample containing about 496 words from a periodical (domain: world affairs)


**CFE** Age and vulnerability. Stevenson, Olive and Key, Michael, Edward Arnold (Publishers) Ltd, Sevenoaks, Kent (1990). Sample containing about 36929 words from a book (domain: social science)

**CFF** Alternative saints. Symonds, Richard,
12. List of works excerpted


CFL The Artist’s and Illustrator’s Magazine. Artist’s & Illustrator’s Mag, London (1991-10). Sample containing about 19509 words from a periodical (domain: arts)

CFM [Adverts from country living]. u.p.. Sample containing about 5669 words of miscellanea (domain: commerce)

CFN [Adverts from The National Trust Magazine]. u.p.. Sample containing about 7521 words of miscellanea (domain: leisure)

CFP [Adverts from Parents]. u.p.. Sample containing about 7344 words of miscellanea (domain: commerce)

CFR [Leaflets advertising goods and products]. u.p.. Sample containing about 23409 words of miscellanea (domain: commerce)

CFS [Adverts from Clothes Show]. u.p.. Sample containing about 37735 words of miscellanea (domain: commerce)

CFT [Goods adverts]. u.p.. Sample containing about 59757 words of miscellanea (domain: commerce)

CFU [Adverts from The Economist]. u.p.. Sample containing about 7328 words of miscellanea (domain: commerce)

CFV [Advertising Standards Authority]. u.p.. Sample containing about 30438 words of miscellanea (domain: commerce)


CG5 Do It Yourself. Link House Magazines Ltd, Milton Keynes (1992-19-19). Sample containing about 28922 words from a periodical (domain: applied science)

CG6 Early language development. Harris, John, Routledge & Kegan Paul plc,
London (1990), 67-170. Sample containing about 38242 words from a book (domain: social science)


CG8 Electronic publishing. John Wiley & Sons Ltd, Chichester (1992-06). Sample containing about 4106 words from a periodical (domain: applied science)

CG9 Electronic publishing. John Wiley & Sons Ltd, Chichester (1992-06). Sample containing about 13194 words from a periodical (domain: applied science)


CGD Family work with elderly people. Froggatt, Alison, Macmillan Publishers Ltd, Basingstoke (1990), 1-107. Sample containing about 38630 words from a book (domain: social science)


CGF Feminism and linguistic theory. Cameron, Deborah, Macmillan Publishers Ltd, Basingstoke (1992), 36-128. Sample containing about 36582 words from a book (domain: arts)


| CHD | The use of orthographic and lexical information for handwriting recognition. Wells, Cynthia Joyce, u.p.. Sample containing about 36503 words of unpublished miscellanea (domain: applied science) |


**CJ4** Steam Railway News. 10A Birkdale Trading Estate, Southport (1992-19-19). Sample containing about 10286 words from a periodical (domain: leisure)

**CJ5** Steam Railway News. 10A Birkdale Trading Estate, Southport (1992-19-19). Sample containing about 8344 words from a periodical (domain: leisure)


**CJ7** Steam Railway News. 10A Birkdale Trading Estate, Southport (1992-19-19). Sample containing about 10949 words from a periodical (domain: leisure)

**CJ8** Steam Railway News. 10A Birkdale Trading Estate, Southport (1992-19-19). Sample containing about 8326 words from a periodical (domain: leisure)


14466 words from a periodical (domain: leisure)  
CJU [Adverts from The Economist]. u.p.. Sample containing about 24987 words of miscellanea (domain: commerce)  
CK0 Hospital circles. Andrews, Lucilla, Severn House, Wallington, Surrey (1986), 8-134. Sample containing about 41076 words from a book (domain: imaginative)  
main: leisure)


CKT The Art Newspaper. Umberto Alle-
12. List of works excerpted


Cl0 The Artist. The Artist Publishing Company Ltd, Tenterden (1993-04), 5-47. Sample containing about 23187 words from a periodical (domain: arts)


CL8 Computergram international. u.p.. Sample containing about 9100 words from a periodical (domain: applied science)

CL9 Computergram international. u.p.. Sample containing about 6727 words from a periodical (domain: applied science)

CLA Computergram international. u.p.. Sample containing about 9131 words from a periodical (domain: applied science)

CLB Computergram international. u.p.. Sample containing about 6880 words from a periodical (domain: applied science)

CLC Computergram international. u.p.. Sample containing about 9127 words from a periodical (domain: applied science)


from a periodical (domain: applied science)


**CLL** [Selection of OUP English Language Teaching promotional leaflets]. Oxford University Press, Oxford (1992/1993). Sample containing about 43505 words of miscellanea (domain: social science)


**CLS** *Film and the working class.* Stead, Peter, Routledge & Kegan Paul plc, London (1989), 1-90. Sample containing about 34205 words from a book (domain: arts)


**CLW** *Frameworks for teaching.* Dale, Roger, Robinson, Alison, Fergusson, R (eds.), Hodder & Stoughton Ltd, Sevenoaks, Kent (1992), 201-299. Sample containing about 38714 words from a book (domain: social science)


**CM0** *The headhunting business.* Jones, Stephanie, Macmillan Publishers Ltd, Basingstoke (1990), 6-97. Sample containing about 40262 words from a book (domain: commerce)

**CM1** *High elves.* King, Bill and Chambers, Andy, Games Workshop, Nottingham (1993), 4-81. Sample containing about 36941 words from a book (domain: imaginative)


12. List of works excerpted


**CMU**  *Special educational needs.*  Evans, Roy (ed.), Blackwell, Oxford (1989), 3-108.  Sample containing about 42788 words from a book (domain: social science)

**CMW**  *Unigram x.*  APT Data Services Ltd. (1993-04/1993-05).  Sample containing about 10765 words from a periodical (domain: applied science)

**CMX**  *Unigram x.*  APT Data Services Ltd. (1993-04/1993-05).  Sample containing about 12484 words from a periodical (domain: applied science)

**CMY**  *Unigram x.*  APT Data Services Ltd. (1993-04/1993-05).  Sample containing about 13218 words from a periodical (domain: applied science)

**CN0**  *Unigram x.*  APT Data Services Ltd. (1993-04/1993-05).  Sample containing about 8458 words from a periodical (domain: applied science)

**CN1**  *Warhammer armies: the Empire.*  Games Workshop, Nottingham (1993), 4-76.  Sample containing about 41212 words from a book (domain: leisure)


**CN4**  *The Artist.*  The Artist Publishing Company Ltd, Tenterden (1992-19-19), 7-49.  Sample containing about 22546 words from a periodical (domain: arts)

**CN5**  *School, work and equality.*  Flude, M., Hales, Margaret and Cosin, B., Hodder & Stoughton Ltd, Sevenoaks, Kent (1989), 1-98.  Sample containing about 41031 words from a book (domain: social science)

**CN6**  *Child care research, policy and practice.*  Kahan, Barbara (ed.), Hodder & Stoughton Ltd, Sevenoaks, Kent (1989), 1-100.  Sample containing about 33186 words from a book (domain: social science)


**CNA**  *British Medical Journal.*  British Medical Association, London (1993-01-16), 1-68.  Sample containing about 14063 words from a periodical (domain: natural sciences)

**CNB**  *Computergram international.*  u.p., Sample containing about 7612 words from a periodical (domain: applied science)

**CNE**  *Computergram international.*  u.p., Sample containing about 9125 words from a periodical (domain: applied science)

**CNF**  *Computergram international.*  u.p., Sample containing about 6808 words from a periodical (domain: applied science)

**CNH**  *Computergram international.*  u.p., Sample containing about 4946 words from a periodical (domain: applied science)

**CNJ**  *Computergram international.*  u.p., Sample containing about 9496 words from a periodical (domain: applied science)
Sample containing about 9319 words from a periodical (domain: applied science)

CNK Computergram international. u.p.
Sample containing about 7455 words from a periodical (domain: applied science)

CNL Computergram international. u.p.
Sample containing about 9654 words from a periodical (domain: applied science)

CNM Computergram international. u.p.
Sample containing about 9227 words from a periodical (domain: applied science)

CNN Computergram international. u.p.
Sample containing about 7201 words from a periodical (domain: applied science)

CNP Computergram international. u.p.
Sample containing about 9607 words from a periodical (domain: applied science)

CNR Computergram international. u.p.
Sample containing about 6756 words from a periodical (domain: applied science)

CNS Computergram international. u.p.
Sample containing about 9331 words from a periodical (domain: applied science)

CNT Computergram international. u.p.
Sample containing about 9066 words from a periodical (domain: applied science)

CNU Computergram international. u.p.
Sample containing about 6623 words from a periodical (domain: applied science)

CNV Computergram international. u.p.
Sample containing about 9051 words from a periodical (domain: applied science)

CNW Computergram international. u.p.
Sample containing about 7287 words from a periodical (domain: applied science)

CNX Computergram international. u.p.
Sample containing about 9483 words from a periodical (domain: applied science)

CNY Computergram international. u.p.
Sample containing about 9266 words from a periodical (domain: applied science)

CP0 Computergram international. u.p.
Sample containing about 7067 words from a periodical (domain: applied science)

CP2 Computergram international. u.p.
Sample containing about 6633 words from a periodical (domain: applied science)

CP3 Computergram international. u.p.
Sample containing about 9239 words from a periodical (domain: applied science)

CP4 Computergram international. u.p.
Sample containing about 9692 words from a periodical (domain: applied science)

CP5 Computergram international. u.p.
Sample containing about 6699 words from a periodical (domain: applied science)

CP6 Computergram international. u.p.
Sample containing about 9379 words from a periodical (domain: applied science)

CP7 Computergram international. u.p.
Sample containing about 6814 words from a periodical (domain: applied science)

CP8 Computergram international. u.p.
Sample containing about 9278 words from a periodical (domain: applied science)

CP9 Computergram international. u.p.
Sample containing about 6884 words from a periodical (domain: applied science)

CPA Computergram international. u.p.
Sample containing about 8704 words from a periodical (domain: applied science)

**CPB** Computergram international. u.p.. Sample containing about 9986 words from a periodical (domain: applied science)

**CPC** Computergram international. u.p.. Sample containing about 8998 words from a periodical (domain: applied science)

**CPD** Computergram international. u.p.. Sample containing about 6503 words from a periodical (domain: applied science)

**CPE** Computergram international. u.p.. Sample containing about 9473 words from a periodical (domain: applied science)

**CPF** Computergram international. u.p.. Sample containing about 6771 words from a periodical (domain: applied science)

**CPG** Computergram international. u.p.. Sample containing about 9296 words from a periodical (domain: applied science)

**CPH** Computergram international. u.p.. Sample containing about 9116 words from a periodical (domain: applied science)

**CPJ** Computergram international. u.p.. Sample containing about 7310 words from a periodical (domain: applied science)

**CPK** Computergram international. u.p.. Sample containing about 9164 words from a periodical (domain: applied science)

**CPL** Computergram international. u.p.. Sample containing about 7045 words from a periodical (domain: applied science)

**CPM** Computergram international. u.p.. Sample containing about 9874 words from a periodical (domain: applied science)

**CPN** Computergram international. u.p.. Sample containing about 9116 words from a periodical (domain: applied science)

**CPP** Computergram international. u.p.. Sample containing about 6978 words from a periodical (domain: applied science)

**CPR** Computergram international. u.p.. Sample containing about 9346 words from a periodical (domain: applied science)

**CPS** Computergram international. u.p.. Sample containing about 9098 words from a periodical (domain: applied science)

**CPT** Computergram international. u.p.. Sample containing about 9931 words from a periodical (domain: applied science)

**CPU** Computergram international. u.p.. Sample containing about 7269 words from a periodical (domain: applied science)

**CPV** Computergram international. u.p.. Sample containing about 9726 words from a periodical (domain: applied science)

**CPW** Computergram international. u.p.. Sample containing about 7143 words from a periodical (domain: applied science)

**CPX** Computergram international. u.p.. Sample containing about 8826 words from a periodical (domain: applied science)

**CPY** Computergram international. u.p.. Sample containing about 6983 words from a periodical (domain: applied science)

**CR0** Computergram international. u.p.. Sample containing about 9226 words from a periodical (domain: applied science)

**CR1** Computergram international. u.p.. Sample containing about 6863 words from a periodical (domain: applied science)

**CR2** Computergram international. u.p.. Sample containing about 9226 words from a periodical (domain: applied science)
Sample containing about 9291 words from a periodical (domain: applied science)

CR3 Computergram international. u.p.. Sample containing about 9131 words from a periodical (domain: applied science)


CRK Modern Catholicism: Vatican II and after. et al and Hasting, Adrian, SPCK, London (1991), 1-82. Sample containing about 39416 words from a book (domain: belief and thought)


CRP Photography. Argus Specialist Publications, Hemel Hempstead (1992-02), 4-48. Sample containing about 13048 words from a periodical (domain: leisure)


12. List of works excerpted


**CS7** Utopia, community care and retreat. Tomlinson, D, Open University Press, Milton Keynes (1991), 5-83. Sample containing about 32413 words from a book (domain: social science)

**CS8** Unigram x. APT Data Services Ltd. (1993-04/1993-05). Sample containing about 7833 words from a periodical (domain: applied science)

**CS9** Unigram x. APT Data Services Ltd. (1993-04/1993-05). Sample containing about 7960 words from a periodical (domain: applied science)

**CSA** Unigram x. APT Data Services Ltd. (1993-04/1993-05). Sample containing about 10523 words from a periodical (domain: applied science)

**CSB** Unigram x. APT Data Services Ltd. (1993-04/1993-05). Sample containing about 9450 words from a periodical (domain: applied science)

**CSC** Unigram x. APT Data Services Ltd. (1993-04/1993-05). Sample containing about 11181 words from a periodical (domain: applied science)

**CSD** Unigram x. APT Data Services Ltd. (1993-04/1993-05). Sample containing about 10417 words from a periodical (domain: applied science)

**CSE** Unigram x. APT Data Services Ltd. (1993-04/1993-05). Sample containing about 10424 words from a periodical (domain: applied science)

**CSF** Unigram x. APT Data Services Ltd. (1993-04/1993-05). Sample containing
<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Publisher</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSG</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 10202 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CSV</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 10820 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CSH</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 11043 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CSW</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 11841 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CSJ</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 10949 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CSX</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 11135 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CSK</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 10787 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CSL</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 10993 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CSM</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 10728 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CSN</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 10549 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CSP</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 10728 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CSR</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 10728 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CSS</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 10728 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CST</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 10728 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CT0</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 8747 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CT1</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 8747 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CT2</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 8747 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CT3</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 8747 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CT4</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 8747 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CT5</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 8747 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CT6</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 8747 words from a periodical (domain: applied science)</td>
</tr>
<tr>
<td>CT7</td>
<td>Unigram x.</td>
<td>APT Data Services Ltd. (1993-04/1993-05).</td>
<td>Sample containing about 8747 words from a periodical (domain: applied science)</td>
</tr>
</tbody>
</table>
about 12749 words from a periodical (domain: applied science)

**CT8** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 8086 words from a periodical (domain: applied science)

**CT9** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 8432 words from a periodical (domain: applied science)

**CTA** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 10299 words from a periodical (domain: applied science)

**CTB** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 8169 words from a periodical (domain: applied science)

**CTC** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 10962 words from a periodical (domain: applied science)

**CTD** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 10756 words from a periodical (domain: applied science)

**CTE** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 8594 words from a periodical (domain: applied science)

**CTF** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 7917 words from a periodical (domain: applied science)

**CTG** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 10521 words from a periodical (domain: applied science)

**CTH** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 11397 words from a periodical (domain: applied science)

**CTJ** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 11342 words from a periodical (domain: applied science)

**CTK** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 10878 words from a periodical (domain: applied science)

**CTL** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 9023 words from a periodical (domain: applied science)

**CTM** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 8042 words from a periodical (domain: applied science)

**CTN** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 10318 words from a periodical (domain: applied science)

**CTP** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 7766 words from a periodical (domain: applied science)

**CTR** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 11006 words from a periodical (domain: applied science)

**CTS** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 7770 words from a periodical (domain: applied science)

**CTT** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 8256 words from a periodical (domain: applied science)

**CTU** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 7332 words from a periodical (domain: applied science)

**CTV** *Unigram x.* APT Data Services Ltd. (1993-04/1993-05). Sample containing about 6928 words from a periodical (domain: applied science)


**CTX** *What Personal Computer.* EMAP Business & Computing, London (1993-04). Sample containing about 49013 words from a periodical (domain: ap-
12. List of works excerpted


CU0 Wisden Cricket Monthly. Wisden Cricket Magazines, Surrey (1992-03). Sample containing about 34660 words from a periodical (domain: leisure)


D8Y Museum Society meeting: Make-do and mend: Lecture/meeting.

D91 Harlow Playhouse public meeting.

D92 Harlow Playhouse public meeting.

D94 Harlow Women’s Institute committee meeting.

D95 Pensioners’ and Trades Union Association meeting.

D96 Pensioners’ and Trades Union Association meeting.

D97 Friends of the Earth: club meeting.

DCH Amnesty International meeting.
DCHPS002
DCHPS003
DCHPS004
DCHPS005
DCHPS006
DCHPS007
Sample containing about 14978 words
speech recorded in public context

DCJ  ‘A’-level history lecture.
DCJPS000
DCJPS001
Sample containing about 5109 words
speech recorded in educational context

DCK  ‘A’-level history lecture.
DCKPS000
Sample containing about 2437 words
speech recorded in educational context

E9N  The East Anglian. East Anglian Daily
Sample containing about 7939 words
from a periodical (domain: world aff-
airs)

E9P  The East Anglian. East Anglian Daily
Sample containing about 24442 words
from a periodical (domain: world aff-
airs)

E9R  The East Anglian. East Anglian Daily
Sample containing about 25252 words
from a periodical (domain: world aff-
airs)

E9S  The East Anglian. East Anglian Daily
Sample containing about 15814 words
from a periodical (domain: world aff-
airs)

E9T  The East Anglian. East Anglian Daily
Sample containing about 19374 words
from a periodical (domain: world aff-
airs)

E9U  The East Anglian. East Anglian Daily
Sample containing about 20351 words
from a periodical (domain: world aff-
airs)

E9V  Crown and nobility 1272-1461. Tuck,
Anthony, Blackwell, Oxford (1986),
84-186. Sample containing about
39125 words from a book (domain: world affairs)

E9W  The Artist’s and Illustrator’s Maga-
zine. Artist’s & Illustrator’s Mag, Lon-
don (1993-19-19), 5-48. Sample contain-
ing about 24586 words from a period-
odical (domain: arts)

E9X  Exploration for metalliferous and re-
lated minerals in Britain. Colman, T B,
British Geological Survey, Nottingham
(1990), 1-37. Sample containing about
21983 words from a book (domain: nat-
ural sciences)

(1993-01-01), 5-45. Sample containing
about 21578 words from a periodical
(domain: arts)

EA0  British Medical Journal. British Med-
ical Association, London (1993-01-16),
77-138. Sample containing about
45013 words from a periodical (do-
main: natural sciences)

EA1  British Medical Journal. British Medi-
cal Association, London (1993-01-16),
155-201. Sample containing about
30638 words from a periodical (do-
main: natural sciences)

EA2  British Medical Journal. British Medi-
cal Association, London (1993-01-16),
347-398. Sample containing about
36342 words from a periodical (do-
main: natural sciences)

EA3  Literacy in theory and practice. Street,
Brian, Cambridge University Press,
Cambridge (1984), 17-94. Sample con-
taining about 37377 words from a book
(domain: social science)

EA4  Community care. Philpot, T (ed.)King,
J and Bowry, K, Reed Publishing
Sample containing about 21147 words
from a periodical (domain: social sci-
ence)

EA5  The girl from Cotton Lane. Bowling,
Harry, Headline Book Publishing plc,
London (1992), 1-118. Sample con-
taining about 36937 words from a book
(domain: imaginative)
| EAA | Managing people at work. Hunt, John, u.p., 1-108. Sample containing about 39568 words from a book (domain: social science) |
| EAP | [New Oxford English Dictionary procedures documents]. u.p.. Sample containing about 6375 words of unpublished miscellanea (domain: commerce) |
| EAR | [New Oxford English Dictionary procedures documents]. u.p.. Sample containing about 4262 words of unpublished miscellanea (domain: commerce) |
12. List of works excerpted

**EB6** Zzap 64!. Newspaper Publications, Ludlow, Shropshire (1992-19-19), 4-68. Sample containing about 31780 words from a periodical (domain: leisure)


**EBR** The sociology of housework. Oakley, A, Blackwell, Oxford (1990), 1-128. Sample containing about 45374 words from a book (domain: social science)


12. List of works excerpted


EC5 [University of Bradford ephemeral]. University of Bradford, Bradford (1992). Sample containing about 11772 words of miscellanea (domain: social science)


EC9 Short breaks – Brighton and Hove 1992. u.p.. Sample containing about 9701 words of miscellanea (domain: leisure)


ECF Citalia Italy complete. CIT (England) Ltd, Croydon (1992). Sample containing about 72927 words of miscellanea (domain: leisure)


ECH Climer and Hill Walker. George Outram & Company Ltd, Glasgow (1991-08). Sample containing about 26797 words from a periodical (domain: leisure)


ECL Devon and Cornwall Seaside Holiday Parks. John Fowler Holidays (1991). Sample containing about 3915 words of
miscellanea (domain: leisure)


**ECS** *Tourism ephemera on Edinburgh.* u.p.. Sample containing about 17944 words of miscellanea (domain: leisure)


**ECX** *Flyer.* Insider Publications Ltd (1991-02). Sample containing about 35202 words from a periodical (domain: leisure)


**ED0** *Cosmos funbreaks by coach to France.* u.p.. Sample containing about 10212 words of miscellanea (domain: leisure)

**ED1** *Germany Switzerland Austria – GTF Tours.* German Tourist Facilities Ltd, London (1992). Sample containing about 14731 words of miscellanea (domain: leisure)


**ED6** *Gramophone.* General Gramophone Pubcs Ltd, Harrow (1992-06). Sample containing about 64637 words from a periodical (domain: arts)


**ED9** *Harpers & Queen.* The National Magazine Company Ltd, London (?). Sample containing about 73820 words from a periodical (domain: leisure)


**EDB** *Health promotion and education leaflets.* u.p.. Sample containing about 2606 words of miscellanea (domain: natural sciences)

Sample containing about 43280 words from a book (domain: social science)


**EDN** *King Solomon’s carpet.* Vine, Barbara, Penguin Group, London (1992), 1-122. Sample containing about 40226 words from a book (domain: imaginative)


**EDR** *The Lifeboat.* Royal National Lifeboat Institution, Poole, Dorset (1991 (Summer)). Sample containing about 21268 words from a periodical (domain: social science)

**EDT** *Management Today.* Haymarket Publishing Group, London (1991-03). Sample containing about 45969 words from a periodical (domain: commerce)

**EDU** *Marxism Today.* u.p., Sample containing about 38426 words from a periodical (domain: world affairs)


**EE0** *[Midland Bank leaflets].* u.p.. Sample containing about 29110 words of miscellanea (domain: commerce)

**EE1** *Military Illustrated.* Military Illustrated Ltd, London (1991-03). Sample containing about 25138 words from a periodical (domain: social science)


**EE6** *The Nottingham Graduate.* University of Nottingham, Nottingham (1992). Sample containing about 8671 words of miscellanea (domain: leisure)

**EE7** *[Official leaflets from Northern Ireland].* u.p.. Sample containing about
9059 words of miscellanea (domain: commerce)


**EED** [Rail travel leaflets]. u.p.. Sample containing about 7520 words of miscellanea (domain: leisure)


**EEH** Holidays in Robin Hood country. u.p.. Sample containing about 5315 words of miscellanea (domain: leisure)

**EEJ** [Royal Mail information leaflets]. Royal Mail, Britain (1991). Sample containing about 13048 words of miscellanea (domain: commerce)


**EEL** [Salford University: miscellaneous documents]. University of Salford, Salford (1991). Sample containing about 18975 words of miscellanea (domain: leisure)


**EEN** Sex, politics and society. Weeks, J, Longman Group UK Ltd, Harlow (1992), 1-95. Sample containing about 37810 words from a book (domain: social science)

**EER** Smallholder. Smallholder Pubs Ltd, King’s Lynn, Norfolk (1991-10). Sample containing about 25599 words from a periodical (domain: leisure)

**EES** Large vocabulary semantic analysis for text recognition. Rose, Tony Gerard, u.p.. Sample containing about 42217 words of unpublished miscellanea (domain: applied science)

**EET** Sovereign worldwide. Sovereign worldwide, Crawley (1989). Sample containing about 27951 words of miscellanea (domain: leisure)

**EEV** Steam Railway News. Lancashire Publications Ltd, Wigan (1991-10). Sample containing about 10390 words from a periodical (domain: leisure)

**EEW** In sunshine or in shadow. Bingham, C, Bantam (Corgi), London (1992), 11-103. Sample containing about 31057 words from a book (domain: imaginative)

**EEX** The story of Swanage. Lewer, David,
Harewood Publications, Bournemouth (1986). Sample containing about 9530 words of miscellanea (domain: world affairs)


EF0 Theology and feminism. Hampson, D, Blackwell, Oxford (1992), 7-96. Sample containing about 38532 words from a book (domain: belief and thought)


EFC [Leaflets on tourism in Wales]. u.p.. Sample containing about 21423 words of miscellanea (domain: leisure)

EFD Is West Belfast working?. West Belfast Economic Forum, Belfast (1990). Sample containing about 32867 words from a periodical (domain: commerce)

EFE Warm hearted weekends. u.p.. Sample containing about 5954 words of miscellanea (domain: leisure)


about 45873 words from a book (domain: world affairs)


**EFS** Truth, dare or promise. Steedman, Carolyn, Virago Press Ltd, London (1985), 103-123. Sample containing about 7822 words from a book (domain: social science)

**EFT** The Lion concise Bible handbook. Alexander, David, Alexander, Pat (eds.), Lion Publishing plc, Tring, Herts (1990), 8-83. Sample containing about 38377 words from a book (domain: belief and thought)


**EUW** Design and development of programs as teaching materials. Fraser, Rosemary, Burkhardt, Hugh and et al, Council for Educational Technology, UK (1982), 13-77. Sample containing about 21182 words from a book (domain: applied science)


**EUY** Fallables. Geesin, Ron, Ron Geesin, UK (1975), 1-101. Sample containing about 10294 words from a book (domain: imaginative)


**EV5** Learning to care on the surgical ward. Merchant, Jane and Attree, Moira, Hodder & Stoughton Ltd, Sevenoaks, Kent (1989), 1-110. Sample containing about
25119 words from a book (domain: social science)


**EVA** *Sleep and dreaming.* Empson, Jacob, Faber & Faber Ltd, London (1989), 3-121. Sample containing about 41017 words from a book (domain: social science)

**EVB** *Teaching spelling.* Torbe, Mike, Ward Lock Education, UK (1988), 5-56. Sample containing about 15786 words from a book (domain: social science)


**EVF** *A career in advertising and public relations.* Whiteaker, Stafford, BT Batsford Ltd, London (1986), 9-100. Sample containing about 33785 words from a book (domain: commerce)

**EVG** *Another time, another season.* Dillon, Anna, Sphere Books Ltd, London (1989), 1-128. Sample containing about 41755 words from a book (domain: imaginative)


**EVK** *Nutshells: constitutional and administrative law.* Hogan, Greer, Sweet & Maxwell Ltd, London (1990), 1-92. Sample containing about 36523 words from a book (domain: social science)


**EVN** *Cue Bruce.* Hockin, Bruce, Redcliffe Press Ltd, UK (1990), 7-78. Sample containing about 35904 words from a book (domain: arts)


**EVV** *Graduated tests in mathematics.* Thorpe, Jim, Ruddock, Graham and Foxman, Derek, Nfer-Nelson Publishing Company Ltd, UK (1989), 1-152. Sample containing about 39587 words from a book (domain: natural sciences)


<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Author(s)</th>
<th>Publisher</th>
<th>Year</th>
<th>Pages</th>
<th>Sample Size</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVY</td>
<td>Manpower solutions</td>
<td>Dean, Derek J, Scutari Projects Ltd, UK</td>
<td>1987</td>
<td>1-60</td>
<td></td>
<td></td>
<td>commerce</td>
</tr>
<tr>
<td>EW1</td>
<td>The age of Balfour and Baldwin 1902-1940</td>
<td>Ramsden, John, Longman Group UK Ltd, Harlow</td>
<td>1978</td>
<td>65-151</td>
<td></td>
<td></td>
<td>world affairs</td>
</tr>
<tr>
<td>EW5</td>
<td>Profitboss</td>
<td>Freemantle, David, Pan Books Ltd, London</td>
<td>1988</td>
<td>17-221</td>
<td></td>
<td></td>
<td>world affairs</td>
</tr>
<tr>
<td>EW6</td>
<td>The quantum world</td>
<td>Polkinghorne, J C, Longman Group UK Ltd, Harlow</td>
<td>1984</td>
<td>1-82</td>
<td></td>
<td></td>
<td>natural sciences</td>
</tr>
<tr>
<td>EW7</td>
<td>Resource-based learning</td>
<td>Beswick, Norman, Heinemann Educational Books Ltd, UK</td>
<td>1984</td>
<td>32-121</td>
<td></td>
<td></td>
<td>commerce</td>
</tr>
<tr>
<td>EW8</td>
<td>Sexual aspects of social work</td>
<td>Stewart, W F R, Woodhead-Faulkner Ltd, UK</td>
<td>1979</td>
<td>13-146</td>
<td></td>
<td></td>
<td>social science</td>
</tr>
<tr>
<td>EW9</td>
<td>Stockport Grammar School 1487-1987</td>
<td>Ball, James, Old Vicarage Publications, Congleton</td>
<td>1987</td>
<td>12-99</td>
<td></td>
<td></td>
<td>world affairs</td>
</tr>
<tr>
<td>EWA</td>
<td>Style in fiction</td>
<td>Short, Michael H and Leech, Geoffrey N, Longman Group UK Ltd, Harlow</td>
<td>1987</td>
<td>11-146</td>
<td></td>
<td></td>
<td>arts</td>
</tr>
<tr>
<td>EWB</td>
<td>Walking the Dales</td>
<td>Harding, Mike, Michael Joseph Ltd, London</td>
<td>1989</td>
<td>51-129</td>
<td></td>
<td></td>
<td>leisure</td>
</tr>
<tr>
<td>EWG</td>
<td>The Americas</td>
<td>Lancaster, A B, Edward Arnold (Publishers) Ltd, Sevenoaks, Kent</td>
<td>1984</td>
<td>1-114</td>
<td></td>
<td></td>
<td>world affairs</td>
</tr>
<tr>
<td>EWH</td>
<td>Armada</td>
<td>Gidley, Charles, Fontana Paperbacks, London</td>
<td>1988</td>
<td>11-120</td>
<td></td>
<td></td>
<td>imaginative</td>
</tr>
<tr>
<td>EWM</td>
<td>Landscape for a good woman [from Truth, dare or promise]</td>
<td>Heron, Liz, Virago Press Ltd, London</td>
<td>1985</td>
<td>153-169</td>
<td></td>
<td></td>
<td>social science</td>
</tr>
<tr>
<td>EWW</td>
<td>Matrices and engineering dynamics</td>
<td>Simpson, A and Collar, A R, Ellis Horwood Ltd, Chichester</td>
<td>1987</td>
<td>11-195</td>
<td></td>
<td></td>
<td>social science</td>
</tr>
<tr>
<td>EWX</td>
<td>Myocardial infarction: acute care and rehabilitation</td>
<td>Carne, S J and Harten-Ash, V J, Duphar Laboratories Ltd, UK</td>
<td>1990</td>
<td>1-51</td>
<td></td>
<td></td>
<td>natural sciences</td>
</tr>
<tr>
<td>EX0</td>
<td>Oriental rugs: a buyer's guide</td>
<td>Al- lane, Lee, Thames &amp; Hudson Ltd, UK</td>
<td>1988</td>
<td>6-108</td>
<td></td>
<td></td>
<td>leisure</td>
</tr>
<tr>
<td>EX1</td>
<td>Motor racing: records, facts and champions</td>
<td>Morrison, Ian, Guinness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>leisure</td>
</tr>
</tbody>
</table>
12. List of works excerpted

Publications, UK (1989), 7-188. Sample containing about 15224 words from a book (domain: leisure)


EX8 Water supplies and the treatment and disposal of effluents. Little, A H, The Textile Institute, UK (1975), 1-31. Sample containing about 11323 words from a book (domain: applied science)

F71 A poet’s response to the pictures of Gauguin: [Picture appreciation lesson].
F71PS000
F71PS001
F71PS002
F71PS003
F71PS004
F71PS005
F71PS006
F71PS007
Sample containing about 6754 words speech recorded in educational context

F72 [Storytelling workshop].
F72PS000
F72PS001
Sample containing about 6867 words speech recorded in educational context

F73 [Ralph Gardner High School: talk between teacher and caretaker].
PS1KU 46, ‘Andrew’, teacher, male
PS1KV 40, caretaker, male
Sample containing about 370 words speech recorded in educational context

F74 [Ralph Gardner High School: art lesson for fourteen-year-olds].
PS1KW 46, ‘Andrew’, teacher, male
PS1KX 35+, teacher, female
PS1KY 14, ‘Michael’, student, male
Sample containing about 929 words speech recorded in educational context

F75 [Mounting a picture: demonstration].
PS1L0 46, ‘Andrew’, teacher, male
Sample containing about 597 words speech recorded in educational context

F76 [School Registration].
PS1L1 46, ‘Andrew’, teacher, AB, male
PS1L2 14, ‘Gillian’, student, DE, female
Sample containing about 161 words speech recorded in educational context

F77 [Etching lesson].
PS1L3 46, ‘Andrew’, teacher, male
PS1L4 45, teacher, female
PS1L5 14, ‘Kevin’, student, male
Sample containing about 4901 words speech recorded in educational context

F78 [Staff meeting].
PS1L6 46, ‘Andrew’, teacher, male
PS1L7 30, teacher, female
Sample containing about 58 words speech recorded in educational context

F7A [Tyneside Cinema Board meeting].
PS1L8 50+, ‘Roger’, male
PS1L9 40+, ‘Peter’, chief executive, male
PS1LA 30+, ‘Susie’, lecturer, female
PS1LB 50+, ‘Geoff’, accountant, male
PS1LC 50+, ‘John’, lecturer, male
PS1LD 40+, ‘Colin’, male
PS1LE 20+, ‘Bryony’, secretary, female
Sample containing about 14014 words speech recorded in business context

F7C [Audio Description Action Group: meeting].
PS1LG 30+, ‘Robert’, arts officer, male
PS1LH 30+, ‘Brian’, theatre manager, male
PS1LJ 50+, ‘Iris’, audio services, female
PS1LK 30+, ‘Jane’, social worker, female
PS1LL  40+, ‘Cathy’, social worker, female
F7CPS000
Sample containing about 13805 words
speech recorded in business context

F7E  [Teachers’ conference: discussing
maths lessons].
PS1LS  40+, ‘Alan’, teacher, male
PS1LT  40+, ‘Cath’, teacher, female
PS1LU  40+, ‘Judith’, teacher, female
PS1LV  30+, ‘Ian’, teacher, male
F7EPS000
Sample containing about 4244 words
speech recorded in educational context

F7F  [Teachers’ conference: discussing
assessment procedures].
PS1LW  50+, ‘Don’, teacher, male
F7FPS000
F7FPS001
F7FPS002
F7FPS003
F7FPS004
Sample containing about 15742 words
speech recorded in educational context

F7G  [Teachers’ conference: discussing
assessment procedures].
PS1M4  40+, ‘Andrew’, teacher, male
PS1M5  30+, ‘Angela’, teacher, female
PS1M6  30+, ‘Paul’, teacher, male
PS1M7  40+, ‘Rod’, teacher, male
PS1M8  50+, ‘Don’, teacher, male
PS1M9  40+, ‘Alan’, teacher, male
PS1MA  30+, ‘Terry’, teacher, male
Sample containing about 5676 words
speech recorded in educational context

F7J  [COHSE/NALGO/NUPE/meeting].
F7JPS000
F7JPS001
F7JPS002
Sample containing about 11154 words
speech recorded in public context

F7K  [Ampleforth Prep School: remedial
English lesson for nine-year-olds].
PS1ME  35+, teacher, female
Sample containing about 1751 words
speech recorded in educational context

F7L  [Ampleforth Prep School: geography
lesson for twelve-year-olds].
PS1MF  50+, teacher, female
Sample containing about 1751 words
speech recorded in educational context

F7M  [Ampleforth Prep School: English
lesson for nine-year-olds].
PS1MG  35+, teacher, male
Sample containing about 19876 words
speech recorded in public context

F7N  [Mark Hall School: religious studies
lesson for eleven-year-olds].
PS1MH  50+, teacher, female
Sample containing about 2058 words
speech recorded in educational context

F7R  [Mark Hall School: science lesson for
fourteen- and fifteen-year-olds].
PS1ML  30+, teacher, female
PS1MM  ‘Steve’, male
Sample containing about 5367 words
speech recorded in educational context

F7T  [Lecture for eighteen-year-old students
on local councils and elections].
PS1MP  30+, ‘Ken’, lecturer, male
F7TPS000
Sample containing about 444 words
speech recorded in educational context

F7U  [Lecture on Isomers].
PS1MR  46, ‘Danny’, lecturer, male
F7UPS000
Sample containing about 11065 words
speech recorded in educational context

F7V  [Personal services: meeting of council-
lors].
F7VPS000
Sample containing about 19876 words
speech recorded in public context

F7W  [Clitheroe magistrates’ court: trials].
Sample containing about 6984 words
speech recorded in public context

F7X  [Clitheroe magistrates’ court: trials].
Sample containing about 11154 words
speech recorded in public context

F7Y  [Harlow Study Centre: oral history in-
terview].
PS1MS  interviewer, female
PS1MT  70+, ‘Mrs Druce’, retired, female
Sample containing about 4832 words
speech recorded in leisure context

F81  [Harlow Study Centre: interview].
PS1MW
PS1MX, male
PS1MY, female
Sample containing about 7762 words speech recorded in leisure context

F82 [Harlow Study Centre: interview].
PSIN0 female
PSIN1 male
PSIN2 female
Sample containing about 5347 words speech recorded in leisure context

F84 [Church of Scotland: Meeting on rules and regulations].
PSIN6 'Hugh', moderator, male
PSIN7 'Mr Malloch', male
PSIN8 'Mr Holt', male
PSIN9 'Mr McKweon', male
F84PS000
Sample containing about 2066 words speech recorded in public context

F85 [Church of Scotland: Report on Baptism].
PSINA 'Hugh', moderator, male
PSINB 'Dr Stuart', male
PSINC 'Ian', male
F85PS000 F85PS001 F85PS002 F85PS003 F85PS004 F85PS005
Sample containing about 10501 words speech recorded in public context

F86 [Church of Scotland: Meeting on rules and regulations].
PSINE 'Hugh', moderator, AB, male
PSINF 'Mr Boyd', male
PSING 'Mr Torrence', male
PSINJ 'Mr McGilvery', male
F86PS000 F86PS001 F86PS002 F86PS003
Sample containing about 9132 words speech recorded in public context

F87 [Church of Scotland: Report on Church funds].
PSINK 'Hugh', moderator, male
F87PS000 F87PS001 F87PS002
Sample containing about 3638 words speech recorded in public context

F88 [Birmingham College of Food: lecture on tourism].
PSINP 30, 'Melvin', lecturer, male
Sample containing about 7817 words speech recorded in educational context

F89 [Birmingham College of Food: student union induction speech].
PSINR 25, 'Joey', student union president, male
Sample containing about 1452 words speech recorded in educational context

F8A [Birmingham College of Food: lecture on food].
PSINS 40+, 'Pat', lecturer, female
Sample containing about 3160 words speech recorded in educational context

F8B [Harlow Rotary Club: talk on engineering].
PSINX 50+, 'Tom Grundy', AB, male
PSINY 50+, 'John', DE, male
PSINZ 60+, 'Donald', male
Sample containing about 5756 words speech recorded in leisure context

F8C [British Red Cross first aid course].
PSIP7 50, 'Audrey', first aid officer, female
Sample containing about 5235 words speech recorded in educational context

F8D [British Red Cross first aid course].
PSIP8 50, 'Audrey', first aid officer, female
PSIP9 40, 'Elaine', lecturer, female
PSIPA 40, 'Denise', lecturer, female
Sample containing about 11556 words speech recorded in educational context

F8E [Newcastle University Department of Marine Biology and Coastal Management: lecture on oceanography].
PSIPB 30+, 'Chris', lecturer, male
Sample containing about 5167 words speech recorded in educational context

F8F [Newcastle University Department of Marine Biology and Coastal Management: tutorial].
PSIPC 40+, tutor, female
F8FPS000 F8FPS001
Sample containing about 643 words speech recorded in educational context

F8G [Newcastle University Department of Marine Biology and Coastal Management: discussion and training video].
Sample containing about 3726 words speech recorded in educational context
F8H  [Newcastle University Department of Marine Biology and Coastal Management: lecture on communication skills and training video].
PS1PD  50+, head of department
Sample containing about 4085 words speech recorded in educational context

F8J  [Newcastle University: lecture on word processing].
PS1PE  30+, lecturer, male
Sample containing about 2066 words speech recorded in educational context

F8L  [Norwich City College: NNEB lecture].
PS1PF  30, ‘Cathy’, lecturer, female
Sample containing about 10567 words speech recorded in educational context

F8M  [Norwich City College: drama course meeting and lecture].
PS1PG  50+, ‘Francis’, lecturer, male
Sample containing about 5757 words speech recorded in educational context

F8N  [Harlow Rotary Club: talk on atomic engineering].
PS1PH  ‘Ian’, lecturer, male
PS1PJ  ‘Don’, male
PS1PK  ‘Frank’, male
PS1PL  ‘Gordon’, male
PS1PM  ‘Jeff’, male
Sample containing about 5436 words speech recorded in leisure context

F8P  [Walsall Local History Centre: talk about Home Guard].
PS1PN  ‘Joyce’, interviewer, female
PS1PP  70, ‘Ernest’, retired, male
Sample containing about 7960 words speech recorded in leisure context

F8R  [Leigh Community Centre: lecture on China].
PS1PR  30, ‘Neil’, lecturer, male
Sample containing about 15819 words speech recorded in educational context

F8S  [Newcastle University: lecture on microbiology].
PS1PS  38, lecturer, male
Sample containing about 5243 words speech recorded in educational context

F8U  [BNC Project: weekly progress meeting].
PS1RA  25, ‘Wendy’, lexicographer, female
PS1R7  27, ‘Clare’, secretary, female
PS1R8  24, ‘Derek’, chemist, male
PS1R9  24, ‘David’, physicist, male
Sample containing about 9511 words speech recorded in business context

F98  Computers and the humanities. Kenny, A, u.p.. Sample containing about 5688 words of unpublished miscellanea (domain: applied science)


F9E  Introduction to the sociology of development. Webster, A, Macmillan Publishers Ltd, Basingstoke (1990), 1-128. Sample containing about 42120 words from a book (domain: social science)


F9G  Beyond the inner city. Byrne, D,
12. List of works excerpted


**F9M** *City psalms.* Zephaniah, B, Bloodaxe books ltd, Newcastle upon Tyne (1992), 11-64. Sample containing about 8494 words from a book (domain: imaginative)


Sample containing about 44551 words from a book (domain: social science)


FAP Murder forestalled. Chester, P, Chivers Press, UK (1990), 7-151. Sample containing about 44565 words from a book (domain: imaginative)


FAT Pillion riders. Taylor, E R, Peter Owen
12. List of works excerpted

(1993), 38-165. Sample containing about 42997 words from a book (domain: imaginative)


**FAY Race, culture and difference.** Donald, J and Rattansi, A, Open University Press, Milton Keynes (1993), 1-97. Sample containing about 42012 words from a book (domain: social science)

**FB0 The radiant way.** Drabble, Margaret, Penguin Group, London (1988), 1-103. Sample containing about 41456 words from a book (domain: imaginative)

**FB1 Rethinking the Russian revolution.** Aceton, Edward, Edward Arnold (Publishers) Ltd, Sevenoaks, Kent (1992), 5-106. Sample containing about 43703 words from a book (domain: world affairs)


**FB8 Cinema Secrets: Special effects.** Millar, D, Quintet Publishing Ltd, London (1990), 7-103. Sample containing about 23773 words from a book (domain: arts)


**FBE What is this thing called science?.**


FBH Community development journal. Oxford University Press, Oxford (1993-07), 199-287. Sample containing about 30448 words from a periodical (domain: social science)


FBS The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 4340 words from a periodical (domain: social science)

FBT The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 6523 words from a periodical (domain: social science)

FBU The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 4931 words from a periodical (domain: social science)

FBV The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 23062 words from a periodical (domain: social science)

FBW The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 12850 words from a periodical (domain: social science)

FBX The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 6391 words from a periodical (domain: social science)

FBY The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 7363 words from a periodical (domain: social science)

FC0 The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 3451 words from a periodical (domain: social science)

FC1 The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 2440 words from a periodical (domain: social science)

FC2 The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 4312 words from a periodical (domain: social science)

FC3 The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 2772 words from a periodical (domain: social science)

FC4 The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 9254
### List of works excerpted

<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Volume</th>
<th>Pages</th>
<th>Words</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>FC5</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 10340 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC6</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 12722 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC7</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 4061 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC8</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 3000 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC9</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 1984 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCA</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 6101 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCB</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 5700 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCC</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 5973 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCD</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 5762 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCE</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 10852 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCF</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 3687 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCG</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 3040 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCH</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 5353 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCJ</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 33093 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCK</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 12831 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCL</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 34864 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCM</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 5692 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCN</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 3949 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCP</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 5581 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCR</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 11097 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCS</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 10317 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCT</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 5663 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCU</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 3422 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCV</td>
<td>The Weekly Law Reports 1992 Volume 3</td>
<td>u.p.</td>
<td>Sample containing about 3774 words from a periodical (domain: social science)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
words from a periodical (domain: social science)

**FCW** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 8166 words from a periodical (domain: social science)

**FCX** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 6355 words from a periodical (domain: social science)

**FCY** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 3220 words from a periodical (domain: social science)

**FD0** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 6761 words from a periodical (domain: social science)

**FD1** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 19129 words from a periodical (domain: social science)

**FD2** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 5857 words from a periodical (domain: social science)

**FD3** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 26954 words from a periodical (domain: social science)

**FD4** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 8488 words from a periodical (domain: social science)

**FD5** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 6378 words from a periodical (domain: social science)

**FD6** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 4244 words from a periodical (domain: social science)

**FD7** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 12670 words from a periodical (domain: social science)

**FD8** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 9353 words from a periodical (domain: social science)

**FD9** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 9997 words from a periodical (domain: social science)

**FDA** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 12239 words from a periodical (domain: social science)

**FDB** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 8149 words from a periodical (domain: social science)

**FDC** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 15291 words from a periodical (domain: social science)

**FDD** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 14261 words from a periodical (domain: social science)

**FDE** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 930 words from a periodical (domain: social science)

**FDF** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 2864 words from a periodical (domain: social science)

**FDG** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 8310 words from a periodical (domain: social science)

**FDH** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 10035 words from a periodical (domain: social science)

**FDJ** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 5516 words from a periodical (domain: social science)

**FDK** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 6600 words from a periodical (domain: social science)

**FDL** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 6206 words from a periodical (domain: social science)
words from a periodical (domain: social science)

**FDM** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 4767 words from a periodical (domain: social science)

**FDN** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 7465 words from a periodical (domain: social science)

**FDP** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 11589 words from a periodical (domain: social science)

**FDR** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 7411 words from a periodical (domain: social science)

**FDS** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 4505 words from a periodical (domain: social science)

**FDT** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 5074 words from a periodical (domain: social science)

**FDU** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 6403 words from a periodical (domain: social science)

**FDV** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 22191 words from a periodical (domain: social science)

**FDW** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 10006 words from a periodical (domain: social science)

**FDX** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 2843 words from a periodical (domain: social science)

**FDY** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 6164 words from a periodical (domain: social science)

**FE0** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 3298 words from a periodical (domain: social science)

**FE1** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 21473 words from a periodical (domain: social science)

**FE2** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 27250 words from a periodical (domain: social science)

**FE3** The Weekly Law Reports 1992 Volume 3. u.p.. Sample containing about 13101 words from a periodical (domain: social science)


**FED** Left side right side. Beaton, Alan, Batsford Academic & Educational, UK (1985), 62-140. Sample containing about 39035 words from a book (domain: social science)

List of works excerpted

FEJ Model financial statements for public and private companies. Stoy Hayward, Butterworth & Company (pub) Ltd, UK (1990), 1-115. Sample containing about 17361 words from a book (domain: commerce)


FEX The vitality diet. Stewart, Alan and Stewart, Maryon, Thorsons Publishers, UK (1990), 7-229. Sample containing about 34333 words from a book (domain: leisure)


FL4 Menopause: television discussion.
FL5 Football: television discussion.
FL6 Eating disorders: television discussion.
FL7 Misogyny: television discussion.
FL8 Scottish Women: discussion about ‘the body beautiful’.
FL9 Scottish Women: discussion about crime.
Sample containing about 5035 words speech recorded in leisure context

**FLA** Scottish Women: discussion about transport.

Sample containing about 4914 words speech recorded in leisure context

**FLB** Scottish Women: discussion about love.

Sample containing about 4569 words speech recorded in leisure context

**FLC** Scottish Women: property, the arts, the press, school.

Sample containing about 4745 words speech recorded in leisure context

**FLD** Families: television discussion.

Sample containing about 4740 words speech recorded in leisure context

**FLE** The Royals: television discussion.

Sample containing about 4779 words speech recorded in leisure context

**FLF** Mental health: television discussion.

Sample containing about 4837 words speech recorded in leisure context

**FLG** Mental health: television discussion.

Sample containing about 4936 words speech recorded in leisure context

**FLH** Animal rights: television discussion.

Sample containing about 4861 words speech recorded in leisure context

**FLK** Young women in Scotland: television discussion.

Sample containing about 5270 words speech recorded in leisure context

**FLL** Traditions: television discussion.

Sample containing about 4904 words speech recorded in leisure context

**FLM** Scottish women: discussion about smoking.

Sample containing about 4837 words speech recorded in leisure context
12. List of works excerpted

Sample containing about 4982 words
speech recorded in leisure context

**FLP** *Scottish women: discussion about weddings.*

Sample containing about 4940 words
speech recorded in leisure context

**FLR** *Scottish women: discussion about sex education.*

Sample containing about 5226 words
speech recorded in leisure context

**FLS** *General Portfolio health and safety meeting.*

Sample containing about 10719 words
speech recorded in educational context

**FLX** *11th year science lesson on chemistry of metal processing.*

Sample containing about 6132 words
speech recorded in educational context

**FM0** *Defensive driving techniques: Lecture/seminar.*

Sample containing about 9342 words
speech recorded in business context

**FM1** *Defensive driving techniques: Lecture/seminar.*

Sample containing about 7508 words
speech recorded in business context

**FM2** *Missprint planning meeting.*

Sample containing about 14349 words
speech recorded in business context

**FM3** *Body Shop presentation: aromatherapy/Body Shop products.*

Sample containing about 3667 words
speech recorded in educational context

**FM4** *Tutorial lesson: GCSE maths tutoring session.*

Sample containing about 11124 words
speech recorded in educational context

**FM5** *Tutorial lesson: GCSE chemistry tutoring session.*

Sample containing about 12889 words
speech recorded in educational context
12. List of works excerpted

FM7 Strangers - talk by PC Bruce: Talk/presentation.
  PSISF ‘Pc Bruce’, police officer, male
  Sample containing about 11056 words speech recorded in educational context

FM8 Strangers - talk by PC Bruce: Talk/presentation.
  PSISG ‘Pc Bruce’, police officer, male
  Sample containing about 1100 words speech recorded in educational context

FM9 Strangers - talk by PC Bruce: Talk/presentation.
  PSISH ‘Pc Bruce’, police officer, male
  Sample containing about 314 words speech recorded in educational context

FMA Strangers - talk by PC Bruce: Talk/presentation.
  PSISJ ‘Pc Bruce’, police officer, male
  Sample containing about 1539 words speech recorded in educational context

FMB Science lesson: year 10.
  PSISK 50, ‘John’, teacher, male
  Sample containing about 3288 words speech recorded in educational context

FMC Music lesson: year 10.
  PSISL 50, ‘John’, teacher, male
  Sample containing about 3724 words speech recorded in educational context

FMD York University Careers Service: careers interview.
  PSISP 50, ‘Martin’, careers advisor, male
  PSISR 20+, ‘Ian’, student, male
  Sample containing about 5553 words speech recorded in educational context

FME Tutorial lesson: GCSE maths.
  PSISS 50+, ‘John’, tutor, male
  PSIST 16, ‘Ian’, student, male
  Sample containing about 10641 words speech recorded in educational context

FMF Tutorial lesson: junior-level spelling and maths.
  PSISU 50+, ‘John’, tutor, male
  PSISV 9, ‘Simon’, student, male
  Sample containing about 5936 words speech recorded in educational context

FMG Tutorial lesson: junior-level spelling and maths.
  PSISW 50+, ‘John’, tutor, male
  PSISX 9, ‘Simon’, student, male

FMGPS000 Sample containing about 13669 words speech recorded in educational context

FMH Tutorial lesson: junior-level spelling and maths.
  PSISY 50+, ‘John’, tutor, male
  PSIT0 9, ‘Simon’, student, male

FMHPS000 Sample containing about 6278 words speech recorded in educational context

FMJ Tutorial lesson: junior-level spelling and maths.
  PSIT1 50+, ‘John’, tutor, male
  PSIT2 9, ‘Simon’, student, male

FMK ‘How we can change the world by our witness’: confirmation preparation lesson.
  PSIT3 42, ‘Elizabeth’, teacher, female
  PSIT4 13, ‘Peter’, student, male

FML What archaeologists do: discussions.
  PSIT5 20+, ‘Toby’, student, male
  PSIT6 48, ‘Andrew’, archaeologist, male
  PSIT7 24, ‘Derek’, transcriber, male
  PSIT8 20+, student, male
  PSIT9 20+, student, female
  PSITA 40+, archaeologist, male
  PSITB 20+, ‘Emma’, archaeologist, female

FMM Tutorial lesson: maths.
  PSITM 50+, ‘Malcolm’, tutor, male
  PSITN 13+, student, female

FMMPS000 Sample containing about 5595 words speech recorded in educational context

FMN Planning and development in York: greenbelt planning - public enquiry.
  PSITP 55, ‘John’, department of the environment adjudicator, male
  PSITR 58, ‘Harry’, deputy chairman, male
  PSIT5 64, ‘George’, barrister, male
  PSITT 46, barrister, male
  PSITU 30, barrister, male
  PSITV 40, barrister, male

FMNPS000 Sample containing about 11236 words speech recorded in public context

FMP Planning and development in York: greenbelt planning - public enquiry.
  PSITW 55, ‘John’, department of the environment adjudicator, male
Sample containing about 15200 words speech recorded in public context

FMR Tutorial lesson: GCSE chemistry.

PS1UB 50, ‘John’, tutor, male
PS1UC 16, ‘Andrew’, student, male

Sample containing about 13083 words speech recorded in educational context

FMS Legal advice: pre-retirement course.

PS1UD 50, solicitor, male

Sample containing about 11906 words speech recorded in public context


FP8 Cases & materials on constitutional & administrative law. Thompson, B, Walsh, B and Allen, M, Blackstone Press, London (1990), 1-70. Sample containing about 35516 words from a book (domain: social science)

FP9 Consumption, identity and style. Tomlinson, Alan (ed.)Ewen, S, Cubitt, S,

FPA *Creed*. Herbert, J, New English Library, Sevenoaks (1991), 129-262. Sample containing about 40796 words from a book (domain: imaginative)


FPW *A shorter history of Greek art*. Robertson, M, Cambridge University Press, Cambridge (1992), 1-89. Sample con-
12. List of works excerpted


- **FPY** *In tune with heaven.* Hodder & Stoughton Ltd, Sevenoaks, Kent (1992), 86-214. Sample containing about 43020 words from a book (domain: belief and thought)


- **FR7** *Evolution and present state of the ocean bill of lading.* Kozolchyk, Boris, u.p.. Sample containing about 23324 words from a book (domain: commerce)


12. List of works excerpted


FRP *The oyster catchers*. Gower, Iris, Bantam (Corgi), London (1992), 158-283. Sample containing about 39981 words from a book (domain: imaginative)


FS7 *Understanding social policy*. Hill, M,


**FSE** *The earth is the Lord’s*. James, W, Orbit (Sphere), London (1992), 57-206. Sample containing about 43645 words from a book (domain: imaginative)

**FSF** *The relic*. Anthony, E, Arrow (Hutchinson), London (1992), 72-180. Sample containing about 35495 words from a book (domain: imaginative)


**FSH** *The fog*. Herbert, J, New English Library, Sevenoaks (1993), 115-229. Sample containing about 36981 words from a book (domain: imaginative)


**FSV** *The Bookseller*. J Whitaker & sons, London (1993-19-19). Sample containing about 29021 words from a periodi-
12. List of works excerpted


FTA  Shield: the international magazine of the BP group. BP (1993). Sample containing about 11719 words from a periodical (domain: commerce)


12. List of works excerpted


**FTY** Dementia: sharpening local plans: priorities for the 90’s. Killeen, Jan (ed.), Scottish Action on Dementia, UK (1990), 1-94. Sample containing about 34686 words from a book (domain: social science)


**FU1** Falling for love. Sharpe, Sue, Virago Press Ltd, London (1987), 1-123. Sample containing about 42461 words from a book (domain: social science)


**FU4** Pamela. Morgan, Fidelis and Havergal, Giles, Amber Lane Press, UK (1987), 5-77. Sample containing about 22686 words from a book (domain: imaginative)

**FU5** Cometary phases. Evans, Christine, Seren Books, UK (1989), 7-106. Sample containing about 11053 words from a book (domain: imaginative)


**FU7** Revolt in Roundhay [excerpt from Truth, Dare or Promise]. Rowbotham, Sheila, Virago Press Ltd, London (1985). Sample containing about 8332 words from a book (domain: social science)


**FUA** Teaching English with video. Allan, Margaret, Longman Group UK Ltd, Harlow (1985), 3-105. Sample containing about 34467 words from a book (domain: social science)


**FUE** Introduction to retirement: pre-retirement course.

**FUF** Introduction to retirement: pre-retirement course.

**FUG** Management training course.

**FUH** Tutorial lesson: junior-level maths.
12. List of works excerpted

Sample containing about 11759 words speech recorded in educational context

**FUJ** British Rail team brief meeting.
PS1UG 55, ‘Dennis’, british rail employee, male
PS1UH 45, ‘Terry’, british rail employee, male
PS1UJ 50, ‘Roger’, british rail employee, male
PS1UK 48, ‘Trevor’, british rail employee, male
PS1UM 54, ‘Hugh’, british rail employee, male
PS1UN 50, ‘Dave’, british rail employee, male
PS1UP 55, ‘Norman’, british rail employee, male
PS1UR 55, ‘Nola’, british rail employee, female

Sample containing about 17918 words speech recorded in educational context

**FUK** British Rail team brief meeting.
PS1US 45, ‘Terry’, british rail employee, male
PS1UT 50, ‘Roger’, british rail employee, male
PS1UU 48, ‘Trevor’, british rail employee, male
PS1UV 52, ‘Jim’, british rail employee, male
PS1UW 54, ‘Hugh’, british rail employee, male
PS1UX 50, ‘Dave’, british rail employee, male
PS1UY 55, ‘Norman’, british rail employee, male
PS1V0 55, ‘Nola’, british rail employee, female

Sample containing about 19174 words speech recorded in business context

**FUL** British Rail team brief meeting.
PS1V2 55, ‘Dennis’, british rail employee, male
PS1V3 45, ‘Terry’, british rail employee, male
PS1V4 50, ‘Roger’, british rail employee, male
PS1V5 48, ‘Trevor’, british rail employee, male
PS1V6 52, ‘Jim’, british rail employee, male
PS1V7 54, ‘Hugh’, british rail employee, male
PS1V8 50, ‘Dave’, british rail employee, male
PS1V9 55, ‘Norman’, british rail employee, male

Sample containing about 15838 words speech recorded in business context

**FUM** Bladder cancer check-up timing: seminar presentation.
PS1VA 45, doctor, male

Sample containing about 1393 words speech recorded in business context

**FUN** Clinical evaluation of urology patients: seminar presentation.
PS1VB 50, doctor, male

Sample containing about 2196 words speech recorded in business context

**FUP** Angiogenesis as an indicator in bladder cancer: seminar presentation.
PS1VC 55, doctor, male

Sample containing about 2577 words speech recorded in business context

**FUR** Response to treatment in bladder cancers: seminar presentation.
PS1VD 65, doctor, male

Sample containing about 1459 words speech recorded in business context

**FUS** Results of prostatectomy survey: seminar presentation.
PS1VE 50, doctor, male

Sample containing about 1340 words speech recorded in business context

**FUT** Presentation on consumer rights.
PS1VF 60, retired trading standards officer, male

Sample containing about 8557 words speech recorded in business context

**FUU** Talk on fire prevention.
PS1VG 55, ‘Jack’, retired fire prevention officer, male

Sample containing about 8040 words speech recorded in business context

**FX5** Radio Forth: radio broadcast.
PS223 ‘David’, disc jockey, Other participants are radio listening phone ins., male

Sample containing about 9671 words speech recorded in leisure context

**FX6** Old Harlow W.
PS224 ‘Margaret’, secretary, female
PS225 ‘Vera’, female
PS226 ‘Gloria’, Is the speaker., female
PS227 ‘Mildred’, female

I. Meeting. Sample containing about 6777 words speech recorded in leisure context

**FX7** General practitioner’s surgery: medical consultation.
PS22S ‘Doctor’, doctor, other participants are doctors patients, AB, male

FX7PS000
FX7PS001

Sample containing about 280 words speech recorded in public context

**FX8** General practitioner’s surgery: medical consultation.
PS22T ‘Doctor’, doctor, other participants are doctors patients, AB, male

FX8PS000

Sample containing about 121 words speech recorded in public context
## 12. List of works excerpted

**FX9 General practitioner’s surgery: medical consultation.**
- **PS22U** ‘Doctor’, doctor, other participants are doctors patients, AB, male
- **FX9PS000**
- **FX9PS001**
  - Sample containing about 326 words speech recorded in public context

**FXB General practitioner’s surgery: medical consultation.**
- **PS22V** ‘Doctor’, doctor, other participants are doctors patients, AB, male
- **FXBPS000**
  - Sample containing about 316 words speech recorded in public context

**FXC General practitioner’s surgery: medical consultation.**
- **PS22W** ‘Doctor’, doctor, other participants are doctors patients, AB, male
- **FXCPS000**
  - Sample containing about 609 words speech recorded in public context

**FXD General practitioner’s surgery: medical consultation.**
- **PS22X** ‘Doctor’, doctor, other participants are doctors patients, AB, male
- **FXDPS000**
  - Sample containing about 750 words speech recorded in public context

**FXE General practitioner’s surgery: medical consultation.**
- **PS22Y** ‘Doctor’, doctor, other participants are doctors patients, AB, male
- **FXEPS000**
  - Sample containing about 109 words speech recorded in public context

**FXF General practitioner’s surgery: medical consultation.**
- **PS230** ‘Doctor’, doctor, other participants are doctors patients, AB, male
- **FXFPS000**
  - Sample containing about 192 words speech recorded in public context

**FXG General practitioner’s surgery: medical consultation.**
- **PS231** ‘Doctor’, doctor, other participants are doctors patients, AB, male
- **FXGPS000**
  - Sample containing about 199 words speech recorded in public context

**FXH General practitioner’s surgery: medical consultation.**
- **PS232** ‘Doctor’, doctor, other participants are doctors patients, AB, male
- **FXHPS000**
- **FXHPS001**
  - Sample containing about 1264 words speech recorded in public context

**FXJ General practitioner’s surgery: medical consultation.**
- **PS233** ‘Doctor’, doctor, other participants are doctors patients, AB, male
- **FXJPS000**
  - Sample containing about 979 words speech recorded in public context

**FXK General practitioner’s surgery: medical consultation.**
- **PS234** ‘Doctor’, doctor, other participants are doctors patients, AB, male
- **FXKPS000**
  - Sample containing about 106 words speech recorded in public context

**FXL General practitioner’s surgery: medical consultation.**
- **PS235** ‘Doctor’, doctor, other participants are doctors patients, AB, male
- **FXLPS000**
- **FXLPS001**
  - Sample containing about 356 words speech recorded in public context

**FXM General practitioner’s surgery: medical consultation.**
- **PS236** ‘Doctor’, doctor, other participants are doctors patients, AB, male
- **FXMPS000**
- **FXMPS001**
- **FXMPS002**
  - Sample containing about 558 words speech recorded in public context

**FXN General practitioner’s surgery: medical consultation.**
- **PS237** ‘Doctor’, doctor, other participants are doctors patients, AB, male
- **FXNPS000**
  - Sample containing about 726 words speech recorded in public context

**FXP Oral history project: interview.**
- **PS238** 80+, ‘John’, crofter, male
- **PS239**, male
  - Sample containing about 1713 words speech recorded in leisure context

**FXR Orchestral society committee meeting.**
- **PS23A** 23, ‘Greg’, doctor, male
- **PS23B** 49, ‘John’, civil servant, male
- **PS23C** 47, ‘Pauline’, teacher, female
- **PS23D** 50, ‘John’, retired teacher, male
12. List of works excerpted

PS23E 44, ‘Julia’, translator, female
PS23F 46, ‘Liz’, teacher, female
PS23G 58, ‘Malcolm’, retired school master, male

Sample containing about 11035 words speech recorded in leisure context

FXT Radio Trent: radio talk show.
PS23H weather forecaster, male
PS23J ‘Rob’, disc jockey, male
PS23K ‘Lisa’, care line worker, female
PS23L ‘Phil’, newsreader, male
PS23M ‘Perry’, australian diplomat, male
PS23N ‘Rob’, reporter, male
PS23P ‘Mick’, council executive, male
PS23R ‘Jenny’, disc jockey, female
PS23S ‘Ken’, reporter, male

Sample containing about 16234 words speech recorded in leisure context

FXU Nottingham Oral History Project: interview.
PS256 male
PS257 79, retired engineering toolmaker, male

Sample containing about 6185 words speech recorded in leisure context

FXV Nottingham Oral History Project: interview.
PS258 male
PS259 79, retired miner, valet, chauffer, male

Sample containing about 5752 words speech recorded in leisure context

FXW Nottingham Oral History Project: interview.
FXWPS000
FXWPS001

Sample containing about 3045 words speech recorded in leisure context

FXX Nottingham Oral History Project: interview.
PS25A
FXXPS000
FXXPS001
FXXPS002
FXXPS003

Sample containing about 9349 words speech recorded in leisure context

FYX Nottingham Oral History Project: interview.
PS25B
FYXPS000

Sample containing about 7433 words speech recorded in leisure context

FY0 Nottingham Oral History Project: interview.
PS25C
FY0PS000

Sample containing about 3728 words speech recorded in leisure context

FY1 Nottingham Oral History Project: interview.
PS25D 30, interviewer, female
PS25E 80, retired miner, male

Sample containing about 11004 words speech recorded in leisure context

FY2 Nottingham Oral History Project: interview.
PS25F
FY2PS000

Sample containing about 3864 words speech recorded in leisure context

FY3 Nottingham Oral History Project: interview.
PS25G
FY3PS000

Sample containing about 3140 words speech recorded in leisure context

FY4 Nottingham Oral History Project: interview.
PS25H
FY4PS000

Sample containing about 7036 words speech recorded in leisure context

FY5 Nottingham Oral History Project: interview.
PS25I
FY5PS000

Sample containing about 17697 words speech recorded in leisure context

FY6 Nottingham Oral History Project: interview.
PS25J
FY6PS000

Sample containing about 8039 words speech recorded in leisure context

FY7 Nottingham Oral History Project: interview.
PS25L
FY7PS000

Sample containing about 7564 words speech recorded in leisure context

FY8 Nottingham Oral History Project: interview.
PS25M nun, social worker
PS25N interviewer

Sample containing about 12703 words speech recorded in leisure context
202

12. List of works excerpted

FY9 Tutorial: GCSE mathematics lesson.

PS25P 50, ‘John’, teacher, male
PS25R 16, ‘Ian’, student, male

Sample containing about 5775 words
speech recorded in educational context

FYA Tutorial.

PS25S 50, ‘John’, teacher, male
PS25T ‘Chris’, student, male

Sample containing about 12019 words
speech recorded in educational context

FYB Methodist church meeting: oral history project interview.

PS25U FYBPS000
FYBPS001
FYBPS002
FYBPS003
FYBPS004
FYBPS005
FYBPS006
FYBPS007

Sample containing about 8978 words
speech recorded in public context

FYD Nottinghamshire Oral History Project: interview.

PS25W 80, ‘Ada’, retired box maker, female
PS25X 30, interviewer, female

Sample containing about 7419 words
speech recorded in leisure context

FYE Nottinghamshire Oral History Project: interview.

PS25Y 80, ‘Ada’, retired box maker, female
PS260 30, interviewer, female

Sample containing about 5824 words
speech recorded in leisure context

FYF Nottinghamshire Oral History Project: business meeting.

PS261 interviewer
PS262 security guard, male

Sample containing about 7353 words
speech recorded in leisure context

FYG Nottinghamshire Oral History Project: business meeting.

PS263 interviewer
PS264 security guard, male

Sample containing about 7174 words
speech recorded in leisure context

FYH Nottinghamshire Oral History Project: talk.

PS265 84, ‘Jack’, retired, male
PS266 , male
PS267 80, ‘Anne’, retired, female

Sample containing about 14311 words
speech recorded in leisure context

FYI Nottinghamshire Oral History Project: talk.

PS268 20, interviewer, male
PS269 70, retired, Ex. Political Activist, male

Sample containing about 11508 words
speech recorded in leisure context

FYK Nottinghamshire Oral History Project: interview.

PS26A interviewer
PS26B police sergeant, male
PS26C police constable, male

Sample containing about 5033 words
speech recorded in leisure context

FYL Nottinghamshire Oral History Project: interview.

PS26D interviewer
PS26E building society manager

Sample containing about 2984 words
speech recorded in leisure context

FYM Nottinghamshire Oral History Project: interview.

PS26F interviewer
PS26G pet shop owner

Sample containing about 5738 words
speech recorded in leisure context

FYP Mothers’ Union general discussion.

PS1NT 60+, ‘chair’, female
PS1NU 70+, ‘Jean’, female
PS1NV 30+, ‘Kate’, doctor, female
FYPPS000

Sample containing about 7434 words
speech recorded in leisure context

FYS Financial management for contractors.


FYW After the asylums. Murphy, E, Faber & Faber Ltd, London (1991), 111-217.
12. List of works excerpted

Sample containing about 37285 words from a book (domain: social science)

FYX Black holes and baby universes. Hawking, Stephen W, Bantam (Corgi), London (1993), 1-139. Sample containing about 39109 words from a book (domain: applied science)


G00 [Miscellaneous articles about desk-top publishing]. Budget, Henry and et al, u.p.. Sample containing about 94394 words of miscellanea (domain: commerce)


204 12. List of works excerpted

a book (domain: world affairs)


**G0K** *Fire prevention in high bay warehouses.* National Materials Handling Centre, Cranfield (1980), 1-82. Sample containing about 16066 words of miscellanea (domain: world affairs)


**G0U** *Organisational analysis.* BPP Management Education Ltd, London (1990-10). Sample containing about 49079 words from a periodical (domain: commerce)


**G0X** *Unexplained laughter.* Ellis, Alice T, Gerald Duckworth & Company Ltd, London (1985), 7-143. Sample containing about 46159 words from a book (domain: imaginative)


about 43042 words from a book (domain: imaginative)


G1W *Taped.* Barnes, Trevor, Coronet
<table>
<thead>
<tr>
<th>Code</th>
<th>Title</th>
<th>Publisher</th>
<th>Sample containing about words from a book (domain: )</th>
</tr>
</thead>
<tbody>
<tr>
<td>G21</td>
<td>The Artist’s and Illustrator’s Magazine. Artist’s &amp; Illustrator’s Mag,</td>
<td>London (1991-10). Sample containing about 21922 words from a periodical (domain: arts)</td>
<td></td>
</tr>
<tr>
<td>G2A</td>
<td>[Collection of estate agents’ property details]. u.p.. Sample containing</td>
<td>about 12765 words of miscellanea (domain: commerce)</td>
<td></td>
</tr>
<tr>
<td>G2G</td>
<td>[JFIT newsletters]. Institute of Electrical Engineers (1990/1991). Sample</td>
<td>containing about 16163 words of miscellanea (domain: applied science)</td>
<td></td>
</tr>
<tr>
<td>G2J</td>
<td>Marxism Today. u.p.. Sample containing about 39021 words from a periodi-</td>
<td>cal (domain: world affairs)</td>
<td></td>
</tr>
<tr>
<td>G2N</td>
<td>[Birmingham City Council: official leaflets]. u.p.. Sample containing</td>
<td>about 12765 words of miscellanea (domain: leisure)</td>
<td></td>
</tr>
</tbody>
</table>
G2P [Official leaflets]. u.p.. Sample containing about 11841 words of miscellanea (domain: commerce)

G2R [Vale of White Horse District Council: official leaflets]. u.p.. Sample containing about 3278 words of miscellanea (domain: commerce)


G31 The Birmingham Magazine. Birmingham University, Birmingham (1990-08). Sample containing about 18810 words from a periodical (domain: social science)


G3E Death of an expert witness. James, P D, Sphere Books Ltd, London (1979),
9-135. Sample containing about 40157 words from a book (domain: imaginative)


G3N *The laws of Occam programming.* Hoare, C A R and Roscoe, A W, Oxford University Computing Laboratory, UK (1986), 1-77. Sample containing about 15932 words from a book (domain: applied science)


G3S *Vets might fly.* Herriot, James, Pan Books Ltd, London (1977), 7-120. Sample containing about 42436 words from a book (domain: imaginative)


G3U *Justice and Peace Group meeting.*

PS1VH 40, ‘Charlie’, tradecraft worker, male
PS1VJ 40, ‘Moira’, tradecraft worker, female
GUPS000
GUPS001
GUPS002
GUPS003
GUPS004
GUPS005
GUPS006

Sample containing about 12715 words speech recorded in business context

G3V *Music lesson: grade V music theory.*

PS1VK 50, ‘Gill’, tutor, female
PS1VL 14, ‘Tom’, student, male

Sample containing about 3907 words speech recorded in educational context

G3W *Music lesson: grade V music theory.*

PS1VM 50, ‘Gill’, tutor, female
PS1VN 14, ‘Tom’, student, male

Sample containing about 3351 words speech recorded in educational context

G3X *Flower-arranging demonstration.*

PS1VP 60, ‘Anna’, retired, female
PS1VR 55, retired, female

Sample containing about 7341 words speech recorded in leisure context

G3Y *Careers guidance: seminar.*

PS1VS 50, ‘Martin’, careers advisor, male
PS1VT 20, ‘Lente’, student, female

Sample containing about 7477 words speech recorded in educational context

G42 *Medical consultation.*

PS1VU 60, doctor, male
PS1VV 50+, female
Sample containing about 614 words
speech recorded in public context
G43 Medical consultation.
PS1VW 60, doctor, male
PS1VX 18, ‘John’, male
Sample containing about 660 words
speech recorded in public context
G44 Medical consultation.
PS1VY 60, doctor, male
PS1W0 65, ‘Isobel’, female
Sample containing about 469 words
speech recorded in public context
G45 Medical consultation.
PS1W1 60, doctor, male
PS1W8 20, ‘Jim’, male
Sample containing about 590 words
speech recorded in public context
G46 Medical consultation.
PS1W3 60, doctor, male
PS1W4 60, ‘Agnes’, female
Sample containing about 1106 words
speech recorded in public context
G47 Medical consultation.
PS1W5 60, doctor, male
PS1W6 60, ‘Cathy’, female
Sample containing about 886 words
speech recorded in public context
G48 Medical consultation.
PS1W7 60, doctor, male
PS1W8 60, female
Sample containing about 366 words
speech recorded in public context
G49 Medical consultation.
PS1W9 60, doctor, male
PS1WA 50, female
Sample containing about 725 words
speech recorded in public context
G4A Medical consultation.
PS1WB 60, doctor, male
Sample containing about 926 words
speech recorded in public context
G4B Medical consultation.
PS1WD 60, doctor, male
Sample containing about 800 words
speech recorded in public context
G4C Medical consultation.
PS1WF 60, doctor, male
G4CP000
Sample containing about 361 words
speech recorded in public context
G4D Medical consultation.
PS1WH 60, doctor, male
G4DPS000
Sample containing about 891 words
speech recorded in public context
G4E Medical consultation.
PS1WK 60, doctor, male
G4EPS000
Sample containing about 499 words
speech recorded in public context
G4F Pre-retirement course: presentation.
PS1WM 75+, course director, male
Sample containing about 5659 words
speech recorded in public context
G4G Pre-retirement course: presentation.
PS1WN 80+, senior fitness instructor, female
Sample containing about 7515 words
speech recorded in public context
G4H Pre-retirement course: presentation.
PS1WP 40, ‘John’, financial advisor, male
Sample containing about 27860 words
speech recorded in public context
G4J Pre-retirement course: presentation.
PS1WR 40, ‘John’, financial advisor, male
G4JPS000
Sample containing about 2443 words
speech recorded in public context
G4K Word-processing tutorial.
PS1WS 50, ‘John’, tutor, male
PS1WT 45, ‘Sue’, secretary, male
PS1WU 16, ‘Sarah’, student, female
Sample containing about 9310 words
speech recorded in educational context
G4N Lockmaking in Willenhall: conversa-
tion with Dick Brown, lockmaker.
PS21S ‘M. Glasson’, interviewer, male
PS21T ‘Dick’, lockmaker, male
G4NPS000
Sample containing about 10311 words
speech recorded in leisure context
G4P Experiences of the leather trade: con-
versation with John Hooper, leather trader.
PS21U ‘J. Hammond’, interviewer, female
PS21V ‘John’, leather trader, male
Sample containing about 2683 words
speech recorded in leisure context
G4R Zeppelin raids, 1916: conversation
with Gilbert Bromley.
PS21W ‘J. Hammond’, interviewer, female
PS21X ‘Gilbert’, leather trader, male
Sample containing about 7785 words speech recorded in leisure context

**G4S** Life in Caldmore: conversation with Bhagan Singh.
PS21Y ‘J. Hammond’, interviewer, female
PS220 ‘Bhagan’, male
Sample containing about 5694 words speech recorded in leisure context

**G4T** Co-operative movement: conversation with Ruth Jackson.
PS221 ‘M. Glasson’, interviewer, male
PS222 ‘Ruth’, female
Sample containing about 11152 words speech recorded in leisure context

**G4U** Sewage - Severn-Trent Water: talk.
PS26L 50, ‘J. Childs’, sewage tmt expert, female
Sample containing about 7124 words speech recorded in leisure context

**G4V** First-year undergraduate tutorial: linguistics.
PS26U ‘Dave’, tutor, male
PS26V ‘Scott’, student, male
PS26W ‘Marilyn’, student, female
PS26R 18, ‘Chris’, student, male
Sample containing about 6846 words speech recorded in educational context

**G4W** First-year undergraduate tutorial: linguistics.
Sample containing about 6308 words speech recorded in educational context

**G4X** Careers meeting.
PS26Y ‘Allan’, manager, male
PS270 ‘Sue’, manager, female
PS271 ‘Liz’, manager, female
PS272 ‘Cynthia’, manager, female
PS273 ‘Jayne’, manager, female
Sample containing about 17119 words speech recorded in business context

**G4Y** Medical consultation.
PS274 doctor, male
PS275 patient
Sample containing about 56 words speech recorded in public context

**G50** Medical consultation.
PS276 doctor, male
PS277 patient
Sample containing about 116 words speech recorded in public context

**G51** Medical consultation.
PS278 doctor, male
PS279 patient
Sample containing about 245 words speech recorded in public context

**G52** Medical consultation.
PS27A doctor, male
PS27B patient
Sample containing about 252 words speech recorded in public context

**G53** Medical consultation.
PS27C doctor, male
PS27D patient
Sample containing about 233 words speech recorded in public context

**G54** Medical consultation.
PS27E doctor, male
PS27F patient
Sample containing about 43 words speech recorded in public context

**G55** Medical consultation.
PS27G doctor, male
PS27H patient
Sample containing about 568 words speech recorded in public context

**G56** Medical consultation.
PS27I doctor, male
PS27J patient
Sample containing about 520 words speech recorded in public context

**G57** Medical consultation.
PS27L doctor, male
PS27M patient
Sample containing about 460 words speech recorded in public context

**G58** Medical consultation.
PS27N doctor, male
PS27P patient
Sample containing about 460 words speech recorded in public context

**G59** Golf club events meeting.
PS27R All participants unknown
PS27S ‘May’, female
PS27T ‘Edith’, female
PS27U ‘Jim’, male
PS27V ‘Duncan’, male
PS27W ‘Agnes’, female
PS27X ‘Margaret’, female
PS27Y ‘Kell’
Sample containing about 7575 words speech recorded in leisure context
**G5A** *Auction rooms.*

Sample containing about 9283 words speech recorded in leisure context

**G5B** *Rotary Club meeting: lecture.*

Sample containing about 4346 words speech recorded in leisure context

**G5C** *Medical consultation.*

Sample containing about 414 words speech recorded in public context

**G5D** *Medical consultation.*

Sample containing about 602 words speech recorded in public context

**G5E** *Medical consultation.*

Sample containing about 517 words speech recorded in public context

**G5F** *Meeting: selection of candidate for election to European Parliament.*

Sample containing about 9802 words speech recorded in public context

**G5G** *Sermons.*

Sample containing about 7564 words speech recorded in public context

**G5J** *Market research interview.*

Sample containing about 11074 words speech recorded in business context

**G5L** *Medical consultation.*

Sample containing about 1680 words speech recorded in public context

**G5M** *Medical consultation.*

Sample containing about 1139 words speech recorded in public context

**G5R** *Medical consultation.*

Sample containing about 1430 words speech recorded in public context

**G5S** *Medical consultation.*

Sample containing about 1277 words speech recorded in public context

**G5T** *Medical consultation.*
PS2BW 35, general practitioner, male
Sample containing about 1719 words speech recorded in public context

G5U Medical consultation.
PS2BY 35, general practitioner, male
PS2CX , male
Sample containing about 1285 words speech recorded in public context

G5V Medical consultation.
PS2C1 35, general practitioner, male
PS2C2 'Meryl', female
PS2C3 , male
Sample containing about 3187 words speech recorded in public context

G5W Medical consultation.
PS2C4 35, general practitioner, male
PS2C5 , male
Sample containing about 1121 words speech recorded in public context

G5X Medical consultation.
PS2C6 35, general practitioner, male
PS2C7 , female
Sample containing about 1267 words speech recorded in public context

G5Y Medical consultation.
PS2C8 35, general practitioner, male
PS2C9 , female
Sample containing about 532 words speech recorded in public context

G60 Medical consultation.
PS2CA 35, general practitioner, male
PS2CB , female
Sample containing about 190 words speech recorded in public context

G61 Tutorial on differentiation.
PS2CC 49, ‘John’, tutor, Not a very strong Merseyside accent., male
PS2CD 30+, female
Sample containing about 3911 words speech recorded in educational context

G62 Oral history project interview.
PS2CE 87, ‘Elizabeth’, retired farmer and housewife, Born in Dalvaine, Angus., female
PS2CF 30+, researcher, Interviewing other 2 participants., male
PS2CG 80+, male
Sample containing about 14780 words speech recorded in leisure context

G63 Oral history project interview.
PS2CH 40+, Interviewing other three participants., male
PS2CJ 60+, ‘Hugh’, retired shipbuilder, male
Sample containing about 8142 words speech recorded in leisure context

G64 Oral history project interview.
PS2CM 92, ‘Eric’, retired lecturer in pharmacy, Born in Girran., male
PS2CN ‘Ann’, Interviewing participant 2., female
Sample containing about 5801 words speech recorded in leisure context


GT4 The Dictionary of National Biography: Missing persons. Nicholls, C S
12. List of works excerpted


12. List of works excerpted


- **GUC** *Corporate resource allocation.* Tomkins, Cyril, Basil Blackwell Ltd, Oxford (1991), 89-207. Sample containing about 42935 words from a book (domain: commerce)


- **GUE** *Calypso’s Island.* Ash, Rosalie, Mills & Boon, Richmond, Surrey (1993), 7-181. Sample containing about 48151 words from a book (domain: imaginative)


- **GUG** *Chung Kuo: The white mountain.* Wingrove, David, New English Library, Sevenoaks (1991), 91-171. Sample containing about 41823 words from a book (domain: imaginative)


- **GUP** *Dead meat.* Barnes, Trevor, New English Library, Sevenoaks (1991), 79-194. Sample containing about 41950 words from a book (domain: imaginative)


12. List of works excerpted

words from a book (domain: imaginative)


GV16 The modern English legal system. Gunn, N J and Bailey, S H, Sweet &


**GVL**  *The night mayor.* Newman, Kim, New English Library, Sevenoaks (1990), 49-185. Sample containing about 43996 words from a book (domain: imaginative)


**GWV**  *The night mayor.* Newman, Kim, New English Library, Sevenoaks (1990), 49-185. Sample containing about 43996 words from a book (domain: imaginative)


London (1990), 151-265. Sample containing about 42289 words from a book (domain: social science)


GX0 [Miscellaneous prayers]. Smith, David K, u.p.. Sample containing about 7243 words of material written to be spoken (domain: belief and thought)

GX1 [Sermons]. Smith, David K, u.p.. Sample containing about 14792 words of material written to be spoken (domain: belief and thought)

GX2 Atomic Energy Authority times and annual review. u.p.. Sample containing about 13576 words of unpublished mis-
cellanea (domain: applied science)

GX3 Competitive edge. AEA Technology, Didcot, Oxfordshire (1993-06). Sample containing about 13863 words from a periodical (domain: applied science)


GX5 LASMO business profile 92. u.p.. Sample containing about 5270 words of unpublished miscellanea (domain: natural sciences)

GX6 Pipeline. u.p.. Sample containing about 7352 words of unpublished miscellanea (domain: natural sciences)

GX7 Pipeline. u.p.. Sample containing about 14078 words of unpublished miscellanea (domain: natural sciences)

GX8 [LASMO plc. u.p.. Annual Reports and Accounts 92]. Sample containing about 14365 words of unpublished miscellanea (domain: natural sciences)

GX9 NewsLine – Royal Bank of Scotland. u.p.. Sample containing about 20283 words from a periodical (domain: social science)

GX10 NewsLine – Royal Bank of Scotland. u.p.. Sample containing about 22611 words from a periodical (domain: social science)

GX11 Where people matter. The Royal Bank of Scotland. Sample containing about 10294 words of unpublished miscellanea (domain: commerce)

GX12 National Library of Scotland – newsletter. u.p.. Sample containing about 9101 words of unpublished miscellanea (domain: applied science)


GX14 [Campaign for the Preservation of Rural Wales: miscellaneous unpublished documents]. u.p.. Sample containing about 99334 words of miscellanea (domain: social science)

GXH [Juniper Green village association minutes]. u.p.. Sample containing about 12093 words of unpublished miscellanea (domain: social science)

GXJ [Collection of leaflets from Lothian Council]. u.p.. Sample containing about 79904 words of unpublished miscellanea (domain: social science)

GXK Notes for my nephews and nieces. Robertson, Angela, u.p.. Sample containing about 20146 words of unpublished miscellanea (domain: social science)

GXK [Personal letters]. u.p.. Sample containing about 7839 words of unpublished miscellanea (domain: leisure)

GXK [Personal letters and other ephemera]. unpublished. Sample containing about 6030 words of unpublished miscellanea (domain: leisure)

GY4 Neighbourhood Watch talk: crime prevention.

PS28B 37, ‘Paul’, police sergeant, male
PS28C 40+, ‘Jan’, housewife, female
PS28D 65+, ‘Selwyn’, retired, male
PS28E 40+, ‘Lyn’, housewife, female
PS28F 50+, housewife, female
PS28G 65+, retired, female
PS28H 60+, ‘Dave’, retired, male
PS28J 70+, retired, male
GY4PS000 Sample containing about 13048 words speech recorded in public context

GY5 Medical consultation.

PS28K doctor
PS28L

Sample containing about 345 words speech recorded in public context

GY6 Medical consultation.

PS28M doctor
PS28N

Sample containing about 1558 words speech recorded in public context

GY7 Medical consultation.

PS28P doctor
PS28R

Sample containing about 1592 words speech recorded in public context

GY8 Medical consultation.

PS28S doctor
PS28T
Sample containing about 1450 words speech recorded in public context

**GY9 Medical consultation.**  
PS28U doctor  
PS28V  
Sample containing about 1281 words speech recorded in public context

**GYA Medical consultation.**  
PS28W doctor  
PS28X  
Sample containing about 363 words speech recorded in public context

**GYB Medical consultation.**  
PS28Y doctor  
PS28Z  
Sample containing about 1016 words speech recorded in public context

**GYC Medical consultation.**  
PS291 doctor  
PS292  
Sample containing about 1459 words speech recorded in public context

**GYD Medical consultation.**  
PS293 doctor  
PS294 , female  
Sample containing about 1724 words speech recorded in public context

**GYE Medical consultation.**  
PS295 doctor  
PS296  
Sample containing about 972 words speech recorded in public context

**GYF Medical consultation.**  
PS297 doctor  
PS299 , female  
Sample containing about 695 words speech recorded in public context

**GYG Medical consultation.**  
PS299 doctor  
PS29A , male  
Sample containing about 896 words speech recorded in public context

**GYH Medical consultation.**  
PS29B doctor  
PS29C  
Sample containing about 503 words speech recorded in public context

**GYJ Medical consultation.**  
PS29D doctor  
PS29E  
Sample containing about 232 words speech recorded in public context

**GYK Oral history project: interview on housing.**  
PS29F worker at law centre  
PS29G  
Sample containing about 6301 words speech recorded in leisure context

**GYL Oral history project: interview on housing.**  
PS29H neighbourhood housing officer  
PS29J  
Sample containing about 946 words speech recorded in leisure context

**GYM Oral history project: interview on housing.**  
PS29K ‘Leona’, unemployed, female  
PS29L ‘Tiffany’, female  
Sample containing about 4502 words speech recorded in leisure context

**GYN Knitwear design and history: Women’s Institute meeting.**  
PS29M 50+, ‘Lewis’, knitwear designer and producer, male  
PS29N  
Sample containing about 6174 words speech recorded in leisure context

**GYP Tutorial maths lesson: GCSE paper 3.**  
PS29P 50, ‘John’, tutor, male  
PS29R 16, ‘Ian’, student, male  
Sample containing about 9324 words speech recorded in educational context

**GYR Chemistry tutorial: GCSE.**  
PS29S 50, ‘John’, tutor, male  
PS29T 16, ‘Andrew’, student, male  
Sample containing about 13208 words speech recorded in educational context

**GYS Oral history project: interview.**  
PS29U 72, ‘William’, farmer, male  
PS29V , male  
PS29W ‘Sarah’, female  
GSYS000  
Sample containing about 4568 words speech recorded in leisure context

**GYT Oral history project: interview.**  
PS29W 72, ‘William’, farmer, male  
PS29X , male  
PS29Y ‘Sarah’, female  
Sample containing about 4427 words speech recorded in leisure context

**GYU Oral history project: interview.**  
PS2A0 76, ‘Laurie’, retired, male  
PS2A1 30+, male
Sample containing about 11701 words speech recorded in leisure context

**GYV** Oral history project, Scotland: interview.
PS2A2 'Bill', engineer, male
PS2A3 'Ken', male
Sample containing about 12196 words speech recorded in leisure context

**GYW** Oral history project: interview.
PS2A4 'Malcolm', male
PS2A5 '2', 'Swinton', retired motor engineer, male
Sample containing about 12930 words speech recorded in leisure context

**GYX** Maths tutorial.
PS2A6 50, 'John', tutor, male
PS2A7 student, male
Sample containing about 13586 words speech recorded in educational context

**GYY** Oral history project: interview.
PS2A8 female
PS2A9 carpet factory worker, female
PS2AA 'Ken', male
Sample containing about 3756 words speech recorded in leisure context

**H00** Oral history project: interview.
PS2AB female
PS2AC carpet factory worker, female
PS2AD 'Ken', male
Sample containing about 1001 words speech recorded in leisure context

**H01** Oral history project: interview.
PS2AE female
PS2AF carpet factory worker, female
PS2AG 'Ken', male
Sample containing about 3060 words speech recorded in leisure context

**H02** Oral history project: interview.
PS2AH female
PS2AJ carpet factory worker, female
PS2AK 'Ken', male
Sample containing about 716 words speech recorded in leisure context

**H03** Oral history project: interview.
PS2AL female
PS2AM 40+, carpet factory worker, female
PS2AN 'Ken', male
Sample containing about 3097 words speech recorded in leisure context

**H04** Oral history project: interview.
PS2AP female
PS2AR office worker in carpet factory, female
PS2AS 'Ken', male
Sample containing about 808 words speech recorded in leisure context

**H05** Oral history project: interview.
PS2AT female
PS2AU carpet factory worker, female
PS2AV 'Ken', male
Sample containing about 3365 words speech recorded in leisure context


**H09** Across the low meadow. Church, D and Gowers, E S, Christine Swift Bookshop, UK (1979), 1-89. Sample containing about 36524 words from a book (domain: world affairs)


**H0B** Coniston copper. Holland, Eric G, Cercone Press, UK (1986), 16-134. Sample containing about 38931 words from a book (domain: world affairs)


**H0F** The green behind the glass. Geras, Adele, Lions Teen Tracks, UK (1989), 1-118. Sample containing about 41764 words from a book (domain: imaginative)
12. List of works excerpted


List of works excerpted

H4A Labour Club public meeting.
PS1XR 35, ‘Steve’, health worker, male
PS1XS 35, ‘John’, council worker, male
PS1XT 56, ‘Jeff’, retired, male
PS1XU 28, ‘ED’, trade union officer, male
PS1XV 40, ‘Graham’, member of parliament, male
PS1XW 65+, ‘Ronald’, retired, male
PS1XX 65+, ‘Ida’, pensioners’ movement representative, female
PS1XY 43, ‘Alan’, trade union representative, male
PS1Y0 35, ‘Gwen’, hosiery worker, female

Sample containing about 15770 words speech recorded in public context

H4B Nottinghamshire Oral History Project: interview.
PS1Y2 70, retired, male
PS1Y3 70, retired, female
PS1Y4 70, retired, female

Sample containing about 9805 words speech recorded in public context

H4C Nottinghamshire Oral History Project: interview.
PS1Y5 72, retired midwife, female
PS1Y6 , female

Sample containing about 16503 words speech recorded in public context

H4D Medical consultations.
PS1Y7 60, general practitioner, male
PS1Y8 , female
PS1Y9 child, female

Sample containing about 701 words speech recorded in public context

H4E Medical consultations.
PS1Y A 60, general practitioner, male
PS1Y B ‘Mary’, female

Sample containing about 884 words speech recorded in public context

H4F Medical consultations.
PS1Y C 60, general practitioner, male
PS1Y D ‘Mary’, female

Sample containing about 932 words speech recorded in public context

H4G Medical consultations.
PS1Y E 60, general practitioner, male
PS1Y F receptionist, female

Sample containing about 188 words speech recorded in public context

H4H Medical consultations.
PS1Y G 60, general practitioner, male
PS1Y H , female

Sample containing about 1139 words speech recorded in public context

H4J Medical consultations.
PS1Y L 60, general practitioner, male
PS1YM , male

Sample containing about 466 words speech recorded in public context

H4K Medical consultations.
PS1Y N 60, general practitioner, male
PS1Y P ‘Alec’, male

Sample containing about 444 words speech recorded in public context

H4L Medical consultations.
PS1Y Q 60, general practitioner, male
PS1YS , female

Sample containing about 24 words speech recorded in public context

H4M Medical consultations.
PS1Y R 60, general practitioner, male
PS1Y S ‘William’, male

Sample containing about 271 words speech recorded in public context

H4P Medical consultations.
PS1Y T 60, general practitioner, male
PS1Y U , female

Sample containing about 502 words speech recorded in public context

H4R Medical consultations.
PS200 60, general practitioner, male
PS201 , male

Sample containing about 203 words speech recorded in public context

H4S Medical consultations.
PS202 60, general practitioner, male
PS203 ‘Bob’, male

Sample containing about 408 words speech recorded in public context

H4T Medical consultations.
PS204 60, general practitioner, male
PS205 ‘Suzanne’, female

Sample containing about 439 words speech recorded in public context

H4U Medical consultations.
PS206 60, general practitioner, male
PS207 , male

Sample containing about 188 words speech recorded in public context
Sample containing about 415 words speech recorded in public context

**H4V Medical consultations.**
- PS208  60, general practitioner, male
- PS209  ‘Robert’, male

Sample containing about 523 words speech recorded in public context

**H4W Medical consultations.**
- PS20A  60, general practitioner, male
- PS20B  , female

Sample containing about 1043 words speech recorded in public context

**H4X Medical consultations.**
- PS20C  60, general practitioner, male
- PS20D  , male

Sample containing about 457 words speech recorded in public context

**H4Y Medical consultations.**
- PS20E  60, general practitioner, male
- PS20F  ‘Mary’, female

Sample containing about 943 words speech recorded in public context

**H50 Medical consultations.**
- PS20G  60, general practitioner, male
- PS20H  , female

Sample containing about 562 words speech recorded in public context

**H51 Medical consultations.**
- PS20J  60, general practitioner, male
- PS20K  ‘Anne’, female

Sample containing about 495 words speech recorded in public context

**H52 Medical consultations.**
- PS20L  60, general practitioner, male
- PS20M  ‘Walter’, male

Sample containing about 89 words speech recorded in public context

**H53 Medical consultations.**
- PS20N  60, general practitioner, male
- PS20O  , male

Sample containing about 365 words speech recorded in public context

**H54 Medical consultations.**
- PS20P  60, general practitioner, male
- PS20Q  ‘Mary’, female

Sample containing about 678 words speech recorded in public context

**H55 Medical consultations.**
- PS20R  60, general practitioner, male
- PS20S  ‘Carol’, female

Sample containing about 204 words speech recorded in public context

**H56 Medical consultations.**
- PS20T  60, general practitioner, male
- PS20U  , female

Sample containing about 473 words speech recorded in public context

**H57 Medical consultations.**
- PS20V  60, general practitioner, male
- PS20W  ‘Steven’, male

Sample containing about 877 words speech recorded in public context

**H58 Medical consultations.**
- PS20X  60, general practitioner, male
- PS20Y  receptionist, female

Sample containing about 50 words speech recorded in public context

**H59 Medical consultations.**
- PS210  60, general practitioner, male
- PS211  40, ‘Alison’, female

Sample containing about 724 words speech recorded in public context

**H5A Medical consultations.**
- PS212  60, general practitioner, male
- PS213  , male

Sample containing about 1059 words speech recorded in public context

**H5B Medical consultations.**
- PS214  60, general practitioner, male
- PS215  ‘William’, male

Sample containing about 52 words speech recorded in public context

**H5C Medical consultations.**
- PS216  60, general practitioner, male
- PS217  , female

Sample containing about 473 words speech recorded in public context

**H5D Careers Service: meeting.**
- PS218  30, ‘Sue’, careers advisor, female
- PS219  40, ‘Keith’, careers advisor, male

Sample containing about 473 words speech recorded in public context

**H5E Careers Service: meeting.**
- PS220  30, ‘Sue’, careers advisor, female
- PS221  40, ‘Bill’, careers advisor, male

Sample containing about 9451 words speech recorded in business context

**H5G Suffolk Sound Archive: interview.**
Sample containing about 16649 words speech recorded in leisure context

Suffolk Sound Archive: interview.

Sample containing about 14477 words speech recorded in leisure context

Suffolk Sound Archive: recording a message.

Sample containing about 836 words speech recorded in public context

Medical consultation.

Sample containing about 433 words speech recorded in public context

Medical consultation.

Sample containing about 283 words speech recorded in public context

Medical consultation.

Sample containing about 336 words speech recorded in public context

Computer advice session.

Sample containing about 13413 words speech recorded in business context


The deep-sea Ophiuroidea of the north Atlantic Ocean. Paterson, Gordon L J,
British Museum (Natural History), UK (1985), 1-120. Sample containing about 39110 words from a book (domain: natural sciences)


H7H Loving and giving. Keane, Molly, Andre Deutsch Ltd, UK (1988), 3-123. Sample containing about 42226 words from a book (domain: imaginative)


H7W Bay of rainbows. James, Dana, Mills & Boon, Richmond, Surrey (1993). Sample containing about 53275 words from a book (domain: imaginative)


H82 The chymical wedding. Clarke, Lindsay, Jonathan Cape, London (1989), 186-301. Sample containing about 46693 words from a book (domain: commerce)


H84 City of dreams. Gill, Anton, Blooms-


**H86** *Conduct of Major Maxim.* Lyall, Gavin, Hodder & Stoughton Ltd, Sevenoaks, Kent (1982), 63-186. Sample containing about 43823 words from a book (domain: imaginative)


**H8B** *The education fact file.* Hales, Margaret, Mackinnon, Donald, Cathcart, Heather and Statham, Hodder & Stoughton Ltd, Sevenoaks, Kent (1991), 4-174. Sample containing about 44598 words from a book (domain: social science)


**H8F** *Forgotten fire.* Mansell, Joanna, Mills & Boon, Richmond, Surrey (1992). Sample containing about 54007 words from a book (domain: imaginative)


**H8N** *Her living image.* Rogers, Jane, Faber & Faber Ltd, London (1990), 55-187. Sample containing about 43187 words from a book (domain: imaginative)


**H8T** *Into the blue.* Goddard, Robert,
12. List of works excerpted


H8U Interpreting the landscape. Aston, Michael, BT Batsford Ltd, London (1992), 21-118. Sample containing about 41805 words from a book (domain: social science)


H8W [Access course essays]. Woodall, Julia R, u.p.. Sample containing about 14990 words of unpublished miscellanea (domain: social science)


12. List of works excerpted


H9T The sea skater. Dunmore, Helen, Bloodaxe books ltd, Newcastle upon Tyne (1986), 12-72. Sample containing about 8326 words from a book (domain: imaginative)


H9X Thank you for having me. Lipman, Maureen, Robson Books Ltd, London (1990), 57-188. Sample containing about 47961 words from a book (domain: arts)


HA3 The colour of magic. Pratchett, Terry, Colin Smythe Ltd, Gerrards Cross (1983), 7-141. Sample containing about 46464 words from a book (domain: imaginative)

HA5  *Viking magic.* Wells, Angela, Mills & Boon, Richmond, Surrey (1993). Sample containing about 55373 words from a book (domain: imaginative)


HA7  *A warning of magic.* Kingston, Kate, Mills & Boon, Richmond, Surrey (1993). Sample containing about 56850 words from a book (domain: imaginative)

HAB  *The Embalmer.* British Institute of Embalmers, Knebworth (1993-12). Sample containing about 12182 words from a periodical (domain: applied science)


HAJ  *[Law Society Publicity].* u.p.. Sample containing about 43794 words of unpublished miscellanea (domain: social science)


HAL  *Scotch whisky.* u.p.. Sample containing about 1676 words of miscellanea (domain: applied science)

HAM  *Don’t take our British spirit for granted.* u.p.. Sample containing about 1496 words of miscellanea (domain: applied science)

HAN  *Scotch whisky.* u.p.. Sample containing about 1362 words of miscellanea (domain: applied science)

HAP  *Wood News.* u.p.. Sample containing about 8666 words of unpublished miscellanea (domain: social science)

HAR  *Wood News.* u.p.. Sample containing about 10403 words of unpublished miscellanea (domain: social science)

HAS  *John Wood Group Employee Report 1992.* u.p.. Sample containing about 4165 words of unpublished miscellanea (domain: social science)

HAT  *The Gusher.* u.p.. Sample containing about 3463 words of unpublished miscellanea (domain: social science)

HAU  *Atomic Energy Authority Times.* The Atomic Energy Authority (1993-09). Sample containing about 10957 words from a periodical (domain: world affairs)

HAV  *Atomic Energy Authority Culham and Harwell News.* u.p.. Sample containing about 3054 words of unpublished miscellanea (domain: social science)

HAW  *Winfrith Journal.* u.p.. Sample containing about 4181 words of unpublished miscellanea (domain: social science)

HAX  *Dounreay Newsletter.* u.p.. Sample containing about 2592 words of unpublished miscellanea (domain: social science)

HAY  *Atomic Energy Authority News for the Process Industries.* u.p.. Sample containing about 4334 words of unpublished miscellanea (domain: social science)
HB0 Atomic Energy Authority Gazette. u.p.. Sample containing about 4670 words of unpublished miscellanea (domain: social science)

HB1 The story of Atomic Energy Authority Technology from national mission to international. u.p. Sample containing about 3654 words of unpublished miscellanea (domain: social science)

HB2 BPXpress. British Petroleum plc. Sample containing about 28905 words from a periodical (domain: applied science)

HB3 [General Accident miscellaneous documents – Claims]. General Accident. Sample containing about 25428 words of unpublished miscellanea (domain: social science)

HB4 [General Accident miscellaneous documents]. General Accident. Sample containing about 29214 words of unpublished miscellanea (domain: social science)

HB5 [General Accident miscellaneous documents]. General Accident. Sample containing about 27941 words of unpublished miscellanea (domain: social science)

HB6 [General Accident miscellaneous documents]. General Accident. Sample containing about 16062 words of unpublished miscellanea (domain: social science)

HB7 Wedgwood product range. u.p.. Sample containing about 678 words of unpublished miscellanea (domain: arts)

HB8 The history of Wedgwood. u.p.. Sample containing about 1256 words of unpublished miscellanea (domain: arts)

HB9 [Wedgwood Group: environmental policy]. u.p.. Sample containing about 2016 words of unpublished miscellanea (domain: arts)


HBB [Waterford Wedgwood reports & accounts 1992]. u.p.. Sample containing about 8303 words of unpublished miscellanea (domain: arts)

HBC Newsline for the Wedgwood Group. u.p.. Sample containing about 13644 words of unpublished miscellanea (domain: arts)

HBD Newsline for the Wedgwood Group. u.p.. Sample containing about 9153 words of unpublished miscellanea (domain: arts)

HBE Newsline for the Wedgwood Group. u.p.. Sample containing about 8438 words of unpublished miscellanea (domain: arts)

HBF Environmental Issues [Rentokil company newsletter]. Rentokil. Sample containing about 16193 words from a periodical (domain: commerce)

HBJ Environmental Issues [Rentokil company newsletter]. Rentokil. Sample containing about 14899 words from a periodical (domain: commerce)

HBK [Scottish Nuclear: miscellaneous documents]. Scottish Nuclear, Scotland. Sample containing about 22335 words of unpublished miscellanea (domain: natural sciences)

HBM [Scotvec (Scottish Vocational Education Council) publications]. u.p.. Sample containing about 46343 words of unpublished miscellanea (domain: social science)

HBN [Scotvec (Scottish Vocational Education Council) advanced courses]. u.p.. Sample containing about 24693 words of unpublished miscellanea (domain: social science)

HBP [Educational certification: miscellaneous documents]. u.p.. Sample containing about 29099 words of unpublished miscellanea (domain: commerce)


HBT Becoming an actuary. u.p.. Sample containing about 2502 words of unpublished miscellanea (domain: commerce)

HBU UK actuarial profession 1992 annual review. u.p.. Sample containing about 3312 words of unpublished miscellanea (domain: commerce)

HBV UK actuarial profession 1991 annual review. u.p.. Sample containing about 2739 words of unpublished miscellanea (domain: commerce)

HBW Institute of Actuaries Education Service: Tutors’ handbook 1990/91. u.p.. Sample containing about 2146 words of unpublished miscellanea (domain: commerce)

HBX AES Tuition Noticeboard. u.p.. Sample containing about 1431 words of unpublished miscellanea (domain: commerce)

HBY Institute of Actuaries: Proceedings. u.p.. Sample containing about 217 words of unpublished miscellanea (domain: commerce)

HC0 CA News. u.p.. Sample containing about 4423 words of unpublished miscellanea (domain: commerce)

HC1 CA News. u.p.. Sample containing about 23913 words of unpublished miscellanea (domain: commerce)

HC2 CA News. u.p.. Sample containing about 24149 words of unpublished miscellanea (domain: commerce)


HC5 Royal Institute of British Architects annual report and accounts. u.p.. Sample containing about 11080 words of unpublished miscellanea (domain: social science)

HC6 Spotlights. u.p.. Sample containing about 3355 words of miscellanea (domain: social science)

HC7 Spotlights. u.p.. Sample containing about 3048 words of miscellanea (domain: social science)

HC8 Spotlights. u.p.. Sample containing about 3657 words of miscellanea (domain: social science)

HC9 Spotlights. u.p.. Sample containing about 1853 words of miscellanea (domain: social science)

HCA Spotlights. u.p.. Sample containing about 4417 words of miscellanea (domain: social science)

HCB Spotlights. u.p.. Sample containing about 3590 words of miscellanea (domain: social science)

HCC Observations: an information exchange for teacher researchers. u.p.. Sample containing about 6996 words of miscellanea (domain: social science)

HCD Observations: an information exchange for teacher researchers. u.p.. Sample containing about 5530 words of miscellanea (domain: social science)

HCE SCRE Newsletter. u.p.. Sample containing about 9666 words of miscellanea (domain: social science)

HCG University of Stirling annual report 1991-1992. u.p.. Sample containing about 5984 words of unpublished miscellanea (domain: social science)

HCH Edit [alumni magazine]: Issue 1 January 1992. u.p.. Sample containing about 8873 words from a periodical (domain: belief and thought)

HCJ Edit [alumni magazine] Issue 2. u.p.
(n.d.), pp3-8. Sample containing about 7404 words from a periodical (domain: belief and thought)

**HCK** *Edit [alumni magazine] Issue 3*. u.p. (n.d.), 12-34. Sample containing about 7349 words from a periodical (domain: belief and thought)

**HCL** [Employment service]. u.p. Sample containing about 14967 words of unpublished miscellanea (domain: social science)

**HCM** *The annual report and accounts of the Law Society May 1992-.* u.p.. Sample containing about 4254 words of unpublished miscellanea (domain: social science)


**HCS** Saddle Mews Residents’ Association. Saddle Mews residents association. Sample containing about 3516 words of unpublished miscellanea (domain: social science)

**HCT** *The Scottish Council: Development and industry*. u.p.. Sample containing about 8359 words of unpublished miscellanea (domain: social science)

**HCU** *IBOA Newsheet*. u.p.. Sample containing about 7021 words of unpublished miscellanea (domain: commerce)

**HCV** *IBOA Newsheet*. u.p.. Sample containing about 6319 words of unpublished miscellanea (domain: commerce)

**HCW** *IBOA Newsheet*. u.p.. Sample containing about 4575 words of unpublished miscellanea (domain: commerce)

**HCX** *IBOA Newsheet*. u.p.. Sample containing about 7641 words of unpublished miscellanea (domain: commerce)

**HCY** *IBOA Newsheet*. u.p.. Sample containing about 5739 words of unpublished miscellanea (domain: commerce)

**HD0** *Minutes: Juniper Green Village Association*. u.p.. Sample containing about 12518 words of unpublished miscellanea (domain: social science)

**HD1** *Minutes: Juniper Green Village Association*. u.p.. Sample containing about 12065 words of unpublished miscellanea (domain: social science)

**HD2** Lothian Regional Council. u.p.. Sample containing about 56447 words of unpublished miscellanea (domain: social science)

**HD3** *Personal letters*. u.p.. Sample containing about 5346 words of unpublished miscellanea (domain: leisure)

**HD4** *Personal letters*. u.p.. Sample containing about 13349 words of unpublished miscellanea (domain: leisure)


**HD6** *Creative writing by schoolgirls: prose and verse*. u.p.. Sample containing about 11134 words of unpublished miscellanea (domain: imaginative)

**HD7** *Creative writing: prose*. Robertson, Angela, u.p., 1-89. Sample containing about 30962 words of unpublished miscellanea (domain: social science)

**HD8** *English Literature Papers*. Robertson, Angela, u.p.. Sample containing about 10800 words of unpublished miscellanea (domain: social science)

**HD9** *Creative writing: verse*. u.p.. Sample containing about 858 words of unpublished miscellanea (domain: imaginative)
HDA [School essays]. u.p.. Sample containing about 9557 words of unpublished miscellanea (domain: leisure)

HDB [Children’s writing]. u.p.. Sample containing about 17724 words of unpublished miscellanea (domain: imaginative)

HDC [Creative writing: prose]. u.p.. Sample containing about 35530 words of unpublished miscellanea (domain: imaginative)

HDD Confirmation meeting on the gift of the Holy Spirit: church meeting.

HDE Enterprise Two Thousand: seminar.

HDF Confirmation meeting on the gift of the Holy Spirit: church meeting.

HDEH Suffolk Sound Archive oral history project: interview.

HDFH Suffolk Sound Archive: interview.

HDI Suffolk Sound Archive: interview.

HDJ Suffolk Sound Archive: interview.

HDK Suffolk Sound Archive: interview.

HDL Suffolk Sound Archive – Ipswich Transport Project: interview.

HDM Suffolk Sound Archive: interview.

HDEH Trade Union Annual Congress: trade union congress.

HDFH Trade Union Annual Congress: trade union congress.

HDFJ Trade Union Annual Congress: trade union congress.

HDR Trade Union Annual Congress: trade union congress.

HDS Trade Union Annual Congress: trade union congress.

HDT Trade Union Annual Congress: trade union congress.
Sample containing about 5936 words
speech recorded in leisure context

HE7 Panorama: documentary.
PS2TW 30+, ‘Christopher Hird’, tv presenter, male
PS2TX 30+, ‘Robert’, information hacker / private investigator, male
PS2TY 30+, bbc interviewer, male
PS2UO 50+, ‘Rodney’, trade union leader, male
PS2UI 50+, banking expert, male
PS2U2 40+, bank representative, male
PS2U3 40+, legal expert, male
PS2U4 20+, ‘Naomi March’, actress, female
PS2U5 30+, ‘Lenny Henry’, comedian/actor, male

Sample containing about 6239 words
speech recorded in leisure context

HE8 Orkney Sound Archive tape OSA 335: interview for oral history project.
PS2UP historian, Interviewer, female
PS2UR 80, ‘Ned’, retired farmer, Interviewee., male

Sample containing about 5854 words
speech recorded in leisure context

HE9 Orkney Sound Archive tape OSA 337: interview for oral history project.
PS2US 20+, historian, Interviewer, female
PS2UT 75+, ‘Davie’, retired surveyor, Interviewee., male
PS2UU 72, ‘Walter’, retired surveyor, Interviewee., male

Sample containing about 6116 words
speech recorded in leisure context

HEA Orkney Sound Archive tape OSA 55: interview for oral history project.
PS2UV 20+, ‘Agnes’, historian, Interviewer, female
PS2UW 50+, ‘David’, Interviewee, male

Sample containing about 3013 words
speech recorded in leisure context

HEC Orkney Sound Archive tape OSA/RO/A: interview.
PS2UX 70+, ‘Kitty’, Interviewee, female
PS2UY radio presenter, Interviewer, female

Sample containing about 2747 words
speech recorded in leisure context

HED Orkney Sound Archive tape OSA/RO/A: interview.
PS2V0 84, ‘Issa’, Interviewee, female
PS2V1 30+, radio presenter, Interviewer, female

Sample containing about 1996 words
speech recorded in leisure context

HEE Orkney Sound Archive tape OSA/RO/A: interview on ‘Sharing Time’ show.
PS2V2 radio presenter, Interviewer, male
PS2V3 30+, ‘Erlend’, oil rig worker, Interviewee, male

Sample containing about 4824 words
speech recorded in leisure context

HEF Orkney Sound Archive tape OSA/RO/B: interviews.
PS2V4 ‘Kieran’, lifeboat winchman, Interviewee, male
PS2V5 radio presenter, Interviewer, female
PS2V6 ‘Tony’, lifeboat doctor, Interviewee, male
PS2V7 ‘Jack’, lifeboat coxswain, Interviewee, male
PS2V8 ‘David’, lifeboat mechanic, Interviewee, male
PS2V9 ‘Billy’, lifeboat crewmember, Interviewee, male
PS2VA ‘Angus’, lifeboat crewmember, Interviewee, male
PS2VB 60+, ‘Jackie’, lifeboat secretary, Interviewee, male

Sample containing about 1733 words
speech recorded in leisure context

HEG Orkney Sound Archive tape OSA/RO/B: interview.
PS2VC 50+, ‘Marcus’, Interviewee. Organizer of ploughing match., male
PS2VD radio presenter, Interviewer., female
PS2VE ‘William’, Interviewee, male
PS2VF radio reporter, Interviewer, male
PS2VG ‘Jim’, Interviewee, male

Sample containing about 2589 words
speech recorded in leisure context

HEH Orkney Sound Archive tape OSA/RO/B: interview.
PS2VI 60+, ‘Sandy’, retired technician/inventor, Interviewee, male

Sample containing about 2368 words
speech recorded in leisure context

HEJ Orkney Sound Archive tape OSA/RO/B: interview.
PS2VJ ‘Mick’, radio presenter, Interviewer, male
PS2VK 40+, ‘Jean’, dairy farmer, Interviewee, female
PS2VL radio presenter, Interviewer, male

Sample containing about 690 words
speech recorded in leisure context

HEK Orkney Sound Archive tape OSA/RO/B: reminiscences.
PS2VM 70+, ‘Edith’, retired farmer, female

Sample containing about 487 words
speech recorded in leisure context

HEL Gwynedd County Council tape 2: interview for oral history project.
PS2VN  historian, Interviewer, male
PS2VP  84, retired district nurse, Interviewee, female
Sample containing about 4505 words speech recorded in leisure context

HEM  Gwynedd County Council tape 4: interview for oral history project.
PS2VR  67, ‘John’, farmer, Interviewee, male
PS2VS  historian, Interviewer, male
Sample containing about 9996 words speech recorded in leisure context

HEN  Gwynedd County Council tape 5: interview for oral history project.
PS2VT  historian, Interviewer, male
PS2VU  30+, ‘Kenneth’, unemployed, Interviewee, male
Sample containing about 9316 words speech recorded in leisure context

HEP  Oral history project: interview.
PS2VV  87, ‘Nellie’, retired teacher, female
PS2VW  interviewer, female
HEPPS000
HEPPS001
Sample containing about 16600 words speech recorded in leisure context

HER  Gwynedd County Council tape 11: interview for oral history project.
PS2VX  79, retired forestry worker, Interviewee, male
PS2VY  historian, Interviewer, male
Sample containing about 7866 words speech recorded in leisure context

HES  Gwynedd County Council tape 12: interview for oral history project.
PS2WO  60+, ‘Evan John’, mechanical engineer, Interviewee, male
PS2W1  historian, Interviewer, male
Sample containing about 7751 words speech recorded in leisure context

HET  Gwynedd County Council tape 13: interview for oral history project.
PS2W2  historian, Interviewer, male
PS2W3  76, ‘Aled’, entrepreneur, Interviewee, male
Sample containing about 7192 words speech recorded in leisure context

HEU  Gwynedd County Council tape 14: interview for oral history project.
PS2W4  historian, Interviewer, male
PS2W5  93, ‘Richard (dick)’, retired seaman, Interviewee, male
Sample containing about 8098 words speech recorded in leisure context

HEV  Gwynedd County Council tape 15: interview for oral history project.
PS2W6  historian, Interviewer, male
PS2W7  92, ‘Harry’, Interviewee, male
Sample containing about 7763 words speech recorded in leisure context

HEW  Racing: the Morning Line: television broadcast.
PS2W8  40+, ‘Derek Thompson’, tv presenter, main presenter of the programme, male
PS2W9  30+, ‘Lesley Graham’, tv presenter/commentator, female
PS2WA  50+, ‘John Oaksey’, tv presenter/commentator, male
PS2WB  40+, ‘Alastair Down’, sports commentator, male
Sample containing about 4586 words speech recorded in educational context

HEX  Central News and Weather: television news broadcast.
PS2WC  30+, newsreader, female
PS2WD  20+, ‘Tracy Barden’, weather forecaster, female
Sample containing about 414 words speech recorded in educational context

PS2WE  40+, ‘Bob Hall’, tv presenter, Link man, male
PS2WF  30+, ‘Phil’, sports reporter, male
PS2WG  30+, ‘Dennis’, sports reporter, male
PS2WH  -82+, ‘Keith’, sports reporter, male
PS2WJ  40+, ‘Tim’, sports reporter, male
PS2WK  40+, sports reporter, male
Sample containing about 938 words speech recorded in educational context

HF0  Central Lobby: television broadcast.
PS2WL  40+, ‘Judy Laybourn’, tv presenter, female
PS2WM  30+, ‘Andrew Fox’, reporter, Football security story, male
PS2WN  40+, ‘Peter Snape’, member of parliament, Football security story, male
PS2WP  30+, ‘John Williams’, researcher, Football security story, male
PS2WR  40+, ‘Bob Jones’, police chief superintendent, Football security story, male
PS2WS  50+, ‘Keith Pearson’, football club secretary, Football security story, male
PS2WT  40+, ‘Richard Faulkner’, football trust representative, Football security story, male
PS2WU  40+, undercover police officer, Football security story, male
PS2WV  40+, police sergeant, Football security story, male
Sample containing about 3576 words speech recorded in educational context
12. List of works excerpted

HF1 Central News: television news broadcast.
PS2X4 30+, ‘Mike Morley’, newsreader, male
PS2X5 50+, ‘Dave Johnson’, police detective superintendent, Story about murder of businessman, male
PS2X6 60+, pensioner, Story about Christmas lights, female
PS2X7 30+, ‘Marie Ashtby’, newsreader, female
PS2X8 ‘Mike Rowbottom’, reporter, Story about attack on Asian family, male
PS2X9 30+, shopkeeper, Crime victim in story about attack on Asian family, male
PS2XA police officer, Story about attack on Asian family, male
PS2XB ‘Jane Saggars’, reporter, Story about murder of businessman, female
PS2XC ‘John Mitchell’, reporter, Story about hit and run, male
Sample containing about 3143 words

HF2 Gwynedd County Council tape 17: interview for oral history project.
PS2Y3 65, ‘Dafydd’, miner, Interviewee. Treasurer of the Festiniog slate dispute, male
PS2Y4 30+, historian, Interviewee, female
Sample containing about 7865 words

HF3 Gwynedd County Council tape 3: interview for oral history project.
PS2Y5 74, retired shopkeeper, Interviewee, female
PS2Y6 historian, Interviewer, male
Sample containing about 9682 words


HGH In search of English: a traveller’s guide. Crystal, David, u.p.. Sample containing about 7420 words of unpublished miscellanea (domain: social science)


HGX *Microprocessor based measurement of plant growth.* Lucas, Keith, u.p.. Sample containing about 27909 words of unpublished miscellanea (domain: applied science)


HH1 *My enemy, my love.* Byrne, Julia, Mills & Boon, Richmond, Surrey (1993). Sample containing about 87955 words from a book (domain: imaginative)


HH3 *New Internationalist.* u.p.. Sample containing about 251456 words from a periodical (domain: world affairs)


HHA *Only two can share.* Murray, Annabel, Mills & Boon, Richmond, Surrey (1993). Sample containing about 52558 words from a book (domain: imaginative)


HHF *Bishop’s Castle Railway Society Journal.* et al and Lucas, Keith, Bishop’s
castle railway society (1990). Sample containing about 5997 words of miscellanea (domain: commerce)


HHY Rapid – ESRC grant abstracts. u.p.. Sample containing about 59581 words of miscellanea (domain: social science)

HJ0 Rapid – ESRC grant abstracts. u.p.. Sample containing about 199735 words of miscellanea (domain: social science)

HJ1 Rapid – ESRC grant abstracts. u.p.. Sample containing about 194081 words of miscellanea (domain: social science)

HJ2 Rapid – ESRC grant abstracts. u.p.. Sample containing about 43679 words of miscellanea (domain: social science)

HJ3 The Belfast Telegraph. u.p.. Sample containing about 165406 words from a periodical (domain: world affairs)

HJ4 The Belfast Telegraph. u.p.. Sample containing about 200828 words from a periodical (domain: world affairs)

HJ5 KPMG MAS engagement manual. KPMG, London (1993). Sample containing about 120652 words of un-
published miscellanea (domain: commerce)

**HJ6** The red triangle [Alumni Newsletter, Napier University]. External Relations Unit, Napier University. Sample containing about 5117 words from a periodical (domain: world affairs)

**HJ7** Journal of the Law Society of Scotland. Law Society’s Hall, Edinburgh (1975-04), 115-150. Sample containing about 32692 words from a periodical (domain: social science)

**HJ8** [Collected letters of the Scottish Sports Council]. u.p.. Sample containing about 12721 words of unpublished miscellanea (domain: leisure)

**HJ9** [Royal Botanic Garden Edinburgh: miscellaneous documents]. u.p.. Sample containing about 31562 words of unpublished miscellanea (domain: social science)

**HJA** [Royal Botanic Garden Edinburgh: miscellaneous documents]. Royal Botanic garden, Edinburgh, Scotland (1985- date marked). Sample containing about 51696 words of unpublished miscellanea (domain: social science)

**HJB** [Edinburgh District Council: assorted press releases]. u.p.. Sample containing about 6341 words of miscellanea (domain: social science)


**HJD** [Unpublished creative writing]. Wilson, Howard, u.p.. Sample containing about 45196 words of unpublished miscellanea (domain: imaginative)

**HJE** Coarse Fisherman. Metrocrest Ltd, UK (1989-10). Sample containing about 24547 words from a periodical (domain: leisure)


**HJH** Her living image. Rogers, Jane, Faber & Faber Ltd, London (1990), 55-171. Sample containing about 46429 words from a book (domain: imaginative)

**HJJ** Looking for an outdoor activity?. British Trust for Conservation. Sample containing about 344 words of miscellanea (domain: leisure)

**HJK** Pampers and the environment (letter). u.p.. Sample containing about 261 words of miscellanea (domain: applied science)

**HJL** [People’s Trust for Endangered Species]. u.p.. Sample containing about 560 words of miscellanea (domain: imaginative)

**HJM** WWF News. World-wide fund for nature, UK (1989). Sample containing about 17030 words of miscellanea (domain: world affairs)

**HJN** [Publisher’s advertising brochure]. Longman UK Ltd. Sample containing about 1039 words of miscellanea (domain: applied science)

**HJQ** [Publisher’s advertising brochure]. Longman. Sample containing about 857 words of miscellanea (domain: applied science)

**HJR** [Publisher’s advertising brochure]. u.p.. A colour atlas of infectious diseases. Sample containing about 350 words of miscellanea (domain: applied science)

**HJS** [Advertising leaflet for Nycomed]. u.p.. Splendid isolation. Sample containing about 148 words of miscellanea (domain: applied science)

**HJT** Research Defence Society. u.p.. Sample containing about 764 words of miscellanea (domain: applied science)

**HJU** RDS: Medical progress for man and animals. u.p.. Sample containing about 295 words of miscellanea (domain: applied science)

**HJV** Executive courses in business English. u.p.. Sample containing about 1438 words of miscellanea (domain: social science)
HJW Learn the language in the country. u.p.. Sample containing about 3816 words of miscellanea (domain: social science)

HJY It’s something you’ve always wanted to do. u.p.. ... Sample containing about 466 words of miscellanea (domain: social science)

HK0 Information sources on courses of higher education. u.p.. Sample containing about 758 words of miscellanea (domain: social science)

HK1 Sources of information on postgraduate courses. u.p.. Sample containing about 574 words of miscellanea (domain: social science)

HK2 What’s on – The Play House. u.p.. Sample containing about 4218 words of miscellanea (domain: arts)

HK3 Free to you. u.p.. Sample containing about 693 words of miscellanea (domain: arts)

HK4 The Simon Community: Christianity at work. u.p.. Sample containing about 828 words of miscellanea (domain: belief and thought)

HK5 Your parish church calendar 1992. u.p.. Sample containing about 301 words of miscellanea (domain: belief and thought)

HK7 Quality champagne specially selected for Tesco. Tesco stores ltd. Sample containing about 685 words of miscellanea (domain: leisure)

HK8 [Tesco guide to low and no alcohol]. Tesco stores ltd. Sample containing about 1950 words of miscellanea (domain: leisure)

HK9 Croxteth Hall & Country Park 1989. u.p.. Sample containing about 480 words of miscellanea (domain: leisure)

HKA [BCP – the no. British Car Parks. 1 car park for Gatwick Airport]. Sample containing about 808 words of miscellanea (domain: leisure)

HKC Freemans. Freemans. Sample containing about 62 words of miscellanea (domain: leisure)

HKD Vevay – Free this superb total colour make-up kit. Vevay. Sample containing about 468 words of miscellanea (domain: leisure)

HKE Young additions. u.p.. Sample containing about 226 words of miscellanea (domain: leisure)

HKF Harlow Sport Centre – general information. u.p.. Sample containing about 863 words of miscellanea (domain: leisure)

HKG Harlow Sport Centre – squash. u.p.. Sample containing about 561 words of miscellanea (domain: leisure)

HKH Bidwells – Trustees of the Hinxton Estate. Bidwells. Sample containing about 138 words of miscellanea (domain: leisure)

HKJ Bidwells – Land adjoining 22 High Street. Bidwells. Sample containing about 108 words of miscellanea (domain: leisure)

HKK Bidwells – The Hinxton Estate. Bidwells. Sample containing about 251 words of miscellanea (domain: leisure)

HKL British Gypsum – Gyproc cove fixing instructions. British Gypsum. Sample containing about 891 words of miscellanea (domain: leisure)

HKM Rockwell – The better alternative to the flat roof. u.p.. Sample containing about 832 words of miscellanea (domain: leisure)

HKN [Letter offering old document]. u.p.. Sample containing about 170 words of unpublished miscellanea (domain: leisure)

HKP Keesings Contemporary Archives. Longman Group UK Ltd, Harlow (1990). Sample containing about 81733 words from a periodical (domain: world affairs)

HKR Keesings Contemporary Archives. Longman Group UK Ltd, Harlow (1990). Sample containing about 79663 words from a periodical (domain: world affairs)

HKS Keesings Contemporary Archives.
Longman Group UK Ltd, Harlow (1990). Sample containing about 78200 words from a periodical (domain: world affairs)

**HKT** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1990). Sample containing about 101940 words from a periodical (domain: world affairs)

**HKU** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1990). Sample containing about 98110 words from a periodical (domain: world affairs)

**HKV** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1990). Sample containing about 102419 words from a periodical (domain: world affairs)

**HKW** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1990). Sample containing about 42380 words from a periodical (domain: world affairs)

**HKX** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1990). Sample containing about 80415 words from a periodical (domain: world affairs)

**HKY** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1990). Sample containing about 80887 words from a periodical (domain: world affairs)

**HL0** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1990). Sample containing about 97487 words from a periodical (domain: world affairs)

**HL1** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1990). Sample containing about 79665 words from a periodical (domain: world affairs)

**HL2** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1990). Sample containing about 44006 words from a periodical (domain: world affairs)


**HL7** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1991). Sample containing about 128070 words from a periodical (domain: world affairs)


**HLA** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1991). Sample containing about 58408 words from a periodical (domain: world affairs)


**HLC** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow
12. List of works excerpted

(1991). Sample containing about 62262 words from a periodical (domain: world affairs)


**HLF** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1992). Sample containing about 61932 words from a periodical (domain: world affairs)

**HLG** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1992). Sample containing about 63859 words from a periodical (domain: world affairs)

**HLH** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1992). Sample containing about 62739 words from a periodical (domain: world affairs)

**HLI** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1992). Sample containing about 61433 words from a periodical (domain: world affairs)

**HLK** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1992). Sample containing about 60670 words from a periodical (domain: world affairs)

**HLJ** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1992). Sample containing about 60735 words from a periodical (domain: world affairs)

**HLK** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1992). Sample containing about 60429 words from a periodical (domain: world affairs)

**HLN** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1992). Sample containing about 53438 words from a periodical (domain: world affairs)

**HLP** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1992). Sample containing about 60953 words from a periodical (domain: world affairs)

**HLS** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1992). Sample containing about 54163 words from a periodical (domain: world affairs)

**HLT** *Keesings Contemporary Archives.* Longman Group UK Ltd, Harlow (1992). Sample containing about 41744 words from a periodical (domain: world affairs)

**HLU** *Trade Union Annual Congress.*

| PS2F7 | ‘Dick’, president of trade union, male |
| PS2F8 | ‘Sheila’, trade unionist, female |
| PS2F9 | ‘Kathy’, trade unionist, female |
| PS2FA | ‘Ethel’, trade unionist, female |
| PS2FB | ‘Kathy’, trade unionist, female |
| PS2FC | ‘Steve’, trade unionist, male |
| PS2FD | ‘John’, trade unionist, male |
| PS2FE | ‘Adrian’, trade unionist, male |

Sample containing about 13058 words speech recorded in business context

**HLW** *Trade Union Annual Congress.*

| PS2K5 | ‘Dick’, president of trade union, male |
| PS2K6 | ‘Bill’, trade unionist, male |
| PS2K7 | ‘Avril’, trade unionist, female |
| PS2K8 | ‘Mel’, trade unionist, male |
| PS2KA | ‘Cliff’, trade unionist, male |
| PS2KB | ‘John’, trade unionist, male |
| HLWPS000 | |
| HLWPS001 | |

Sample containing about 11704 words speech recorded in business context

**HLX** *Trade Union Annual Congress.*

| PS2KL | ‘Dick’, president of trade union, male |
| PS2KM | ‘Dave’, trade unionist, male |
| PS2KN | ‘Nick’, trade unionist, male |
| PS2KP | ‘Frank’, trade unionist, male |
12. List of works excerpted

Sample containing about 2425 words speech recorded in business context

**HLY Trade Union Annual Congress.**

- PS2KR ‘Dick’, president of trade union, male
- PS2KS ‘Neil’, trade unionist, male
- PS2KT ‘Derek’, trade unionist, male
- PS2KV ‘Len’, trade unionist, male
- PS2KW ‘Owen’, trade unionist, male
- PS2KX ‘Peter’, trade unionist, male
- PS2KY ‘Hywel’, trade unionist, male

Sample containing about 13912 words speech recorded in business context

**HM2 Camera club meeting.**

- PS2MT ‘Jean’, female
- PS2MU ‘Mark’, male
- PS2MV ‘George’, male

Sample containing about 13523 words speech recorded in leisure context

**HM4 London Talkback Radio: radio broadcast.**

- PS2N3 ‘Douglas’, radio presenter, male
- PS2N4 ‘Pam’, sports presenter, female
- PS2N5 ‘Dave’, radio presenter, male
- PS2N6 ‘Frances’, female
- PS2N7 ‘Judy’, astrologer, female
- PS2N8 ‘Humphrey’, male
- PS2N9 ‘David’, male
- PS2NA ‘Bob’, male
- PS2NB ‘Madeline’, writer, female

Sample containing about 12143 words speech recorded in leisure context

**HM5 London Talkback Radio: radio broadcast.**

- PS2NL ‘Douglas’, radio presenter, male
- PS2NM ‘Pam’, sports presenter, female

Sample containing about 6896 words speech recorded in leisure context

**HM6 Pearson: annual general meeting and extraordinary general meeting.**

- PS2P1 ‘Michael’, director, male
- PS2P2 ‘David’, male
- PS2P3 ‘Simon’, director, male
- PS2P4 ‘John’, male

Sample containing about 6341 words speech recorded in business context

**HM7 Pearson interim results: meeting.**

- PS2P7 ‘Michael’, director, male
- PS2P8 ‘Frank’, managing director, male
- PS2P9 ‘James’, finance director, male

Sample containing about 3759 words speech recorded in business context

**HMA BBC Radio Nottingham: radio broadcast.**

- PS329 ‘Geoff Owen’, radio presenter, male
- PS32A ‘Alison Ford’, newsreader, female
- PS32B ‘Jonathan Clays’, reporter, male
- PS32C reporter, female
- PS32D 14+, Pre-recorded phone call, female
- PS32E ‘Mick Cox’, police detective superintendent, male
- PS32F ‘Nigel Bell’, reporter, male
- PS32G ‘Louise’, radio presenter, Action Line presenter, female
- PS32H ‘David’, education and training development officer, male

Sample containing about 15734 words speech recorded in leisure context

**HMD BBC Radio Nottingham: radio broadcast.**

- PS365 ‘Geoff’, radio presenter, male
- PS366 Caller on ’Lost for Lunchtime’ programme, female
- PS367 ‘John’, weather forecaster, male
- PS368 10+, schoolchild, Giving clue for ‘Kids’ County’ game, male
- PS369 10+, schoolchild, Giving clue for ‘Kids’ County’ game, male
- PS36A 10+, schoolchild, Giving clue for ‘Kids’ County’ game, male
- PS36B 10+, schoolchild, Giving clue for ‘Kids’ County’ game, male
- PS36C 10+, schoolchild, Giving clue for ‘Kids’ County’ game, male
- PS36D ‘Elizabeth’, Kids’ County game caller, female

Sample containing about 12499 words speech recorded in leisure context

**HMG Public Eye: documentary.**

- PS39M 30+, ‘Phil Parry’, tv presenter, male
- PS39N , male
- PS39P Miner’s widow, female
- PS39R mines rescue worker, male
- PS39S , male
- PS39T ‘Chris’, retired mines inspector, male
- PS39U , male
- PS39V ‘Len’, miner, male
- PS39W mines inspector, male

Sample containing about 4349 words speech recorded in leisure context

**HMH The Money Programme - part 1: documentary.**

- PS3AF tv presenter, Voiceover at start of programme, male
- PS3AG investor in business, male
- PS3AH tv presenter, male
- PS3AJ actor, male
- PS3AK business manager, male
- PS3AL business expert, female
12. List of works excerpted

PS3AM  ‘Ron’, company director, male
PS3AN  business investor, Might be speaker, male
HMHP S000

Sample containing about 3601 words speech recorded in leisure context

HMJ  *The Money Programme - part 2: documentary.*
PS3AR  tv presenter, female
PS3AS  ‘Paul’, lawyer, male
PS3AT  lawyer, male
PS3AU  ‘Keith’, departmental head, general motors, male
PS3AV  lawyer, male
PS3AW  i c i representative, male
PS3AX  lawyer, male
PS3AY  mercedes representative, male

Sample containing about 1875 words speech recorded in leisure context

HMK  *The Money Programme - part 3: documentary.*
PS3B1  tv presenter, male
PS3B2  , male
PS3B3  bell atlantic representative, male

Sample containing about 557 words speech recorded in leisure context

HML  *Gwynedd County Council tape 16: interview for oral history project.*
PS3B4  72, ‘Elizabeth’, Interviewee, female
PS3B5  historian, Interviewer., male

Sample containing about 8887 words speech recorded in leisure context

HMM  *Gwynedd County Council tape 19: interview for oral history project.*
PS3B6  60+, ‘Robert’, retired miner, male
PS3B7  historian, Speaks with a stutter, male

Sample containing about 7140 words speech recorded in leisure context

HMN  *The Central Match - Live: television sports broadcast.*
PS3B8  40+, ‘Tony Francis, ’ton”, tv presenter/sports commentator, male
PS3B9  ‘Gary Newbon’, tv presenter/sports commentator, male
PS3BA  50+, ‘Ron Atkinson’, football manager, sports commentator, male
PS3BB  50+, ‘Jimmy Greaves, ’greavesey”, tv presenter/sports commentator, male
PS3BC  ‘Alan Parry’, sports commentator, male
PS3BD  ‘Brian Little’, football manager, male
PS3BE  40+, ‘Frank Clark’, football manager, male
PS3BF  ‘John Helm’, sports commentator, male

Sample containing about 18640 words speech recorded in leisure context

HMP  *Jonathan Cowap Morning Show: radio broadcast.*
PS3BN  ‘Jonathan Cowap’, radio presenter, male
PS3BP  ‘Judy’, Phone-in caller, female
PS3BR  radio presenter, female
PS3BS  ‘Mr Mcmillan Scott’, euro member of parliament, male
PS3BT  ‘John Reynoldson’, county planning officer, male
PS3BU  ‘Peter’, minerals planning consultant, male
PS3BV  ‘Jason Cullen’, radio presenter, A A Roadwatch report, male
PS3BW  ‘Terry’, Phone-in caller, male
PS3BX  ‘Stewart’, Phone-in caller, male

Sample containing about 9029 words speech recorded in leisure context


HNN  *A good clean fight.* Robinson, Derek, Harvill, London (1993), 300-400. Sample containing about 38526 words from a book (domain: imaginative)


HNS  *The kiosk on the brink.* McKendrick, Jamie, Oxford University Press, Ox-
ford (1993), 11-52. Sample containing about 5346 words from a book (domain: imaginative)


**HNU Plato's ladder.** Romer, Stephen, Oxford University Press, Oxford (1992), 8-78. Sample containing about 7510 words from a book (domain: imaginative)


**HNX The sirocco room.** McKendrick, Jamie, Oxford University Press, Oxford (1991), 1-49. Sample containing about 6554 words from a book (domain: imaginative)


**HP0 The rich are with you always.** Macdonald, Malcolm, Coronet Books, London (1978), 187-304. Sample containing about 46926 words from a book (domain: imaginative)


**HP4 [Wimpey newsletter].** Wimpey. Sample containing about 34668 words of unpublished miscellanea (domain: applied science)

**HP5 [BP annual report].** u.p.. Sample containing about 11524 words of unpublished miscellanea (domain: applied science)

**HP6 [Scottish Amicable newsletter].** Scottish Amicable. Sample containing about 25420 words of unpublished miscellanea (domain: social science)

**HP7 [Johnson Matthey Annual Report 1993].** u.p.. Sample containing about 12675 words of unpublished miscellanea (domain: commerce)

**HP8 Assay – magazine of the Johnson Matthey Group.** Reteif, France and Godwin, Ian D, u.p.. Sample containing about 15539 words of unpublished miscellanea (domain: commerce)

**HP9 P&O Services.** u.p.. Sample containing about 3269 words of unpublished miscellanea (domain: commerce)

**HPA BNFL News.** British Nuclear Fuels plc, Manchester (1992 March - May). Sample containing about 10705 words from a periodical (domain: applied science)

**HPB BNFL News.** British Nuclear Fuels plc, Manchester (1992 March - May). Sample containing about 10157 words from a periodical (domain: applied science)

**HPC BNFL News.** British Nuclear Fuels plc, Manchester (1992 March - May). Sample containing about 10157 words from a periodical (domain: applied science)

**HPD Scotvec reports.** u.p.. Sample containing about 26800 words of unpublished miscellanea (domain: social science)

**HPE WISE: Women into Science & Engi-
neering. u.p.. Sample containing about 3387 words of unpublished miscellanea (domain: applied science)

**HPF** *The Engineering Council Guidelines on Risk Issues*. u.p.. Sample containing about 14690 words of unpublished miscellanea (domain: applied science)

**HPG** *[School essays on literature]*. u.p.. Sample containing about 26627 words of unpublished miscellanea (domain: arts)

**HPH** *[Christian Aid leaflets]*. u.p.. Sample containing about 21437 words of unpublished miscellanea (domain: social science)

**HPJ** *[GDA annual report]*. u.p.. Sample containing about 5090 words of unpublished miscellanea (domain: commerce)

**HPK** *[Parish Council letters]*. u.p.. Sample containing about 28081 words of unpublished miscellanea (domain: social science)

**HPL** *[Lothian Council: unpublished documents]*. u.p.. Sample containing about 19411 words of unpublished miscellanea (domain: social science)

**HPM** *Subjective risk and memory for driving situations [Unpublished PhD thesis]*. u.p.. Sample containing about 50069 words of unpublished miscellanea (domain: social science)

**HPN** *The purpose and use of the research [Unpublished PhD thesis]*. u.p.. Sample containing about 50262 words of unpublished miscellanea (domain: social science)

**HPP** *[Spokes: unpublished letters & articles]*. u.p.. Sample containing about 43325 words of unpublished miscellanea (domain: social science)

**HPR** *[Unpublished short stories]*. u.p.. Sample containing about 7421 words of unpublished miscellanea (domain: imaginative)

**HPS** *[Unpublished essays]*. u.p.. Sample containing about 9984 words of unpublished miscellanea (domain: world affairs)


HR5 The hunt by night. Mahon, Derek, Oxford University Press, Oxford (1986), 9-63. Sample containing about 7246 words from a book (domain: imaginative)

HR6 The dean it was that died. Whitehead, Barbara, Constable & Company Ltd, London (1991), 55-177. Sample containing about 42715 words from a book (domain: imaginative)


HRM Urban sociology. Dickens, Peter, Harvester & Wheatsheaf, Hemel Hempstead (1990), 28-152. Sample containing about 43788 words from a book (domain: social science)

HRN Nature. Macmillan Magazines Ltd,
12. List of works excerpted


**HRT** [Link – the house journal of the Pauls group]. u.p.. Sample containing about 48653 words of unpublished miscellanea (domain: arts)

**HRU** Guinness Today. u.p.. Sample containing about 7493 words of unpublished miscellanea (domain: arts)

**HRV** Guinness Globe. u.p.. Sample containing about 7128 words of unpublished miscellanea (domain: arts)

**HRW** Dawson International plc: Notice of annual general meeting. u.p.. Sample containing about 4185 words of unpublished miscellanea (domain: arts)

**HRX** Dawson International plc: Annual report '93. u.p.. Sample containing about 6019 words of unpublished miscellanea (domain: arts)

**HRY** Glenpatrick News. u.p.. Sample containing about 11455 words of unpublished miscellanea (domain: leisure)

**HS0** Glenpatrick News. u.p.. Sample containing about 10947 words of unpublished miscellanea (domain: leisure)

**HS1** Glenpatrick News. u.p.. Sample containing about 12012 words of unpublished miscellanea (domain: leisure)

**HS2** Glenpatrick News. u.p.. Sample containing about 11646 words of unpublished miscellanea (domain: leisure)

**HS3** Glenpatrick News. u.p.. Sample containing about 13171 words of unpublished miscellanea (domain: leisure)

**HS4** Glenpatrick News. u.p.. Sample containing about 2057 words of unpublished miscellanea (domain: leisure)

**HS7** Special religious education division. u.p.. Sample containing about 7992 words of unpublished miscellanea (domain: belief and thought)


**HS9** A guide to the new simplified divorce. u.p.. Sample containing about 3345 words of unpublished miscellanea (domain: social science)

**HSA** [Unpublished creative writing]. u.p.. Sample containing about 15332 words of unpublished miscellanea (domain: imaginative)

**HSB** [Unpublished university lecture notes]. u.p.. Sample containing about 12537 words of unpublished miscellanea (domain: natural sciences)

**HSC** [Unpublished university essays]. u.p.. Sample containing about 37352 words of unpublished miscellanea (domain: world affairs)


words from a periodical (domain: arts)

**HSR**: *Exciting new titles in parasitology from Chapman and Hall*. Chapman & Hall, London. Sample containing about 859 words of miscellanea (domain: natural sciences)

**HSS**: *Man on the moon – 20th anniversary special offer*. u.p.. Sample containing about 675 words of miscellanea (domain: natural sciences)

**HST**: *Olympus Scientific Products News*. Olympus Scientific Products. Sample containing about 2529 words of miscellanea (domain: applied science)

**HSU**: *The Queen’s awards for export and technology 1990*. Department of Trade & Industry (1990). Sample containing about 689 words of miscellanea (domain: applied science)

**HSW**: *Superfile – database manager for mailing lists*. Southdata Ltd.. Sample containing about 539 words of miscellanea (domain: applied science)

**HSX**: *Alex – advanced OCR*. Southdata Ltd.. Sample containing about 311 words of miscellanea (domain: applied science)

**HSY**: *CompuAdd*. CompuAdd. The catalogue. Sample containing about 2915 words of miscellanea (domain: applied science)

**HT0**: *Reading your meter*. Eastern Electricity. Sample containing about 254 words of miscellanea (domain: applied science)

**HT2**: *Satisfy your desire for a living flame gas fire*. British Gas Eastern. Sample containing about 418 words of miscellanea (domain: applied science)

**HT3**: *Let E. Eastern Electricity. T. safeguard your future*. Sample containing about 617 words of miscellanea (domain: applied science)

**HT4**: *Keep warm – keep well*. Department of health/Central office for information. Sample containing about 4884 words of miscellanea (domain: applied science)


**HT6**: *Tesco cares: a guide to the world around us*. Tesco stores ltd. Sample containing about 2107 words of miscellanea (domain: applied science)

**HT7**: *CFCs and the ozone layer*. Tesco stores ltd. Sample containing about 703 words of miscellanea (domain: applied science)

**HT8**: *Phosphates and the water supply*. Tesco stores ltd. Sample containing about 458 words of miscellanea (domain: applied science)

**HT9**: *Paper and the environment*. Tesco stores ltd. Sample containing about 584 words of miscellanea (domain: applied science)

**HTA**: *Recycling and the environment*. Tesco stores ltd. Sample containing about 856
words of miscellanea (domain: applied science)

**HTC** The Embalmer. British Institute of Embalmers, Knebworth (1993-12). Sample containing about 11503 words from a periodical (domain: applied science)


**HTF** Bureaucracy and political power. Smith, B C, Wheatsheaf Books Ltd, Brighton (1988), 40-175. Sample containing about 42584 words from a book (domain: world affairs)


**HTN** A little lower than the angels. McCaughrean, Geraldine, Oxford University Press, Oxford (1987), 13-133. Sample containing about 37691 words from a book (domain: imaginative)


**HTV** Sociology of the global system. Sklair, Leslie, Harvester & Wheatsheaf, Hemel Hempstead (1991), 001-126. Sample containing about 45926 words from a book (domain: social science)


**HU0** The white rose murder. Clynes, Michael, Headline Book Publishing plc, London (1992), 73-209. Sample containing about 42518 words from a book
12. List of works excerpted

HU1  *The Embalmer*. British Institute of Embalmers, Knebworth (1993-12), 3-35. Sample containing about 19268 words from a periodical (domain: applied science)


HU3  *Gut: Journal of Gastroenterology and Hepatology*. Alan, R N (ed.), British Medical Association, London. Sample containing about 176734 words from a periodical (domain: applied science)

HU4  *Gut: Journal of Gastroenterology and Hepatology*. Alan, R N (ed.), British Medical Association, London. Sample containing about 167165 words from a periodical (domain: applied science)

HU5  *Annual report*. The general council of the bar, England (1992), 1-25. Sample containing about 15215 words of unpublished miscellanea (domain: commerce)


HU9  *Bishops’ Conference of Scotland*. u.p.. Sample containing about 18293 words of unpublished miscellanea (domain: belief and thought)

HUA  *Short stories*. Potter, Alistair, u.p.. Sample containing about 13283 words of unpublished miscellanea (domain: imaginative)

HUB  *Schoolboys’ essays*. u.p.. Sample containing about 21358 words of unpublished miscellanea (domain: social science)


HUCPS000 Sample containing about 13033 words speech recorded in business context


HUEPS000 Sample containing about 12880 words speech recorded in business context


HUEPS001 Sample containing about 13483 words speech recorded in business context

HUF  *King’s College London: political philosophy lecture*. PS2MY 30+, ‘Wolff’, lecturer, male

HUFPS000 Sample containing about 17579 words speech recorded in educational context

HUG  *Hatfield Airfield air traffic control*. PS2PA 47, air traffic controller, male PS2PB 46, air traffic controller, male PS2PC 42, air traffic controller, female

HUGHPS000 Sample containing about 1484 words speech recorded in public context

HUH  *London School of Economics: lecture*. PS2PD 45, lecturer, male
Sample containing about 7793 words speech recorded in educational context

**HUI** London School of Economics: lecture.
PS2PE 45, 'Chris', lecturer, male
PS2PF ‘Andrea’, student, female
PS2PG ‘Mike’, student, male
PS2PH ‘Kirsty’, student, female
PS2PJ ‘Mary-jane’, student, female
PS2PK ‘John’, student, male
PS2PL ‘Emma’, student, female
PS2PM ‘Anne-marie’, student, female
PS2PN ‘Katherine’, student, female

Sample containing about 15927 words speech recorded in educational context

**HUK** London School of Economics: lecture on the psychoanalytical study of society.
PS2PU 45, ‘Chris’, lecturer, male
PS2PV ‘Joy’, student, female
PS2PW ‘Tom’, student, male
PS2PX ‘Dean’, student, male
PS2PY ‘Heather’, student, female
PS2R0 ‘Duncan’, student, male
PS2R1 ‘Oliver’, student, male
PS2R2 ‘Sue’, student, female

Sample containing about 14078 words speech recorded in educational context

**HUL** London School of Economics: lecture on the psychoanalytical study of society.
PS2R3 45, ‘Chris’, lecturer, male

Sample containing about 13461 words speech recorded in educational context

**HUM** London School of Economics: lecture on the psychoanalytical study of society.
PS2R4 45, ‘Chris’, lecturer, male

Sample containing about 15806 words speech recorded in educational context

**HUN** London School of Economics: lecture on the psychoanalytical study of society.
PS2R5 45, ‘Chris’, lecturer, male

Sample containing about 16624 words speech recorded in educational context

**HUP** Pearson preliminary results: analysts meeting.
PS2Y7 ‘Michael’, male
PS2X8 ‘David’, chairman, male
PS2Y9 ‘Frank’, chief executive, male
PS2YA ‘James’, finance director, male
PS2YB ‘Robert’, male
PS2YC ‘Ken’, male
PS2YD ‘Charles’, male
PS2YE ‘Lorna’, female
PS2YF ‘Stuart’, chief executive officer, male

Sample containing about 10032 words speech recorded in business context

**HUR** Christie’s auction rooms: auction of mechanical music.
PS308 auctioneer

Sample containing about 14247 words speech recorded in leisure context

**HUS** Christie’s auction rooms: auction of oriental ceramics and works of art.
PS309 auctioneer

Sample containing about 11749 words speech recorded in leisure context

**HUT** Pastoral parish council meeting.

Sample containing about 7640 words speech recorded in public context

**HUU** Trent Law School: lecture.
PS30A 35, lecturer, female

Sample containing about 6649 words speech recorded in educational context

**HUV** James Whale Phone-in: radio broadcast.
PS3C8 40+, ‘James Whale’, radio presenter, male
PS3C9 ‘Danny’, radio technician, male
PS3CA ‘Douglas’, Phone-in caller, male
PS3CB 60+, ‘Sam’, Phone-in caller, male
PS3CC ‘Melanie’, civil servant, Phone-in caller, female
PS3CD ‘Alf’, Phone-in caller, male
PS3CE ‘Danny’, radio presenter, A A Roadwatch travel news, male
PS3CF ‘David’, nspcc employee, Phone-in caller, male
PS3CG ‘Michael’, male

Sample containing about 17806 words speech recorded in leisure context

**HUW** On the Record: television broadcast.
PS3CM 50+, ‘John Humphries’, tv presenter, male
PS3CN 50+, ‘John Hume’, member of parliament (sdp leader), male
PS3CP british rail announcer, male
PS3CR 30+, ‘Michael Gold’, television reporter, male
PS3CS ‘Michael Gold’, male
PS3CT 40+, ‘Protestor’, female
PS3CU 30+, ‘Protestor’, male
PS3CV 50+, ‘Protestor’, female
PS3CW 20+, waiter, male

Sample containing about 10377 words speech recorded in leisure context

**HUX** Gwynedd County Council tape 18: interview for oral history project.

Sample containing about 8285 words speech recorded in leisure context
12. List of works excerpted

**HU Y** Gwynedd County Council tape 20: interview for oral history project.

PS3D9 historian, Interviewer, male

PS3DA 28, ‘David’, quarry worker, Interviewee, male

Sample containing about 8549 words speech recorded in leisure context

**HV0** James Whale Phone-in: radio broadcast.

PS3DB 40+, ‘James Whale’, radio presenter, male

PS3DC 30+, ‘Alan’, phone-in caller, male

PS3DD Recorded voice of American spokesman, male

PS3DE 30+, ‘Karen’, radio studio technician, female

PS3DF 30+, ‘Danny Savage’, radio presenter, A Roadwatch report, male

PS3DG 30+, ‘Noel’, n.u.t. press officer, male

PS3DH 30+, ‘Rita’, Phone-in caller, female

PS3DJ 20+, ‘Rachael’, guide for guided tour company, female

PS3DK 20+, radio studio technician, male

Sample containing about 16683 words speech recorded in leisure context

**HV1** Central Weekend Live - part 1: television broadcast.

PS3E3 30+, ‘Nicky Campbell’, tv presenter, male

PS3E4 40+, ‘Sue Jay’, tv presenter, female

PS3E5 30+, ‘Oliver James’, consultant psychologist, Slight lisp on R, male

PS3E6 40+, ‘Dr Vernon Coleman’, medical consultant, male

PS3E7 40+, ‘Dr Cosmo Hallstrom’, medical consultant, male

PS3E8 40+, ‘Donna’, journalist, female

PS3E9 40+, ‘Dr Simon Fradd’, general practitioner, male


PS3EB 40+, ‘Peter’, male

Sample containing about 5439 words speech recorded in leisure context

**HV2** Central Weekend Live - part 2: television broadcast.

PS3EE 30+, ‘Nicky Campbell’, tv presenter, male

PS3EF 50+, ‘John Taylor’, member of parliament (ulster unionist) for strangford, male

PS3EG 40+, ‘Clare Short’, member of parliament (labour), female

PS3EH 60+, ‘Tim Pat’, historian, ira expert, male

PS3EJ 40+, ‘Marten Dillon’, author, Author of The Dirty War, about the Irish conflict, male

PS3EK 50+, ‘Sir William Ross’, member of parliament for east londonderry, male

PS3EL 50+, ‘Ray’, ex-soldier, male

PS3EM 50+, ‘Michael’, lieutenant-colonel, counter-terrorism expert, male

PS3EN 40+, ‘Mary’, Representing the Troops Out movement, female

Sample containing about 5297 words speech recorded in leisure context

**HV3** Central Weekend Live - part 3: television broadcast.

PS3EP 40+, ‘Ed Doolan’, tv presenter, male

PS3ER 30+, ‘Beth’, animal rights defender, female

PS3ES 40+, ‘Michael’, representative of circus society, male

PS3ET 30+, ‘Diane’, representative of circus watch, female

PS3EU 30+, ‘Gavin’, circus trainer/performer, male

PS3EV 50+, ‘Pat’, representative of animal welfare alliance, female

PS3EW 40+, ‘Dave’, circus owner, male

PS3EX 50+, ‘Dr Marthe’, animal behaviourist, Slight lisp on R, female

PS3EY 40+, ‘Rusty’, clown, male

Sample containing about 3459 words speech recorded in leisure context

**HV4** Inaugural address of new president of National Association of Round Tables of Great Britain and Northern Ireland.

PS3F2 ‘Dyfrig James’, businessman, male

Sample containing about 862 words speech recorded in leisure context

**HV5** Orkney Sound Archive tape OSA/RO/D: radio broadcast.

PS3F3 60+, ‘Mary’, Native of West Mainland, female

PS3F4 30+, radio presenter, Interviewer, female

Sample containing about 3852 words speech recorded in leisure context

**HV6** Orkney Sound Archive tape OSA/RO/D: radio broadcast.

PS3F5 ‘Mansie Flaws’, radio presenter, Interviewer, female

PS3F6 70+, retired ferryman, Native of Wyre, Orkney., male

Sample containing about 3825 words speech recorded in leisure context

**HV7** Orkney Sound Archive tape OSA/RO/C: radio broadcast.

PS3F7 10+, school child, male

PS3F8 radio presenter, Interviewer, male

PS3F9 10+, school child, male

PS3FA 10+, school child, male

PS3FB 10+, school child, male

PS3FC 10+, ‘Gary’, school child, male

PS3FD ‘Dennis’, postmaster, male

PS3FE farmer, male
Sample containing about 1409 words speech recorded in leisure context

HV8 Orkney Sound Archive tape
OSA/RO/C: radio broadcast.
PS3FF school child, female
PS3FG radio presenter, Interviewer, female
PS3FH school child, female
PS3FJ 70+, female
PS3FK 70+, female

Sample containing about 779 words speech recorded in leisure context

HV9 Orkney Sound Archive tape
OSA/RO/C: radio broadcast.
PS3FL 60+, ‘David’, Native of Hoy., male
PS3FN ‘Liz Davies’, radio presenter, Interviewer, female
PS3FP 17+, student, female
PS3FR 17+, ‘Jenny’, student, female
PS3FS 17+, ‘Lucy’, student, female

Sample containing about 929 words speech recorded in leisure context

HVA Orkney Sound Archive tape
OSA/RO/C: radio broadcast.
PS3FM 30+, ‘Fiona’, local councillor, artist, Interviewee, female
PS3FN ‘Liz Davies’, radio presenter, Interviewer, female
PS3FP 17+, student, female
PS3FR 17+, ‘Jenny’, student, female
PS3FS 17+, ‘Lucy’, student, female

Sample containing about 1643 words speech recorded in leisure context

HVB Orkney Sound Archive tape
OSA/RO/C: radio broadcast.
PS3FT radio presenter, Interviewer, female
PS3FU ‘Stewart’, school child, male
PS3FW ‘Carol’, school child, male
PS3FX ‘Nathan’, school child, male
PS3FY ‘Louise’, school child, female
PS3G0 ‘Nettie’, school child, female

Sample containing about 1643 words speech recorded in leisure context

HVC Central Weekend Live - part 1: television broadcast.
PS3G1 30+, ‘Nicky Campbell’, tv presenter, male
PS3G2 40+, ‘Ed Doolan’, tv presenter, male
PS3G3 30+, ‘Tony’, male
PS3G4 20+, ‘Alaina’, female
PS3G5 50+, ‘Theresa Gorman’, member of parliament (conservative), female
PS3G6 20+, ‘Lindsay’, female
PS3G7 20+, ‘John’, male
PS3G8 50+, ‘Sir Ivan Lawrence’, member of parliament (conservative), male
PS3G9 30+, ‘Andrew’, male

Sample containing about 6278 words speech recorded in leisure context

HVD Central Weekend Live - part 2: television broadcast.
PS3GL 30+, ‘Nicky Campbell’, tv presenter, male
PS3GM 50+, ‘Ronnie’, army brigadier, male
PS3GN 30+, ‘Andrew’, ex-soldier, male
PS3GP 20+, ‘Richard’, ex-soldier, male
PS3GR 50+, ‘David Lightbown’, member of parliament (conservative), male
PS3GS 30+, ‘Tim’, ex-soldier, male
PS3GT 40+, ‘Paul’, Slight lisp on S, male
PS3GU 30+, ‘Alain’, rem in army training regiment, male
PS3GV 30+, ex-soldier (awol), male

Sample containing about 5159 words speech recorded in leisure context

HVE Central Weekend Live - part 3: television broadcast.
PS3GX 40+, ‘Sue Jay’, tv presenter, female
PS3GY 40+, ‘Barbara’, journalist, Australian?, female
PS3H0 50+, ‘Tony’, male
PS3H1 50+, ‘Vince’, male
PS3H2 60+, ‘Lance’, male
PS3H3 30+, ‘Alan’, male
PS3H4 40+, ‘John’, male
PS3H5 40+, ‘Deniel’, stylist, male
PS3H6 50+, ‘Joshua’, stylist, male

Sample containing about 3572 words speech recorded in leisure context

HVF EIP meeting at Strensall Village Hall, day 1, morning session: public county council meeting.
PS3HD 50+, ‘Eric Barnett’, department of the environment adjudicator, Chairing the meeting., male
PS3HE 40+, ‘Mr P Davies’, legal representative, Representing North Yorkshire County Council, male
PS3HF 50+, ‘Stuart Cowley’, legal representative, Representing North Yorkshire County Council, male
PS3HG 40+, ‘R Whittaker’, senior inspector, female
PS3HH 35+, ‘Roy Donson’, legal representative, Representing House Builders Federation, male
PS3HJ 50+, ‘M Rees’, department of the environment representative, female
PS3HK 30+, ‘Malcolm Spittle’, north yorkshire county council representative, male
PS3HL 40+, ‘Steven Gregson’, legal representative, Representing Barton Wilmore, male
PS3HM 30+, ‘Mrs Long’, north yorkshire county council representative, female

Sample containing about 21960 words speech recorded in public context

HVG EIP meeting at Strensall Village Hall, day 1, afternoon session: public county council meeting.
PS3JO 50+, ‘Eric Barnett’, department of the environment adjudicator, Chairperson, male
12. List of works excerpted

PS3J1 50+, ‘Margaret Rees’, department of the environment representative, female
PS3J2 40+, ‘R Whittaker’, senior inspector, female
PS3J3 30+, ‘Dave Girt’, leeds city council representative, male
PS3J4 50+, ‘W Caulfield (bill)’, bradford district council representative, male
PS3J5 35+, ‘Roy Donson’, house builders federation representative, male
PS3J6 40+, ‘Stephen Grigson’, solicitor, male
PS3J7 30+, ‘J Cunnane’, solicitor, Representing Sam Smith’s Tadcaster Brewery, male
PS3J8 40+, ‘P Sedgewick’, solicitor, male
PS3J9 Sample containing about 15663 words speech recorded in public context

HVK EIP meeting at Strensall Village Hall, day 4 session: public county council meeting.
PS3JH 50+, ‘Eric Barnett’, department of the environment adjudicator, Chairing the meeting, male
PS3JJ 40+, ‘Mr P Davies’, legal representative, Representing North Yorkshire County Council, male
PS3JK 40+, ‘D Whittaker’, senior inspector, female
PS3JL 35+, ‘Roy Donson’, house builders federation representative, male
PS3JM 30+, ‘R Thomas’, solicitor, male
PS3JN ‘Mr Wincup’, department of the environment representative, male
PS3JP ‘M Courcier’, solicitor, male
PS3JV ‘Mr Brighton’, male
PS3JS ‘Mr Brooke’, male

Sample containing about 24634 words speech recorded in public context

HVJ EIP meeting at Strensall Village Hall, day 3, morning session: public county council planning meeting.
PS3J3 50+, ‘Mr E Barnett’, department of the environment adjudicator, Chairing the enquiry, male
PS3J4 40+, ‘Mr Peter Davies’, north yorkshire county council legal representative, male
PS3J5 40+, ‘Miss D Whittaker’, senior inspector, female
PS3J6 30+, ‘Mr David Curtis’, york city council legal representative, male
PS3J7 25+, ‘Mr Ian Smith’, ryedale district council legal representative, male
PS3J8 40+, ‘Mr Steven Grigson’, solicitor, from Barton Willmore, male
PS3J9 30+, ‘Mr Malcolm Spittle’, north yorkshire county council legal representative, male
PS3JA 30+, ‘Mrs Long’, north yorkshire county council legal representative, female
PS3JB 35+, ‘Mr Roy Donson’, house builders federation representative, male

Sample containing about 23224 words speech recorded in public context


12. List of works excerpted

---


HWF *Lifespan computer manuals.* Corporate, u.p.. Sample containing about 221635 words of unpublished miscellanea (domain: applied science)


HWM *The jewel that was ours.* Dexter, Colin, Pan Books Ltd, London (1992), 89-254. Sample containing about 46954 words from a book (domain: imaginative)


HWS *Gut: Journal of Gastroenterology and Hepatology.* Alan, R N (ed.), British Medical Association, London. Sample containing about 170292 words from a periodical (domain: applied science)

HWT *Gut: Journal of Gastroenterology and Hepatology.* Alan, R N (ed.), British Medical Association, London. Sample containing about 59868 words from a periodical (domain: applied science)


HX0 *Scotvec update.* Scotvec, Glasgow (1992/1919). Sample containing about 16564 words of unpublished miscellanea (domain: social science)


HX3 *Hotel, Catering & Institutional Management Association.* u.p.. Sample containing about 18426 words of unpublished miscellanea (domain: commerce)

HX4 *Hotel, Catering & Institutional Management Association.* u.p.. Sample containing about 32091 words of un-
List of works excerpted


HX6 Christian Aid. Christian Aid, Edinburgh, Scotland. Sample containing about 965 words of unpublished miscellanea (domain: social science)


HX9 [Unpublished academic thesis]. Wilson, Howard, u.p.. Sample containing about 40822 words of unpublished miscellanea (domain: applied science)

HXA [Articles by a management consultant]. u.p.. Sample containing about 7261 words of unpublished miscellanea (domain: commerce)

HXB [Solicitors’ information packs]. u.p.. Sample containing about 38263 words of unpublished miscellanea (domain: social science)


HXD Introduction to computer law. Bainbridge, David I, Longman Group UK Ltd, Harlow (1993), 24-121. Sample containing about 45312 words from a book (domain: social science)


HXH How to write essays, dissertations & theses in literary studies. Fabb, Nigel and Durant, Alan, Longman Group UK Ltd, Harlow (1993), 9-140. Sample containing about 46432 words from a book (domain: arts)


HXN [article from] European economic integration. European economic integration. McDonald, Frank, Dearden, Stephen (eds.)Kemp, John, Longman
12. List of works excerpted

Group UK Ltd, Harlow (1992), 59-80. Sample containing about 11521 words from a book (domain: world affairs)


**HXS** *The fabliau in English.* Hines, John, Longman Group UK Ltd, Harlow (1993), 1-139. Sample containing about 44895 words from a book (domain: arts)


**HXU** *De Gaulle.* Shennan, Andrew, Longman Group UK Ltd, Harlow (1993), 16-139. Sample containing about 44281 words from a book (domain: world affairs)


**HXW** *Land law.* Lim, Hilary, Longman Group UK Ltd, Harlow (1992), 24-121. Sample containing about 46865 words from a book (domain: social science)


**HY3** *Markets and dealers.* Davis, Phil, Longman Group UK Ltd, Harlow (1992), 111-134. Sample containing about 11132 words from a book (domain: commerce)


**HY6** *English medieval mystics: games of faith.* Glasscoe, Marion, Longman Group UK Ltd, Harlow (1993), 1-159. Sample containing about 46423 words from a book (domain: arts)


**HY8** *The special relationship.* Bartlett, C J, Longman Group UK Ltd, Harlow (1992), 1-124. Sample containing about 46090 words from a book (domain:
List of works excerpted

**HY9** Politics under the later Stuarts. Harris, Tim, Longman Group UK Ltd, Harlow (1993), 117-239. Sample containing about 46659 words from a book (domain: world affairs)

**HYA** The teaching of drama in the primary school. Woolland, Brian, Longman Group UK Ltd, Harlow (1993), 34-138. Sample containing about 38921 words from a book (domain: social science)


**HYC** Suffolk Sound Archive oral history project: interview.

- **PS228** ‘Joyce’, female
- **PS229** ‘Jack’, retired head lad, male

Sample containing about 12152 words speech recorded in leisure context

**HYD** King’s College London: philosophy discussion class.

- **PS2N2** 30, ‘Brennan’, lecturer, male

Sample containing about 8958 words speech recorded in educational context

**HYE** Pearson preliminary results: presentation and press and analysts’ questions.

- **PS2YH** ‘Michael’, chief executive officer, male
- **PS2YJ** ‘Frank’, chairman, male
- **PS2YK** ‘James’, male
- **PS2YL** ‘Mark’, reporter, male
- **PS2YM** ‘Jason’, reporter, male
- **PS2YN** ‘Richard’, reporter, male
- **PS2YP** ‘Raymond’, reporter, male
- **PS2YR** ‘William’, reporter, male
- **PS2YS** ‘Roger’, reporter, male

Sample containing about 12023 words speech recorded in business context

**HYF** Pearson analyst presentation: analysts’ meeting.

- **PS302** ‘Michael’, male
- **PS303** ‘Frank’, chief executive, male
- **PS304** ‘James’, financial director, male
- **PS305** ‘David’, male
- **PS306** ‘Mark’, male
- **PS307** ‘Mark’, male

Sample containing about 6368 words speech recorded in business context

**HYG** Town council meeting. Sample containing about 9767 words speech recorded in public context

**HYH** Trent Law School: tutorial.

- **PS30B** 35, ‘John’, lecturer, male

Sample containing about 3480 words speech recorded in educational context

**HYJ** Town council grants meeting.

- **PS3C0** 40+, ‘Graham’, farmer and developer, male
- **PS3C1** 40+, ‘Paul’, homeopath, male
- **PS3C2** 40+, accountant, female
- **PS3C3** 80+, farmer, male
- **PS3C4** 60+, farmer, male
- **PS3C5** 50+, ‘Trevor’, builder, male

Sample containing about 15696 words speech recorded in public context

**HYK** Town council planning meeting.

- **PS3CH** 40+, ‘Graham’, farmer and developer, male
- **PS3CJ** 80+, farmer, male
- **PS3CK** 40+, builder, male

Sample containing about 7767 words speech recorded in public context

**HYL** Nottingham University Economics Department: tutorial.

- **PS3K6** 30, ‘Lloyd’, lecturer, male
- **PS3K7** ‘Matthew’, student, male

Sample containing about 4947 words speech recorded in educational context

**HYM** Nottingham University Economics Department: agricultural economics lecture.

- **PS3K8** 30, ‘Chris’, lecturer, male

Sample containing about 4490 words speech recorded in educational context

**HYN** Nottingham University Economics Department: agricultural economics tutorial.

- **PS3K9** 30, ‘Tim’, lecturer, male

Sample containing about 5248 words speech recorded in educational context

**HYP** Nottingham University Economics Department: tutorial.

- **PS3KA** 30, ‘Lloyd’, lecturer, male
- **PS3KB** ‘Nadine’, student, female
- **PS3KC** ‘Lyn’, student, female
- **PS3KD** student, male
- **PS3KE** student, male

Sample containing about 6147 words speech recorded in educational context
HYR  Nottingham University Economics Department: lecture.
    PS3KF  lecturer
    Sample containing about 4523 words speech recorded in educational context

HYS  Nottingham University Economics Department: tutorial.
    PS3KG  30, ‘Tim’, lecturer, male
    HYSPS000
    Sample containing about 6173 words speech recorded in educational context

HYT  Nottingham University Economics Department: lecture.
    PS3KH  lecturer
    HYTPS000
    Sample containing about 5789 words speech recorded in educational context

HYU  Commercial Union training session.
    PS3KJ  ‘Dawn’, trainer, female
    PS3KK  trainer, male
    PS3KL  ‘Elaine’, insurance company employee, female
    PS3KM  ‘John’, insurance company employee, male
    Sample containing about 10605 words speech recorded in educational context

HYV  Commercial Union training session.
    PS3KN  ‘Martin’, trainer, male
    PS3KP  trainer, female
    PS3KR  ‘Andy’, insurance company employee, male
    PS3KS  ‘Margaret’, insurance company employee, female
    PS3KT  ‘Mark’, insurance company employee, male
    PS3KU  ‘John’, insurance company employee, male
    PS3KV  ‘Marie’, insurance company employee, female
    PS3KW  ‘Nora’, insurance company employee, female
    PS3KX  ‘Irene’, insurance company employee, female
    Sample containing about 9890 words speech recorded in educational context

HYW  Commercial Union training session.
    PS3KY  trainer, male
    Sample containing about 3242 words speech recorded in educational context

HYX  Shropshire County Council: committee meeting.
    PS3L0  ‘Jerry’, councillor, male
    Sample containing about 20051 words speech recorded in public context

HYY  British Association of Industrial Editors - Edinburgh Internal Communications Group: informal meeting.
    PS3L1  32, ‘Peter’, ex bp public affairs, male
    PS3L2  47, bnc worker, female
    Sample containing about 6809 words speech recorded in educational context


J0V  Electronic information resources and the historian. Ross, Seamus and Higgs, Edward, u.p.. Sample containing about 79435 words of unpublished miscellanea (domain: applied science)


JOY  A Martian sends a postcard home. Raine, Craig, Oxford University Press, Oxford (1979), 1-46. Sample contain-
12. List of works excerpted


J1C [Leeds United e-mail list]. u.p.. Sample containing about 41810 words of unpublished miscellanea (domain: leisure)

J1D [Leeds United e-mail list]. u.p.. Sample containing about 7660 words of unpublished miscellanea (domain: leisure)

J1E [Leeds United e-mail list]. u.p.. Sample containing about 29320 words of unpublished miscellanea (domain: leisure)

J1F [Leeds United e-mail list]. u.p.. Sample containing about 28956 words of unpublished miscellanea (domain: leisure)

J1G [Leeds United e-mail list]. u.p.. Sample containing about 44965 words of unpublished miscellanea (domain: leisure)

J1H [Leeds United e-mail list]. u.p.. Sample containing about 48669 words of unpublished miscellanea (domain: leisure)

J1J [Leeds United e-mail list]. u.p.. Sample containing about 19787 words of unpublished miscellanea (domain: leisure)


J1L Speeches by Tony Hall. Hall, Tony, u.p.. Sample containing about 8298 words of material written to be spoken (domain: social science)

J1M Scottish TV – news scripts. u.p.. Sample containing about 38983 words of material written to be spoken (domain:
12. List of works excerpted

J1N **Scottish TV – sports news scripts.** u.p.. Sample containing about 2969 words of material written to be spoken (domain: leisure)

J1P **The cost of default.** The Law Society, Great Britain (1992-10), 1-38. Sample containing about 10225 words of unpublished miscellanea (domain: commerce)


J1S **The Scottish Council for Development and Industry annual report.** The Scottish council for development & industry, Scotland (1990/1919), 1-24. Sample containing about 5092 words of unpublished miscellanea (domain: social science)


J1V **The economic significance of scotch whisky.** The Scotch whisky association, Scotland, Edinburgh (1992), 1-16. Sample containing about 5230 words of unpublished miscellanea (domain: applied science)

J1W **Scotch and tax.** The Scotch whisky association, Scotland, Edinburgh (1992), 1-13. Sample containing about 3394 words of unpublished miscellanea (domain: applied science)

J1X **Wavelength.** P & O, London (1993). Sample containing about 7539 words from a periodical (domain: social science)

J1Y **Wavelength.** P & O, London (1993). Sample containing about 9195 words from a periodical (domain: social science)

J20 **Guinness plc.** u.p.. Sample containing about 2375 words of unpublished miscellanea (domain: leisure)

J21 **United Distillers.** u.p.. Sample containing about 2798 words of unpublished miscellanea (domain: applied science)

J22 **Guinness brewing worldwide.** u.p.. Sample containing about 1636 words of unpublished miscellanea (domain: leisure)

J23 **Alcohol & society: some questions and answers.** u.p.. Sample containing about 3926 words of unpublished miscellanea (domain: applied science)

J24 **Guinness brewing worldwide.** u.p.. Sample containing about 4912 words of unpublished miscellanea (domain: applied science)

J25 **Guinness.** u.p.. Sample containing about 6954 words of unpublished miscellanea (domain: applied science)

J26 **[Prospectus].** The Henley Research Centre, Henley, 1-13. Sample containing about 4641 words of unpublished miscellanea (domain: social science)

J27 **Short courses in religious and moral education.** u.p.. Sample containing about 15302 words of unpublished miscellanea (domain: social science)


J2A **Regulations for programmes of the HCIMA professional qualifications.** HCIMA, England, London (1990), 1-33. Sample containing about 7941 words of unpublished miscellanea (domain: social science)

containing about 25943 words of miscellanea (domain: social science)

**J2C** *University of Oxford*. u.p.. Sample containing about 9335 words of unpublished miscellanea (domain: social science)

**J2D** *Research on animals in Oxford*. u.p.. Sample containing about 7597 words of unpublished miscellanea (domain: natural sciences)

**J2E** *Will we ever learn*. u.p.. Sample containing about 4436 words of unpublished miscellanea (domain: world affairs)

**J2F** *[Unpublished short stories]*. Benn, June, u.p.. Sample containing about 7616 words of unpublished miscellanea (domain: imaginative)


**J2N** *The Environment Digest*. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 12728 words from a periodical (domain: applied science)

**J2P** *The Environment Digest*. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 12095 words from a periodical (domain: applied science)

**J2R** *The Environment Digest*. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 15657 words from a periodical (domain: applied science)

**J2S** *The Environment Digest*. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 12522 words from a periodical (domain: applied science)

**J2T** *The Environment Digest*. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 14190 words from a periodical (domain: applied science)

**J2U** *The Environment Digest*. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 12943 words from a periodical (domain: applied science)

**J2V** *The Environment Digest*. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 14010 words from a periodical (domain: applied science)

**J2W** *The Environment Digest*. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 14600 words from a periodical (domain: applied science)

**J2X** *The Environment Digest*. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 13628 words from a periodical (domain: applied science)

**J2Y** *The Environment Digest*. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 18345 words from a periodical (domain: applied science)

| J31 | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 12845 words from a periodical (domain: applied science) |
| J32 | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 13544 words from a periodical (domain: applied science) |
| J33 | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 13749 words from a periodical (domain: applied science) |
| J34 | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 13314 words from a periodical (domain: applied science) |
| J35 | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 13947 words from a periodical (domain: applied science) |
| J36 | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 13984 words from a periodical (domain: applied science) |
| J37 | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 18308 words from a periodical (domain: applied science) |
| J38 | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 14145 words from a periodical (domain: applied science) |
| J39 | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 12051 words from a periodical (domain: applied science) |
| J3A | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 12452 words from a periodical (domain: applied science) |
| J3B | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 13954 words from a periodical (domain: applied science) |
| J3C | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 14172 words from a periodical (domain: applied science) |
| J3D | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 13949 words from a periodical (domain: applied science) |
| J3E | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 12090 words from a periodical (domain: applied science) |
| J3F | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 14306 words from a periodical (domain: applied science) |
| J3G | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 13586 words from a periodical (domain: applied science) |
| J3H | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 14433 words from a periodical (domain: applied science) |
| J3J | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 17390 words from a periodical (domain: applied science) |
| J3K | The Environment Digest. Wright, Martin (ed.), Environment Publishers Ltd. Sample containing about 11807 words from a periodical (domain: applied science) |
12. List of works excerpted

Sample containing about 10452 words from a periodical (domain: applied science)

J3M Garrick Society annual general meeting.
P5M7 ‘Keith’, bbc employee, AB, male
P5M8 24, ‘Tracy’, DE, female
P5M9 ‘David’, DE, male
P5MA ‘Graham’, male
P5MB ‘Joyce’, female
P5MC ‘Pam’, female
P5MD ‘Linda’, female
P5ME ‘Bert’, male
P5MF ‘Kenneth’, male
Sample containing about 5879 words speech recorded in business context

J3N St Helens Aquarist Society annual general meeting.
P5MG 50+, ‘Bert’, draughtsman, male
P5MH 21+, housewife, female
P5MJ 21+, ‘Emily’, housewife, female
P5MK 35, ‘Robert’, factory worker, male
P5ML ‘Bill’, retired, male
Sample containing about 6229 words speech recorded in business context

J3P Shropshire County Council Social Services Committee: committee meeting.
P5MM , male
P5MN , male
P5MP ‘Mike’, male
P5MR ‘Malcolm’, male
P5MS , female
P5MT ‘Tom’, male
P5MU , male
P5MV ‘George’, male
J3PPS000
Sample containing about 21605 words speech recorded in public context

J3R Shropshire County Council Social Services Committee: committee meeting.
P5N6 , male
P5N7 ‘Mike’, male
P5N8 ‘Mike’, male
P5NN , female
J3RPS000 J3RPS001 J3RPS002
Sample containing about 13141 words speech recorded in public context

J3S Shropshire County Council Public Protection Committee: committee meeting.
P5SNG , male
P5SNH ‘Jennifer’, female
P5SNJ , male
P5SNK ‘Dave’, male
P5SNL , male
P5SNM , male
P5SNN major, male
P5SNP ‘George’, chief fire officer, male
P5SNR colonel, male
Sample containing about 10168 words speech recorded in public context

J3T Shropshire County Council Public Protection Committee: committee meeting.
P5SNS , male
P5SNT ‘Jennifer’, female
P5SNU ‘Dave’, male
P5SNV ‘George’, chief fire officer, male
P5SNW colonel, male
J3TPS000
Sample containing about 10451 words speech recorded in public context

J3U Friends Provident Insurance training session.
P5SPO sales person, male
P5SP1 sales person, female
P5SP2 ‘Robert’, salesman, male
P5SP3 ‘Gareth’, salesman, male
P5SP4 ‘Bill’, instructor, male
P5SP5 ‘Gill’, instructor, female
P5SP6 ‘Roger’, male
P5SP7 ‘Danny’, male
Sample containing about 7627 words speech recorded in public context

J3V Friends Provident Insurance training session.
P5SP8 sales person, female
P5SP9 sales person, male
P5SPA salesman, male
P5SPB salesman, male
Sample containing about 5831 words speech recorded in public context

J3W Royal Yachting Association annual general meeting.
P5SPC ‘Rod’, racing coach, male
P5SPD ‘Peter’, male
P5SPE ‘John’, male
P5SPF ‘Chris’, male
P5SPG
P5SPH ‘Frank’, male
P5SPJ ‘Richard’, male
P5SPK ‘Bob’, male
Sample containing about 10488 words speech recorded in public context

J3X Royal Yachting Association instruction video: sailing.
P5SPN , female
P5SPP , female
P5SRR , male
P5SPS , male
P5SPT , male
PS3PU, female
PS3PV, 'Suzanne', female
PS3PW, male
PS3PX, 'John', national sailing coach, male

Sample containing about 7301 words speech recorded in business context

J3Y Gardeners' Question Time: radio programme.
PS3R1, 'Stefan', radio presenter and gardening expert, male
PS3R2, radio presenter, female
PS3R3, 'Fred', gardening expert, male
PS3R4, 'Daphne', gardening expert, female
PS3R5, 'Walter', gardening expert, male
PS3R6, 'Frank', gardening expert, male

Sample containing about 6611 words speech recorded in leisure context

J40 British Association of Industrial Editors - Edinburgh Internal Communications Group: informal meeting.
PS3R7, 32, 'Peter', ex bp public affairs, male
PS3R8, 47, 'Anne', bnc worker, female
PS3R9, 52, 'Susan', scottish office executive, female
PS3RA, 38, 'Tukuse', japanese teacher, female

Sample containing about 6642 words speech recorded in business context

J41 West Sussex Council: Strategic Planning Committee.
PS3RB, male
PS3RC, 'Harold', male
PS3RD, 'Chris', male
PS3RE, 'Alan', male
PS3RF, female
PS3RG, male
PS3RH, male
PS3RJ, male
PS3RK, 'Jim', male

Sample containing about 12888 words speech recorded in public context

J42 West Sussex Council: Strategic Planning Committee meeting.
PS3RR, male
PS3RS, 'Neil', male
PS3RT, 'Harold', male
PS3RU, male
PS3RV, 'Cliff', female
PS3RW, male
PS3RX, female
PS3RY, male

Sample containing about 6599 words speech recorded in public context

J43 West Sussex Council: Social Services Committee meeting. Sample containing about 8077 words speech recorded in public context

J44 West Sussex Council: Highways Committee meeting.
PS3S0, female

Sample containing about 9590 words speech recorded in public context

J45 West Sussex Council: Highways Committee meeting.
PS3S1, female

Sample containing about 3569 words speech recorded in public context


J59 [Letter concerning control of industrial major accident hazards]. u.p.. Sample containing about 507 words of unpublished miscellanea (domain: world affairs)

J5A Against all odds. Greenpeace. Sample containing about 672 words of miscellanea (domain: world affairs)

J5B Ecover bio-degradable household cleaning products. Ecover. Sample containing about 1048 words of mis-
12. List of works excerpted

cellanea (domain: leisure)

JSC Why Bob Geldof thinks you should read the New Internationalist. New Internationalist. Sample containing about 554 words of miscellanea (domain: world affairs)

JSD World peace. Greenpeace. ... Sample containing about 470 words of miscellanea (domain: world affairs)

JSE 1992: aren’t you sick of it?. Communist party. Sample containing about 589 words of miscellanea (domain: commerce)

JSF Thank God someone’s making waves. Greenpeace. Sample containing about 320 words of miscellanea (domain: world affairs)

JSG Do you want to walk in safety in the countryside?. u.p.. Sample containing about 607 words of miscellanea (domain: world affairs)

JSH Your Communist Euro candidate. Communist party. Sample containing about 711 words of miscellanea (domain: world affairs)

JSJ Facing the future. Communist party. Sample containing about 398 words of miscellanea (domain: world affairs)

JSK The more the Government does, the less you’re free to do. National front. u.p.. Sample containing about 141 words of miscellanea (domain: world affairs)

JSL [Fund-raising tandem trek sponsor form]. BDAFSA. Sample containing about 263 words of unpublished miscellanea (domain: world affairs)

JSM The London whale walk. Greenpeace. Sample containing about 495 words of miscellanea (domain: world affairs)

JSN ANC catalogue. African National Congress. Sample containing about 439 words of miscellanea (domain: world affairs)

JSP [Correspondence concerning property purchase]. u.p.. Sample containing about 839 words of unpublished miscellanea (domain: social science)


12. List of works excerpted

Sample containing about 13309 words from a book (domain: social science)


J78 Media law. Robertson, Geoffrey, Longman Group UK Ltd, Harlow (1990), 1-91. Sample containing about 44653 words from a book (domain: social science)


12. List of works excerpted

from a periodical (domain: arts)

**J7J** *Language and Literature Volume 1 No.*
Culpeper, Jonathan, Longman Group UK Ltd, Harlow (1992), 141-143. 2 1992. Sample containing about 1346 words from a periodical (domain: arts)

**J7K** *Language and Literature Volume 1 No.*
Bex, Tony, Longman Group UK Ltd, Harlow (1992), 143-145. 2 1992. Sample containing about 1651 words from a periodical (domain: arts)

**J7L** *Language and Literature Volume 1 No.*
Bellard-Thomson, Carol, Longman Group UK Ltd, Harlow (1992), 146-147. 2 1992. Sample containing about 1247 words from a periodical (domain: arts)

**J7M** *Language and Literature Volume 1 No.*
Semino, Elena, Longman Group UK Ltd, Harlow (1992), 147-150. 2 1992. Sample containing about 1571 words from a periodical (domain: arts)

**J7P** *Language and Literature Volume 2 No.*
Nash, Walter, Longman Group UK Ltd, Harlow (1993), 19-34. 1 Sample containing about 7135 words from a periodical (domain: arts)

**J7R** *Language and Literature Volume 2 No.*
Black, Elizabeth, Longman Group UK Ltd, Harlow (1993), 37-47. 1 1993. Sample containing about 5354 words from a periodical (domain: arts)

**J7S** *Language and Literature Volume 2 No.*

**J7T** *Language and Literature Volume 2 No.*

**J7U** *Language and Literature Volume 2 No.*
Butler, Lance St John, Longman Group UK Ltd, Harlow (1993), 70-72. 1 1993. Sample containing about 1490 words from a periodical (domain: arts)

**J7V** *Language and Literature Volume 2 No.*
McRae, John, Longman Group UK Ltd, Harlow (1993), 75-76. 1 1993. Sample containing about 753 words from a periodical (domain: arts)

**J7W** *Language and Literature Volume 2 No.*

**J7X** *Language and Literature Volume 2 No.*

**J80** *Language and Literature Volume 2 No.*
Wales, Katie, Longman Group UK Ltd, Harlow (1993), 137-139. 2 1993. Sample containing about 1226 words from a periodical (domain: arts)

**J81** *Language and Literature Volume 2 No.*
Green, Keith, Longman Group UK Ltd, Harlow (1993), 138-139. 2 1993. Sample containing about 578 words from a periodical (domain: arts)

**J82** *Language and Literature Volume 2 No.*
Cunico, Sonia, Longman Group UK Ltd, Harlow (1993), 141-143. 2 1993. Sample containing about 1613 words from a periodical (domain: arts)

**J83** *Language and Literature Volume 2 No.*

**J84** *Language and Literature Volume 2 No.*

**J85** *Language and Literature Volume 2 No.*
Clark, Billy, Longman Group UK Ltd, Harlow (1993), 151-153. 2 1993. Sample containing about 1485 words from a
12. List of works excerpted

periodical (domain: arts)


J8B *Audio Description Action Group: meeting.*

PS3S2 30+, ‘Robert’, arts officer, male
PS3S3 30+, ‘Brian’, theatre manager, male
PS3S4 50+, ‘Iris’, audio services, female
PS3S5 30+, ‘Jane’, social worker, female
PS3S6 40+, ‘Cathy’, social worker, female

Sample containing about 13750 words speech recorded in business context

J8D *Teachers’ conference: discussing tutorial programme.*

PS3S7 40+, ‘Rod’, male
J8DPS000
J8DPS001
J8DPS002
J8DPS003
J8DPS004

Sample containing about 20496 words speech recorded in educational context

J8F *Harlow Record Office - recollections of Harlow: interview.*

PS1MU, female
PS1MV 74, retired, female

Sample containing about 6769 words speech recorded in leisure context

J8G *Life in Harlow: interview by Harlow Study Centre.*

PS1N3, female
PS1N4 electrical retailer, male
PS1N5, female

Sample containing about 11906 words speech recorded in leisure context

J8J *Drugs: television discussion.*

J8JPS000
J8JPS001
J8JPS002
J8JPS003
J8JPS004
J8JPS005
J8JPS006
J8JPS007
J8JPS008

Sample containing about 4965 words speech recorded in leisure context

J8K *Bioenergetics: lecture.*

PS3SE 40+, lecturer, male

Sample containing about 6648 words speech recorded in educational context

J8M *Albert Gunter: sermon.*

PS1RE ‘Albert’, minister, male

Sample containing about 2667 words speech recorded in public context

J8Y *Albert Gunter: sermon.*

PS1RY ‘Albert Gunter’, minister, male

Sample containing about 6706 words speech recorded in public context

J90 *Albert Gunter: sermon.*

PS1S0 ‘Albert Gunter’, minister, male

Sample containing about 5418 words speech recorded in public context

J91 *Maths tutorial.*

PS3SR 50, ‘John’, tutor, male
PS3SS 16, ‘Ruben’, student, male

Sample containing about 12156 words speech recorded in educational context

J92 *University of York Student Union Entertainments Committee meeting.*

PS3ST ‘Thomas’, student, male
PS3SU ‘Andy’, student, male
PS3SV ‘Dave’, student, male
PS3SW 23, ‘Pete’, student, male
PS3SX student, male
PS3SY ‘John’, student, male
PS3ST0 19, ‘Becky’, student, female
PS3ST1 19, ‘Chris’, student, male
PS3ST2 student, male

Sample containing about 7419 words speech recorded in business context

J94 *Nottingham University, Economics Dept.*

PS3TJ lecturer, male
PS3TK ‘Adrian’, student, male
PS3TL ‘Helen’, student, female
Sample containing about 5028 words speech recorded in educational context

**J95**  *Nottingham University, Economics Dept.*
*PS3TN* lecturer, male

Sample containing about 4366 words speech recorded in educational context

**J97**  *British Market Research Bureau monthly meeting.*
*PS3U2* ‘John’, managing director, male
*PS3U3* ‘Stella’, director, female
*PS3U5* ‘Richard’, administrator, male
*PS3U6* ‘Alan’, director, male
*PS3U7* ‘Rita’, director, female
*PS3UA* ‘Caroline’, consultant, female

Sample containing about 10869 words speech recorded in business context

**J98**  *Herts County Council: committee meeting.*
*PS3UD*
*J98PS000*  
*J98PS001*

Sample containing about 425 words speech recorded in public context

**J99**  *Herts County Council: committee meeting.*
*PS3UE*
*J99PS000*  
*J99PS001*  
*J99PS002*

Sample containing about 399 words speech recorded in public context

**J9A**  *Bushey Writing Group: poetry and prose writing.*
*PS3UF* ‘Fred’, male
*PS3UG* ‘Janet’, female
*PS3UH* ‘Kahi’, female
*PS3UJ* ‘Peter’, male
*PS3UK* ‘Ann’, female
*PS3UL* ‘Cybil’, female

Sample containing about 12750 words speech recorded in educational context

**J9B**  *Lincolnshire County Council: board meeting.*
*PS3UM* ‘Graham’, chairman, All participants are work colleagues., male
*PS3UN* ‘Paul’, male
*PS3UP* ‘Weller’, councillor
*PS3UR* ‘Morley’, councillor
*PS3US* ‘Wyle’, councillor
*PS3UW* ‘Taylor’, councillor

Sample containing about 14970 words speech recorded in public context

**J9C**  *Lincolnshire County Council: board meeting.*
*PS3UY* ‘Graham’, chairman, All participants are work colleagues., male
*PS3V1* ‘Weller’, councillor
*PS3V3* ‘Wyle’, councillor
*PS3V5* ‘Noel’, male
*PS3V7* ‘Taylor’, councillor

Sample containing about 3982 words speech recorded in public context

**J9D**  *Lincolnshire: Board meeting.*
*PS3V9* chairman, All participants are work colleagues., male
*PS3VB* ‘Graham’, male
*PS3VC* ‘Barrel’, councillor, male
*PS3VH* ‘Michael’, male
*PS3VD* ‘Dodsworth’, male
*PS3VE* ‘Foster’, female
*PS3VF* ‘Golding’, female
*PS3VG* ‘Chris’, male
*PS3VH* ‘Hill’, councillor, male

Sample containing about 15137 words speech recorded in public context

**J9E**  *Lincolnshire: Board meeting.*
*PS3VU* chairman, All participants are work colleagues., male

Sample containing about 430 words speech recorded in public context

**J9F**  *Bradford Metropolitan Council: meeting.*
Sample containing about 3728 words speech recorded in public context

**J9G**  *Bradford Metropolitan Council: meeting.*
Sample containing about 6086 words speech recorded in public context

**J9H**  *Bradford Metropolitan Council: meeting.*
Sample containing about 4195 words speech recorded in public context

**J9I**  *Bradford Metropolitan Council: meeting.*
Sample containing about 5465 words speech recorded in public context

**J9J**  *Bradford Metropolitan Council: meeting.*
Sample containing about 7058 words speech recorded in public context

**J9K**  *Bradford Metropolitan Council: meeting.*
Sample containing about 6959 words speech recorded in public context

**J9L**  *Bradford Metropolitan Council: meeting.*
Sample containing about 7058 words speech recorded in public context

**J9M**  *Leicestershire County Council: council chambers.*
*PS3WK* chairman, All participants are work colleagues., male
*PS3WL* ‘Roberts’, councillor, male
*PS3WM* ‘Ryan’, councillor, male
*PS3WN* ‘Clements’, councillor
*PS3WP* ‘Parsons’, councillor
Sample containing about 17200 words speech recorded in public context

**J9N**  
*MFI interim figures presentation: press announcement.*

Sample containing about 5002 words speech recorded in business context

**J9P**  
*Charity committee meeting.*

Sample containing about 16246 words speech recorded in public context

**J9R**  
*Technical meeting.*

Sample containing about 2637 words speech recorded in public context

**J9S**  
*EIP meeting at Strensall Village Hall, day 5, evening session: public county council planning meeting.*

Sample containing about 27270 words speech recorded in public context

**J9V**  
*EIP meeting at Strensall Village Hall, day 7, morning session: public county council planning meeting.*

Sample containing about 11025 words speech recorded in public context

**J9T**  
*EIP meeting at Strensall Village Hall, day 5, morning session: public county council planning meeting.*
Sample containing about 18554 words
speech recorded in public context

**J9X**  
Job interview.  
PS40G  54, ‘Rod’, professional interviewer, male  
PS40H  46, ‘Ron’, salesman, Interviewee, male  
Sample containing about 11886 words
speech recorded in business context

**J9Y**  
Job interview.  
PS40J  54, ‘Rod’, professional interviewer, male  
PS40K  46, ‘Ron’, salesman, Interviewee, male  
Sample containing about 10917 words
speech recorded in business context

**JA0**  
Interview.  
PS40L  54, ‘Rod’, professional interviewer, male  
PS40M  33, ‘Patrick’, salesperson, Interviewee, male  
Sample containing about 9141 words
speech recorded in business context

**JA1**  
Interview.  
PS40N  54, ‘Rod’, professional interviewer, male  
PS40O  38, ‘Lynda’, salesperson, Interviewee, female  
Sample containing about 8150 words
speech recorded in business context

**JA2**  
Interview.  
PS40P  54, ‘Rod’, professional interviewer, male  
PS40Q  53, ‘Trevor’, salesperson, Interviewee, male  
Sample containing about 10857 words
speech recorded in business context

**JA3**  
Interview.  
PS40R  54, ‘Rod’, professional interviewer, male  
PS40S  46, ‘Richard’, salesperson, Interviewee, male  
Sample containing about 7970 words
speech recorded in business context

**JA4**  
Interview.  
PS40T  54, ‘Rod’, professional interviewer, male  
PS40U  46, ‘Richard’, salesperson, Interviewee, male  
Sample containing about 13568 words
speech recorded in business context

**JA5**  
Southwell Parish Council: meeting of parish council for small town.  
PS40X  30+, ‘Lynn’, Chairperson, female  
PS40Y  30+, ‘Steve’, Clerk, male  
PS410  45+, ‘David’, male  
PS411  45+, ‘Roger’, businessman, male  
PS412  45+, ‘Christine’, female  
PS413  60+, ‘Madeleine’, female  
PS414  50+, ‘Stuart’, male  
PS415  50+, ‘Peter’, male  
Sample containing about 9639 words
speech recorded in public context

**JA6**  
Meeting.  
PS418  50+, ‘Tom’, male  
PS419  25+, ‘Adrian’, male  
PS41A  60+, ‘David’, male  
Sample containing about 4907 words
speech recorded in business context

**JA8**  
French conversation class.  
PS41E  45, ‘Mrs J Long’, schoolteacher, female  
PS41F  9, ‘Phillip’, school pupil, male  
PS41G  9, ‘Kieran’, school pupil, male  
PS41H  9, ‘Richard’, school pupil, male  
PS41J  9, ‘Christopher’, school pupil, male  
PS41K  9, ‘Katrina’, school pupil, female  
PS41L  9, ‘Josephine’, school pupil, female  
PS41M  9, ‘Stacey’, school pupil, female  
PS41N  9, ‘Lee’, school pupil, male  
Sample containing about 2373 words
speech recorded in educational context

**JA9**  
Local Government Ombudsman office: planning meeting.  
PS41Y  30+, ‘Anita’, team leader, female  
PS420  26, ‘Graham’, assistant administrator, male  
PS421  30+, ‘Rob’, team leader, male  
PS422  40+, ‘Richard’, team leader, male  
PS423  50+, transport organizer, female  
PS424  50+, ‘David’, administrator, Chairing meeting, male  
PS425  40+, ‘Marlene’, team leader, female  
PS426  20+, ‘Karen’, female  
Sample containing about 14128 words
speech recorded in business context

**JAA**  
Science demonstration lesson.  
PS427  34, ‘Mr H A Redfern’, schoolteacher, male  
PS428  40+, ‘Mrs Trinder’, schoolteacher, or non-teaching assistant, female  
PS429  9, ‘Chris’, school pupil, male  
PS42A  9, ‘Louise’, school pupil, female  
PS42B  9, ‘Jennifer’, school pupil, female  
PS42C  9, ‘Phillip’, school pupil, male  
PS42D  9, ‘Jody’, school pupil, female  
PS42E  9, ‘Alan’, school pupil, male  
PS42F  9, ‘Steven’, school pupil, male  
Sample containing about 4687 words
speech recorded in educational context

**JAB**  
Classroom history lead session.  
PS42V  45+, ‘P Lynch’, schoolteacher, female  
PS42W  9, ‘Kieran’, school pupil, male  
PS42X  9, ‘Daisy’, school pupil, female  
PS42Y  9, ‘Jennifer’, school pupil, female  
PS430  9, ‘Katrina’, school pupil, female  
PS431  9, ‘Philip’, school pupil, male  
PS432  9, ‘Adam’, school pupil, male  
PS433  51, ‘Mr Rose’, retired music teacher, Making recording, male  
Sample containing about 2235 words
speech recorded in educational context
JAC  
EIP meeting at Strensall, day 6, morning session: public county council planning meeting.

PS434 50+, ‘Mr E Barnett’, department of the environment adjudicator, male
PS435 50+, ‘Mr Ken Williamson’, legal representative, for North Yorkshire County Council, male
PS436 40+, ‘Mr David Potter’, legal representative, for North Yorkshire County Council, male
PS437 40+, ‘Miss D Whittaker’, department of the environment senior inspector, female
PS438 50+, ‘Mr E Barnett’, department of the environment adjudicator, male
PS439 30+, ‘Mr David Curtis’, legal representative, for York City Council, male
PS440 35+, ‘Mr Les Saunders’, legal representative, for Department of the Environment, male
PS441 40+, ‘Professor David Lock’, legal representative, for David Lock Associates, male
PS442 20+, ‘Miss Fiona Firth’, legal representative, for Montague Evans, female
PS443 30+, ‘Mr Dave Girt’, legal representative, for Leeds City Council, male

Sample containing about 10866 words speech recorded in public context

JAD  
EIP meeting at Strensall, day 6, afternoon session: public county council planning meeting.

PS43K 50+, ‘Mr E Barnett’, department of the environment adjudicator, male
PS43L 30+, ‘Mr Joe Cunnane’, legal representative, for J C Cunnane Associates, male
PS43M 40+, ‘Professor David Lock’, legal representative, for David Lock Associates, male
PS43N 40+, ‘Miss D Whittaker’, department of the environment senior inspector, female
PS43P 40+, ‘Mr David Potter’, legal representative, for North Yorkshire County Council, male
PS43Q 30+, ‘Mr Terry Heselton’, legal representative, for Selby District Council, male
PS43R 30+, ‘Mr David Allenby’, legal representative, for Harrogate District Council, male
PS43S 40+, ‘Mr Laycock’, legal representative, for Harrogate Civic Society, male
PS43U 40+, ‘Mr Steel’, legal representative, for Richmondshire District Council, male

Sample containing about 21570 words speech recorded in public context

JAE  
Interview.

PS44S 54, ‘Rod’, interviewer, male
PS44T 50, ‘Maggie’, salesperson, Interviewee, female

Sample containing about 5819 words speech recorded in business context

JJ6  
Nottingham University: lecture.

PS44V 50, ‘Maggie’, salesperson, Interviewee, female

Sample containing about 7834 words speech recorded in public context

JHE  
Bradford Metropolitan Council: debate.

Sample containing about 8346 words speech recorded in public context

AJB  
Hertfordshire County Council: Environment Committee meeting.

PS44Y 50, ‘Mr Terry Heselton’, legal representative, for Selby District Council, male
PS44Z 50, ‘Mr David Allenby’, legal representative, for Harrogate District Council, male
PS450 50, ‘Mr Steel’, legal representative, for Richmondshire District Council, male

Sample containing about 9455 words speech recorded in public context

JJC  
Estate agency: interview.

PS45K 54, ‘Rod’, interviewer, male
PS45L 50, ‘Maggie’, salesperson, Interviewee, female

Sample containing about 1066 words speech recorded in business context

JJD  
Bradford Metropolitan Council: debate.

Sample containing about 8346 words speech recorded in public context

JJE  
Bradford Metropolitan Council: debate.

Sample containing about 7834 words speech recorded in public context
JJF Bradford Metropolitan Council: debate. Sample containing about 3960 words speech recorded in public context

JJG Wiltshire County Council: committee meeting.
PS45P ‘Cole’, no further information provided, male
PS45R no further information provided, male
PS45S ‘Coleman’, no further information provided, male
PS45T ‘Small’, no further information provided, male
PS45U ‘Chalk’, no further information provided, male
PS45V ‘Pager’, no further information provided, male
PS45W ‘Okinson’, no further information provided, male
PS45X ‘Caplan’, no further information provided, male
PS45Y no further information provided, male
Sample containing about 19668 words speech recorded in public context

JHH Tarmac Construction Ltd training session.
PS46G further information not given, female
Sample containing about 14851 words speech recorded in business context

JJJ Bristol University: training session.
PS46H no further respondent information given
Sample containing about 70 words speech recorded in educational context

JJL Student seminar on absolute egalitarianism in China.
PS46K ‘Philip’, male
Sample containing about 15919 words speech recorded in educational context

JLN Bristol University history department: seminar.
PS46M no further information given
JJNPS000 Sample containing about 14229 words speech recorded in educational context

JJP Friends Provident Insurance: training session.
PS46N saleswoman, female
PS46R salesman, male
PS46T ‘Martin’, instructor, male
PS46U ‘Maggie’, instructor, female
Sample containing about 11473 words speech recorded in business context

JJR Bacons College: lesson.
PS46V lecturer, no further information given
PS46W pupil, no further information given, male
PS46X ‘Louisa’, pupil, female
PS46Y pupil, no further information given, female
PS470 pupil, no further information given, female
PS471 ‘Tisha’, pupil, no further information given, female
PS472 ‘Daniel’, pupil, male
PS473 ‘Stuart’, pupil, no further information given, male
PS474 ‘Clair’, pupil, female
Sample containing about 2954 words speech recorded in educational context

JJS Bacons College: lesson.
PS47D lecturer, no further information given, male
PS47E ‘Berkam’, pupil, male
PS47F teacher, no further information given, female
PS47G no further information given, female
PS47H ‘David’, pupil, no further information given, male
PS47J pupil, female
PS47K teacher, male
PS47L pupil
PS47M teacher
Sample containing about 10288 words speech recorded in educational context

JJT Whiteside v. Sample containing about 12387 words speech recorded in public context

JJU Whiteside v.
PS47X ‘Hidden’, judge
PS47Y barrister, AB, male
Sample containing about 13120 words speech recorded in public context

JJV O’Halloran v.
PS480 ‘Drake’, judge, AB, male
PS481 ‘Cox’, female
PS482 ‘Bate’, male
PS483 ‘Elaine’, female
PS484 ‘Clive’, police inspector, male
PS485 ‘Michael’, detective constable, male
PS486 ‘Denser’, male
PS487 ‘Michael’, police superintendent, male
Sample containing about 12295 words speech recorded in public context

JJW O’Halloran v.
PS488 ‘Michael’, police superintendent, male
PS489 ‘Drake’, judge, AB, male
PS48A ‘Denser’, barrister, male
PS48B ‘Isles’, police sergeant, male
PS48C ‘John’, police sergeant, male
Sample containing about 11531 words speech recorded in public context

JJX Hearing.
PS48D judge, no further information, male
PS48E counsel, no further information given
PS48F  no further information given, male
Sample containing about 6396 words
speech recorded in public context

JYY  High Court of Justice: hearing.
PS48G  no further respondent information given
JYYPS000
JYYPS001
Sample containing about 4963 words
speech recorded in public context

JK0  Royal Courts: hearing.
PS48H  no further information given
JK0PS000
JK0PS001
JK0PS002
JK0PS003
Sample containing about 13037 words
speech recorded in public context

JK1  Scientific research - the whole truth?: seminar.
PS48L  60, ‘Rolf’, student, male
PS48P  ‘Colin’, tutor, male
Sample containing about 8802 words
speech recorded in educational context

JK2  The anthem of nations: seminar.
PS48R  ‘Colin’, tutor, male
Sample containing about 7104 words
speech recorded in educational context

JK5  Trinity School: classroom interaction.
PS495  ‘Dexter’, teacher, male
PS496  ‘James’, AB, male
PS497  ‘Graham’, student, male
PS498  ‘Martin’, student, female
PS499  ‘Paul’, student, male
PS49A  ‘Craig’, student, male
PS49B  ‘Lisa’, student, female
PS49C  ‘Stephen’, student, male
PS49D  ‘Daniel’, student, male
Sample containing about 10998 words
speech recorded in educational context

JK6  Classroom interaction.
PS49U  ‘Rachael’, pupil, female
PS49V  ‘Evelyn’, pupil, female
PS49W  ‘Claire’, pupil, female
PS49X  ‘Catherine’, pupil, female
PS49Y  ‘Lawrence’, pupil, male
PS4A0  ‘Nicholas’, pupil, male
PS4A1  ‘Michael’, pupil, male
PS4A2  ‘Edward’, pupil, male
PS4A3  ‘David’, pupil, male
Sample containing about 3964 words
speech recorded in educational context

JK7  Abbey Life introduction course: training.
PS4AD  30, ‘Carol’, trainer, Quiz style session with scores recorded on black-
board.Occasional use of viewgraph projections and flip chart in demonstration answers., female
PS4AE  ‘David’, trainer, male
PS4AF  ‘John’, trainee, male
PS4AG  ‘Gerald’, trainee, male
PS4AH  ‘Sue’, trainee, female
PS4AJ  ‘Janet’, trainee, female
PS4AK  ‘June’, trainee, female
PS4AL  ‘Philip’, trainee, male
PS4AM  ‘Ilias’, trainee, male
Sample containing about 10708 words
speech recorded in business context

JK8  Abbey Life introduction course: training.
PS4AS  30, ‘Carol’, trainer, Training session includes a five minute written exercise and thirty minute video lecture., female
PS4AT  ‘David’, training sessions, male
PS4AU  ‘John’, trainee, male
PS4AV  ‘Gerald’, trainee, male
PS4AW  ‘Sue’, trainee, female
PS4AX  ‘Janet’, janet, female
PS4B0  ‘Philip’, trainee, male
PS4B1  ‘Ilias’, trainee, male
Sample containing about 9396 words
speech recorded in business context

JK9  Abbey Life introduction course: training.
PS4B7  30, ‘Carol’, trainer, Occasional use of viewgraph projections and flip chart in demonstration answers., female
PS4BB  ‘David’, training sessions, male
PS4B9  ‘John’, trainee, male
PS4BA  ‘Gerald’, trainee, male
PS4BC  ‘Sue’, trainee, female
PS4BE  ‘Janet’, janet, female
PS4BF  ‘Philip’, trainee, male
PS4BG  ‘Ilias’, trainee, male
Sample containing about 10352 words
speech recorded in business context

JN6  Air UK sales meeting.
PS4BK  ‘Gary’, sales executive, AB, male
PS4BL  ‘Brett’, sales executive, DE, male
PS4BM  ‘Stuart’, sales executive, DE, male
PS4BN  ‘Mike’, sales executive, male
Sample containing about 16742 words
speech recorded in business context

JN7  Air UK sales meeting.
PS4BP  sales executive, male
PS4BR  sales executive, male
PS4BS  ‘Stuart’, sales executive, male
PS4BT  ‘Mike’, sales executive, male
PS4BU  ‘Brett’, sales executive, male
Sample containing about 6215 words
speech recorded in business context

JN8  Northern Development Company: meeting.
PS4BV  no further information given
JNBPS000

Sample containing about 1437 words speech recorded in business context

JNB  Leicestershire County Council: debate.
PS4C4  no further information given, male
PS4C5  Taylor’, no further information given, male
PS4C6  ‘Setchfield’, no further information given, female
PS4C7  ‘Beale’, no further information given, female
PS4C8  ‘Beale’, no further information given, male
PS4C9  ‘Joranpucher’, no further information given, male
PS4CA  ‘Bury’, no further information given, female
PS4CB  ‘Willmott’, no further information given, male
PS4CC  ‘Howard’, no further information given, male

Sample containing about 17325 words speech recorded in public context

JND  Training session.
PS4DB  no further information given, male
PS4DC  ‘Kathy’, no further information given, female
PS4DD  ‘Lee’, no further information given, male
PS4DE  ‘Marcus’, no further information given, male
PS4DF  ‘Greg’, no further information given, male
PS4DG  ‘Ian’, no further information given, male
PS4DH  ‘Lisa’, no further information given, female
PS4DJ  ‘Neil’, no further information given, male

Sample containing about 17325 words speech recorded in public context

JNE  High Court of Justice: hearing.
PS4DK  ‘Beaumont’, police officer, male
PS4DL  judge, male
PS4DM  ‘Bates’, lawyer, no further information given, male
PS4DN  ‘Stuart’, police officer, male
PS4DP  ‘Louis’, police officer, male
PS4DR  ‘Gary’, police officer, male
PS4DS  Brendan’, police officer, male
PS4DT  lawyer, male
PS4DU  court official, male

Sample containing about 11349 words speech recorded in public context

JNF  Save The Children: meeting.
PS4DW  ‘Bill’, news reporter, male
PS4DX  ‘Mike’, chairman, male
PS4DY  ‘Michael’, director, male
PS4E0  ‘Caroline’, no further information given, female
PS4E1  ‘Princess Anne’, princess, female
PS4E2  ‘Patricia’, actress, female
PS4E3  ‘Sally’, chair, female
PS4E4  no further information given, male
PS4E5  no further information given, male

Sample containing about 10379 words speech recorded in public context

JNG  Save The Children: meeting.
PS4EG  ‘Sally’, chair, female
PS4EH  ‘Janet’, shop manageress, female
PS4EJ  ‘Kathy’, area rep, female
PS4EK  ‘Jacquie’, shop leader, female
PS4EL  ‘Elise’, no further information given, male
PS4EM  ‘Ismar’, student, no further information given, female
PS4EN  ‘Nicholas’, director general, male
PS4EP  ‘Mike’, chair, male
PS4ER  ‘Mark’, no further information given, male

Sample containing about 10008 words speech recorded in public context

JNH  Annual general meeting.
PS4ES  no further information given
JNHPS000
JNHPS001
JNHPS002
JNHPS003
JNHPS004

Sample containing about 5265 words speech recorded in business context

JNJ  Annual general meeting.
PS4ET  no further information given
JNJPS000
JNJPS001
JNJPS002
JNJPS003
JNJPS004
JNJPS005
JNJPS006
JNJPS007

Sample containing about 5493 words speech recorded in business context

JNK  Annual general meeting.
PS4EU  ‘President’, no further information given, male
PS4EV  treasurer, no further information given
PS4EW  ‘Judy’, no further information given, female
PS4EX  ‘Samantha’, no further information given, female
PS4EY  ‘Simon’, no further information given, male

Sample containing about 13766 words speech recorded in business context

JNL  National Council for Voluntary Organisations annual general meeting. Sample containing about 5819 words speech recorded in business context
JNM  House of Commons Select Committee for Defence: meeting.
PS4F0  ‘Perry’, male
PS4F1  ‘Michael’, male
PS4F2  ‘Gordon’
PS4F3  ‘John’, male
PS4F4  ‘Nicholas’, male
Sample containing about 8317 words speech recorded in public context

JNN  House of Commons Select Committee for Defence: meeting.
PS4F5  ‘Gordon’, director general, AB, male
PS4F6  ‘White’, group captain, AB, male
PS4F7  ‘Bruce’, male
PS4F8  ‘Peter’, male
PS4F9  ‘Michael’, male
PS4FA  ‘Neville’
PS4FB  ‘Campbell’, male
PS4FC  ‘Evans’, male
PS4FD  ‘Winston’, member of parliament, AB, male
Sample containing about 10858 words speech recorded in public context

JNP  House of Commons Select Committee for Social Security: committee meeting.
PS4FF  chairman, no further information given, male
PS4FG  ‘Ken’, no further information given, male
PS4FH  ‘Malcolm’, no further information given, male
PS4FJ  ‘John’, no further information given, male
PS4FK  ‘Guy’, no further information given, female
PS4FL  ‘Ron’, no further information given, male
PS4FM  ‘Jim’, no further information given, male
PS4FN  ‘Michael’, no further information given, male
PS4FP  ‘Jane’, no further information given, female
Sample containing about 9823 words speech recorded in public context

JNR  Seminar presentation at conference.
PS4FW  50+, medical consultant, Chairperson., male
PS4FX  40+, medical consultant, Presenting seminar, male
PS4FY  medical consultant, Delegate, male
PS4G0  medical consultant, Delegate
PS4G1  medical consultant, Delegate, male
PS4G2  medical consultant, Delegate, male
Sample containing about 1657 words speech recorded in educational context

JNS  Consultation with solicitor.
PS4G3  45, ‘Neville’, solicitor, male
PS4G4  40, ‘Geoffrey’, unemployed, Client, male
Sample containing about 4763 words speech recorded in business context

JNT  Consultation with solicitor.
PS4G5  45, ‘Neville’, solicitor, male
PS4G6  20, ‘Sarah’, Client and main speaker., female
PS4G7  50+, Speaks very little., female
Sample containing about 1504 words speech recorded in business context

JNU  Consultation with solicitor.
PS4G8  45, ‘Neville’, solicitor, male
PS4G9  30, Client, male
Sample containing about 1353 words speech recorded in business context

JNY  Sales pitch.
PS4GE  51, ‘John’, retired music teacher., Householder, male
PS4GF  51, ‘Gill’, music teacher, Householder, female
PS4GG  35, ‘J Marsden’, salesperson, male
Sample containing about 7085 words speech recorded in business context

JNY  Sales pitch.
PS4GH  51, ‘John’, retired music teacher, Householder, male
PS4GJ  25+, ‘Markus’, salesperson, male
PS4GK  51, ‘Gill’, music teacher, Householder, female
Sample containing about 2811 words speech recorded in business context

JP0  Meeting.
PS4GL  ‘Mr Poole’, male
PS4GM  ‘Mr Fremantle’, male
PS4GN  ‘Jane Warner’, female
PS4GP  ‘Mr Ingall’, male
PS4GR  ‘Mr Rowan’, male
Sample containing about 13241 words speech recorded in business context

JP1  Meeting.
PS4GS  ‘John’, manager, male
PS4GT  ‘Shane’, manager, male
Sample containing about 5699 words speech recorded in business context

JP2  Meeting.
PS4GU  ‘S Williams’, male
PS4GV  ‘T Fremantle’, male
Sample containing about 587 words speech recorded in business context

**JP3**  Meeting.
*PS4GW* ‘John’, manager, male  
*PS4GX* ‘Shane’, manager, male

Sample containing about 2428 words speech recorded in business context

**JP4**  Tutorial lesson.
*PS4GY* 50, ‘John’, tutor, male  
*PS4H0* 18, ‘Sara’, nursery nurse, female

Sample containing about 11185 words speech recorded in educational context

**JP5**  Tutorial lesson.
*PS4H1* 50, ‘John’, tutor, male  
*PS4H2* 16, ‘Ruben’, school student, male

Sample containing about 3304 words speech recorded in educational context

**JP6**  Computers lecture.
*PS4H3* 35+, lecturer, male

Sample containing about 5359 words speech recorded in educational context

**JP7**  York Green Party business meeting: local politics.
*PS4H4* 32, ‘Helen’, Secretary, female  
*PS4H5* 28, ‘Andy’, Newsletter editor, male  
*PS4H6* 26, ‘Ste (yes, Ste)’, Coordinator, male  
*PS4H7* 30+, ‘Bob’, male  
*PS4H8* 40+, ‘Andy’, male

Sample containing about 13388 words speech recorded in business context

**JP8**  Training session.
*PS4HA*  lecturer, no further information given

Sample containing about 2721 words speech recorded in educational context

**JS7**  Herts County Council Environment Committee meeting: debate.
*PS4HB* councillor, DE, male  
*PS4HC* ‘Charles’, councillor, male  
*JS7PS000*  
*JS7PS001*  
*JS7PS002*  
*JS7PS003*  
*JS7PS004*  
*JS7PS005*  
*JS7PS006*

Sample containing about 13359 words speech recorded in public context

**JS9**  Wiltshire County Council: committee meeting.
*PS4JA* ‘Coleman’, male  
*PS4JB* ‘Small’, male  
*PS4JC* ‘Chalk’, male  
*PS4JD* ‘Okinson’, male  
*JS9PS000*  
*JS9PS001*  
*JS9PS002*  
*JS9PS003*  
*JS9PS004*

Sample containing about 21009 words speech recorded in public context

**JSA**  Tarmac Construction: training session.
*PS4JX* trainer, male  
*PS4JY* ‘Bob’, male  
*PS4K0* ‘Tony’, male  
*PS4K1* ‘Jeff’, male  
*PS4K2* ‘Mike’, male  
*PS4K3* ‘Sarah’, female  
*PS4K4* ‘Rob’, male  
*PS4K5* ‘Gordon’, male  
*PS4K6* ‘John’, male

Sample containing about 19412 words speech recorded in public context

**JSC**  Royal Courts: hearing.
*PS4KC* judge, no further information given  
*PS4KD* solicitor, no further information given  
*PS4KF* usher, no further information given  
*PS4KG* solicitor, no further information given

Sample containing about 8101 words speech recorded in public context

**JSD**  Coopers and Lybrand: seminar.
*PS4KH* no further information given

Sample containing about 4607 words speech recorded in business context

**JSE**  .
*PS4KJ* ‘Andrew’, no further information given, male  
*PS4KK* no further information given  
*PS4KL* no further information given  
*PS4KM* no further information given, female  
*PS4KN* ‘Paula’, no further information given, female  
*PS4KP* ‘Alex’, no further information given, male  
*PS4KR* ‘Tony’, no further information given, male

Sainsbury: seminar. Sample containing about 3758 words speech recorded in educational context
12. List of works excerpted

**JSF** House of Commons debate.
- PS4KS ‘Betty’, speaker of the house of commons, AB, female
- PS4KT ‘Riddock’, member of parliament, male
- PS4KU ‘Winnick’, member of parliament, male
- PS4KV ‘Mead’, member of parliament, female
- PS4KW ‘Gilfore’, member of parliament, female
- PS4KX ‘Wilson’, member of parliament, male
- PS4KY ‘Banks’, member of parliament, male
- PS4LJ ‘Mickey’, member of parliament, male

Sample containing about 21868 words
speech recorded in public context

**JSG** House of Commons debate.
- PS4LA deputy speaker of the house of commons, female
- PS4LC ‘Taylor’, member of parliament, male
- PS4LD ‘Jenkins’, member of parliament, male
- PS4LE ‘Shepherd’, member of parliament, male
- PS4LF ‘Taylor’, member of parliament, male
- PS4LG ‘Holm’, member of parliament, male
- PS4LH ‘Morgan’, member of parliament, male
- PS4LI ‘Betty’, speaker of the house of commons, female

Sample containing about 20140 words
speech recorded in public context

**JSH** House of Commons debate.
- PS4LP ‘Betty’, speaker of the house of commons, female
- PS4LR ‘Gillan’, member of parliament, female
- PS4LS ‘Jones’, member of parliament, male
- PS4LT ‘Heal’, member of parliament, male
- PS4LU ‘Lloyd’, member of parliament, male
- PS4LV ‘James’, member of parliament, male
- PS4LW ‘Palsey’, member of parliament, male
- PS4LX ‘Savers’, member of parliament, male
- PS4LY ‘Brazier’, member of parliament, male

Sample containing about 12000 words
speech recorded in public context

**JSJ** House of Lords: debate.
- PS4N7 ‘Montgomery’, member of parliament, male
- PS4N8 ‘Maloy’, member of parliament, male
- PS4N9 ‘Payton’, male
- PS4NA ‘Macintosh’, member of parliament, male

Sample containing about 18302 words
speech recorded in public context

**JSK** Durham University: politics lecture.
Sample containing about 9750 words
speech recorded in educational context

**JSL** Durham University: politics lecture.
Sample containing about 6183 words
speech recorded in educational context

**JSM** Durham University: politics lecture.
Sample containing about 6233 words
speech recorded in educational context

**JSN** TNT Express: Team focus training session.
- PS4NE 58, ‘Peter’, trainer, male
- PS4NF 30, ‘Whitwerth’, female
- PS4NG 30, ‘Mary’, telesales person, female
- PS4NH 30, ‘Mark’, sales executive, male
- PS4NJ 30, ‘Jack’, sales executive, male
- PS4NK 30, ‘Jane’, telesales person, female
- PS4NL 30, ‘Linda’, telesales person, female
- PS4NM 30, ‘Dave’, telesales person, male
- PS4NN 30, ‘Matt’, sales executive

Sample containing about 17206 words
speech recorded in business context

**JST** St Dominic’s: training session.
- PS4R4 teacher, male
- PS4R5 , female
- PS4R6 ‘Claire’, female
- PS4R7 ‘Catherine’, female
- PS4R8 ‘Michael’, male
- PS4R9 ‘Anne’, no further respondent information given, female

Sample containing about 6304 words
speech recorded in educational context

**JSU** St Dominic’s: training session.
- PS4RA teacher, no further information given, female
- PS4RB , female
- PS4RC ‘Michelle’, student, female
- PS4RD ‘Elizabeth’, student, female
- PS4RE ‘Martin’, student, male
- PS4RF ‘Lucy’, student, female
- PS4RG ‘Stacey’, student, female
- PS4RH ‘Laura’, student, female
- PS4RJ ‘Cassandra’, student, female

Sample containing about 4696 words
speech recorded in educational context

**JSV** St Dominic’s: training session.
- PS4RK teacher, no further information given, male
- PS4RL ‘Alison’, student, female
- PS4RM student, female
- PS4RN ‘Francesca’, student, female
- PS4RP ‘Richard’, student, male
- PS4RS ‘Margaret’, student, female

Sample containing about 4197 words
speech recorded in educational context

**JSY** Lecture on statistics.
- PS4SF 31, ‘Barker’, lecturer, male

Sample containing about 10568 words
speech recorded in educational context

**JTO** Aston University: social science lecture.
- PS4SG 31, ‘Barker’, lecturer, (background chatter, with, male

Sample containing about 9040 words
speech recorded in educational context
JT1 Aston University: social science lecture.
PS4SH 31, ‘Barker’, lecturer, no further respondent information given, male

Sample containing about 10147 words speech recorded in educational context

JT2 John Ruskin Sixth Form College: lecture.
PS4SJ tutor, male
PS4SK ‘Jim’, sound recordist, male
PS4SL ‘Darren’, student, male
PS4SM ‘Tanya’, student, female
PS4SN ‘Zakia’, student, male
PS4SP ‘Sam’, student, female
PS4SR ‘Ben’, student, male
PS4SS ‘Anthony’, student, male
PS4ST ‘Nihal’, student, male

Sample containing about 8653 words speech recorded in educational context

JT3 Abbey Life: training session.
PS4SX ‘Graham’, business trainer, male
PS4SY ‘Mike’, insurance representative, male
PS4T0 ‘David’, male
JT3PS000
JT3PS001

Sample containing about 13318 words speech recorded in business context

JT4 John Ruskin Sixth Form College: lecture.
PS4T1 tutor, male
PS4T2 ‘Martin’, sound recordist, male
PS4T3 ‘JO’, student, female
PS4T4 ‘Vicky’, student, female
PS4T5 ‘Lynn’, student, female
PS4T6 ‘Darren’, student, male
PS4T7 ‘David’, student, male
PS4T8 ‘Assad’, student, male
PS4T9 ‘Hannah’, student, female

Sample containing about 11919 words speech recorded in business context

JT5 Abbey Life: training session. Sample containing about 10918 words speech recorded in educational context

JT7 [Local Council meeting].
PS4TK ‘Evans’, councillor, male
PS4TL ‘Rosensteel’, councillor, male
PS4TM ‘Morgan’, councillor, female
PS4TN head of council, male
PS4TP ‘Hoskins’, councillor, female
PS4TR ‘Edwards’, councillor, male
PS4TS ‘Jones’, councillor, male
PS4TT ‘Hilliard’, councillor, male
PS4TU ‘Coleman’, councillor, female

Sample containing about 12104 words speech recorded in public context

JT8 [Local Council meeting].
PS4U2 head of council, male
PS4U3 ‘Price’, councillor, male
PS4U4 ‘Evans’, councillor, female
PS4U5 ‘Thomson’, councillor, female
PS4U6 ‘Howarth’, councillor, male
PS4U7 ‘Gloucester’, councillor, female
PS4U8 ‘Smith’, councillor, male
PS4U9 ‘Nevis’, councillor, male
PS4UA ‘Evans’, councillor, female

Sample containing about 12282 words speech recorded in public context

JT9 Meeting at Southwell Racecourse.
PS4UJ 30, ‘Wayne’, general manager of racecourse, male
PS4UK 40+, ‘Mike’, steward at racecourse, male
PS4UL 28, ‘Mandy’, secretary at racecourse, female

Sample containing about 1975 words speech recorded in business context

JTA Managerial Meeting.
PS4UN manager, male
PS4UP ‘Russell’, manager, male

Sample containing about 10958 words speech recorded in business context

JTB Managerial Meeting.
PS4UR 30+, ‘Pete?’, manager, male
PS4US 30+, male
PS4UT 30+, ‘Ray’, male
PS4UU 30+, ‘Lynne’, female
PS4UV, male
PS4UW ‘Stewart’, male
PS4UX, male
PS4UY, male
PS4UV, male

Sample containing about 11990 words speech recorded in business context

JTC Managerial Meeting.
PS4V1, male
PS4V2, male
PS4V3 ‘Ray’, male
PS4V4 30+, ‘Lynne’, female
PS4V5, male
PS4V6, male
PS4V7, male
PS4V8, male
PS4V9, male

Sample containing about 11213 words speech recorded in business context

JTD Edingley Parish Council meeting.
PS4VA 45+, ‘Arnold Bower’, councillor, Chairperson, male
PS4VB 55+, ‘Pat Beeyer’, secretary, female
PS4VC 60+, ‘John Watts’, male
PS4VD 65+, ‘Kevin Topham’, councillor, Vice chairperson, male
PS4VE 50+, ‘Irene Young’, female
PS4VF 30+, ‘Dale Ross’, male
PS4VG 45+, ‘Anne Tyler’, female
PS4VH 60+, female
PS4VJ 50+, ‘Derek Walker’, grant and development officer, male
Sample containing about 8859 words speech recorded in public context

JTE Induction of new voluntary workers at Barley Hall.

PS4VP 55+, ‘Charles’, historical consultant, In charge of Barley Hall project (loosely speaking), male
PS4VR 23, ‘Clare’, student and volunteer guide, female
PS4VS Passer-by, female
PS4VT 34, ‘Mike’, volunteer personnel coordinator, male
PS4VU ‘Linda’, shop assistant, female
PS4VV 60+, ‘Elsie’, volunteer guide, female
PS4VW prospective volunteer, female
PS4VX prospective volunteer, male
PS4VY , female
Sample containing about 14696 words speech recorded in public context

JTF Worsley Liberal Democrats: annual general meeting.

PS4W1 55+, ‘Mr A. Kerry Holt’, Chairperson, male
PS4W2 61, ‘Mr Bob Boyd’, Secretary, male
PS4W3 50+, ‘Mr Peter Brown’, Treasurer, male
PS4W4 75+, ‘Mr Arthur Smith’, President, male
PS4W5 ‘Mr Aled Owen’, Question from audience, male
PS4W6 50+, Question from audience, male
PS4W7 70+, ‘Mr Tom Hilton’, Membership secretary, male
PS4W8 50+, Question from audience, male
PS4W9 60+, ‘John’, Question from audience, male
Sample containing about 6481 words speech recorded in public context

JWA Hertfordshire County Council: meeting.

PS4WB council chairman, AB, male
PS4WC ‘Christopher’, councillor, DE, male
PS4WD ‘Jarvis’, councillor, DE, male
PS4WE councillor, male
PS4WF ‘John’, councillor, male
PS4WG ‘Martin’, councillor, male
PS4WH ‘Bob’, labour councillor, male
PS4WJ ‘Mr Eldrum’, councillor, male
PS4WK ‘Pauline’, conservative councillor, female
Sample containing about 13254 words speech recorded in public context


JXP New Internationalist [advertising leaflet]. New Internationalist. Sample containing about 695 words from a periodical (domain: world affairs)


12. List of works excerpted

Sample containing about 58490 words from a book (domain: imaginative)


JY1 His woman. Steele, Jessica, Mills & Boon, Richmond, Surrey (1991). Sample containing about 45851 words from a book (domain: imaginative)


JY6 One shining summer. Wilder, Quinn, Mills & Boon, Richmond, Surrey (1993). Sample containing about 52001 words from a book (domain: imaginative)


JYF West of Bohemia. Steele, Jessica, Mills & Boon, Richmond, Surrey (1993). Sample containing about 60036 words from a book (domain: imaginative)

JYL Tarmac Construction Ltd: training session.

PS6N1 lecturer, AB, male
PS6N2 trainee, DE, male
PS6N3 trainee, DE, male
PS6N4 ‘Alan’, trainee, male
12. List of works excerpted

PS6N5 'Terry', trainee, male
PS6N6 'Chris', trainee, male
PS6N7 'Aveline', trainee, female
PS6N8 trainee, male
PS6N9 'Tom', trainee, male

Sample containing about 12166 words
speech recorded in business context

JYM TNT training session: seminar.
PS6ND 30, trainer, female
PS6NE 20, telesales person, female
PS6NF 20, telesales person, female
PS6NG 20, telesales person, female
PS6NH 20, telesales person, female
PS6NJ 20, telesales person, female
PS6NK 20, telesales person, female
PS6NL 20, 'Turner', telesales person, female

Sample containing about 11849 words
speech recorded in business context

JYN Aston University psychology department: lecture.
PS6NM 31, psychology lecturer, male
PS6NN 20+, student, talks about gay and lesbian
issues, male
PS6NP 20+, student, talks about work problems,
male
PS6NR 20+, student, talks about the dreams of
blind people, male
PS6NS 20+, student, talks about questionnaires,
female
PS6NT 19, student, female

Sample containing about 18744 words
speech recorded in educational context

K1B [Central television news scripts]. u.p..
Sample containing about 43747 words
of material written to be spoken (domain: leisure)

K1C [Central television news scripts]. u.p..
Sample containing about 45382 words
of material written to be spoken (domain: leisure)

K1D [Central television news scripts]. u.p..
Sample containing about 37393 words
of material written to be spoken (domain: leisure)

K1E [Central television news scripts]. u.p..
Sample containing about 46525 words
of material written to be spoken (domain: leisure)

K1F [Central television news scripts]. u.p..
Sample containing about 40382 words
of material written to be spoken (domain: leisure)

K1G [Central television news scripts]. u.p..
Sample containing about 45420 words
of material written to be spoken (domain: leisure)

K1H [Central television news scripts]. u.p..
Sample containing about 40364 words
of material written to be spoken (domain: leisure)

K1J [Central television news scripts]. u.p..
Sample containing about 43372 words
of material written to be spoken (domain: leisure)

K1K [Central television news scripts]. u.p..
Sample containing about 42987 words
of material written to be spoken (domain: leisure)

K1L [Central television news scripts]. u.p..
Sample containing about 42655 words
of material written to be spoken (domain: leisure)

K1M [Central television news scripts].
Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about
44443 words of material written to be spoken (domain: leisure)

K1N [Central television news scripts].
Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about
41718 words of material written to be spoken (domain: leisure)

K1P [Central television news scripts].
Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about
41985 words of material written to be spoken (domain: leisure)

K1R [Central television news scripts].
Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about
45357 words of material written to be spoken (domain: leisure)

K1S [Central television news scripts].
Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about
44870 words of material written to be spoken (domain: leisure)

K1T [Central television news scripts].
Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about
42076 words of material written to be spoken (domain: leisure)
K1U [Central television news scripts]. Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about 42638 words of material written to be spoken (domain: leisure)

K1V [Central television news scripts]. Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about 43882 words of material written to be spoken (domain: leisure)

K1W [Central television news scripts]. Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about 41078 words of material written to be spoken (domain: leisure)

K1X [Central television news scripts]. Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about 43950 words of material written to be spoken (domain: leisure)

K1Y [Central television news scripts]. Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about 46520 words of material written to be spoken (domain: leisure)

K20 [Central television news scripts]. Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about 41783 words of material written to be spoken (domain: leisure)

K21 [Central television news scripts]. Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about 43137 words of material written to be spoken (domain: leisure)

K22 [Central television news scripts]. Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about 43169 words of material written to be spoken (domain: leisure)

K23 [Central television news scripts]. Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about 46596 words of material written to be spoken (domain: leisure)

K24 [Central television news scripts]. Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about 44589 words of material written to be spoken (domain: leisure)

K25 [Central television news scripts]. Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about 48680 words of material written to be spoken (domain: leisure)

K26 [Central television news scripts]. Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about 43306 words of material written to be spoken (domain: leisure)

K27 [Central television news scripts]. Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about 44477 words of material written to be spoken (domain: leisure)

K28 [Central television news scripts]. Boileau, John (ed.), Central TV, Abingdon (1993). Sample containing about 7456 words of material written to be spoken (domain: leisure)

K29 Belfast Telegraph: Applied Science pages. u.p.. Sample containing about 2055 words from a periodical (domain: applied science)

K2A Belfast Telegraph: Arts section. u.p.. Sample containing about 8047 words from a periodical (domain: arts)

K2B Belfast Telegraph: Religious affairs stories. u.p.. Sample containing about 2180 words from a periodical (domain: belief and thought)

K2C Belfast Telegraph: Business section. u.p.. Sample containing about 3546 words from a periodical (domain: commerce)

K2D Belfast Telegraph: Leisure pages. u.p.. Sample containing about 74848 words from a periodical (domain: leisure)

K2E Belfast Telegraph: News and features. u.p.. Sample containing about 5685 words from a periodical (domain: social science)

K2F Belfast Telegraph: Foreign news pages. u.p.. Sample containing about 33077 words from a periodical (domain: world affairs)
12. List of works excerpted

K2G Belfast Telegraph: Applied Science pages. u.p.. Sample containing about 1955 words from a periodical (domain: applied science)

K2H Belfast Telegraph: Arts section. u.p.. Sample containing about 2067 words from a periodical (domain: arts)

K2J Belfast Telegraph: Religious affairs stories. u.p.. Sample containing about 422 words from a periodical (domain: belief and thought)

K2K Belfast Telegraph: Business section. u.p.. Sample containing about 3962 words from a periodical (domain: commerce)

K2L Belfast Telegraph: Leisure pages. u.p.. Sample containing about 18678 words from a periodical (domain: leisure)

K2M Belfast Telegraph: News and features. u.p.. Sample containing about 1718 words from a periodical (domain: social science)

K2N Belfast Telegraph: Foreign news pages. u.p.. Sample containing about 21493 words from a periodical (domain: world affairs)

K2P Belfast Telegraph: Applied Science pages. u.p.. Sample containing about 2752 words from a periodical (domain: applied science)

K2R Belfast Telegraph: Arts section. u.p.. Sample containing about 14426 words from a periodical (domain: arts)

K2S Belfast Telegraph: Religious affairs stories. u.p.. Sample containing about 194 words from a periodical (domain: belief and thought)

K2T Belfast Telegraph: Business section. u.p.. Sample containing about 10181 words from a periodical (domain: commerce)

K2U Belfast Telegraph: Leisure pages. u.p.. Sample containing about 34980 words from a periodical (domain: leisure)

K2V Belfast Telegraph: News and features. u.p.. Sample containing about 7886 words from a periodical (domain: social science)

K2W Belfast Telegraph: Foreign news pages. u.p.. Sample containing about 35696 words from a periodical (domain: world affairs)

K2X Belfast Telegraph: Applied Science pages. u.p.. Sample containing about 2562 words from a periodical (domain: applied science)

K2Y Belfast Telegraph: Arts section. u.p.. Sample containing about 7416 words from a periodical (domain: arts)

K30 Belfast Telegraph: Religious affairs stories. u.p.. Sample containing about 172 words from a periodical (domain: belief and thought)

K31 Belfast Telegraph: Business section. u.p.. Sample containing about 5940 words from a periodical (domain: commerce)

K32 Belfast Telegraph: Leisure pages. u.p.. Sample containing about 64437 words from a periodical (domain: leisure)

K33 Belfast Telegraph: Applied Science pages. u.p.. Sample containing about 524 words from a periodical (domain: natural sciences)

K34 Belfast Telegraph: News and features. u.p.. Sample containing about 7017 words from a periodical (domain: social science)

K35 Belfast Telegraph: Foreign news pages. u.p.. Sample containing about 39397 words from a periodical (domain: world affairs)

K36 Liverpool Daily Post and Echo. u.p.. Sample containing about 685 words from a periodical (domain: applied science)

K37 Liverpool Daily Post and Echo: Arts section. u.p.. Sample containing about 8531 words from a periodical (domain: arts)

K38 Liverpool Daily Post and Echo: Religious affairs stories. u.p.. Sample containing about 1279 words from a periodical (domain: belief and thought)

K39 Liverpool Daily Post and Echo: Business section. u.p.. Sample contain-
ing about 7529 words from a periodical (domain: commerce)

K3A Liverpool Daily Post and Echo: Leisure pages. u.p.. Sample containing about 26774 words from a periodical (domain: leisure)

K3B Liverpool Daily Post and Echo: News and features. u.p.. Sample containing about 1420 words from a periodical (domain: social science)

K3C Liverpool Daily Post and Echo: Foreign news pages. u.p.. Sample containing about 38633 words from a periodical (domain: world affairs)

K3D Liverpool Daily Post and Echo: Applied Science pages. u.p.. Sample containing about 590 words from a periodical (domain: applied science)

K3E Liverpool Daily Post and Echo: Arts section. u.p.. Sample containing about 5727 words from a periodical (domain: arts)

K3F Liverpool Daily Post and Echo: Religious affairs stories. u.p.. Sample containing about 215 words from a periodical (domain: belief and thought)

K3G Liverpool Daily Post and Echo: Business section. u.p.. Sample containing about 7948 words from a periodical (domain: commerce)

K3H Liverpool Daily Post and Echo: Leisure pages. u.p.. Sample containing about 29635 words from a periodical (domain: leisure)

K3J Liverpool Daily Post and Echo: News and features. u.p.. Sample containing about 8980 words from a periodical (domain: social science)

K3K Liverpool Daily Post and Echo: Foreign news pages. u.p.. Sample containing about 36109 words from a periodical (domain: world affairs)

K3L Liverpool Daily Post and Echo: Applied Science pages. u.p.. Sample containing about 1427 words from a periodical (domain: applied science)

K3M Liverpool Daily Post and Echo: Arts section. u.p.. Sample containing about 3179 words from a periodical (domain: arts)

K3N Liverpool Daily Post and Echo: Business section. u.p.. Sample containing about 2734 words from a periodical (domain: commerce)

K3P Liverpool Daily Post and Echo: Leisure pages. u.p.. Sample containing about 13516 words from a periodical (domain: leisure)

K3R Liverpool Daily Post and Echo: Applied Science pages. u.p.. Sample containing about 150 words from a periodical (domain: natural sciences)

K3S Liverpool Daily Post and Echo: News and features. u.p.. Sample containing about 1876 words from a periodical (domain: social science)

K3T Liverpool Daily Post and Echo: Foreign news pages. u.p.. Sample containing about 23028 words from a periodical (domain: world affairs)

K3U Liverpool Daily Post and Echo: Applied Science pages. u.p.. Sample containing about 1009 words from a periodical (domain: applied science)

K3V Liverpool Daily Post and Echo: Arts section. u.p.. Sample containing about 2976 words from a periodical (domain: arts)

K3W Liverpool Daily Post and Echo: Business section. u.p.. Sample containing about 4487 words from a periodical (domain: commerce)

K3X Liverpool Daily Post and Echo: Leisure pages. u.p.. Sample containing about 29847 words from a periodical (domain: leisure)

K3Y Liverpool Daily Post and Echo: Applied Science pages. u.p.. Sample containing about 352 words from a periodical (domain: natural sciences)

K40 Liverpool Daily Post and Echo: News and features. u.p.. Sample containing about 3941 words from a periodical (domain: social science)

K41 Liverpool Daily Post and Echo: Foreign news pages. u.p.. Sample contain-
12. List of works excerpted

ing about 24533 words from a periodical (domain: world affairs)

K42 Liverpool Daily Post and Echo: Applied Science pages. u.p.. Sample containing about 536 words from a periodical (domain: applied science)

K43 Liverpool Daily Post and Echo: Arts section. u.p.. Sample containing about 2265 words from a periodical (domain: arts)

K44 Liverpool Daily Post and Echo: Business section. u.p.. Sample containing about 1651 words from a periodical (domain: commerce)

K45 Liverpool Daily Post and Echo: Leisure pages. u.p.. Sample containing about 25042 words from a periodical (domain: leisure)

K46 Liverpool Daily Post and Echo: News and features. u.p.. Sample containing about 610 words from a periodical (domain: social science)

K47 Liverpool Daily Post and Echo: Foreign news pages. u.p.. Sample containing about 18648 words from a periodical (domain: world affairs)

K48 Liverpool Daily Post and Echo: Applied Science pages. u.p.. Sample containing about 2790 words from a periodical (domain: applied science)

K49 Liverpool Daily Post and Echo: Arts section. u.p.. Sample containing about 6219 words from a periodical (domain: arts)

K4A Liverpool Daily Post and Echo: Religious affairs stories. u.p.. Sample containing about 452 words from a periodical (domain: belief and thought)

K4B Liverpool Daily Post and Echo: Business section. u.p.. Sample containing about 11550 words from a periodical (domain: commerce)

K4C Liverpool Daily Post and Echo: Leisure pages. u.p.. Sample containing about 28557 words from a periodical (domain: leisure)

K4D Liverpool Daily Post and Echo: News and features. u.p.. Sample containing about 7048 words from a periodical (domain: social science)

K4E Liverpool Daily Post and Echo: Foreign news pages. u.p.. Sample containing about 22060 words from a periodical (domain: world affairs)

K4F Liverpool Daily Post and Echo: Applied Science pages. u.p.. Sample containing about 2143 words from a periodical (domain: applied science)

K4G Liverpool Daily Post and Echo: Arts section. u.p.. Sample containing about 4323 words from a periodical (domain: arts)

K4H Liverpool Daily Post and Echo: Business section. u.p.. Sample containing about 11858 words from a periodical (domain: commerce)

K4I Liverpool Daily Post and Echo: Leisure pages. u.p.. Sample containing about 32022 words from a periodical (domain: leisure)

K4K Liverpool Daily Post and Echo: Applied Science pages. u.p.. Sample containing about 486 words from a periodical (domain: natural sciences)

K4L Liverpool Daily Post and Echo: News and features. u.p.. Sample containing about 10200 words from a periodical (domain: social science)

K4M Liverpool Daily Post and Echo: Foreign news pages. u.p.. Sample containing about 29967 words from a periodical (domain: world affairs)

K4N Northern Echo: Applied Science pages. u.p.. Sample containing about 6557 words from a periodical (domain: applied science)

K4P Northern Echo: Arts section. u.p.. Sample containing about 29967 words from a periodical (domain: arts)

K4Q Northern Echo: Religious affairs stories. u.p.. Sample containing about 1133 words from a periodical (domain: belief and thought)

K4S Northern Echo: Business section. u.p.. Sample containing about 37751 words from a periodical (domain: commerce)
<table>
<thead>
<tr>
<th>Code</th>
<th>Publication</th>
<th>Domain</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>K4T</td>
<td><em>Northern Echo: Leisure pages.</em></td>
<td>Leisure</td>
<td>about 211901</td>
</tr>
<tr>
<td>K4U</td>
<td><em>Northern Echo: Applied Science pages.</em></td>
<td>Natural Sciences</td>
<td>about 2021</td>
</tr>
<tr>
<td>K4V</td>
<td><em>Northern Echo: News and features.</em></td>
<td>Social Science</td>
<td>about 62560</td>
</tr>
<tr>
<td>K4W</td>
<td><em>Northern Echo: Foreign news pages.</em></td>
<td>World Affairs</td>
<td>about 211369</td>
</tr>
<tr>
<td>K4X</td>
<td><em>Northern Echo: Applied Science pages.</em></td>
<td>Applied Science</td>
<td>about 2572</td>
</tr>
<tr>
<td>K4Y</td>
<td><em>Northern Echo: Arts section.</em></td>
<td>Arts</td>
<td>about 27982</td>
</tr>
<tr>
<td>K50</td>
<td><em>Northern Echo: Religious affairs stories.</em></td>
<td>Belief and Thought</td>
<td>about 5958</td>
</tr>
<tr>
<td>K51</td>
<td><em>Northern Echo: Business section.</em></td>
<td>Commerce</td>
<td>about 44278</td>
</tr>
<tr>
<td>K52</td>
<td><em>Northern Echo: Leisure pages.</em></td>
<td>Leisure</td>
<td>about 172180</td>
</tr>
<tr>
<td>K53</td>
<td><em>Northern Echo: Applied Science pages.</em></td>
<td>Natural Sciences</td>
<td>about 1026</td>
</tr>
<tr>
<td>K54</td>
<td><em>Northern Echo: News and features.</em></td>
<td>Social Science</td>
<td>about 132902</td>
</tr>
<tr>
<td>K55</td>
<td><em>Northern Echo: Foreign news pages.</em></td>
<td>World Affairs</td>
<td>about 175048</td>
</tr>
<tr>
<td>K57</td>
<td><em>The Scotsman: Arts section.</em></td>
<td>Arts</td>
<td>about 48246</td>
</tr>
<tr>
<td>K58</td>
<td><em>The Scotsman: Religious affairs stories.</em></td>
<td>Belief and Thought</td>
<td>about 3994</td>
</tr>
<tr>
<td>K59</td>
<td><em>The Scotsman: Business section.</em></td>
<td>Commerce</td>
<td>about 116432</td>
</tr>
<tr>
<td>K5A</td>
<td><em>The Scotsman: Leisure pages.</em></td>
<td>Leisure</td>
<td>about 146882</td>
</tr>
<tr>
<td>K5B</td>
<td><em>The Scotsman: Applied Science pages.</em></td>
<td>Natural Sciences</td>
<td>about 384</td>
</tr>
<tr>
<td>K5C</td>
<td><em>The Scotsman: News and features.</em></td>
<td>Social Science</td>
<td>about 54129</td>
</tr>
<tr>
<td>K5D</td>
<td><em>The Scotsman: Foreign news pages.</em></td>
<td>Social Science</td>
<td>about 286598</td>
</tr>
<tr>
<td>K5F</td>
<td><em>The Scotsman: Arts section.</em></td>
<td>Applied Science</td>
<td>about 42973</td>
</tr>
<tr>
<td>K5G</td>
<td><em>The Scotsman: Religious affairs stories.</em></td>
<td>Belief and Thought</td>
<td>about 2323</td>
</tr>
<tr>
<td>K5H</td>
<td><em>The Scotsman: Business section.</em></td>
<td>Commerce</td>
<td>about 121751</td>
</tr>
<tr>
<td>K5J</td>
<td><em>The Scotsman: Leisure pages.</em></td>
<td>Commerce</td>
<td>about 116334</td>
</tr>
<tr>
<td>K5K</td>
<td><em>The Scotsman: Applied Science pages.</em></td>
<td>Natural Sciences</td>
<td>about 954</td>
</tr>
<tr>
<td>K5L</td>
<td><em>The Scotsman: News and features.</em></td>
<td>Social Science</td>
<td>about 46813</td>
</tr>
</tbody>
</table>
words from a periodical (domain: social sciences)

K5M The Scotsman: Foreign news pages. u.p.. Sample containing about 270447 words from a periodical (domain: world affairs)


K60 Appreciation of literature: lecture and discussion.

PS5AX 30+, ‘Liz’, lecturer, AB, female
Sample containing about 17568 words speech recorded in educational context

K61 Walsall Local Studies Centre: interview.
PS5AY ‘Joyce’, female
PS5B0 ‘Edward’, male
Sample containing about 11127 words speech recorded in leisure context

K62 Walsall Local Studies Centre: interview.
PS5B1 , male
PS5B2 retired, female
Sample containing about 7733 words speech recorded in leisure context

K63 Walsall Local Studies Centre: interview.
PS5B3 ‘Joyce’, female
PS5B4 ‘Gilbert’, retired, male
Sample containing about 8696 words speech recorded in leisure context

K64 Suffolk Sound Archive: interview.
PS5B5
PS5B6 ‘Doris’, retired children’s nurse, female
Sample containing about 4604 words speech recorded in leisure context

K65 Suffolk Sound Archive: interview for the Museum of East Anglian Life.
PS5B7 , female
PS5B8 67, ‘Harry’, retired butcher, male
PS5B9 ‘Jessie’, retired, female
Sample containing about 10889 words speech recorded in leisure context

K66 Suffolk Sound Archive: opening a restored control tower and airfield as a memorial to the Three-Ninetieth Bomb Group. Sample containing about 6635 words speech recorded in leisure context

K67 Suffolk Sound Archive: dedication of museum.
PS5BA ‘Horace’, retired ground executive officer usaf, male
Sample containing about 2963 words speech recorded in leisure context

K68 Suffolk Sound Archive: interview.
PS5BB ‘Yvonne’, female
PS5BC ‘Sidney’, retired police officer, male
Sample containing about 6608 words speech recorded in leisure context

K69 Suffolk Sound Archive: interview.
Sample containing about 8149 words speech recorded in leisure context

K6A On the Record: television broadcast.

PS5BD ‘Joyce’, female
PS5BE ‘Melville’, retired fire officer, male

Sample containing about 10478 words speech recorded in leisure context

K6B The Money Programme: television broadcast.

PS5BT ‘John ?’, tv presenter, male
PS5BG ‘Michael Heseltine’, cabinet minister (conservative), male
PS5BJ ‘David Walter’, reporter, male
PS5BK ‘Rhodes ?’, member of parliament, male
PS5BL ‘Anders’, swedish transport official, male
PS5BM ‘Bill ?', professor, male
PS5BN ‘Stig ?', director general of swedish railways, male
PS5BP ‘Tim ?', member of parliament (conservative), male

Sample containing about 6261 words speech recorded in leisure context

K6C Six O’Clock News: television broadcast.

PS5CK ‘Nicholas Witchell’, newsreader, male
PS5CL ‘Neil Bennet’, reporter, male
PS5CM ‘David’, solicitor, male
PS5CN ‘Paddy’, male
PS5CP ‘Billy’, male
PS5CR ‘Moira Stewart’, newsreader, female
PS5CS ‘Nick’, reporter, male
PS5CT ‘Max’, newspaper editor, male
PS5CU ‘Anthony’, retired chairman of the bar, male

Sample containing about 4771 words speech recorded in leisure context

K6D Six O’Clock News: television broadcast.

PS5DW ‘Nicholas Witchell’, newsreader, male
PS5DX ‘David Schuckman’, reporter, male
PS5DY ‘Malcolm Rifkind’, defence secretary, male
PS5E0 ‘David’, member of parliament, male
PS5E1 ‘Nicholas’, member of parliament, male
PS5E2 ‘John Menzies’, member of parliament, male
PS5E3 ‘Moira Stewart’, newsreader, female
PS5E4 ‘Iain Carson’, reporter, male

PS5E5 , male

Sample containing about 4724 words speech recorded in leisure context

K6E Six O’Clock News: television broadcast.

PS5F0 ‘Nicholas Witchell’, newsreader, male
PS5F1 ‘Neil Bennet’, reporter, male
PS5F2 ‘Raymond’, police constable, male
PS5F3 ‘Anna Ford’, newsreader, female
PS5F4 ‘Tiona Holden’, reporter, female
PS5F5 ‘Austin’, male
PS5F6 ‘Michael’, solicitor, male
PS5F7 tutor, male
PS5F8 ‘Polly Toynbee’, reporter, female

Sample containing about 4762 words speech recorded in leisure context

K6F Six O’Clock News: television broadcast.

PS5G0 ‘Anna Ford’, newsreader, female
PS5G1 ‘John Sergeant’, reporter, male
PS5G2 ‘Robin Cook’, member of parliament (labour), male
PS5G3 ‘Malcolm (rifkind?)’, member of parliament, male
PS5G4 ‘Tim Eggar?’, member of parliament, male
PS5G5 ‘Nicholas’, member of parliament, male
PS5G6 ‘Jenny’, newsreader, female
PS5G7 ‘Stephen Evans’, reporter, male
PS5G8 miner, male

Sample containing about 4974 words speech recorded in leisure context

K6G Six O’Clock News: television broadcast.

PS5H0 ‘Anna Ford’, newsreader, female
PS5H1 ‘Laurie Margolis’, reporter, male
PS5H2 ‘Maureen’, female
PS5H3 ‘Paul’, metropolitan police commissioner, male
PS5H4 ‘Barry’, police constable, male
PS5H5 ‘Ivan’, male
PS5H6 ‘Fred’, teacher, male
PS5H7 ‘Jenny’, newsreader, female
PS5H8 ‘Neil Bennet’, reporter, male

Sample containing about 4700 words speech recorded in leisure context

K6H Six O’Clock News: television broadcast.

PS5K0 ‘Nicholas Witchell’, newsreader, male
PS5K1 ‘Gerry Baker’, reporter, male
PS5K2 shop manager, male
PS5K3 ‘Derek’, company spokesperson, male
PS5K4 ‘Stephen’, member of parliament, male
PS5K5 ‘Harriet Harman’, member of parliament, female
PS5K6 ‘Simon’, member of parliament, male
PS5K7 ‘Moira Stewart’, newsreader, female
PS5K8 ‘Laurie Margolis’, reporter, male
Sample containing about 4832 words speech recorded in educational context

K6J Tutorial lesson.
P5L1U 50, ‘John’, maths and science tutor, male
K6JP000

Sample containing about 11915 words speech recorded in educational context

K6K Oral history project: interview.
P5L1V 50, ‘John’, maths and science tutor, male
P5L1W ‘Mabel’, singer, female
K6JP000

Sample containing about 3840 words speech recorded in leisure context

K6L Oral history project: interview.
P5L1X ‘Margaret’, historian, female
P5L1Y ‘Jo Ellen’, historian, female
P5M0 70+, ‘Nan’, retired french polisher, female
K6JP000

Sample containing about 7308 words speech recorded in leisure context

K6M Oral history project: interview.
P5M1 historian, Interviewing, male
K6MPS000

Sample containing about 10636 words speech recorded in leisure context

K6N Oral history project: interview.
P5M2 40+, ‘Ian’, historian, Interviewing speaker, male
K6NPS000
K6NPS001
K6NPS002

Sample containing about 7951 words speech recorded in leisure context

K6P Orkney Library Sound Archive tape OLSA 123: interview for oral history project.
P5M3 65, ‘Peter’, retired crofter, From North Ronaldsay. Being interviewed by, male
K6P7PS000

Sample containing about 6729 words speech recorded in leisure context

K6Q Orkney Sound Archive tape OSA 303: interview for oral history project.
P5M4 historian, Interviewing, female
K6RPS000
K6RPS001

Sample containing about 6729 words speech recorded in leisure context

K6R Orkney Sound Archive tape OSA 303: interview for oral history project.
P5M5 historian, Interviewing, female
K6SPS000
K6SPS001

Sample containing about 4332 words speech recorded in leisure context

K6T Orkney Sound Archive tape OSA 305: interview for oral history project.
P5M6 ‘Maimes’, Being interviewed by, female
K6TPS000

Sample containing about 5836 words speech recorded in leisure context

K6U Gwynedd County Council tape 1: interview for oral history project.
P5M7 historian, Interviewing, female
K6UPS000

Sample containing about 8752 words speech recorded in leisure context

K6V Tarmac Construction Ltd: training session.
P5M8 no further information given

Sample containing about 16478 words speech recorded in business context

K6W Nottingham Constabulary: meeting.
P5M9 47, ‘Storer’, no further information given, male
P5M0 ‘Collishan’, no further information given, male
P5MB 38, ‘Jeffery’, sergeant, male
P5MC 38, ‘Mills’, sergeant, male
P5MD 47, ‘Stone’, inspector, male
P5ME 48, ‘Williams’, inspector, male
P5MF 47, ‘Wright’, no further information given, male

Sample containing about 16389 words speech recorded in business context

K6X Nottingham Constabulary: meeting.
P5M1 47, ‘Storer’, no further information given, male
P5MK ‘Collishan’, no further information given, male
P5ML 38, ‘Jeffery’, sergeant, male
P5MM 38, ‘Mills’, sergeant, male
P5MN 47, ‘Stone’, inspector, male
P5MP 48, ‘Williams’, inspector, male
P5MS 40, ‘Hadfield’, sergeant, male
P5MT 47, ‘Smith’, sergeant, male

Sample containing about 3392 words speech recorded in business context

K6Y Friends Provident Insurance: training session.
P5M2 ‘Joan’, saleswoman, female
P5MV ‘Robert’, salesman, male
P5MW ‘Bill’, salesman, male
P5MX ‘Ron’, salesman, male
P5N0 ‘Maggie’, instructor, no further information given, female
P5N1 ‘Martin’, no further information given, female
Sample containing about 15713 words speech recorded in business context

**K70 Friends Provident Insurance: training session.**

PS5N3 ‘Joan’, saleswoman, female  
PS5N5 ‘Bill’, salesman, male  
PS5N6 ‘Ron’, salesman, male  
PS5N8 ‘Maggie’, instructor, no further information given, female  
PS5N9 ‘Martin’, no further information given, female  
PS5NA no further information given, female

Sample containing about 13585 words speech recorded in business context

**K71 Friends Provident Insurance: training session.**

PS5NB ‘Joan’, saleswoman, female  
PS5NC ‘Robert’, salesman, male  
PS5ND ‘Bill’, salesman, male  
PS5NE ‘Ron’, salesman, male  
PS5NF instructor, no further information given, male

Sample containing about 11048 words speech recorded in business context

**K73 Royal Courts: hearing.**

PS5NN ‘Green’, solicitor, male  
PS5NP ‘Saville’, judge, male  
PS5NR ‘Duffy’, solicitor, male  
PS5NS court official, male

Sample containing about 14018 words speech recorded in public context

**K74 J.**

PS5NT trainer, no further information given, female  
PS5NU ‘Kenny’, no further information given, male  
PS5NV ‘Heidi’, no further information given, female  
PS5NW ‘Ray’, no further information given, male  
PS5NX ‘Alex’, no further information given, male  
PS5NY ‘Gary’, no further information given, male  
PS5P0 ‘Ian’, no further information given, male  
PS5P1 ‘Rachel’, no further information given, female  
PS5P2 ‘Dee’, no further information given, female

Sainsbury: training session. Sample containing about 12429 words speech recorded in business context

**K75 J.**

PS5PA trainer, no further information given, female

Sample containing about 9504 words speech recorded in business context

**K76 House of Lords: hearing.**

PS5PB ‘Motterstone’, member of parliament, no further information given, male  
PS5PC ‘Rippon’, member of parliament, male  
PS5PD ‘Macintosh’, member of parliament, male  
PS5PE ‘Boyd-carpenter’, male

Sample containing about 9482 words speech recorded in public context

**K77 House of Commons: hearing.**

PS5PF no further information given, male  
PS5PG ‘Ken’, no further information given, male  
PS5PH ‘Jimmy’, no further information given, male  
PS5PJ ‘David’, no further information given, male  
PS5PK ‘Peter’, no further information given, male  
PS5PM ‘Clifford’, no further information given, male  
PS5PN ‘Jane’, no further information given, female  
PS5PP ‘Adams’, no further information given, male

Sample containing about 14059 words speech recorded in public context

**K78 Durham University: lecture.** Sample containing about 10813 words speech recorded in educational context

**K79 Confederation of British Industries [presentation].**

PS5PS 40, ‘Sudhir’, analyst, male  
PS5PT 40, ‘Richard’, analyst, male

Sample containing about 4318 words speech recorded in business context

**K7D Oral history project: interview.**

PS5PU 54, ‘Tommy’, Interviewee, AB, male  
PS5PV historian, Interviewer, DE, male

Sample containing about 4002 words speech recorded in educational context

**K7E Oral history project: interview.**

PS5PW historian, Interviewer, male  
PS5PX 81, ‘George’, retired lighthouse keeper, Interviewee, male

Sample containing about 4001 words speech recorded in educational context

**K7F Tutorial lesson.**

PS5PY 50, ‘John’, tutor, tutor, male  
PS5R0 16, ‘Sarah’, school student, student, female

Sample containing about 6155 words speech recorded in educational context
12. List of works excerpted

K7G Oral history project: interview.
PS5R1 71, ‘Thomas’, retired baker, interviewee, male
PS5R2 historian, interviewer, male
Sample containing about 7048 words speech recorded in educational context


K96 Community Care. u.p.. Sample containing about 21593 words from a periodical (domain: social science)


K98 [Sainsbury’s magazines]. u.p.. Sample containing about 9654 words from a periodical (domain: social science)

K99 JS Journal. J Sainsbury. Sample containing about 13031 words from a periodical (domain: social science)

K9A JS Journal. J Sainsbury. Sample containing about 13416 words from a periodical (domain: social science)

K9B Courtaulds News. u.p.. Sample containing about 16799 words of unpublished miscellanea (domain: commerce)
12. List of works excerpted

K9C  Courtaulds News.  u.p..  Sample containing about 17375 words of unpublished miscellanea (domain: commerce)

K9D  Courtaulds News.  u.p..  Sample containing about 15360 words of unpublished miscellanea (domain: commerce)

K9E  Courtaulds News.  u.p..  Sample containing about 5891 words of unpublished miscellanea (domain: commerce)

K9F  ICI Chemicals & Polymers: Environmental Issues.  u.p..  Sample containing about 5307 words of unpublished miscellanea (domain: commerce)

K9G  ICI Chemicals & Polymers: Environmental Issues.  u.p..  Sample containing about 4799 words of unpublished miscellanea (domain: commerce)

K9H  ICI Chemicals & Polymers: Environmental Issues.  u.p..  Sample containing about 6443 words of unpublished miscellanea (domain: commerce)

K9J  ICI Innovation.  ICI.  Sample containing about 24051 words from a periodical (domain: commerce)

K9K  Q: The Quality Magazine of ICI Chemicals & Polymers.  u.p..  Sample containing about 11589 words from a periodical (domain: commerce)

K9L  Q: The Quality Magazine of ICI Chemicals & Polymers.  ICI.  Sample containing about 10084 words from a periodical (domain: commerce)

K9M  What’s Cooking.  u.p..  Sample containing about 5304 words of unpublished miscellanea (domain: applied science)

K9N  What’s Cooking.  u.p..  Sample containing about 4151 words of unpublished miscellanea (domain: applied science)

K9P  What’s Cooking.  u.p..  Sample containing about 4825 words of unpublished miscellanea (domain: applied science)

K9R  CCG People staff newspaper.  u.p..  Sample containing about 4102 words of unpublished miscellanea (domain: applied science)

K9S  [British Gas South Western leaflets].  u.p..  Sample containing about 13601 words of unpublished miscellanea (domain: commerce)

K9T  National Library of Scotland – newsletter.  u.p..  Sample containing about 2643 words of unpublished miscellanea (domain: applied science)

K9U  National Library of Scotland – newsletter.  u.p..  Sample containing about 8513 words of unpublished miscellanea (domain: applied science)

K9V  Transport.  u.p..  Sample containing about 14211 words from a periodical (domain: commerce)

K9W  Transport.  u.p..  Sample containing about 12608 words from a periodical (domain: commerce)

K9X  [Unpublished school essays].  u.p..  Sample containing about 13956 words of unpublished miscellanea (domain: social science)

K9Y  [School welcome pack].  u.p..  Sample containing about 2475 words of unpublished miscellanea (domain: social science)

KA0  [School regulations and other papers].  u.p..  Sample containing about 2346 words of unpublished miscellanea (domain: social science)

KA1  [School essays on literature].  u.p..  Sample containing about 46614 words of unpublished miscellanea (domain: arts)

KA2  [Unpublished creative writing].  u.p..  Sample containing about 8602 words of unpublished miscellanea (domain: imaginative)

KA3  [School essays etc.  u.p..  ]  Sample containing about 5309 words of unpublished miscellanea (domain: social science)

KA4  Campus.  Stirling University.  Sample containing about 9491 words of a periodical (domain: world affairs)

KA5  [Scottish Christian Aid committee documents].  u.p..  Sample containing about
12. List of works excerpted

8991 words of unpublished miscellanea (domain: belief and thought)


KAJ The Medau Society. u.p.. Sample containing about 28890 words of unpublished miscellanea (domain: leisure)

KAK [Labour Party pamphlet on Education] Will we ever learn? Sample containing about 4452 words of unpublished miscellanea (domain: social science). u.p..


KAM [Autobiographical notes written for the archives of the British Academy]. Leech, Geoffrey, u.p.. Sample containing about 4570 words of unpublished miscellanea (domain: social science)

KAN [Report to a sponsor on lecture tour of India]. u.p. (1976). Sample containing about 2611 words of unpublished miscellanea (domain: social science)

KAP Christian Aid. u.p.. Sample containing about 15083 words of unpublished miscellanea (domain: social science)

KAR [Personal letters]. u.p.. Sample containing about 8950 words of unpublished miscellanea (domain: leisure)

KAS [Unpublished poems]. u.p.. Sample containing about 3246 words of unpublished miscellanea (domain: imaginative)

KAT [Creative writing]. u.p.. Sample containing about 3676 words of unpublished miscellanea (domain: imaginative)

KAU [Personal letters]. u.p.. Sample containing about 1908 words of unpublished miscellanea (domain: leisure)

KAV [Schoolchildren's creative writing]. u.p.. Sample containing about 4466 words of unpublished miscellanea (domain: imaginative)

KAX [Creative writing – poems]. u.p.. Sample containing about 1920 words
of unpublished miscellanea (domain: imaginative)

KAY [School essays on literature]. u.p.. Sample containing about 30176 words of unpublished miscellanea (domain: arts)

KB0 19 conversations recorded by ‘Margaret’ (PS002) between 13 and 16 March 1992 with 9 interlocutors, totalling 3901 s-units, 42462 words, and 6 hours 28 minutes 0 seconds of recordings.

PS002 ‘Margaret’, 60, retired, Lancashire, female
PS003 ‘Eva’, 50, farmer’s wife, Welsh, female
PS004 ‘Philip’, 25, farmer’s son, male
PS005 ‘Ida’, 40, shop assistant, Welsh, female
PS006 ‘Bob’, 50, minister, London, male
PS007 ‘Alan’, 79, retired (lecturer and missionary), AB, male
PS008 ‘John’, 40, moderator, Welsh, male
PS009 ‘Gethyn’, 40, minister, Welsh, male
PS010 ‘Noel’, 60, minister, Welsh, male

KB1 17 conversations recorded by ‘Albert’ (PS01A) between 1 and 6 February 1992 with 9 interlocutors, totalling 5856 s-units, 39285 words, and over 3 hours 6 minutes 40 seconds of recordings.

PS01A ‘Albert’, 53, unemployed, Central Northern England, DE, male
PS01B ‘June’, 47, forecourt attendant, Central Northern England, female
PS01C ‘Ada’, 70, retired, Central Northern England, female
PS01D ‘Corinne’, 18, student, Central Northern England, female
PS01E ‘Colin’, 39, unemployed, Central Northern England, male
PS01F ‘Karen’, 38, housewife, Central Northern England, female
PS01G ‘Sasha’, 5, student, Central Northern England, female
PS01H ‘Rose’, 67, retired, Central Northern England, female
PS01J ‘None’, 46+, shopkeeper, Scottish, female

KB2 25 conversations recorded by ‘Alec’ (PS01T) between 31 January and 7 February 1992 with 5 interlocutors, totalling 5729 s-units, 35089 words (duration not recorded).

PS01T ‘Alec’, 68, retired, Central Northern England, DE, male
PS01U ‘Ann’, 45, saleswoman, Central Northern England, female
PS01V ‘Joyce’, 63, retired, Central Northern England, DE, female
PS01W ‘Joan’, retired, Central Northern England, female

KB3 8 conversations recorded by ‘Alison’ (PS147) between 30 January and 4 February 1992 with 4 interlocutors, totalling 1666 s-units, 10492 words, and 2 hours 20 minutes 32 seconds of recordings.

PS147 ‘Alison’, 43, housewife, Scottish, AB, female
PS148 ‘Grant’, 7, student (state primary), Scottish, AB, male
PS149 ‘Bob’, 53, deputy prison governor, Scottish, AB, male
PS14A ‘Annabel’, 2, student (private pre), Scottish, AB, female

KB4 1 conversation recorded by ‘Alan2’ (PS1CP) [dates unknown] with 5 interlocutors, totalling 174 s-units, 828 words, and 8 minutes 0 seconds of recordings.

PS1CP ‘Alan’, 55, driver, London, DE, male
PS1CR ‘Mandy’, London, female
PS1CS ‘Darren’, 24, unemployed, London, DE, male
PS1CU ‘Gary’, London, male
PS1CV ‘Tracey’, female

KB5 11 conversations recorded by ‘Andrew’ (PS028) on 30 May 1991 with 6 interlocutors, totalling 452 s-units, 5541 words, and 58 minutes 17 seconds of recordings.

PS028 ‘Andrew’, 16, student, London, C2, male
PS02S ‘Suzanne’, 17, student, London, female
PS02Y ‘Val’, 46, systems analyst, Northern England, female
PS02T ‘John’, 17, student, London, male
PS02A ‘Sanjay’, 18, student, London, male
PS02B ‘Jon’, 18, student, London, male

KB6 8 conversations recorded by ‘Angela’ (PS029) between 2 and 6 December 1991 with 7 interlocutors, totalling 2360 s-units, 12953 words, and 1 hour 22 minutes 59 seconds of recordings.

PS029 ‘Angela’, 29, out work (pt), Central South-west England, DE, female
PS02A ‘Chris’, 25, engineer, Central South-west England, DE, male
PS02B ‘Angela’, 32, housewife, London, female
PS02C ‘Angela’, 32, housewife, London, female
PS02D ‘Zoe’, 28, housewife, Central South-west England, female
PS02E ‘Sue’, 29, hairdresser, Central South-west England, female
PS02F ‘Ben’, 3, student (state pre), DE, male
| PS02G | 'Ann', 46, team leader, North-east England, C2, female |
| PS02H | 'Stuart', 33, factory operative, East Anglia, DE, male |
| PS02J | 'Brian', 45, paper-man, Northern England, male |
| PS02K | 'Jill', 40, estate agent, East Anglia, female |
| PS02L | 'Jeanie', 63, retired, Scottish, female |
| PS02M | 'Sue', 40, care assistant, Scottish, female |
| PS02N | 'Shirley', 45, publican, Midlands, female |
| PS02P | 'Alan', 50, Scottish, male |
| PS02R | 'Terry', Scottish, male |
| PS02S | 'Bob', 46, publican, Central Midlands, male |
| PS02T | 'Harry', 50, male |
| PS02U | 'John', 60, retired (policeman), London, male |
| PS02V | 'Martin', 23, marketing person, Northern England, male |
| PS02X | 'Helen', 22, housewife, East Anglia, female |
| PS02Y | 'Jonathan', 17, Youth Training Scheme trainee, East Anglia, male |
| PS030 | 'Christian', 19, process worker, male |
| PS031 | 'Richard', 2, pre-school, male |
| PS032 | 'Anna Marie', 1, pre-school, female |
| PS033 | 'Bernie', 26, building labourer, male |
| PS034 | 'Bill', 61, team leader, male |
| PS035 | 'Jean', 57, process worker, female |
| PS036 | 'Steve', 26, process worker, male |
| PS037 | 'Deidre', 44, senior team leader, female |
| PS038 | 'John', 56, process worker, male |
| PS039 | 'Dawn', 50, process worker, female |
| PS040 | 'Tina', 27, team leader, East Anglia, female |
| PS041 | 'Val', 36, Queen's counsel, female |
| PS042 | 'Trevor', 40, process worker, male |
| PS043 | 'Brenda', 44, female |
| PS044 | 'Lynda', 21, female |
| PS045 | 'Neville', 50, male |
| PS046 | 'Maxine', 25, process worker, Northern England, female |
| PS047 | 'Merlin', 36, process worker, East Anglia, male |
| PS048 | 'Grace', female |
| PS049 | 'None' |

| PS14B | 'Ann', 53, registered childminder, Central Northern England, AB, female |
| PS14C | 'James', 57, charge nurse, North-east England, AB, male |
| PS14D | 'Rebecca', 9, student, North-east England, female |
| PS14E | 'Amy', 8, student, North-east England, female |
| PS14F | 'Matthew', 8, student, North-east England, male |
| PS14H | 'Richard', 6, student, North-east England, male |
| PS14K | 'Martin', 6, student, Northern England, male |
| PS14L | 'Dorothy', 36, probation officer, Central Midlands, female |
| PS14M | 'James', 2, pre-school, North-east England, male |
| PS14N | 'Bryony', 2, pre-school, North-east England, female |
| PS14P | 'Richard', 2, pre-school, North-east England, male |
| PS14R | 'Isabelle', 53, shop assistant, North-east England, female |
| PS15B | 'Sally', 10, student, North-east England, female |
| PS15C | 'Andrea', 37, teacher, North-east England, female |
| PS15D | 'Sid', 35+, teacher, North-east England, male |
| PS15G | 'Marilyn', 42, secretary, Northern England, female |
| PS15S | 'Bill', 55, clerk to town council, North-east England, male |
| PS15U | 'Betty', 58, housewife, North-east England, female |
| PS15V | 'Paul', 27, unemployed, North-east England, male |
| PS166 | 'Jackie', 38, clerical officer, North-east England, female |
| PS167 | 'Pat', 39, shop assistant, North-east England, female |
| PS168 | 'Chris', 9, student, North-east England, male |
| PS169 | 'Margaret', 86, retired, North-east England, female |
| PS17F | 'Eve', 7, student, North-east England, female |
| PS17G | 'Maggie', 30, childminder, Northern England, female |
| PS17L | 'Arthur', 70, retired, North-east England, male |
| PS17M | 'Tom', 70, retired, Irish, male |
| PS17N | 'Rosie', 35, childminder, Northern England, female |
| PS18D | 'Louise', 29, housewife, North-east England, male |
| PS18F | 'Erin', 7, student, North-east England, female |
| PS18G | 'Kathy', 32, housewife, Northern England, female |
| PS18H | 'Martin', 6, student, North-east England, male |
| PS18S | 'None' |
| PS172 | 'Evelyn' |
| PS17G | 'Maggie', 30, childminder, Northern England, female |
| PS17L | 'Arthur', 70, retired, Northern England, male |
| PS17M | 'Tom', 70, retired, Irish, male |
| PS17N | 'Rosie', 35, childminder, Northern England, female |
| PS18D | 'Louise', 29, housewife, North-east England, male |
| PS50S | 'Jean', 35, housewife, Northern England, female |
| PS6NU | 'Alison', 24, housewife, Northern England, female |

| KB7 | 60 conversations recorded by ‘Ann’ (PS02G) between 28 November and 5 December 1991 with 35 interlocutors, totalling 16243 s-units, 103997 words, and over 13 hours 5 minutes 36 seconds of recordings. |
| KB8 | 34 conversations recorded by ‘Ann2’ (PS14B) on 3 April 1992 with 37 interlocutors, totalling 12018 s-units, 78798 words, and over 8 hours 19 minutes 57 seconds of recordings. |
| KB9 | 44 conversations recorded by ‘Annette’ |
(PS1CX) between 1 and ?? February 1992 with 6 interlocutors, totalling 5497 s-units, 31762 words, and over 2 minutes 5 seconds of recordings.

PS1CX  'Annette', 44, administrative assistant, Lancashire, C1, female
PS1CY  'Teresa', 19, stable hand, Lancashire, C1, female
PS1D0  'David', 26, engineer, Lancashire, male
PS1D1  'Tracy', 24, housewife, Lancashire, female
PS1D2  'Donald', 71, retired, Northern England, male
PS1KS  'Pat', 48, Lancashire, female

KBA 17 conversations recorded by 'Anthony' (PS1DA) between 30 November and 4 December 1991 with 8 interlocutors, totalling 1192 s-units, 5272 words, and 1 hour 5 minutes 22 seconds of recordings.

KBA000  'None'
PS1DA  'Anthony', 62, retired, Central South-west England, C2, male
PS1DB  'Chris', 61, housewife, Central South-west England, female
PS1DC  'Chris', 26, warehouse op, Central South-west England, male
PS1DD  'Kevin', 31, warehouse op, Central South-west England, male
PS1DF  'Achmed', 30, warehouse op, Indian Sub-continent, male
PS1DG  'Achmed', 30, warehouse op, Indian Sub-continent, male

KBB 41 conversations recorded by 'Arthur' (PS03S) between 10 and 13 January 1992 with 7 interlocutors, totalling 11521 s-units, 76309 words, and over 8 hours 52 minutes 5 seconds of recordings.

PS03S 'None'
PS03T  'Evelyn', 75, retired, Central Midlands, C2, female
PS03U  'Tom', 54, draughtsman, London, male
PS03V  'Jackie', 51, catering supervisor, Central Midlands, female
PS03W  'Les', 75, retired, Central Midlands, male
PS03X  'Kath', 73, retired, Central Midlands, female
PS03Y  'Vince', 73, retired, North-east England, male

KBC 14 conversations recorded by 'Audrey' (PS1A9) between 2 and 9 April 1992 with 9 interlocutors, totalling 6341 s-units, 31337 words, and 3 hours 38 minutes 41 seconds of recordings.

PS1A9  'Audrey', 61, housewife, Lancashire, AB, female
PS1AA  'Gordon', 61, teacher, Lancashire, AB, male
PS1AB  'Margaret', 45, nurse, Lancashire, female
PS1AC  'Joan', 50+, clerk, Central Northern England, female
PS1AD  'Kevin', 29, computer engineer, Northern England, male
PS1AE  'Carl', 31, pharmacist, Northern England, AB, male
PS1AF  'None'
PS1AG  'Elaine', 28, housewife, Northern England, female
PS1AH  'Iris', 60, housewife, Lancashire, female

KBE 27 conversations recorded by 'Betty' (PS04B) between 28 November and ?? December 1991 with 9 interlocutors, totalling 10543 s-units, 62321 words, and over 5 hours 40 minutes 22 seconds of recordings.

PS04B 'None'
PS04C  'Googie', 76, housewife, Central South-west England, female
PS04D  'Rose', 64, housewife, Central South-west England, female
PS04E  'Wendy', 33, housewife, Central South-west England, female
PS04F  'Julie', 25, production worker, Central South-west England, female
PS04G  'Maggie', 32, shop assistant, Central South-west England, female
PS04H  'Sally', 23, production worker, Central South-west England, female
PS04I  'David', 55, disabled unemployed, Central South-west England, male
PS04J  'Edna', 76, housewife, Central South-west England, female

KBF 103 conversations recorded by 'Brenda' (PS04U) between 28 November and ?? December 1991 with 6 interlocutors, totalling 13578 s-units,
105262 words, and over 11 hours 59 minutes 46 seconds of recordings.

PS04U ‘Brenda’, 34, housewife, London, C2, female
PS04V ‘Dave’, 37, hgv driver, London, C2, male
PS04W ‘Lee’, 15, student (state secondary), London, C2, male
PS04X ‘Carla’, 11, student (state primary), London, C2, female
PS04Y ‘Jean’, 62, housewife, Home Counties, female
PS050 ‘Nigel’, 23, plumber, London, male

KBG 10 conversations recorded by ‘Carl’ (PS051) between 21 and 27 February 1992 with 9 interlocutors, totalling 4559 s-units, 28812 words, and 3 hours 30 minutes 23 seconds of recordings.

PS051 ‘Carl’, 34, teacher, North-east England, C1, male
PS052 ‘Susan’, 33, teacher (pt), Home Counties, C1, female
PS053 ‘Christopher’, 4, pre-school, C1, male
PS054 ‘Rebecca’, 1, pre-school, C1, female
PS055 ‘Andrew’, 7, student (primary), C1, male
PS056 ‘Robert’, 9, student (secondary), C1, male
PS057 ‘Peter’, 58, lorry driver, East Anglia, male
PS058 ‘Alan’, 65, retired, East Anglia, male
PS059 ‘Pat’, 62, retired, Home Counties, male

KBH 63 conversations recorded by ‘Carol’ (PS05B) between 30 May and 5 June 1991 with 8 interlocutors, totalling 7122 s-units, 47995 words, and over 6 hours 42 minutes 9 seconds of recordings.

PS05B ‘Carol’, 36, administrator (pt), Home Counties, C1, female
PS05C ‘Joelle’, 18, au pair, European (French), DE, female
PS05D ‘Adam’, 36, textile agent, Home Counties, C1, male
PS05E ‘Charlotte’, 2, pre-school, C1, female
PS05F ‘Pauline’, 30+, shop owner, Home Counties, female
PS05H ‘Ron’, 60+, retired, Home Counties, male
PS05J ‘Pat’, 60+, retired, Home Counties, male
PS05K ‘Toby’, 5, male

KBK 228 conversations recorded by ‘Chris’ (PS05X) between 31 May and 1 June 1991 with 8 interlocutors, totalling 7432 s-units, 53920 words, and over 5 hours 39 minutes 52 seconds of recordings.

PS05X ‘Chris’, 51, export merchant, Home Counties, AB, male
PS101 ‘Noreine’, 54, teacher, London, AB, female
PS102 ‘Lynda’, 44, export assistant, London, female
PS103 ‘Susan’, 30, cleaner, London, female
PS104 ‘Bill’, 30+, plastics company director, London, male
PS106 ‘Tony’, 60+, retired, London, male
PS107 ‘David’, 45, property manager, London, male
PS108 ‘Evelyn’, 60+, retired, London, female

KBL 40 conversations recorded by ‘Cherrilyn’ (PS06A) between 21 and 26 February 1992 with 9 interlocutors, totalling 5508 s-units, 27294 words (duration not recorded).

PS06A ‘Cherrilyn’, 43, nursing auxiliary, Lower South-west England, C1, female
PS06B ‘Jessica’, 13, student (state secondary), Lower South-west England, C1, female
PS06C ‘None’
PS06D ‘Sophie’, 21, student, Lower South-west England, C1, female
PS06E ‘Kate’, 12, student, Lower South-west England, female
PS06F ‘Brian’, 43, mortgage broker, Lower South-west England, C1, male
PS06G ‘Fiona’, 33, matron of nursing home, Central South-west England, female
PS06H ‘Lorraine’, 16, care assistant, Home Counties, female
PS06J ‘None’, 70, Lower South-west England, male

KBM 13 conversations recorded by ‘Chris2’ (PS1BL) on 2 April 1992 with 7 interlocutors, totalling 2837 s-units, 19723 words, and 2 hours 22 minutes 47 seconds of recordings.

PS1BL ‘Chris’, 15, student, North-west Midlands, AB, male
PS1BM ‘Lynne’, 45, bookkeeper, North-west Midlands, AB, female
PS1BN ‘David’, 16, student, North-west Midlands, male
PS1BP ‘Steven’, 16, student, North-west Midlands, male
12. List of works excerpted

PS1BR ‘Simon’, 15, student, North-west Midlands, male
PS6P6 ‘Rhiannon’, 16, student, North-west Midlands, female
PS6P7 ‘Robert’, 48, tax inspector, North-west Midlands, AB, male

KBN 18 conversations recorded by ‘Clare’ (PS05Y) between 12 and 15 March 1992 with 6 interlocutors, totalling 1981 s-units, 10609 words, and 2 hours 36 minutes 21 seconds of recordings.

PS05Y ‘Clare’, 15, student, Welsh, C2, female
PS060 ‘Melissa’, 16, student, Welsh, female
PS061 ‘William’, 52, carpenter, Welsh, C2, male
PS062 ‘Kim’, 9, student, Welsh, female
PS063 ‘Helen’, 29, housewife, Welsh, female
PS064 ‘Marilyn’, 50, housewife, Welsh, C2, female

KBP 15 conversations recorded by ‘Clarence’ (PS065) between 13 and 19 March 1992 with 4 interlocutors, totalling 5039 s-units, 27179 words, and 2 hours 23 minutes 42 seconds of recordings.

PS065 ‘Clarence’, 65, retired, Lancashire, DE, male
PS066 ‘Nina’, 67, retired, Lancashire, DE, female
PS067 ‘Nev’, 72, retired, North-east Midlands, male
PS068 ‘Lil’, 70, retired, Lancashire, female

KBR 26 conversations recorded by ‘Colin’ (PS069) between 20 and 26 February 1991 with 5 interlocutors, totalling 2220 s-units, 11864 words (duration not recorded).

PS069 ‘Colin’, British Rail employee, Home Counties, male
PS10D ‘Sheila’, housekeeper, Home Counties, female
PS10E ‘Nigel’, electrician, Home Counties, male
PS10F ‘Neil’, unemployed, Home Counties, male
PS10G ‘Joanne’, Home Counties, female

KBS 12 conversations recorded by ‘Dad’ (PS070) [dates unknown] with 5 interlocutors, totalling 1314 s-units, 6972 words (duration not recorded).

PS070 ‘Dad’, 79, retired, Upper South-west England, male
PS07A ‘Mum’, 81, retired, Upper South-west England, female
PS07B ‘Doreen’, 50, housekeeper, Midlands, female
PS07C ‘Jane’, 30, hairdresser, Midlands, female
PS07D ‘Adam’, 8, student, Midlands, male

KBT 4 conversations recorded by ‘Dean’ (PS07E) on 1 February 1992 with 5 interlocutors, totalling 583 s-units, 2910 words, and 17 minutes 0 seconds of recordings.

PS07E ‘Dean’, 32, miner chargehand, Central Northern England, C2, male
PS07F ‘John’, 40, unemployed, Central Northern England, DE, male
PS07G ‘David’, 28, unemployed, Central Northern England, DE, male
PS07H ‘Michael’, 38, sales representative, Central Northern England, male
PS07J ‘Eddie’, 37, lorry driver, Central Northern England, male

KBU 2 conversations recorded by ‘Denise’ (PS18E) between 6 and 7 April 1992 with 8 interlocutors, totalling 1829 s-units, 11022 words, and 1 hour 3 minutes 14 seconds of recordings.

PS18E ‘Denise’, 25, student nurse, Central Northern England, AB, female
PS18H ‘None’
PS18J ‘None’
PS18K ‘None’
PS18L ‘None’
PS18M ‘None’
PS18N ‘None’

KBV 19 conversations recorded by ‘Donald’ (PS07K) between 13 and 20 March 1992 with 6 interlocutors, totalling 993 s-units, 4916 words, and over 2 hours 18 minutes 56 seconds of recordings.

PS07K ‘Donald’, 33, telecommunication technician, North-west Midlands, C1, male
PS07L ‘Karen’, 32, housewife, North-east Midlands, C1, female
PS07M ‘Lake’, 6, student, North-east Midlands, C1, male
PS07N ‘Gary’, 10, student, North-east Midlands, C1, male
PS07P ‘Marie’, 1, pre-school, North-east Midlands, C1, female

KBW 62 conversations recorded by ‘Dorothy’ (PS087) between 13 and 20 March 1992 with 25 interlocutors, totalling 19706 s-units, 115332 words, and over 13 hours 0 minutes 42 seconds of recordings.

PS087 ‘Dorothy’, 34, teacher (pt), North-west Midlands, AB, female
PS088 ‘Tim’, 3, student (private pre), North-east Midlands, AB, male
PS089 ‘Christopher’, 5, student (state primary), North-east Midlands, AB, male
PS08A ‘Andrew’, 36, purchasing manager, North-east Midlands, AB, male
PS08B ‘Elinor’, 14, student, North-east Midlands, female
12. List of works excerpted

PS08C  ‘Jane’, 40, housewife, North-east Midlands, female
PS08D  ‘Diane’, 38, housewife, North-east Midlands, female
PS08R  ‘Sue’, 40, teacher, North-east Midlands, female
PS08S  ‘Margaret’, 60+, retired, North-east Midlands, female
PS08T  ‘Roger’, 45, teacher, male
PS08U  ‘June’, 60, teacher, Canadian, female
PS08V  ‘Jean’, 60, retired, North-east Midlands, female
PS08W  ‘None’, 40, secretary, North-east Midlands, female
PS08X  ‘Jane’, 24, teacher, North-east Midlands, female
PS08Y  ‘Dawn’, 30, housewife, North-east Midlands, female
PS08Z  ‘Yvonne’, 30+, housewife, North-east Midlands, female
PS09A  ‘Pat’, 30+, teacher, female
PS09B  ‘Gwen’, 50, teacher/nursery owner, North-east Midlands, female
PS09C  ‘Lily’, 40, nursery assistant, North-east Midlands, female
PS09D  ‘Marjorie’, 70, retired, North-west Midlands, female
PS09E  ‘Ann’, 35, teacher, female
PS09F  ‘Rosemary’, 50, housewife, Home Counties, female
PS09G  ‘Mary’, 45, drawing plans, North-east Midlands, female
PS09H  ‘None’ teacher, Home Counties, female

KBX  11 conversations recorded by ‘Donald2’ (PS1DW) between 17 and 10 January 1992 with 6 interlocutors, totalling 1967 s-units, 28273 words (duration not recorded).

KBP00  ‘None’ North-west Midlands
KBP01  ‘None’ North-west Midlands
KBP02  ‘Donald’, 56, taxi driver, North-west Midlands, C2, male
KBP03  ‘Anna’, 60, housewife, European (Russian), female
KBP04  ‘Mary’, 60, housewife, European (Russian), female
KBP05  ‘Barry’, 50 North-west Midlands, male

KBY  11 conversations recorded by ‘Elizabeth’ (PS08X) between 11 and 17 January 1992 with 5 interlocutors, totalling 1038 s-units, 6316 words (duration not recorded).

KBP06  ‘Elisabeth’, 19, student, North-west Midlands, C1, female
KBP07  ‘None’, 18, student, North-west Midlands, male
KBP08  ‘Matthew’, 23, manager, North-west Midlands, C1, male
KBP09  ‘Anne’, 48, clerk (pt), North-west Midlands, C1, female
KBP10  ‘Mike’, 53, computer operator, North-west Midlands, C1, male

KC0  14 conversations recorded by ‘Enid’ (PS08Y) between 21 and 27 February 1992 with 10 interlocutors, totalling 8028 s-units, 45900 words, and 4 hours 12 seconds of recordings.

KC1  32 conversations recorded by ‘Frank’ (PS09E) between 21 and 28 February 1992 with 9 interlocutors, totalling 3193 s-units, 20607 words, and 3 hours 22 minutes 23 seconds of recordings.

KC2  28 conversations recorded by ‘Fred’ (PS09T) between 13 and 20 March 1992 with 9 interlocutors, totalling 6625 s-units, 44626 words, and 6 hours 35 minutes 19 seconds of recordings.
12. List of works excerpted

PS0A2 ‘Sandra’, 39, housewife, North-east Midlands, female

KC3 15 conversations recorded by ‘Frederick’ (PS0A8) between 10 and 15 January 1992 with 10 interlocutors, totalling 3930 s-units, 36931 words (duration not recorded).

PS0A8 ‘Frederick’, 66, retired (social worker), Central Midlands, AB, male

PS0A9 ‘Janet’, 49, health visitor, Central Midlands, female

PS0AA ‘Peter’, 49, chemist, Home Counties, male

PS0AB ‘Joan’, 68, housewife, Central Midlands, female

PS0AC ‘Vicki’, 40, teacher, South Midlands, female

PS0AD ‘Heidi’, 15, schoolgirl, Central Midlands, female

PS0AE ‘Yvonne’, 40, clerk, Central Midlands, female

PS0AF ‘Trish’, 30, physiotherapist, Central Midlands, female

PS0AG ‘Eileen’, 61, cleaner, Irish, female

KC4 25 conversations recorded by ‘Fred2’ (PS1E4) between 31 May and 5 June 1991 with 9 interlocutors, totalling 3050 s-units, 23715 words, and 3 hours 15 minutes 48 seconds of recordings.

PS1E4 ‘Fred’, 78, retired, North-west Midlands, AB, male

PS1E5 ‘None’

PS1E6 ‘None’

PS1E7 ‘None’

PS1E8 ‘None’

PS1E9 ‘None’

PS1EA ‘None’

PS1EB ‘None’

PS1EC ‘None’

KC5 16 conversations recorded by ‘Gail’ (PS0AJ) between 2 and 4 December 1991 with 9 interlocutors, totalling 3620 s-units, 22248 words, and over 2 hours 2 minutes 3 seconds of recordings.

PS0AJ ‘Gail’, 30, housewife, Central South-west England, DE, female

PS0AK ‘None’

PS0AL ‘Alison’, 28, housewife, Central South-west England, DE, female

PS0AM ‘None’

PS0AN ‘Natasha’, 7, student (state primary), Central South-west England, female

PS0AP ‘None’

PS0AR ‘None’

PS0AS ‘None’

PS0AT ‘None’

KC6 25 conversations recorded by ‘Gavin’ (PS0BA) between 11 and 17 January 1992 with 10 interlocutors, totalling 2413 s-units, 14962 words (duration not recorded).

PS0BA ‘Gavin’, 25, photographer, Merseyside, C1, male

PS0BB ‘Sue’, 24, sales assistant, Central Northern England, C1, female

PS0BC ‘Richard’, 24, journalist, Midlands, male

PS0BD ‘Terry’, 26, hairdresser, Central Northern England, male

PS0BE ‘Lindsey’, 24, hairdresser, Central Northern England, female

PS0BF ‘Saranne’, 23, journalist, Midlands, female

PS0BG ‘Adrian’, 25, journalist, Midlands, male

PS0BH ‘Steve’, 25, journalist, Midlands, male

PS0BJ ‘Nick’, 20, office junior, Midlands, male

KC7 7 conversations recorded by ‘Gill’ (PS0BK) on 15 January 1992 with 5 interlocutors, totalling 1640 s-units, 15279 words (duration not recorded).

PS0BK ‘Gill’, 21, waitress, Home Counties, DE, female

PS0BL ‘Jemma’, 22, unemployed, Home Counties, DE, female

PS0BS ‘Nancy’, 18, business student, London, female

PS0BT ‘Jim’, 21, student, Home Counties, male

KC8 11 conversations recorded by ‘Gillian’ (PS0BY) between 29 November and 6 December 1991 with 6 interlocutors, totalling 1822 s-units, 18878 words, and 2 hours 42 minutes 29 seconds of recordings.

PS0BY ‘Gillian’, 32, housewife, East Anglia, AB, female

PS0C1 ‘Robert’, 34, computer programmer, East Anglia, AB, male

PS0C4 ‘None’

PS0C5 ‘Dennis’, 60, retired, London, male

PS0C6 ‘Iris’, 61, retired, London, female

PS0C7 ‘Bonnie’, 40+, housewife, East Anglia, female

KC9 46 conversations recorded by ‘Ginny’ (PS0CG) between 28 November and 6 December 1991 with 9 interlocutors, totalling 7182 s-units, 55039 words, and 6 hours 12 minutes 49 seconds of recordings.

PS0CG ‘Ginny’, 65, housewife, North-east Midlands, AB, female

PS0CH ‘Keith’, 63, design engineer, London, AB, male

PS0CJ ‘Dinda’, 25, customs controller, Home Counties, AB, female

PS0CK ‘Vicky’, 40+, shop assistant, London, female
12. List of works excerpted

PS0CL  ‘None’
PS0CM  ‘Jackie’s shop assistant, female’
PS0CN  ‘Turan’, 50, hair salon owner, Turkish, male
PS0CP  ‘None’
PS0CR  ‘None’

KCA 22 conversations recorded by ‘Gordon’ (PS0DL) between 10 and 13 January 1992 with 8 interlocutors, totalling 3033 s-units, 21431 words (duration not recorded).
PS0DL  ‘Gordon’, 32, unemployed, Welsh, DE, male
PS0DM  ‘Debbie’, 29, housewife, Welsh, DE, female
PS0DN  ‘Hayley’, 29, housewife, Welsh, female
PS0DP  ‘Lyn’, 32, housewife, Welsh, female
PS0DR  ‘Tom’, 56, factory worker, Welsh, male
PS0DT  ‘Sean’, 6, student (state primary), Welsh, DE, male
PS0DU  ‘Kirsty’, 3, student (state pre), Welsh, DE, female

KCB 13 conversations recorded by ‘Graeme’ (PS0DX) between 22 and 26 February 1992 with 7 interlocutors, totalling 2686 s-units, 15326 words (duration not recorded).
PS0DX  ‘Graeme’, 62, company secretary, New Zealand, AB, male
PS0DY  ‘Ann’, 58, housewife, AB, female
PS0E2  ‘Sarah’, 23, sales representative, female
PS0E3  ‘Antony’, 25, car salesman, male
PS0E4  ‘Joyce’, 60+, housewife, female
PS0E5  ‘Hazel’, 53, coffee shop assistant, female
PS0E6  ‘Zara’, 57, coffee shop owner, female

KCC 2 conversations recorded by ‘Hazel’ (PS0FS) on 11 January 1992 with 2 interlocutors, totalling 563 s-units, 5311 words (duration not recorded).
PS0FS  ‘Hazel’, 39, vdu clerk (pt), Welsh, C1, female
PS0F6  ‘Cerys’, 7, schoolgirl (state primary), Welsh, C1, female
PS0F7  ‘Marie’, 66, housewife, Welsh, female

KCD 107 conversations recorded by ‘Helen’ (PS0E8) between 31 May and 1 June 1991 with 7 interlocutors, totalling 5384 s-units, 33104 words, and 5 hours 19 minutes 52 seconds of recordings.
PS0E8  ‘Helen’, 40, laboratory technician, Home Counties, AB, female
PS0E9  ‘Clare’, 11, student (state primary), Home Counties, AB, female
PS0EA  ‘Amy’, 7, student (state primary), Home Counties, AB, female
PS1KL  ‘None’

KCE 24 conversations recorded by ‘Helena’ (PS0EB) between 12 and 20 March 1992 with 9 interlocutors, totalling 7370 s-units, 50776 words, and 5 hours 47 minutes 8 seconds of recordings.
PS0EB  ‘Helena’, 16, student, North-east Midlands, C2, female
PS0EC  ‘Emma’, 16, student, Upper South-west England, female
PS0ED  ‘Sheila’, 41, driving instructor, North-east Midlands, C2, female
PS0EE  ‘David’, 17, student, North-east Midlands, C2, male
PS0EG  ‘Joanne’, 13, student, Upper South-west England, female
PS0EH  ‘Andy’, 18, insurance clerk, Upper South-west England, male
PS0EJ  ‘Scott’, 19, student, Upper South-west England, male
PS0EK  ‘Susan’, 16, student, Upper South-west England, female

KCF 30 conversations recorded by ‘Herbert’ (PS1EM) between 13 and 19 March 1992 with 9 interlocutors, totalling 3755 s-units, 21898 words, and 3 hours 51 minutes 57 seconds of recordings.
PS1EM  ‘Herbert’, 56, landscape gardener, Welsh, C2, male
PS1EN  ‘Iris’, 54, home care assistant, Welsh, C2, female
PS1EP  ‘Alison’, 21, landscape gardener, Welsh, C2, female
PS1ER  ‘Stuart’, 22, landscape gardener, Welsh, C2, female
PS1ES  ‘Gordon’, 72, retired, Welsh, male
PS1ET  ‘Mary’, 76, housewife, Welsh, female
PS1EU  ‘Mike’, 40, driver, Welsh, male
PS1EV  ‘Sally’, 80, retired, Welsh, female
PS1EW  ‘Eddie’, 72, retired, Welsh, male

KCG 20 conversations recorded by ‘Jane’ (PS19L) between 2 and 9 April 1992 with 10 interlocutors, totalling 2998 s-units, 28227 words, and 3 hours 10 minutes 52 seconds of recordings.
PS19L  ‘Jane’, 33, housewife, North-west Midlands, C2, female
PS19M  ‘John’, 38, hgv driver, North-west Midlands, C2, male
PS19N  ‘Stacey’, 7, student (state primary), North-west Midlands, C2, female
PS19P  ‘Chris’, 41, barmaid, North-west Midlands, female
PS19R  ‘Madge’, 65+, housewife, North-west Midlands, female
PS19S ‘Margaret’, 65+, housewife, North-west Midlands, female
PS19T ‘Bill’, 65+, retired, North-west Midlands, male
PS19U ‘Lynn’, 30, housewife, North-west Midlands, female
PS19V ‘Mary’, 64, housewife, North-west Midlands, female

KCH 19 conversations recorded by ‘Jane2’ (PS1BS) between 4 and 10 April 1992 with 7 interlocutors, totalling 6654 s-units, 37385 words, and over 4 hours 46 minutes 52 seconds of recordings.

PS1BS ‘Jane’, 40, optician/student, Humberside, AB, female
PS1BT ‘Phillip’, 46, chartered engineer, Humberside, AB, male
PS1BU ‘Christopher’, 9, student (state primary), Humberside, AB, male
PS1BV ‘David’, 6, student (state primary), Humberside, AB, male
PS1BW ‘Julie’, 27, administrative officer, Central Northern England, female
PS1BX ‘Susan’, 24, teacher, Central Northern England, female
PS6R1 ‘Janet’, 42, clerk, Central Northern England, female

KCM 11 conversations recorded by ‘Jonathan’ (PS0FE) between 15 and 17 January 1992 with 8 interlocutors, totalling 1089 s-units, 7154 words (duration not recorded).

PS0FE ‘Jonathan’, 17, care assistant, Welsh, C2, male
PS11J ‘Corale’, 23, take-away worker, Welsh, C2, female
PS11L ‘Becky’, 7, student, Welsh, female
PS11M ‘Philip’, 25, inspector, Welsh, male
PS11N ‘Julie’, 23, factory worker, Welsh, female
PS11P ‘Rachael’, 3, pre-school, Welsh, female
PS11R ‘Jeff’, 47, panel beater, Welsh, C2, male

KCN 55 conversations recorded by ‘Josephine’ (PS0FF) between 12 and 20 March 1992 with 8 interlocutors, totalling 7580 s-units, 45451 words, and 7 hours 59 minutes 4 seconds of recordings.

PS0FF ‘Josephine’, 54, housewife, Welsh, C1, female
PS0FG ‘Tina’, 84, retired, Welsh, C1, female
PS0FH ‘Danny’, 57, disabled unemployed, Welsh, C1, male
PS0FJ ‘Susan’, 32, housewife, Welsh, female
PS0FK ‘Steve’, 34, ship’s captain, North-east Midlands, male
PS0FL ‘Mark’, 25, assistant manager, Welsh, C1, male
PS0FM ‘Ricky’, 9, student, Welsh, male
PS0FN ‘Sally’, 30, housewife, Welsh, female

KCP 26 conversations recorded by ‘Joy’ (PS0GM) between 13 and 19 March 1992 with 9 interlocutors, totalling 10914 s-units, 77331 words, and 7 hours 4 minutes 13 seconds of recordings.

PS0GM ‘Joy’, 46, housewife, South Midlands, C1, female
PS0GN ‘Connie’, 75, housewife, Home Counties, female
PS0GP ‘Norman’, 79, retired, Midlands, male
PS0GR ‘John’, 72, retired, North-east England, male
PS0GS ‘Ivy’, 69, retired, Home Counties, female
PS0GT ‘Rudy’, 61, warehouse manager, West Indian, C1, male
12. List of works excerpted

PS0GU 'Wendy', 33, housewife, Midlands, female
PS0GV 'Penny', 36, clerk, Midlands, female
PS0GW 'Tracey', 29, housewife, Midlands, female

KCR 2 conversations recorded by ‘Joyce’ (PS1CA) on 4 April 1992 with 4 interlocutors, totalling 577 s-units, 3696 words, and 24 minutes 54 seconds of recordings.
PS1CA 'Malcolm', 43, retired (police officer), Central Northern England, C1, male
PS1CB 'Doreen', 43, housewife, Central Northern England, female
PS1CC 'Judy', 41, clerk, Central Northern England, C1, female
PS1CD 'Joanne', 16, student, Central Northern England, C1, female

PS0GV 'None'

KCU 9 conversations recorded by ‘Julie’ (PS0GF) between 20 and 22 February 1992 with 6 interlocutors, totalling 10696 s-units, 49751 words (duration not recorded).
PS0GF 'Julie', 24, housewife, Lower South-west England, DE, female
PS0GG 'Gary', 25, storesperson, Lower South-west England, DE, male
PS0GH 'Vicki', 2, pre-school, Lower South-west England, DE, female
PS0GJ 'Shelly', 20, housewife, Lower South-west England, female
PS0GK 'Phil', 25, cold store worker, Lower South-west England, male
PS0GL 'Sammy', 4, pre-school, Lower South-west England, male

PS0GA 'Sue', 47, housewife, East Anglia, female
PS0GE 'Ashley', 10, student, East Anglia, male

PS0GF 'None'

KCV 50 conversations recorded by ‘Katherine’ (PS0H7) between 2 and 5 June 1991 with 3 interlocutors, totalling 5727 s-units, 32714 words, and over 4 hours 26 minutes 59 seconds of recordings.
PS0H7 'Katherine', 57, housewife, European (German), AB, female
PS125 'Patrick', 56, company secretary, AB, male
PS126 'Stefan', 25, student, European (German), male

PS0FU 'Wendy', 30, housewife, East Anglia, female
PS0FX 'Sarah', 47, hairdresser, East Anglia, female
PS0GA 'Val', 44, shopkeeper, London, female
PS0GB 'Katt', 43, unemployed, London, male
PS0GH 'None'

KCT 29 conversations recorded by ‘June’ (PS0FP) between 29 November and 5 December 1991 with 17 interlocutors, totalling 14213 s-units, 97005 words, and 9 hours 18 minutes 49 seconds of recordings.
PS0FP 'June', 40, dinner lady (pt), East Anglia, C2, female
PS0FR 'Jonathan', 10, student (state primary), East Anglia, C2, male
PS0FS 'Geoffrey', 40, chargehand, East Anglia, C2, male
PS0FU 'Wendy', 30, housewife, East Anglia, female
PS0FX 'Val', 44, shopkeeper, London, female
PS0GA 'Katt', 43, unemployed, London, male
PS0GE 'Ashley', 10, student, East Anglia, male

PS0GF 'None'

KCW 19 conversations recorded by ‘Kathleen’ (PS0H8) between 15 and 17 January 1992 with 10 interlocutors, totalling 4977 s-units, 23839 words (duration not recorded).
PS0H8 'Kathleen', 15, student (state secondary), Central Midlands, AB, female
PS127 'Christine', 44, teacher, Northern England, AB, female
PS128 'Mary', 13, student (state secondary), Central Midlands, AB, female
PS129 'Susan', 11, student (state secondary), Central Midlands, AB, female
PS12A 'Jackie', 37, dinner-lady, Central Midlands, female
PS12B 'Claire', 11, student, Central Midlands, female
PS12C 'Lorna', 14, student, Central South-west England, female
PS12D 'Helen', 14, student, Central South-west England, female
PS12E 'Amanda', 15, student, Midlands, female

PS0GH 'None'

KCX 25 conversations recorded by ‘Kathleen2’ (PS1FC) between 30 January and 7 February 1992 with 9 interlocutors, totalling 9024 s-units, 60332 words (duration not recorded).
PS125 'Patrick', 56, company secretary, AB, male
PS126 'Stefan', 25, student, European (German), male
PS127 'Christine', 44, teacher, Northern England, AB, female
PS128 'Mary', 13, student (state secondary), Central Midlands, AB, female
PS129 'Susan', 11, student (state secondary), Central Midlands, AB, female
PS12A 'Jackie', 37, dinner-lady, Central Midlands, female
PS12B 'Claire', 11, student, Central Midlands, female
PS12C 'Lorna', 14, student, Central South-west England, female
PS12D 'Helen', 14, student, Central South-west England, female
PS12E 'Amanda', 15, student, Midlands, female

PS0GA 'Sue', 47, housewife, East Anglia, female
PS0GE 'Ashley', 10, student, East Anglia, male

PS0GF 'None'

KCV 50 conversations recorded by ‘Katharine’ (PS0H7) between 2 and 5 June 1991 with 3 interlocutors, totalling 5727 s-units, 32714 words, and over 4 hours 26 minutes 59 seconds of recordings.
PS0H7 'Katherine', 57, housewife, European (German), AB, female
PS125 'Patrick', 56, company secretary, AB, male
PS126 'Stefan', 25, student, European (German), male

PS0FU 'Wendy', 30, housewife, East Anglia, female
PS0FX 'Sarah', 47, hairdresser, East Anglia, female
PS0GA 'Val', 44, shopkeeper, London, female
PS0GB 'Katt', 43, unemployed, London, male
PS0GE 'Ashley', 10, student, East Anglia, male

PS0GF 'None'

KCW 19 conversations recorded by ‘Kathleen’ (PS0H8) between 15 and 17 January 1992 with 10 interlocutors, totalling 4977 s-units, 23839 words (duration not recorded).
PS0H8 'Kathleen', 15, student (state secondary), Central Midlands, AB, female
PS127 'Christine', 44, teacher, Northern England, AB, female
PS128 'Mary', 13, student (state secondary), Central Midlands, AB, female
PS129 'Susan', 11, student (state secondary), Central Midlands, AB, female
PS12A 'Jackie', 37, dinner-lady, Central Midlands, female
PS12B 'Claire', 11, student, Central Midlands, female
PS12C 'Lorna', 14, student, Central South-west England, female
PS12D 'Helen', 14, student, Central South-west England, female
PS12E 'Amanda', 15, student, Midlands, female

PS0GH 'None'

KCX 25 conversations recorded by ‘Kathleen2’ (PS1FC) between 30 January and 7 February 1992 with 9 interlocutors, totalling 9024 s-units, 60332 words (duration not recorded).
PS125 'Patrick', 56, company secretary, AB, male
PS126 'Stefan', 25, student, European (German), male

PS0FU 'Wendy', 30, housewife, East Anglia, female
PS0FX 'Sarah', 47, hairdresser, East Anglia, female
PS0GA 'Val', 44, shopkeeper, London, female
PS0GB 'Katt', 43, unemployed, London, male
PS0GE 'Ashley', 10, student, East Anglia, male

PS0GF 'None'

KCV 50 conversations recorded by ‘Katharine’ (PS0H7) between 2 and 5 June 1991 with 3 interlocutors, totalling 5727 s-units, 32714 words, and over 4 hours 26 minutes 59 seconds of recordings.
PS0H7 'Katherine', 57, housewife, European (German), AB, female
PS125 'Patrick', 56, company secretary, AB, male
PS126 'Stefan', 25, student, European (German), male

PS0FU 'Wendy', 30, housewife, East Anglia, female
PS0FX 'Sarah', 47, hairdresser, East Anglia, female
PS0GA 'Val', 44, shopkeeper, London, female
PS0GB 'Katt', 43, unemployed, London, male
PS0GE 'Ashley', 10, student, East Anglia, male

PS0GF 'None'

KCW 19 conversations recorded by ‘Kathleen’ (PS0H8) between 15 and 17 January 1992 with 10 interlocutors, totalling 4977 s-units, 23839 words (duration not recorded).
PS0H8 'Kathleen', 15, student (state secondary), Central Midlands, AB, female
PS127 'Christine', 44, teacher, Northern England, AB, female
PS128 'Mary', 13, student (state secondary), Central Midlands, AB, female
PS129 'Susan', 11, student (state secondary), Central Midlands, AB, female
PS12A 'Jackie', 37, dinner-lady, Central Midlands, female
PS12B 'Claire', 11, student, Central Midlands, female
PS12C 'Lorna', 14, student, Central South-west England, female
PS12D 'Helen', 14, student, Central South-west England, female
PS12E 'Amanda', 15, student, Midlands, female

PS0GH 'None'

KCX 25 conversations recorded by ‘Kathleen2’ (PS1FC) between 30 January and 7 February 1992 with 9 interlocutors, totalling 9024 s-units, 60332 words (duration not recorded).
words, and over 6 hours 20 minutes 37 seconds of recordings.

PS1FC  ‘Kathleen’, 37, crossing warden, Central Northern England, C2, female
PS1FD  ‘Fló’, 54, youth worker, North-east England, female
PS1FE  ‘Maggie’, 42, housewife, London, female
PS1FF  ‘None’, 40+, taxi driver, Central Northern England, male
PS1FG  ‘Alice’, 40+, smallholder, Central Northern England, female
PS1FH  ‘Steve’, 39, electrician, Central Northern England, C2, male
PS1FI  ‘Krista’, 2, pre-school, Central Northern England, female
PS1FK  ‘Enid’, 30+, Central Northern England, female

KCY 11 conversations recorded by ‘Keith’ (PS0H9) between 10 and 15 January 1992 with 10 interlocutors, totalling 2501 s-units, 24051 words (duration not recorded).

PS0H9  ‘Keith’, 36, telecommunication engineer, Central Midlands, male
PS0HA  ‘Richard’, 33, telecommunication engineer, Central Midlands, male
PS0HB  ‘Russell’, 34, bakery worker, Central Midlands, male
PS0HC  ‘Michael’, 35, engineer, Central Midlands, male
PS0HD  ‘Kathleen’, 42, nurse, Central Midlands, female
PS0HE  ‘Teresa’, 42, engineer, Central Midlands, male
PS0HF  ‘Joseph’, 72, retired, Central Midlands, male
PS0HG  ‘Eileen’, 74, housewife, Central Midlands, female
PS0HH  ‘Dave’, 45, engineer, North-east Midlands, male

KD0 106 conversations recorded by ‘Kevin’ (PS0HM) between 29 November and 5 December 1991 with 14 interlocutors, totalling 13948 s-units, 77692 words, and 10 hours 39 minutes 22 seconds of recordings.

PS0HM  ‘Kevin’, 41, draughtsman, London, C1, male
PS0HN  ‘Paul’, 12, student (state secondary), London, C1, male
PS0HP  ‘Ruth’, 40, teacher, C1, female
PS0HR  ‘Michelle’, 29, local government officer, European (French), female
PS0HS  ‘Eric’, 74, retired, London, male
PS0HT  ‘Adrian’, 40, salesman, London, male
PS0HU  ‘Karen’, 41, secretary, London, female
PS0HV  ‘Andrew’, 33, local government officer, London, male
PS0HW  ‘Lisa’, 13, student, London, female
PS0HX  ‘Babs (aka Mutty)’, 70+, retired, Lower South-west England, female
PS0HY  ‘Joy’, 70+, retired, London, female
PS0J0  ‘Michael’, 15, student, London, male
PS1KN  ‘None’

KD1 14 conversations recorded by ‘Larna’ (PS0JA) between 20 and 24 February 1992 with 8 interlocutors, totalling 5183 s-units, 40487 words, and 4 hours 52 minutes 21 seconds of recordings.

PS0JA  ‘Larna’, 25, housewife, C2, female
PS0JB  ‘Charlotte’, 4, student (state pre), C2, female
PS0JC  ‘Pauline’, 29, housewife, female
PS0JD  ‘Vicky’, female
PS0JE  ‘Alex’, 25, secretary, female
PS0JF  ‘Anthony’, 3, pre-school, C2, male
PS0JG  ‘David’, 45, lorry driver, male

KD2 22 conversations recorded by ‘Linda’ (PS0JI) between 30 January and 2 February 1992 with 9 interlocutors, totalling 3761 s-units, 19822 words (duration not recorded).

PS0JI  ‘Linda’, 20, trainee typist, Central Northern England, C1, female
PS0J2  ‘Alice’, 60+, retired, Central Northern England, female
PS0J3  ‘Dave’, 45, disabled unemployed, Central Northern England, DE, male
PS0J4  ‘Ivy’, 51, housewife, Central Northern England, DE, female
PS0J5  ‘David’, 21, plasterer, Central Northern England, DE, male
PS0J6  ‘Brian’, 25, plasterer, Central Northern England, DE, male
PS0J7  ‘Tracey’, 23, trainee typist, Central Northern England, C1, female
PS0J8  ‘Peter’, 50, salesman, Central Northern England, male
PS0J9  ‘Margaret’, 45, shop assistant, Central Northern England, female

KD3 40 conversations recorded by ‘Lisa’ (PS0JJ) between 12 and 20 March 1992 with 7 interlocutors, totalling 5991 s-units, 33516 words, and 6 hours 30 minutes 17 seconds of recordings.

PS0JJ  ‘Lisa’, 30, housewife, South Midlands, C2, female
PS0JK  ‘Peter’, 0, pre-school, C2, male
PS0JL  ‘Melvin’, 29, self-employed panel beater, C2, male
PS0JP  ‘Diane’, 27, housewife, South Midlands, female
PS0JR  ‘Anne’, 52, moggot farm manager, South Midlands, female
PS0JS  ‘Cheryl’, 27, housewife, female
PS0JT  ‘Ben’, 2, pre-school, male

KD4 7 conversations recorded by ‘Margaret’ (PS0JW) between 13 and 14 January 1992 with 5 interlocutors, totalling
12. List of works excerpted

887 s-units, 7068 words (duration not recorded).

PS0JW ‘Margaret’, 34, teacher (pt), Midlands, AB, female
PS13B ‘Sheila’, 40+, teacher, Midlands, female
PS13A ‘Ben’, 5, student (state primary), Midlands, AB, male
PS13B ‘Katie’, 3, student (state pre), Midlands, AB, female
PS13C ‘Adrian’, 35, environmental health officer, Midlands, AB, male

PS14 55 conversations recorded by ‘Mark’ (PS0JX) between [date unknown] and 26 February 1992 with 25 interlocutors, totalling 9977 s-units, 54036 words, and over 6 hours 21 minutes 48 seconds of recordings.

PS0JX ‘Mark’, 27, technician, C1, male
PS13B ‘Sue’, 27, nurse, C1, female
PS0K0 ‘Dad’, 60, retired, male
PS0K1 ‘Mum’, 50, nurse, European (Dutch), female
PS0K2 ‘Isobel’, 29, secretary, female
PS0K3 ‘Derek’, 60+, male
PS0K4 ‘Pauline’, 60+, female
PS0K5 ‘Spencer’, 30, manager, male
PS0K6 ‘Sally’, 30, teacher, female
PS0K7 ‘Norman’, 28, technician, male
PS0K8 ‘Bedge’, 26, engineer, male
PS0K9 ‘Yun’, 30+, technician, Chinese, male
PS0KA ‘Ian’, 50+, engineer, male
PS0KB ‘Justin’, 20+, technician, male
PS0KC ‘John’, 30+, technician, male
PS0KD ‘Paul’, 20+, technician, male
PS0KE ‘Carol’, 28, clerk, female
PS0KF ‘Dave’, 25, technician, male
PS0KG ‘Daphney’, 40+, secretary, female
PS0KH ‘Collin’, 40+, engineer, male
PS0KJ ‘Andy’, 20+, engineer, Scottish, male
PS0KK ‘Frank’, 40+, manager, European (German), male
PS0KL ‘Paul’, 30+, technician, male
PS0KM ‘George’, 40+, engineer, male

KD6 19 conversations recorded by ‘Martin’ (PS0KN) between 11 and 16 January 1992 with 8 interlocutors, totalling 4721 s-units, 39227 words (duration not recorded).

PS0KN ‘Martin’, 17, student, Midlands, DE, male
PS13K ‘Lynn’, 37, nurse (pt), Midlands, DE, female
PS13L ‘Robert’, 3, student (state pre), Midlands, DE, male
PS13M ‘Dave’, 40, unemployed, Central Southwest England, DE, male
PS13N ‘Geoff’, 13, student (state secondary), Midlands, DE, male
PS13P ‘Rich’, 17, student, Midlands, male
PS13R ‘Scott’, 19, mechanic, Midlands, male
PS13S ‘Sarah’, 20, clerk, Midlands, female

KD7 24 conversations recorded by ‘Matt’ (PS0KP) between 20 February and ?? March 1992 with 6 interlocutors, totalling 3809 s-units, 19206 words (duration not recorded).

PS0KP ‘Matt’, 35, financial advisor, Scottish, C1, male
PS0KR ‘Jan’, 35, building society branch assistant, C1, female
PS0KS ‘Matthew’, 13, student (state secondary), C1, male
PS0KT ‘Laura’, 10, student (state primary), C1, female
PS0KU ‘Christopher’, 7, student (state primary), C1, male
PS0KV ‘Christopher’, male

KD8 31 conversations recorded by ‘Martine’ (PS0LK) between 12 and 20 March 1992 with 10 interlocutors, totalling 10787 s-units, 76445 words, and over 7 hours 15 minutes 1 second of recordings.

PS0LK ‘Martine’, 25, senior technician, Welsh, C1, female
PS0LL ‘Mike’, 28, construction technician, Welsh, C1, female
PS0LM ‘Mirielle’, 55, housewife, Welsh, female
PS0LN ‘None’, 45, pub landlord, Home Counties, male
PS0LP ‘Harold’, 58, engineer, Welsh, male
PS0LR ‘Nora’, 76, housewife, Welsh, female
PS0LS ‘Wll’, 45, civil engineer, Merseyside, male
PS0LT ‘Michael’, 40, technical director, Home Counties, male
PS0LU ‘Jim’, 27, technician, Home Counties, male

KD9 17 conversations recorded by ‘Mark2’ (PS1G2) between [date unknown] and ?? April 1992 with 13 interlocutors, totalling 2788 s-units, 12902 words (duration not recorded).

PS1G2 ‘Mark’, 17, unemployed, Irish, C2, male
PS1G3 ‘Kerry’, 36, housewife, Irish, C2, female
PS1G4 ‘Angela’, 10, student (state primary), Irish, C2, female
PS1G5 ‘Michael’, 12, student (state secondary), Irish, C2, male
PS1G6 ‘Albert’, 16, barman, Irish, male
PS1G7 ‘Robert’, 19, shipyard employee, Irish, male
PS1G8 ‘Leigh’, 17, postman, Irish, male
PS1G9 ‘Debbie’, 15, student, Irish, female
PS1GA ‘Anne’, 16, student, Irish, female
PS1GB ‘Jonston’, 37, Irish, male
PS1GC ‘Ormo’, 21, Irish, male
PS1GD ‘Julianne’, 19, Irish, female
PS6TK ‘None’
KDA 132 conversations recorded by ‘Mark3’ (PS1GE) between 29 November and ?? December 1991 with 11 interlocutors, totalling 8807 s-units, 71717 words, and over 11 hours 8 minutes 38 seconds of recordings.

PS1GE ‘Mark’, 25, aircraft engineer, Home Counties, C2, male
PS1GF ‘Stuart’, 22, aircraft engineer, Home Counties, DE, male
PS1GG ‘Mick’, 21, aircraft engineer, Home Counties, male
PS1GH ‘Chris’, 21, aircraft engineer, Home Counties, male
PS1GI ‘Karl’, 24, aircraft engineer, Welsh, male
PS1GK ‘Nigel’, 26, aircraft engineer, Home Counties, male
PS1GL ‘Graham’, 17, trainee aircraft engineer, Home Counties, male
PS1GM ‘Barry’, 42, leading hand aircraft engineer, Home Counties, male
PS1GN ‘Mick’, 32, sweeper, Home Counties, male
PS1GP ‘Damian’, 21, aircraft engineer, Home Counties, male

KDB 27 conversations recorded by ‘Mikila’ (PS0KY) on 28 November 1991 with 10 interlocutors, totalling 2759 s-units, 16597 words, and 1 hour 55 minutes 44 seconds of recordings.

PS0KY ‘Mikila’, 17, student, Home Counties, C1, female
PS0L0 ‘None’
PS0L1 ‘None’
PS0L2 ‘None’
PS0L3 ‘None’
PS0L4 ‘None’
PS0L5 ‘Kathryn’, 17, student, Home Counties, female
PS0L6 ‘Tabitha’, 17, student, Home Counties, female
PS0L7 ‘Maggi’, 17, student, Home Counties, female

KDC 5 conversations recorded by ‘Michael’ (PS19Y) on 3 February 1992 with 3 interlocutors, totalling 94 s-units, 1062 words, and 9 minutes 34 seconds of recordings.

PS19Y ‘Michael’, 18, sales person, Lancashire, C2, male
PS1A4 ‘Elliot’, 22, inland revenue employee, Central Northern England, male
PS1A5 ‘Jason’, 17, unemployed, Lancashire, male
PS1A6 ‘Kenny’, 23, scientist, Scottish, male

KDD 13 conversations recorded by ‘Murray’ (PS0KW) on 21 February 1992 with 1 interlocutor, totalling 240 s-units, 2206 words (duration not recorded).

PS0KW ‘Murray’, 74, retired (machine operator), Canada, C2, male
PS0KW ‘Vera’, 70, retired, C2, female

KDE 110 conversations recorded by ‘Nicola’ (PS0M4) between 3 and 5 June 1991 with 5 interlocutors, totalling 4112 s-units, 24488 words, and 3 hours 42 minutes 15 seconds of recordings.

PS0M4 ‘Nicola’, 33, housewife, Home Counties, C2, female
PS0M5 ‘Oliver’, 3, student (private pre), London, C2, male
PS13T ‘Bill’, 34, plumber, London, C2, male
PS13U ‘Linda’, 34, housewife, London, female
PS13X ‘Derek’, 63, retired, London, male

KDF 8 conversations recorded by ‘Pamela’ (PS0M6) between 31 January and 4 February 1992 with 4 interlocutors, totalling 349 s-units, 1359 words, and 16 minutes 59 seconds of recordings.

PS0M6 ‘Pamela’, 32, sales assistant, Lancashire, C2, female
PS0M7 ‘Margaret’, 58, market researcher, Scottish, female
PS0M8 ‘Marcus’, 1, pre-school, Lancashire, C2, male
PS0M9 ‘Paul’, 38, bus driver (psv), Lancashire, C2, male

KDG 18 conversations recorded by ‘Patricia’ (PS0MA) between 14 and 16 April 1992 with 10 interlocutors, totalling 2707 s-units, 12125 words (duration not recorded).

PS0MA ‘Patricia’, 31, child minder, Irish, C1, female
PS0MB ‘None’
PS0MC ‘None’
PS0MD ‘David’, 28, cleaning supervisor, Irish, C1, male
PS0ME ‘None’
PS0MF ‘Richard’, 23, shop manager, Irish, male
PS0MG ‘None’
PS0MH ‘None’
PS0MJ ‘None’

KDH 13 conversations recorded by ‘Paul’ (PS0MX) between 10 and 16 April 1992 with 5 interlocutors, totalling 2264 s-units, 8207 words (duration not recorded).

PS0MX ‘Paul’, 32, baker, Irish, C2, male
PS0MY ‘Francis’, 12, student (state secondary), Irish, C2, male
PS0N0 ‘Lemar’, 8, student (state primary), Irish, C2, male
PS0N1 ‘Nora’, 32, stitcher, Irish, C2, female
KDJ 6 conversations recorded by ‘Pauline’ (PS0N3) between 21 and 24 February 1992 with 8 interlocutors, totalling 1668 s-units, 16234 words, and over 1 hour 49 minutes 24 seconds of recordings.

PS0N3 ‘Pauline’, 26, unemployed, North-east England, C1, female
PS0N4 ‘Bob’, 53, sales assistant, Scottish, C1, male
PS0N5 ‘Tracey’, 26, telephonist, Scottish, female
PS0N6 ‘David’, 2, pre-school, male
PS0N7 ‘Kathy’, 55, canvasser, South Midlands, female
PS0N8 ‘Kay’, 24, canvasser, North-east Midlands, female
PS0N9 ‘Bill’, 55, canvasser, Southern (?), male
PS0NA ‘Michelle’, North-east Midlands, female

KDK 18 conversations recorded by ‘Paul2’ (PS1H4) between 29 November and 5 December 1991 with 8 interlocutors, totalling 556 s-units, 3462 words, and 34 minutes 16 seconds of recordings.

PS1H4 ‘Paul’, 34, aircraft dispatcher, London, C1, male
PS1H5 ‘Sarah’, 8, student (state primary), London, C1, female
PS1H6 ‘Lisa’, 6, student (state primary), London, C1, female
PS1H7 ‘Caroline’, 10, student (state primary), London, C1, female
PS1H8 ‘Pam’, 34, aircraft dispatcher, Scottish, female
PS1H9 ‘Ursula’, 32, aircraft dispatcher, London, female
PS1HA ‘Sue’, 32, aircraft dispatcher, London, female

KDL 30 conversations recorded by ‘Rachel’ (PS0NX) between 30 November and 5 December 1991 with 7 interlocutors, totalling 1445 s-units, 8658 words, and 1 hour 34 minutes 25 seconds of recordings.

PS0NX ‘Rachel’, 16, student (private secondary), London, AB, female
PS0NY ‘Barbara’, 65, retired, Upper South-west England, female
PS0P0 ‘Jenny’, 44, doctor, Upper South-west England, AB, female
PS0P1 ‘None’
PS0P3 ‘None’, 48, shop assistant, London, female
PS0P4 ‘None’, 35, shop assistant, London, female

KDM 103 conversations recorded by ‘Raymond’ (PS0PN) between 13 and 19 March 1992 with 46 interlocutors, totalling 17138 s-units, 108739 words, and over 1 hour 50 minutes 5 seconds of recordings.

PS0PN ‘Raymond’, 57, retired, North-west Midlands, C1, male
PS0PP ‘Margaret’, 55, retired, London, C1, female
PS0PR ‘Brian’, 61, technician, North-west Midlands, male
PS0PS ‘Shirley’, 50, housewife, North-west Midlands, female
PS0PT ‘Brian’, 50, photographer, North-west Midlands, male
PS0PU ‘Eric’, 54, retired, North-west Midlands, male
PS0PV ‘Chris’, 45, housewife, North-west Midlands, female
PS0PW ‘John’, 46, driver, Merseyside, male
PS0PX ‘Paula’, 70, housewife, London, female
PS0PY ‘Gwyneth’, 66, housewife, North-west Midlands, female
PS0R0 ‘Bruce’, 35, technician, Merseyside, male
PS0R1 ‘Joan’, 57, tutor, North-west Midlands, female
PS0R2 ‘Joanna’, 31, upholsterer, Home Counties, female
PS0R3 ‘Arthur’, 54, driver, Merseyside, male
PS0R4 ‘Paul’, 36, policeman, North-west Midlands, male
PS0R5 ‘Cathy’, 27, laboratory technician, North-west Midlands, female
PS0R6 ‘Nik’, 11, boys brigade, Welsh, male
PS0R7 ‘Dorothy’, 60, machinist, Welsh, female
PS0R8 ‘Laura’, 6, student, North-west Midlands, female
PS0R9 ‘Scott’, 5, student, North-west Midlands, male
PS0RA ‘Emily’, 8, student, North-west Midlands, female
PS0RB ‘Carrie’, 31, housewife, North-west Midlands, female
PS0RC ‘Ken’, 61, retired, North-west Midlands, male
PS0RD ‘Norman’, 61, joiner, North-west Midlands, male
PS0RE ‘Molly’, 57, housewife, North-west Midlands, female
PS0RF ‘John’, 60, schoolmaster, North-west Midlands, male
PS0RG ‘George’, 45, tutor, Central Midlands, male
PS0RH ‘Nanette’, 43, housewife, European (Dutch), female
PS0RJ ‘Victoria’, 37, housewife, United States, female
PS0RK ‘Arthur’, 61, storeman, South Midlands, male
PS0RL ‘Glyn’, 40, auction worker, Welsh, male
PS0RM ‘Hughes’, 63, chemist, Central Midlands, male
PS0RN ‘Hayley’, 27, chemist, North-west Midlands, female
PS0RP ‘Hilary’, 62, housewife, North-west Midlands, female
312

PS0RR ‘Phyllis’, 42, secretary, North-west Midlands, female
PS0RS ‘Louise’, 22, typist, North-west Midlands, female
PS0RT ‘Tim’, 61, dentist, North-west Midlands, male
PS0RU ‘Ted’, 62, retired, Merseyside, male
PS0RV ‘Pat’, 57, housewife, London, female
PS0RW ‘Mike’, 48, restorer, North-west Midlands, male
PS0RX ‘Tony’, 27, remover, Merseyside, male
PS0RY ‘Alice’, 64, housewife, North-west Midlands, female
PS0S0 ‘Brian’, 45, salesman, North-west Midlands, male
PS0S1 ‘Mary’, 61, housewife, North-west Midlands, female
PS0S2 ‘Teresa’, 42, housewife, North-west Midlands, female

KDN 90 conversations recorded by ‘Raymond2’ (PS1HH) between 15 and 17 April 1992 with 10 interlocutors, totalling 7019 s-units, 43638 words, and over 41 minutes 46 seconds of recordings.

KDP 70 conversations recorded by ‘Richard’ (PS0NB) between 17 and 18 May 1991 with 9 interlocutors, totalling 3210 s-units, 17061 words, and 3 hours 18 minutes 13 seconds of recordings.

KDS 14 conversations recorded by ‘Rosemary’ (PS0NR) between 14 and 16 April 1991 with 6 interlocutors, totalling 2252 s-units, 14407 words (duration not recorded).

PS0GR ‘Raymond’, 53, unemployed, Irish, DE, male
PS0HG ‘Jean’, 45, housewife, Irish, DE, female
PS0HI ‘Paul’, 34, window cleaner, Irish, male
PS0HJ ‘Kylie’, 3, pre-school, Irish, female
PS0HK ‘Kelly Ann’, 10, student, Irish, male
PS0HL ‘Stephen’, 11, student, Irish, male
PS0HM ‘John’, 39, painter, Irish, male
PS0HN ‘Raymond’, 20, painter, Irish, DE, male

KDU 18 conversations recorded by ‘Sam’ (PS0S3) between 15 and 21 April 1992 with 7 interlocutors, totalling 793 s-units, 8038 words, and over 8 minutes 2 seconds of recordings.

KDT 3 conversations recorded by ‘Robert’ (PS1CE) between 2 and 6 April 1992 with 4 interlocutors, totalling 419 s-units, 3098 words, and 1 hour 4 minutes 47 seconds of recordings.

KDV 20 conversations recorded by ‘Sandra’ (PS0SB) between 10 and 17 January 1992 with 9 interlocutors, totalling 5457 s-units, 27333 words, and 3 hours 31 minutes 40 seconds of recordings.

KDTPS000 ‘None’
PS1CE ‘Robert’, 34, self employed general contractor, Central Northern England, C2, male
PS1CF ‘Ellen’, 32, office manager, Central Northern England, C2, female
PS1CG ‘Graham’, 33, gardener, Central Northern England, male

KDU 18 conversations recorded by ‘Sam’ (PS0S3) between 15 and 21 April 1992 with 7 interlocutors, totalling 793 s-units, 8038 words, and over 8 minutes 2 seconds of recordings.

PS0S3 ‘Sam’, 58, company director, Irish, AB, male
PS0S4 ‘Georgie’, 56, clerk, Irish, male
PS0S5 ‘Betty’, 60, Irish, female
PS0S6 ‘Diana’, 51, housewife, Irish, AB, female
PS0S8 ‘Shirley’, 26, bank official, Irish, AB, female
PS0S9 ‘Bill’, 61, retired, Irish, male
PS0SA ‘Margaret’, 55, typist (pt), Irish, female

PS0S1 ‘Jan’, 38, teacher, North-west Midlands, AB, female
PS0S2 ‘Teresa’, 42, housewife, North-west Midlands, female

KDV 20 conversations recorded by ‘Sandra’ (PS0SB) between 10 and 17 January 1992 with 9 interlocutors, totalling 5457 s-units, 27333 words, and 3 hours 31 minutes 40 seconds of recordings.

PS0SB ‘Sandra’, 38, ancillary nurse, Central Midlands, C1, female
PS0SC ‘Kyle’, 9, student (state primary), Central Midlands, C1, male
PS0SD ‘Margaret’, 55, ancillary nurse, Central Midlands, female
PS0SE ‘Penny’, 10+, student, Central Midlands, female
PS0SF ‘R’, 50+, headmistress, Central Midlands, female
PS0SG ‘Tony’, 44, engineer, Central Midlands, C1, male
KDW 27 conversations recorded by 'Sandra2' (PS1C1) between 9 and 16 January 1992 with 25 interlocutors, totalling 8669 s-units, 58393 words (duration not recorded).

KDY 18 conversations recorded by 'Sidney' (PS0SV) [dates unknown] with 8 interlocutors, totalling 1829 s-units, 13318 words, and over 53 minutes 3 seconds of recordings.

KE0 17 conversations recorded by 'Simmone' (PS0SW) between 20 and 27 February 1992 with 9 interlocutors, totalling 5150 s-units, 26574 words, and over 2 hours 15 minutes 23 seconds of recordings.

KE1 26 conversations recorded by 'Stephen' (PS0TU) between 10 and 16 April 1992 with 9 interlocutors, totalling 3949 s-units, 19903 words (duration not recorded).
KE2 153 conversations recorded by ‘Terence’ (PS0W2) between 20 and 27 February 1992 with 10 interlocutors, totalling 10080 s-units, 77961 words, and over 12 hours 49 minutes 22 seconds of recordings.

PS0W2 ‘Terence’, 70, retired (headteacher), East Anglia, AB, male
PS0W3 ‘Richard’, 44, fireman, Lower South-west England, male
PS0W4 ‘Margaret’, 70, retired, Irish, AB, female
PS0W5 ‘Lucy’, 13, student, Lower South-west England, female
PS0W6 ‘Holly’, 13, student, Lower South-west England, female
PS0W7 ‘Adrian’, 13, student, Lower South-west England, male
PS0W8 ‘Danielle’, 13, student, Lower South-west England, female
PS0W9 ‘Christine’, 40, housewife, Lower South-west England, female
PS0WA ‘Mima’, 50, housewife, Lower South-west England, female

KE3 150 conversations recorded by ‘Tony’ (PS0V4) between 28 November and 5 December 1991 with 9 interlocutors, totalling 9314 s-units, 62395 words, and 9 hours 36 minutes 58 seconds of recordings.

PS0V4 ‘Tony’, 37, civil service, Central South-west England, C1, male
PS0V5 ‘Jackie’, 35, civil servant (pt), Central South-west England, C1, female
PS0V6 ‘Christine’, 11, student (state secondary), Central South-west England, C1, female
PS0V7 ‘Rene’, 60+, retired, Central South-west England, female
PS0V8 ‘Mum’, 59, shop assistant, Central South-west England, female
PS0V9 ‘Vicky’, 29, sales assistant, Central South-west England, female
PS0VA ‘Dot’, 40, civil servant, Central South-west England, female
PS0VB ‘Martin’, 37, civil servant, Central South-west England, male
PS0VC ‘Margaret’, 50, civil servant, Central Northern England, female

KE4 22 conversations recorded by ‘Valerie’ (PS0WN) between 30 January and ?? ?? 1992 with 7 interlocutors, totalling 3280 s-units, 15170 words, and 1 hour 55 minutes 21 seconds of recordings.

PS0WN ‘Valerie’, 36, staff nurse (pt), Scottish, C1, female
PS0WP ‘Peter’, 34, sales representative, Scottish, C1, male
PS0WR ‘Jackie’, 8, student (state primary), Scottish, C1, female
PS0WS ‘David’, 10, student (state primary), Scottish, C1, male

PS0WT ‘Dawn’, 11, student, Scottish, female
PS0WU ‘None’, 40+, dentist, Scottish, female
PS0WW ‘None’, 50+, telephone engineer, Scottish, male
PS0WX ‘Dougie’, 37, sales representative, Scottish, male

KE5 4 conversations recorded by ‘Wayne’ (PS0X2) between 20 and 22 February 1992 with 6 interlocutors, totalling 575 s-units, 4800 words (duration not recorded).

PS0X2 ‘Wayne’, 20, unemployed, Central South-west England, DE, male
PS0X3 ‘Michael’, 47, factory operative, Central South-west England, DE, male
PS0X4 ‘Brenda’, 49, factory operative, Central South-west England, DE, female
PS0X5 ‘Martin’, 22, forklift truck driver, Central South-west England, male
PS0X6 ‘Gary’, 24, factory worker, Central South-west England, male
PS0X7 ‘Sam’, 17, student, Central South-west England, female

KE6 59 conversations recorded by ‘Wendy’ (PS0X8) between 21 and 28 February 1992 with 8 interlocutors, totalling 10803 s-units, 70883 words, and 10 hours 46 minutes 58 seconds of recordings.

PS0X8 ‘Wendy’, 54, machine minder, Central South-west England, C2, female
PS0X9 ‘Bev’, 25, unemployed, Central South-west England, C2, female
PS0XA ‘Michael’, 55, production engineer, Central South-west England, C2, male
PS0XB ‘Beth’, 30, office worker, Scottish, female
PS0XC ‘Jill’, 40, office worker, Central South-west England, female
PS0XD ‘Sandra’, 45, machine minder, Home Counties, female
PS0XE ‘Doreen’, 62, machine minder, London, female
PS0XG ‘Dee’, 49, packer, Central South-west England, female

KEB Walsall Local Studies Centre: interview.

PS6MY ‘Joyce’, AB, female
PS6N0 retired, DE, male

Sample containing about 6466 words speech recorded in leisure context

KGH BBC Radio Nottingham: radio broadcast.

PS38 ‘Geoff’, radio presenter, AB, male
PS39 ‘Sue’, Phone-in caller, DE, female
PS38A ‘Teresa’, weather forecaster, DE, female
PS38B 10+, schoolchild, Giving clue for Kids’ County quiz, male
12. List of works excerpted

PS38C 10+, schoolchild, Giving clue for Kids’ County quiz, male
PS38D 10+, schoolchild, Giving clue for Kids’ County quiz, male
PS38E 10+, schoolchild, Giving clue for Kids’ County quiz, male
PS38F 10+, schoolchild, Giving clue for Kids’ County quiz, male
PS38G ‘Trudy’, Quiz phone-in caller, female

Sample containing about 15814 words speech recorded in leisure context

KGS Team Focus - training course (TNT Express Ltd).

PS4NU 58, ‘P. Finister’, trainer, male
PS4NV 30, ‘S. Whitworth’, manager, female
PS4NW 30, ‘M. Bonner’, telesales person, female
PS4NX 30, ‘M. Brome’, sales executive, male
PS4NY 50, ‘J. Hartley’, sales executive, male
PS4P0 30, ‘J. Walton’, telesales person, female
PS4P1 30, ‘L. Birbeck’, telesales person, female
PS4P2 50, ‘D. Inman’, telesales person, male
PS4P3 30, ‘M. Dighton’, sales executive, male

Sample containing about 9330 words speech recorded in business context

KGT Team Focus - training course (TNT Express Ltd).

PS4P8 23, ‘Justin’, sales executive, male
PS4PA 30, ‘B. Singh’, sales executive, male
PS4PB 20+, ‘J. Jowor’, sales executive, male
PS4PC 40+, ‘J. Gibson’, sales executive, female
PS4PD 18, ‘J. Fowles’, sales executive, female
PS4PE 30+, ‘P. Summers’, sales executive, female
PS4PF 20+, ‘P. Ford’, sales executive, male
PS4PG 30+, ‘A. Bond’, sales executive, male

Sample containing about 7368 words speech recorded in business context

KGS Team Focus - training course (TNT Express Ltd).

PS4PS 20+, ‘J. Worrow’, sales executive, male
PS4PT 40+, ‘J. Gibson’, sales executive, female
PS4PV 30+, ‘P. Summers’, sales executive, female
PS4PW 20+, ‘P. Ford’, sales executive, male
PS4PX 30+, ‘A. Bond’, sales executive, male

Sample containing about 16990 words speech recorded in business context

KGM Leicestershire County Council: debate.

PS4DA ‘Beale’, this is a vote where most people say only ‘for’ or ‘against’, therefore they are not listed, male

Sample containing about 16159 words speech recorded in public context

KGN Student seminar on Hunan report.

PS46J , male

Sample containing about 16876 words speech recorded in educational context

KGP Black sheep of the family: seminar.

PS48T 60, ‘Joyce’, student, female
PS48Y 50, ‘Jan’, student, female

Sample containing about 9483 words speech recorded in educational context

KGR Adult education seminar: Justice for all.

PS4DV ‘Frank’, tutor, male

Sample containing about 12112 words speech recorded in educational context

KGS Team Focus - training course (TNT Express Ltd).

PS4NU 58, ‘P. Finister’, trainer, male
PS4NV 30, ‘S. Whitworth’, manager, female
PS4NW 30, ‘M. Bonner’, telesales person, female
PS4NX 30, ‘M. Brome’, sales executive, male
PS4NY 50, ‘J. Hartley’, sales executive, male
PS4P0 30, ‘J. Walton’, telesales person, female
PS4P1 30, ‘L. Birbeck’, telesales person, female
PS4P2 50, ‘D. Inman’, telesales person, male
PS4P3 30, ‘M. Dighton’, sales executive, male

Sample containing about 9330 words speech recorded in business context

KGT Team Focus - training course (TNT Express Ltd).

PS4P8 23, ‘Justin’, sales executive, male
PS4PA 30, ‘B. Singh’, sales executive, male
PS4PB 20+, ‘J. Jowor’, sales executive, male
PS4PC 40+, ‘J. Gibson’, sales executive, female
PS4PD 18, ‘J. Fowles’, sales executive, female
PS4PE 30+, ‘P. Summers’, sales executive, female
PS4PF 20+, ‘P. Ford’, sales executive, male
PS4PG 30+, ‘A. Bond’, sales executive, male

Sample containing about 7368 words speech recorded in business context

KGU Team Focus - training course (TNT Express Ltd).

PS4PN 23, ‘Justin’, sales executive, male
PS4PP 20+, sales executive, female
PS4PR 30, ‘B. Singh’, sales executive, male
PS4PS 20+, ‘J. Jowor’, sales executive, male
PS4PT 40+, ‘J. Gibson’, sales executive, female
PS4PV 18, ‘J. Fowles’, sales executive, female
PS4PV 30+, ‘P. Summers’, sales executive, female
PS4PW 20+, ‘P. Ford’, sales executive, male
PS4PX 30+, ‘A. Bond’, sales executive, male

Sample containing about 16990 words speech recorded in business context

KGV St Dominic’s: lesson.

PS4S0 teacher, no further information given, male
PS4S4 ‘Richard’, student, male
PS4S5 ‘David’, student, male
PS4S6 ‘Margaret’, student, female
PS4S7 ‘Brian’, student, male
PS4S8 ‘Catherine’, student, female

Sample containing about 1735 words speech recorded in educational context

KGW Lecture on child sex abuse.

PS4SD 31, ‘Brown’, lecturer, male

Sample containing about 18451 words speech recorded in educational context

KGX St Luke’s School Council meeting.

PS4TF ‘Hilliard’, councillor, male
12. List of works excerpted

Sample containing about 12269 words speech recorded in public context

**KJS**  *Call Nick Ross - radio phone-in debating programme: live broadcast.*

Sample containing about 9645 words speech recorded in leisure context


Sample containing about 8030 words speech recorded in leisure context

**KLU**  *Albert Gunter: sermon.* Sample containing about 5359 words speech recorded in public context

**KLV**  *General Portfolio management meeting.* Sample containing about 5374 words speech recorded in educational context

**KMW**  *Word-processing tutorial.* Sample containing about 15947 words speech recorded in business context

**KX**  *Careers service meeting.* Sample containing about 14553 words speech recorded in educational context

**KLE**  *Albert Gunter: sermon.* Sample containing about 4266 words speech recorded in public context
PS21L  ‘Bill’, careers advisor, male
PS21M  ‘Sylvia’, careers advisor, female
PS21N  ‘Kevin’, careers advisor, male
PS21P  ‘Cath’, careers advisor, female
PS21R  ‘Diane’, careers advisor, female

Sample containing about 10307 words

speech recorded in business context

KLY  Trade Union Annual Congress.

PS2HY  ‘Dick’, president of trade union, male
PS2J0  ‘Mick’, trade unionist, male
PS2J1  ‘Duncan’, trade unionist, male
PS2J2  ‘John’, trade unionist, male
PS2J3  ‘Sally’, trade unionist, female
PS2J4  ‘Frank’, trade unionist, male
PS2J5  ‘Avril’, trade unionist, female
PS2J6  ‘William’, trade unionist, male
PS2J7  ‘Paul’, trade unionist, male

Sample containing about 13768 words

speech recorded in business context

KM0  Trade Union Annual Congress.

PS2L2  ‘Dick’, president of trade union, male
PS2L3  ‘Tom’, trade unionist, male
PS2L4  ‘Nigel’, trade unionist, male
PS2L5  ‘Theresa’, trade unionist, female
PS2L6  ‘Maureen’, trade unionist, female
PS2L7  ‘Kevin’, trade unionist, male
PS2L8  ‘Sheila’, trade unionist, female
PS2L9  ‘Dorothy’, trade unionist, female
PS2LA  ‘Joanne’, trade unionist, female

Sample containing about 12853 words

speech recorded in business context

KM1  Trade Union Annual Congress.

PS2LL  ‘Denise’, trade unionist, female
PS2LM  ‘Chris’, trade unionist, male
PS2LN  ‘Duncan’, secretary of trades council, male
PS2LP  ‘Mike’, trade unionist, male
PS2LR  ‘Ron’, trade unionist, male
PS2LS  ‘ED’, trade unionist, male
PS2LT  ‘John’, trade unionist, male

Sample containing about 12143 words

speech recorded in business context

KM2  Dennis McCarthy Show: radio broadcast.

PS30C  30+, ‘Rob Tomkinson’, newsreader, male
PS30D  30+, ‘David White’, director of social services, male
PS30E  ‘Mallory Gelder’, reporter
PS30F  head of un operations in bosnia, male
PS30G  50+, ‘Dennis McCarthy’, radio presenter, male
PS30H  40+, ‘Ken’, Phone-in caller, male
PS30I  40+, ‘Angela’, Phone-in caller, female
PS30J  ‘John’, Phone-in caller, male
PS30K  40+, ‘John’, Phone-in caller, male

Sample containing about 13855 words

speech recorded in leisure context


PS312  ‘Geoff’, radio presenter, male
PS313  14+, ‘Lesley’, Phone-in caller, female
PS314  ‘ Sue’, Phone-in caller, female
PS315  ‘June’, Phone-in caller, female
PS316  ‘Sharron’, Phone-in caller, female
PS317  ‘Annie Smith’, reporter, female
PS318  ‘Alison Ford’, newsreader, female
PS319  ‘Nigel Bell’, reporter, male
PS31A  ‘Ray Hilton’, union branch leader, male

Sample containing about 11809 words

speech recorded in leisure context

KM4  British Market Research Bureau monthly meeting.

PS3TP  ‘John’, managing director, male
PS3TR  ‘Stella’, director, female
PS3TS  ‘John’, departmental chairman, male
PS3TT  ‘Richard’, administrator, male
PS3TU  ‘Alan’, director, male
PS3TV  ‘Rita’, director, female
PS3TX  ‘Elizabeth’, personnel director, female

Sample containing about 18325 words

speech recorded in business context

KM5  Abbey Life training session: employee training.

PS4BW  25, ‘Conan’, salesman, male
PS4BX  32, ‘Karen’, salesperson, female
PS4BY  53, ‘Bob’, salesman, male
PS4C0  47, ‘Peter’, male
KM5PS000

Sample containing about 12679 words

speech recorded in business context

KM6  Bristol University history department: employee training.

PS46L  no further information given

Sample containing about 14440 words

speech recorded in educational context

KM7  EIP meeting at Strensall Village Hall, day 7, afternoon session: public county council planning meeting.

PS3YY  50+, ‘Mr F Barnett’, department of the environment adjudicator, male
PS400  40+, ‘Mr David P Lock’, solicitor, from David Lock Associates, male
PS401  40+, ‘Miss D Whittaker’, department of the environment senior inspector, female
PS402  40+, ‘Mr D Potter’, legal representative, for North Yorkshire County Council, male
PS403  30+, ‘Mr D Allenby’, legal representative, for Harrogate Borough Council, male
PS404  50+, ‘Mr Terry Heselton’, legal representative, for Selby District Council, male
PS405  35+, ‘Mr Les J Saunders’, legal representative, for the Department of the Environment, male
PS406  45+, ‘Mr F Broughton’, legal representative, for the Ministry of Agriculture, male
12. List of works excerpted

PS407  45+, ‘Mr P Earle’, legal representative, for Richmondshire District Council, male
Sample containing about 16930 words speech recorded in public context

KM8  Halam Parish Council meeting.
PS6SW  60+, ‘E Bust’, Chairperson, male
PS6SX  30+, ‘P Rickett’, Secretary/clerk, male
PS6SY  55+, ‘Glynis Herbert’, female
PS6T0  55+, ‘R Brown’, male
PS6T1  60+, ‘T Hallam’, male
PS6T2  55+, ‘H Rickett’, male
PS6T3  65+, ‘D Kemp’, male
PS6T4  30+, ‘S Bust’, male
PS6T5  51, ‘John Rose’, retired music teacher, Recording the meeting, male
Sample containing about 11407 words speech recorded in public context

KN1  Royal Courts: hearing.
PS4KB  judge, no further information given, male
KN1PS000
KN1PS001
KN1PS002
KN1PS003
Sample containing about 6708 words speech recorded in public context

KN2  BBC Radio Nottingham: radio broadcast.
PS37G  ‘Martin Fisher’, radio presenter, male
PS37H  football commentator, male
PS37J  ‘Mick Walker’, football manager, male
PS37K  ‘Colin Slater’, football commentator, male
PS37L  football commentator, male
PS37M  football commentator, male
PS37N  ‘Roy Bailey’, football club chairman, male
PS37P  football commentator, male
PS37R  football commentator, male
Sample containing about 15005 words speech recorded in leisure context

KN3  Suffolk County Council Highways and Transport Committee meeting.
PS3SD  40, ‘Guy’, councillor, AB, male
Sample containing about 19356 words speech recorded in public context

KN6  Albert Gunter: sermon.
PS6SU  ‘Albert Gunter’, minister, AB, male
Sample containing about 6359 words speech recorded in public context

KN7  Albert Gunter: sermon.
PS1RG  ‘Albert Gunter’, minister, male
KN7PS000
Sample containing about 6868 words speech recorded in public context

KN8  Albert Gunter: sermon.
PS1RH  ‘Albert Gunter’, minister, male
Sample containing about 5452 words speech recorded in public context

KN9  Albert Gunter: sermon.
PS1RJ  ‘Albert Gunter’, minister, male
Sample containing about 4863 words speech recorded in public context

KNA  Albert Gunter: sermon.
PS1RK  ‘Albert Gunter’, minister, male
Sample containing about 4031 words speech recorded in public context

KNB  Albert Gunter: sermon.
PS1RL  ‘Albert Gunter’, minister, male
Sample containing about 3582 words speech recorded in public context

KNC  Guppy’s Enterprise Club - (invited speaker): lecture/seminar.
PS6LY  50, ‘Neal’, removal man, AB, male
Sample containing about 20576 words speech recorded in leisure context

KND  Maths tutorial.
PS6M0  50, ‘Malcolm’, tutor, male
PS6M1  15, student, female
Sample containing about 4423 words speech recorded in educational context

KNE  Maths tutorial.
PS6M2  50, ‘Malcolm’, tutor, male
PS6M3  15, student, female
Sample containing about 5634 words speech recorded in educational context

KNF  Talk by WPC.
PS26H  ‘Ted’, Main speaker, male
PS26J
PS26K  ‘Heather’, female
PS6SV
KNFPS000
Sample containing about 10320 words speech recorded in public context

KNG  Medical consultation.
PS6M4  general practitioner, male
PS6M5
Sample containing about 568 words speech recorded in public context

KNH  Medical consultation.
PS6M6  general practitioner, male
PS6M7
Sample containing about 520 words speech recorded in public context

KNJ  Medical consultation.
PS6M8  general practitioner, male
PS6M9
| List of works excerpted |  |
|------------------------|--|--|
| **KNK** Medical consultation. |  |
| PS6MA general practitioner, male |  |
| PS6MB |  |
| Sample containing about 460 words |  |
| **KNR** 3 conversations recorded by ‘206’ |  |
| (PS4XN) [dates unknown] with 5 interlocutors, totalling 455 s-units, 2979 words (duration not recorded). |  |
| PS4XP ‘Rachel’, 19, student, Merseyside, female |  |
| PS4XR ‘Jill’, 21, student, East Anglia, female |  |
| PS4XS ‘Sarah’, 19, student, Lower South-west England, C2, female |  |
| PS4XT ‘Lee’, 23, student, Lower South-west England, male |  |
| **KNS** 5 conversations recorded by ‘712’ |  |
| (PS4XU) [dates unknown] with 5 interlocutors, totalling 518 s-units, 4933 words, and over 8 minutes 37 seconds of recordings. |  |
| PS4XU ‘None’student, C1, male |  |
| PS4XV ‘Barry’, 14, student, male |  |
| PS5AP ‘None’ |  |
| PS5AR ‘None’ |  |
| PS5AS ‘None’ |  |
| **KNT** 2 conversations recorded by ‘715’ |  |
| (PS4XW) [dates unknown] with 4 interlocutors, totalling 44 s-units, 291 words, and 2 minutes 31 seconds of recordings. |  |
| KNTPS000 ‘None’ |  |
| PS4XW ‘None’, 13, student, C1, male |  |
| PS4XX ‘Sarah’, 13, student, female |  |
| **KNU** 1 conversation recorded by ‘716’ |  |
| (PS4Y1) [dates unknown] with 2 interlocutors, totalling 11 s-units, 65 words, and 1 minute 22 seconds of recordings. |  |
| PS4Y1 ‘None’, 13, student, AB |  |
| PS4Y2 ‘Mother’, 32, lawyer, AB, female |  |
| **KNV** 12 conversations recorded by ‘717’ |  |
| (PS4Y3) [dates unknown] with 9 interlocutors, totalling 1244 s-units, 7356 words, and over 1 hour 49 minutes 18 seconds of recordings. |  |
| KNVPS000 ‘None’ |  |
| PS4Y3 ‘None’student, C2, male |  |
| PS4Y4 ‘Julian’, 48, teacher, male |  |
| PS4Y5 ‘Marc’, 12, student, male |  |
| PS4Y6 ‘Abbey’, 13, student, male |  |
| PS4Y7 ‘Ado’, 16, teacher (pt), C2, female |  |
| PS4Y8 ‘Dinah’, 41, secretary, C2, female |  |
| PS4Y9 ‘Selassie’, 11, student, C2, male |  |
| **KNW** 3 conversations recorded by ‘727’ |  |
| (PS4YL) [dates unknown] with 4 interlocutors, totalling 117 s-units, 1120 words (duration not recorded). |  |
| PS4YL ‘None’student, C1 |  |
| PS4YM ‘Richard’, 50, teacher, male |  |
| PS4YN ‘Kanti’, 48, civil servant, C1, male |  |
| PS4YP ‘Shashain’, 8, student, C1, male |  |
| **KNY** 9 conversations recorded by ‘Alex’ |  |
| (PS4YX) between [date unknown] and ?? ?? 1993 with 7 interlocutors, totalling 2413 s-units, 13924 words, and over 56 minutes 54 seconds of recordings. |  |
| KNYPS000 ‘None’ |  |
| PS4YX ‘Alex’, 14, student, London, AB, male |  |
| PS4YY ‘Marc’, 14, student, London, male |  |
| PS500 ‘Daniel’, 14, student, male |  |
| PS501 ‘Emma’, 13, student, London, female |  |
| PS6U5 ‘None’ |  |
| PS6U6 ‘None’ |  |
| **KP0** 8 conversations recorded by ‘Alistair’ |  |
| (PS50D) [dates unknown] with 6 interlocutors, totalling 1081 s-units, 7387 words (duration not recorded). |  |
| PS50D ‘Alistair’, 15, student, AB, male |  |
| PS50E ‘James’, 15, student, male |  |
| PS50F ‘Jimmy’, 15, student, male |  |
| PS50G ‘Rupert’, 16, student, male |  |
| PS50J ‘Ben’, 15, student, male |  |
| **KP1** 28 conversations recorded by ‘Arthur2’ |  |
| (PS50T) between 27 February and 2 March 1992 with 7 interlocutors, totalling 9863 s-units, 66500 words (duration not recorded). |  |
| PS50T ‘Arthur’, 44, teacher, Merseyside, C1, male |  |
| PS50U ‘A.’, 70, housewife, North-west Midlands, female |  |
| PS50V ‘Paula’, 43, teacher, North-west Midlands, C1, female |  |
| PS50W ‘Anthony’, 16, student, Northern England, C1, male |  |
| PS50X ‘Paul’, 14, student, Northern England, C1, male |  |
| PS50Y ‘Philip’, 35 North-west Midlands, male |  |
| PS510 ‘John’, 72, retired (engineer), North-west Midlands, male |  |
| **KP2** 11 conversations recorded by ‘Carla’ |  |
| (PS513) [dates unknown] with 8 interlocutors, totalling 1078 s-units, 8570 words, and 1 hour 23 minutes 38 seconds of recordings. |  |
| PS513 ‘Carla’s student, AB, female |  |
| PS514 ‘Pappa’, 13, student, female |  |
| PS515 ‘Gabriel’ teacher, female |  |
12. List of works excerpted

KP3 17 conversations recorded by 'Caroline' (PS51F) on ?? ?? 1993 with 9 interlocutors, totalling 2901 s-units, 17457 words, and over 2 hours 22 minutes 47 seconds of recordings.

KP4 17 conversations recorded by 'Cassie' (PS51S) between [date unknown] and ?? ?? 1993 with 9 interlocutors, totalling 4375 s-units, 32640 words, and over 4 hours 50 minutes 18 seconds of recordings.

KP5 4 conversations recorded by 'Catherine' (PS527) on 16 October 1993 with 3 interlocutors, totalling 4101 s-units, 21997 words, and 1 hour 37 minutes 51 seconds of recordings.

KP6 10 conversations recorded by 'Catrina' (PS52C) on ?? ?? 1993 with 9 interlocutors, totalling 3658 s-units, 33704 words, and over 1 hour 20 minutes 19 seconds of recordings.

KP7 4 conversations recorded by 'Chris3' (PS52N) on 3 April 1992 with 9 interlocutors, totalling 374 s-units, 1806 words, and over 17 minutes 36 seconds of recordings.

KP8 7 conversations recorded by 'Christopher' (PS52T) between 30 January and 8 February 1992 with 6 interlocutors, totalling 3884 s-units, 20442 words, and 2 hours 1 minute 42 seconds of recordings.

KP9 10 conversations recorded by 'Craig' (PS532) [dates unknown] with 9 interlocutors, totalling 1211 s-units, 6496 words (duration not recorded).

KPA 20 conversations recorded by 'Danmy' (PS53C) [dates unknown] with 9 inter-
locutors, totalling 3734 s-units, 26892 words (duration not recorded).

KPAPS000 ‘None’
PS53C ‘Danny’, 13, student, C1, male
PS53D ‘Andrew’, 13, student, male
PS53E ‘Hearn’, teacher, male
PS53F ‘Andrew’, 13, student, male
PS53G ‘Nick’, 13, student, male
PS53H ‘Dhillo’, 13, student, male
PS53J ‘Daniel’, 13, student, male
PS53K ‘Taeke’, 14, student, male

KPB 5 conversations recorded by ‘Eddie’ (PS540) [dates unknown] with 3 interlocutors, totalling 620 s-units, 2738 words (duration not recorded).

KPBPS000 ‘None’
PS540 ‘Eddie’, 15, student, C1, male

KPC 1 conversation recorded by ‘Frances’ (PS543) between 2 April 1991 and [date unknown] with 4 interlocutors, totalling 467 s-units, 1997 words (duration not recorded).

PS543 ‘Frances’, 31, sales assistant, Northern England, AB, female
PS544 ‘Francis’, 6, student (private primary), Northern England, AB, male
PS545 ‘Kaley’, 5, student (private primary), Northern England, AB, female
PS546 ‘Brett’, 1, pre-school, Northern England, AB, male

KPD 5 conversations recorded by ‘Gary’ (PS549) on 30 January 1992 with 5 interlocutors, totalling 806 s-units, 4826 words, and 42 minutes 17 seconds of recordings.

PS549 ‘Gary’, 36, shunter, Scottish, C2, male
PS54A ‘Laura’, 11, student, Scottish, C2, female
PS54B ‘Karen’, 8, student, Scottish, C2, female
PS54C ‘Lilas’, 33, shop assistant, Scottish, C2, female
PS54D ‘Jake’, 73, retired, Scottish, male
PS54E ‘Lilas’, 71, retired, Scottish, female

KPE 17 conversations recorded by ‘Grace’ (PS54G) [dates unknown] with 7 interlocutors, totalling 3199 s-units, 15481 words, and over 1 hour 57 minutes 59 seconds of recordings.

KPEPS000 ‘None’
KPEPS001 ‘None’
PS54G ‘Grace’, 15, student, London, female
PS54H ‘Chantel’, 15, student, London, female
PS54J ‘Fiona’, 16, student, female
PS54K ‘Ian’, 15, student, London, male
PS54L ‘Samantha’, 18, student, female
PS6U2 ‘None’

KPF 22 conversations recorded by ‘Jock’ (PS54T) [dates unknown] with 6 interlocutors, totalling 538 s-units, 3817 words (duration not recorded).

PS54T ‘Jock’, 16, student, AB, male
PS54U ‘James’, 16, student, male
PS54V ‘William’, 13, student, AB, male
PS54Y ‘Rupert’, 16, student, male
PS550 ‘Julian’, 16, student, male

KPG 39 conversations recorded by ‘Josie’ (PS555) [dates unknown] with 34 interlocutors, totalling 6802 s-units, 42330 words, and over 59 minutes 42 seconds of recordings.

PS555 ‘Josie’, 14, student, London, C2, female
PS556 ‘Shelley’, 15, student, London, female
PS557 ‘Kerry’, 15, student, London, female
PS558 ‘Grace’, 15, student, London, female
PS559 ‘Daniel’, 16, student, London, male
PS55A ‘Tirlo’, 13, student, London, male
PS55B ‘Sean’, 12, student, London, male
PS55C ‘Petro’, 12, student, London, male
PS55D ‘Raphael’, 10, student, London, male
PS55E ‘Vergina’, 17, student, London, female
PS55F ‘Tina’, 13, student, London, female
PS55G ‘Greta’ housewife, female
PS55H ‘Jane’ housewife, London, female
PS55J ‘Madigan’ teacher, female
PS55K ‘Alice’ housewife, London, female
PS55L ‘Kate’, 14, student, London, female
PS55M ‘James’, 15, student, London, male
PS55N ‘Peter’, 15, student, London, male
PS55P ‘Chris’ housewife, female
PS55R ‘Ozzi’, 16, student, male
PS55S ‘Wesley’, 15, student, London, male
PS55AG ‘None’ teacher, London, female
PS55AH ‘Cassie’, 15, student, London, female
PS55AJ ‘None’
PS6R4 ‘Jessica’, 12, student, London, female
PS6R5 ‘Andrew’, 15, student, London, male
PS6R6 ‘Karen’, 16, student, London, female
PS6R7 ‘None’, London, male
PS6TS ‘None’
PS6TT ‘None’
PS6TU ‘None’
PS6TY ‘None’

KPH 7 conversations recorded by ‘Kath’ (PS55T) on ?? ?? 1993 with 4 interlocutors, totalling 1692 s-units, 11244 words, and 1 hour 0 minutes 26 seconds of recordings.

PS55T ‘Kath’, 17, student, Home Counties, C2, female
PS55U ‘Claire’, 17, student, Home Counties, female
PS55V ‘Roxy’, 17, student, Home Counties, female
PS6U9 ‘None’
KPJ 7 conversations recorded by ‘Katriane’ (PS560) on 31 January 1992 with 3 interlocutors, totalling 575 s-units, 3559 words, and 25 minutes 53 seconds of recordings.

PS560 ‘Katriane’, 24, catering manager, Scottish, C1, female
PS561 ‘Nina’, 32, canteen assistant, Scottish, female
PS562 ‘Sandy’, 32, catering manager, Scottish, male

KPK 4 conversations recorded by ‘Kitty’ (PS563) between 16 and 19 October 1993 with 9 interlocutors, totalling 1962 s-units, 4901 words, and 1 hour 56 minutes 46 seconds of recordings.

KPKPS000 ‘None’
KPKPS001 ‘None’
KPKPS002 ‘None’
KPKPS003 ‘None’
KPKPS004 ‘None’
KPKPS005 ‘None’
PS563 ‘Kitty’, 23, student, Home Counties, C1, female
PS564 ‘Martin’, 27, athletic events organiser, male
PS565 ‘Steve’, 28, financial adviser, male

KPL 4 conversations recorded by ‘Leon’ (PS56D) [dates unknown] with 4 interlocutors, totalling 866 s-units, 5577 words (duration not recorded).

PS56D ‘Leon’, 13, student, C1, male
PS56E ‘Sarah’, 13, student, female
PS56F ‘Hugo’, 13, student, male

KPM 5 conversations recorded by ‘Madge’ (PS56M) between 8 and 9 April 1992 with 8 interlocutors, totalling 1399 s-units, 9081 words, and 56 minutes 17 seconds of recordings.

KPMPS000 ‘None’
KPMPS001 ‘None’
KPMPS002 ‘None’
KPMPS003 ‘None’
KPMPS004 ‘None’
KPMPS005 ‘None’
KPMPS006 ‘None’
PS56M ‘Madge’, 73, retired, Central Northern England, C1, female

KPN 6 conversations recorded by ‘Marion’ (PS571) [dates unknown] with 9 interlocutors, totalling 599 s-units, 3734 words (duration not recorded).

PS571 ‘Marion’, 22, student, London, AB, female
PS573 ‘Cathi’, 20, student, London, female
PS574 ‘Angie’, 20, student, Central Northern England, female
PS575 ‘None’, 45, lecturer, United States, female
PS576 ‘Rachel’, 20, student, London, female
PS577 ‘Lucy’, 20, student, London, female
PS578 ‘Alice’, 21, student, London, female

KPP 8 conversations recorded by ‘Matthew’ (PS57A) [dates unknown] with 7 interlocutors, totalling 1319 s-units, 7616 words (duration not recorded).

PS57A ‘Matthew’, 13, student, AB, male
PS57B ‘Josh’, 17, student, male
PS57C ‘Ryan’, 18, student, male
PS57D ‘Lara’, 17, student, female
PS57E ‘Rob’, 19, student, male
PS57F ‘Alex’, 13, student, male
PS57G ‘Aaron’, 13, student, male

KPR 5 conversations recorded by ‘Monica’ (PS57L) on 19 October 1993 with 4 interlocutors, totalling 1980 s-units, 9604 words, and over 1 hour 26 minutes 56 seconds of recordings.

PS57L ‘Monica’, 21, student, Home Counties, C1, female
PS57M ‘Andy’, 21, security guard, Home Counties, male
PS57N ‘Nick’, 21, student, Home Counties, male

KPS 1 conversation recorded by ‘Pamela2’ (PS57P) [dates unknown] with 2 interlocutors, totalling 13 s-units, 66 words (duration not recorded).

PS57P ‘Pamela’, 16, trainee warehouse assistant, Scottish, DE, female
PS57R ‘Tracy’, 12, student (state secondary), Scottish, DE, female
PS57S ‘Jill’, 14, student, Scottish, female

KPT 11 conversations recorded by ‘Peter’ (PS57T) [dates unknown] with 7 interlocutors, totalling 1349 s-units, 7008 words, and 51 minutes 5 seconds of recordings.

KPTPS000 ‘None’
KPTPS001 ‘None’
PS57T ‘Peter’, 14, student, London, AB, male
PS57U ‘Tom’, 15, student, London, male
PS57V ‘Muhammad’, 15, student, London, male
PS57W ‘Alex’, 15, student, London, male
PS6RV ‘Danny’, 15, student, London, male

KPU 4 conversations recorded by ‘Rachel2’ (PS582) on 9 October 1993 with 4 interlocutors, totalling 2955 s-units, 16953 words, and over 1 hour 30 minutes 58 seconds of recordings.

PS582 ‘Rachel’, 27, student, Home Counties, C1, female
PS583 ‘Anne-Marie’, 29, secretary, Home Counties, C1, female
PS584 ‘Michael’, 36, p.a., Irish, male
PS585 ‘Gearoid’, 40, p.a., Irish, male

KPV 13 conversations recorded by ‘Rebecca’ (PS586) between 15 and 19 October 1993 with 9 interlocutors, totalling 7959 s-units, 67920 words, and over 5 hours 15 minutes 27 seconds of recordings.

KPWS 000 ‘None’
KPVS 001 ‘None’
KPVS 002 ‘None’
PS586 ‘Rebecca’, 19, student, Home Counties, AB, female

PS5AF ‘Ann’, 19, student, female
PS6RW ‘Anne’; lecturer, female
PS6RX ‘Kitty’, 20, student, female
PS6RY ‘Inga’, 22, student, European (Serbo-Croat), female

KPV 6 conversations recorded by ‘Robert’ (PS58H) [dates unknown] with 6 interlocutors, totalling 1023 s-units, 6745 words, and 15 minutes 16 seconds of recordings.

PS58H ‘Robert’, 15, student, Unidentifiable, C2, male
PS58J ‘Obina’, 15, student, Unidentifiable, male
PS6SM ‘Orgady’, 18, student, female
PS6SN ‘Amanda’, 15, student, C2, female
PS6SP ‘Mum’; receptionist, C2, female

KPVX 7 conversations recorded by ‘Robin’ (PS58K) [dates unknown] with 6 interlocutors, totalling 1126 s-units, 5165 words, and 15 minutes 16 seconds of recordings.

PS58K ‘Robin’, 14, student, AB, male
PS58L ‘Jones’; teacher, male
PS58M ‘Zoe’, 13, student, female
PS58N ‘Ben’, 14, student, male
PS58P ‘Oliver’, 13, student, male
PS58Q ‘Jenny’, 13, student, female

KPY 8 conversations recorded by ‘Sarah’ (PS58U) [dates unknown] with 9 interlocutors, totalling 1080 s-units, 7772 words, and over 1 hour 3 minutes 30 seconds of recordings.

PS58U ‘Sarah’, 13, student, C1, female
PS58V ‘Charitra’, 13, student, female
PS58W ‘Keat-Yee’, 13, student, female
PS58X ‘Roosevelt’, 13, student, male
PS58Y ‘Elizabeth’, 37, housewife, C1, female
PS590 ‘Kate’, 13, student, female
PS591 ‘Emily’, 10, student, C1, female
PS59AN ‘Amy’, 59, retired, female

KR 12 conversations recorded by ‘Sheila’ (PS59B) between 31 January and 4 February 1992 with 9 interlocutors, totalling 2658 s-units, 18900 words, and over 1 hour 22 minutes 19 seconds of recordings.

KRPS 000 ‘None’
KRPS 001 ‘None’
PS59B ‘Sheila’, 52, housewife, North-east England, C1, female
PS59C ‘Wendy’, 50, housewife, Central Northern England, female
PS59D ‘Sharon’, 23, housewife, Central Northern England, female
PS59E ‘Edna’, 49, housewife, North-east England, female
PS59F ‘Paul’, 18, serviceman, Central Northern England, male
PS59G ‘Michael’, 5, student, Central Northern England, male
PS59H ‘Sammy Jo’, 3, student, Central Northern England, female

KR1 6 conversations recorded by ‘Skonev’ (PS59U) [dates unknown] with 7 interlocutors, totalling 728 s-units, 5091 words, and over 8 minutes 56 seconds of recordings.

PS59U ‘Skonev’, 12, student, Home Counties, C1, male
PS59V ‘Blake’, 13, student, male
PS59W ‘Antony’, 13, student, male
PS59X ‘Roland’, 45, lecturer, C1, male
PS59Y ‘Caelia’, 10, student, C1, female
PS59Z ‘Patrick’, 13, student, male

KR2 13 conversations recorded by ‘Terry’ (PS5A1) [dates unknown] with 13 interlocutors, totalling 1655 s-units, 7446 words, and over 13 minutes 18 seconds of recordings.

PS5A1 ‘Terry’, 14, student, London, C2, male
PS5A2 ‘Clifton’, 14, student, male
PS5A3 ‘JO’, 14, student, female
PS5A4 ‘Mother’, 39, housewife, C2, female
PS5A5 ‘Michael’, 1, pre-school, C2, male
PS5A6 ‘Nick’, 14, student, male
PS5A7 ‘Georgina’, 16, student, C2, female
PS5A8 ‘Agi’, 19, student, male
PS5A9 ‘None’, male
PS5AA ‘Damion’, 14, student, male
PS5AB ‘Bunnie’, 14, student, female
PS5AT ‘None’
PS6U4 ‘None’

KRE Ideas in Action programmes: radio broadcast.

PS5R3 ‘A’, AB, male
PS5R4 ‘PD’, DE, male
PS5R5 ‘JT’, DE, female
PS5R6 ‘DG’, male
PS5R7 ‘B’, male
PS5R8 ‘FH’, male
12. List of works excerpted

---

**KRF** Ideas in Action programmes (02): radio broadcast.

- PS5RI ‘PL’, male
- PS5RA ‘CT’, male
- PS5RB ‘JD’, male

Sample containing about 14753 words speech recorded in educational context

**KRL** Bill Heine radio phone-in (02).

- PS5RL ‘PS’, male
- PS5RM ‘PS’, male
- PS5RN ‘PS’, male

Sample containing about 8733 words speech recorded in leisure context

---

**KRG** Ideas in Action programmes (03): radio broadcast.

- PS5ST ‘A’, female
- PS5SB ‘BB’, male
- PS5SA ‘A’, male

Sample containing about 29163 words speech recorded in educational context

---

**KRH** Ideas in Action programmes (04): radio broadcast.

- PS5TH ‘A’, male
- PS5TA ‘GC’, male
- PS5TD ‘GM’, male

Sample containing about 41865 words speech recorded in educational context

---

**KRN** St Aldates in the Civil War [talk].

- PS62F ‘RK’

Sample containing about 6033 words speech recorded in educational context

---

**KRP** Environmental Health Officers’ conference: lecture.

- PS62G ‘A’, male
- PS62H ‘B’, male
- PS62J ‘C’, male

Sample containing about 36816 words speech recorded in public context

---

**KRR** Britain and Europe - European music: radio programme.

- PS63E ‘A’, male
- PS63F ‘MH’

Sample containing about 2142 words speech recorded in educational context

---

**KRS** Britain and Europe - European art: radio programme.

- PS63G ‘A’, male
12. List of works excerpted

Sample containing about 2554 words speech recorded in educational context

**KRT**  *Fox FM News: radio programme.*

PS63H  'NL'
PS631  'A', male  PS63G  'F', male  PS63H  'RJ', male  PS63J  'JB', female  PS63K  'G', male  PS63L  'JM', male  PS63M  'H', male  PS63N  'LR', male  PS63P  'MU', male  PS63R  'HN', male  PS63S  'I', male  PS63T  'DW', female  PS63U  'RH', male  PS63V  'EA', male  PS63W  'JF', male  PS63X  'MM', male  PS63Y  'DF', male  PS640  'NH', male  PS641  'IG', male  PS642  'NC', male  PS643  'MH', male  PS644  'TN', male  PS645  'CG', male  PS646  'LS', male  PS647  'DH', male  PS648  'Zippy', male  PS649  'CB', female  PS64A  'A', female  PS64B  'RR', male  PS64C  'ML', male  PS64D  'AC', female  PS64E  'JZ', female  PS64F  'D', female  PS64G  'RB', male  PS64H  'DS', male  PS64J  'KG', male  PS64K  'AH', female  PS64L  'LM', male  PS64M  'RG', female  PS64N  'E', male  PS64O  'CB', female  PS64P  'AH', male  PS64Q  'BH', male  PS64R  'P', male  PS64S  'AK', female  PS64X  'Q', male  PS64Y  'Bungle', male  PS650  'Jeffrey', male  PS651  'JB', male  PS652  'SH', male  PS653  'SK', male  PS654  'MS', male  PS655  'PS', male  PS656  'RV', female  PS657  'AA', male  PS658  'MH', male  PS659  'JE', female  PS660  'EL', female  PS661  'NY', female  PS662  'JJ', male  PS663  'JW', male  PS664  'NA', male  PS665  'PT', male
12. List of works excerpted

PS67H 'MF', male
PS67J 'PG', male
PS67K 'FD', male
PS67L 'TK', male
PS67M 'TP', male
PS67N 'NM', male
PS67P 'JS', male
PS67R 'GD', male
PS67S 'MR', male
PS67T 'PJ', male
PS67U 'AT', male
PS67V 'GW', male
PS67W 'RG', male
PS67X 'MJ', male
PS67Y 'AP', male
PS680 'PA ', male
PS681 'FJ', male
PS682 'VB', female
PS683 'RK', male
PS684 'B', female
PS685 'GB', male
PS686 'JT', female
PS687 'JK', female
PS688 'HC', female
PS689 'RA ', male
PS68A 'AB', female
PS68B 'PK', male
PS68C 'AG', female
PS68D 'MW', male
PS68E 'MH', female
PS68F 'SP', female
PS68G 'BJ', male
PS68H 'BW', female
PS68I 'PJ', male
PS68J 'PH', male
PS68K 'WT', male
PS68L 'KP', male
PS68M 'PL', male
PS68N 'CW', male
PS68P 'MB', female
PS68R 'KM', male
PS68S 'K', female
PS68T 'SS', female
PS68U 'AJ', male
PS68V 'BG', male
PS68W 'GF', male
PS68X 'RS', male
PS68Y 'NF', female
PS68Z 'CL', female
PS690 'DP', male
PS691 'JS', male
PS692 'PB', male
PS693 'MF', male
PS694 'IJ', male
PS695 'WH', male
PS696 'AK', male
PS697 'C', female
PS698 'JC', male
PS699 'C', female
PS69A 'J', male
PS69B 'AL', male
PS69C 'J', male
PS69D 'BH', male
PS69E 'LC', male
PS69F 'VF', male
PS69G 'LR', male
PS69H 'TT', male
PS69I 'LM', female
<table>
<thead>
<tr>
<th>Code</th>
<th>First Initial</th>
<th>Last Name</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS6BM</td>
<td>'MA'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6BN</td>
<td>'JG'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6BP</td>
<td>'M'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6BS</td>
<td>'JO'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6BT</td>
<td>'LB'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6BU</td>
<td>'DD'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6BV</td>
<td>'AM'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6BW</td>
<td>'NO'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6BX</td>
<td>'DJ'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6BY</td>
<td>'N'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6C0</td>
<td>'EH'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6C1</td>
<td>'BM'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6C2</td>
<td>'O'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6C3</td>
<td>'JH'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6C4</td>
<td>'R'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6C5</td>
<td>'JS'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6C6</td>
<td>'AT'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6C7</td>
<td>'DL'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6C8</td>
<td>'KH'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6C9</td>
<td>'BS'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6CA</td>
<td>'SC'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6CB</td>
<td>'LT'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6CC</td>
<td>'DM'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6CD</td>
<td>'NA'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6CE</td>
<td>'J'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6CF</td>
<td>'ID'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6CG</td>
<td>'CB'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6CH</td>
<td>'RD'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6CI</td>
<td>'RO'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6CJ</td>
<td>'GH'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6CL</td>
<td>'CM'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6CM</td>
<td>'LT'</td>
<td>male</td>
<td></td>
</tr>
</tbody>
</table>

Sample containing about 155729 words speech recorded in educational context.

**KRU Independent Television News.**

<table>
<thead>
<tr>
<th>Code</th>
<th>First Initial</th>
<th>Last Name</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS6DN</td>
<td>'VW'</td>
<td>AB, female</td>
<td></td>
</tr>
<tr>
<td>PS6CP</td>
<td>'T'</td>
<td>DE, male</td>
<td></td>
</tr>
<tr>
<td>PS6CR</td>
<td>'A'</td>
<td>DE, male</td>
<td></td>
</tr>
<tr>
<td>PS6CS</td>
<td>'DJ'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6CT</td>
<td>'RM'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6CU</td>
<td>'B'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6CV</td>
<td>'WM'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6CW</td>
<td>'VA'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6CX</td>
<td>'BN'</td>
<td>female</td>
<td></td>
</tr>
</tbody>
</table>

Sample containing about 17774 words speech recorded in educational context.

**KRV Independent Television News (02).**

<table>
<thead>
<tr>
<th>Code</th>
<th>First Initial</th>
<th>Last Name</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS6FJ</td>
<td>'T'</td>
<td>DE, male</td>
<td></td>
</tr>
<tr>
<td>PS6FE</td>
<td>'B'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6FF</td>
<td>'C'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6FG</td>
<td>'D'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6FH</td>
<td>'E'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6FI</td>
<td>'F'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6FK</td>
<td>'G'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6FL</td>
<td>'RP'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6FM</td>
<td>'MW'</td>
<td>male</td>
<td></td>
</tr>
</tbody>
</table>

Sample containing about 1495 words speech recorded in educational context.

**KRW Interviews with educationalists.**

Sample containing about 11726 words speech recorded in educational context.

**KRX The Land Army: lecture.**

<table>
<thead>
<tr>
<th>Code</th>
<th>First Initial</th>
<th>Last Name</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS6FY</td>
<td>'C'</td>
<td>AB, female</td>
<td></td>
</tr>
<tr>
<td>PS6G0</td>
<td>'DG'</td>
<td>DE, female</td>
<td></td>
</tr>
<tr>
<td>PS6G1</td>
<td>'A'</td>
<td>DE, female</td>
<td></td>
</tr>
<tr>
<td>PS6G2</td>
<td>'B'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6G3</td>
<td>'C'</td>
<td>male</td>
<td></td>
</tr>
<tr>
<td>PS6G4</td>
<td>'D'</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sample containing about 5587 words speech recorded in educational context.

**KRY OUP Electronic Publishing Group: business meeting.**

<table>
<thead>
<tr>
<th>Code</th>
<th>First Initial</th>
<th>Last Name</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS6G5</td>
<td>'AE'</td>
<td>AB</td>
<td></td>
</tr>
<tr>
<td>PS6G6</td>
<td>'RM'</td>
<td>DE</td>
<td></td>
</tr>
<tr>
<td>PS6G7</td>
<td>'JC'</td>
<td>DE</td>
<td></td>
</tr>
<tr>
<td>PS6G8</td>
<td>'TB'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6G9</td>
<td>'SM'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6GA</td>
<td>'PM'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6GB</td>
<td>'TL'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6GC</td>
<td>'A'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6GD</td>
<td>'B'</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sample containing about 6776 words speech recorded in business context.

**KS0 Oxford City Council Health and Environmental Protection Committee meeting.**

<table>
<thead>
<tr>
<th>Code</th>
<th>First Initial</th>
<th>Last Name</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS6H7</td>
<td>'Chair'</td>
<td>AB, female</td>
<td></td>
</tr>
<tr>
<td>PS6H8</td>
<td>'G'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6H9</td>
<td>'Chair2'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6HA</td>
<td>'T'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6HB</td>
<td>'H'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6HC</td>
<td>'Foe'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6HD</td>
<td>'B'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6HE</td>
<td>'A'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6HF</td>
<td>'El'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6HG</td>
<td>'BP'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6HH</td>
<td>'C'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6HI</td>
<td>'D'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6HK</td>
<td>'E'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6HL</td>
<td>'D'</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sample containing about 11314 words speech recorded in public context.

**KS1 Oxford City Council Health and Environmental Protection Committee meeting (02).**

<table>
<thead>
<tr>
<th>Code</th>
<th>First Initial</th>
<th>Last Name</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS6HM</td>
<td>'Chair'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6HN</td>
<td>'G'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6HP</td>
<td>'S'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6HR</td>
<td>'A'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6HS</td>
<td>'M'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6HT</td>
<td>'T'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6HU</td>
<td>'D'</td>
<td>female</td>
<td></td>
</tr>
<tr>
<td>PS6HV</td>
<td>'T'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS6HW</td>
<td>'R'</td>
<td>female</td>
<td></td>
</tr>
</tbody>
</table>
12. List of works excerpted

PS6HX 'B', female
PS6HY 'M', female
PS6J0 'J', male
PS6J1 'AS', female
PS6J2 'O', female
PS6J3 'GD', male
PS6J4 'C', female
PS6J5 'TW', male
PS6J6 'A', male
PS6J7 'LK', female
PS6J8 'NK', female
PS6J9 'L', female
PS6JA 'KD', male
PS6JB 'KP', female
PS6JC 'D', male
PS6JD 'AR', female
PS6JE 'P', male
PS6JF 'GD', female
PS6JG 'CM', female
PS6JH 'EB', male
PS6JJ 'C', male
PS6JK 'K', female
PS6JL 'TF', male
PS6JM 'N', female
PS6JN 'TR'
PS6JP 'PL', male
PS6JR 'RB'
PS6JS 'JC'
PS6JT 'A', female
PS6JU 'BH', male
PS6JV 'TC', male
PS6JW 'CB', male
PS6JX 'CC', male
PS6JY 'FS', female
PS6K0 'H', female
PS6K1 'J', female
PS6K2 'M', female
PS6K3 'H', female
PS6K4 'J', female
PS6K5 'O', female
PS6K6 'E'

Sample containing about 21373 words speech recorded in public context

KS2 Interview with Roger Black.
PS6JP 'PL', male
PS6JR 'RB'

Sample containing about 13700 words speech recorded in educational context

KS3 Lecture on philosophy.
PS6JS 'JC'

Sample containing about 9796 words speech recorded in educational context

KS4 South East Arts Face the Media course: lecture.
PS6JT 'A', female
PS6JU 'BH', male
PS6JV 'TC', male
PS6JW 'CB', male
PS6JX 'CC', male
PS6JY 'FS', female
PS6K0 'RM'

Sample containing about 3572 words speech recorded in educational context

KS5 Minute talks with South East Arts administrators.
PS6K1 'A', male
PS6K2 'FS', female
PS6K3 'B', male
PS6K4 'C', male
PS6K5 'BH', male
PS6K6 'D'

Sample containing about 2677 words speech recorded in educational context

KS6 Telephone system training course.
PS6K7 'SJ', female
PS6K8 'G', female
PS6K9 'M', female
PS6KA 'H', female
PS6KB 'J', female
PS6KC 'K', female
PS6KD 'F', female
PS6KE 'B', male
PS6KF 'T', female
PS6KG 'E', female
PS6KH 'L', female
PS6KJ 'L', male
PS6KK 'AR', female
PS6KL 'A', female
PS6KM 'W', female
PS6KN 'D', female
PS6KP 'E'

Sample containing about 11314 words speech recorded in educational context

PS6KR 'PB', male
PS6KS 'HF', male
PS6KT 'NH', male
PS6KU 'A', male
PS6KV 'JS', male
PS6KW 'NQ', male
PS6KX 'AW', male
PS6KY 'RB', male
PS6L0 'BH', male
PS6L1 'DT', male
PS6L2 'BW', male
PS6L3 'AB', male
PS6L4 'GC', male
PS6L5 'CV', male
PS6L6 'SG', male
PS6L7 'MW', male
PS6L8 'B', male
PS6L9 'C', male
PS6LA 'D', male
PS6LB 'PS', male
PS6LC 'CR', male
PS6LD 'BC', male
PS6LE 'SW', male
PS6LF 'RR', male
PS6LG 'KV', male
PS6LH 'TP', male
PS6LJ 'AM', male
PS6LK 'KC', male
PS6LL 'EJ', male
PS6LM 'NR', male
PS6LN 'GB', male
PS6LP 'TW', male
PS6LR 'TB', male
PS6LS 'AT', male
PS6LT 'MM', male
PS6LU 'AH', male
PS6LY 'RT', male
PS6LW 'AG', male
PS6LX 'ND'

Sample containing about 42602 words speech recorded in leisure context

KS8 Belfast Festival at Queen’s 4-23
12. List of works excerpted

November 1991. u.p.. Sample containing about 24833 words of miscellanea (domain: arts)


KSN 9 conversations recorded by ‘Anthony2’ (PS6NV) [dates unknown] with 4 interlocutors, totalling 2421 s-units, 13743 words (duration not recorded).
- PS6NV ‘Anthony’, 13, student, C1, male
- PS6NW ‘Kevin’, 13, student, male
- PS6NX ‘Patrick’, 10, student, C1, male

KSP 2 conversations recorded by ‘Barry2’ (PS6NY) [dates unknown] with 3 interlocutors, totalling 306 s-units, 1427 words, and over 6 minutes 26 seconds of recordings.
- PS6NY ‘Barry’, 14, student, C2, male
- PS6PD ‘Peter’, 14, student, male
- PS6PI ‘Kevin’, 12, student, male

KSR 7 conversations recorded by ‘Clare2’ (PS6PC) between 12 and 16 October 1993 with 12 interlocutors, totalling 1676 s-units, 9863 words, and over 20 minutes 27 seconds of recordings.
- PS6PC ‘Clare’, 21, student, London, AB, female
- PS6PE ‘Catherine’, 24, lettings negotiator, London, female
- PS6PF ‘Mark’, 20, student, London, male
- PS6PH ‘Dave’, 51, bank inspector, London, male
- PS6PM ‘Chris’, 29, carpenter, London, male
- PS6PN ‘Brian’, 40, stockbroker, London, male
- PS6PP ‘Brian’, 50, banker, male
- PS6PR ‘Jeff’, 50, London, male
- PS6TR ‘None’

KSS 16 conversations recorded by ‘June2’ (PS6R8) between 2 and 8 April 1992 with 9 interlocutors, totalling 5147 s-units, 34975 words, and over 1 hour 19 minutes 18 seconds of recordings.
- PS6R8 ‘June’, 53, housewife, North-west Midlands, DE, female
- PS6R9 ‘Arthur’, 45, unemployed, Lancashire, DE, male
- PS6RA ‘Richard’, 21, radar operator, Northern England, male
- PS6RB ‘Angela’, 19, care assistant, Lancashire, DE, female
- PS6RC ‘Peggy’, 72, salvation army, Lancashire, female
- PS6RD ‘Ernest’, 75, salvation army, Northern England, male
- PS6RF ‘David’, 27, unemployed, Northern England, male

KST 12 conversations recorded by ‘Margaret2’ (PS6RG) between 20 and 27 February 1992 with 7 interlocutors, totalling 5346 s-units, 31800 words, and 2 hours 47 minutes 0 seconds of recordings.
- PS6RG ‘Margaret’, 50, housewife, South-west, AB, female
- PS6TD ‘None’
- PS6TE ‘None’
- PS6TF ‘None’
- PS6TG ‘None’
- PS6TH ‘None’
- PS6TJ ‘None’

KSU 2 conversations recorded by ‘Michael2’ (PS6RM) [dates unknown] with 4 interlocutors, totalling 494 s-units, 2237 words (duration not recorded).
- PS6RM ‘Michael’, 13, student, C2, male
- PS6RN ‘Christine’, 45, credit controller, C2, female
- PS6RP ‘Leslie’, 45, unemployed, female
- PS6RR ‘Steve’, 21, unemployed, C2, male

KSV 25 conversations recorded by ‘Richard2’ (PS1BY) between 21 and 27 February 1992 with 8 interlocutors, totalling 6013 s-units, 38297 words (duration not recorded).
- PS1BY ‘Richard’, 16, student, AB, male
- PS1C0 ‘None’
- PS1K2 ‘None’
- PS1K3 ‘None’
- PS1K4 ‘None’
- PS1K5 ‘None’
- PS1K6 ‘None’
- PS1K7 ‘None’

KSW 6 conversations recorded by ‘Richard4’ (PS6SG) [dates unknown] with 5 interlocutors, totalling 1098 s-units, 6020 words (duration not recorded).
- PS6SG ‘Richard’, 13, student, male
- PS6SH ‘Frances’, 13, student, female
- PS6SJ ‘Lisa’, 13, student, female
- PS6SK ‘Ben’, male
- PS6SL ‘Billy’, male