# Creating a Web-Based Archive of Dialect Recordings

**Jonathan Robinson**
(Curator of English Accents and Dialects, The British Library)

***Abstract:*** *This paper documents an initiative that has brought together over 600 recorded testimonies from two different audio collections for extensive public access via the World Wide Web. Within the umbrella of the NOF-funded Collect Britain project, The Way We Speak combines recordings from two unique and important audio collections held by the British Library Sound Archive. Short extracts from sound recordings from the Survey of English Dialects (© University of Leeds) have been matched to extracts of interviews from the massive oral history archive, The Millennium Memory Bank (© BBC) from an identical or similar geographic location. In addition the site features multiple samples of speakers from urban areas within the MMB. Users can chart the evolution of spoken English over the last fifty years, and the site documents elements both of change and of continuity, celebrating the rich variety of spoken English, both in the past and indeed in the present. In total 681 recordings are now live at http://www.collectbritain.co.uk/collections/dialects/ and form part of a web site aimed at giving a glimpse of the variety of holdings within the British Library, all thematically linked by the concept of 'place'.*

## The British Library Sound Archive

The British Library (BL) is one of the world's finest and largest libraries. With a collection that ranges from books, periodicals and manuscripts to stamps, photographs and music, it provides a resource that attracts some twelve million researchers every year. The British Institute of Recorded Sound, founded in 1955, became part of the BL in 1983. Re-named the National Sound Archive, its remit was (and still is) to collect, preserve and give access to its wide range of holdings. Now simply known as the Sound Archive it holds over 550,000 hours of recorded sound on wax cylinder, 78 rpm coarse-groove disc, open reel magnetic tape, vinyl, audio cassette, CD, mini disc and other digital carriers. The collections come from all over the world and cover the entire range of recorded sound from music, drama and literature, to oral history and wildlife sounds. Examples include Royal Shakespeare Company productions since 1964; BBC Radio Drama since 1930; industrial and

environmental recordings; sound effects for film and theatre; historic and rare spoken voice recordings; and representative samples of music published in the UK and worldwide by voluntary deposit. The big story for 2004 was the rediscovery of the original Edison wax cylinder of Florence Nightingale recorded in July 1890. This is now in the care of the Sound Archive, and the creation of a new dubbing of the recording has been possible.

Information about individual recordings in the collections is available via the Sound Archive's online catalogue at http://www.cadensa.bl.uk. The on-site Viewing and Listening service at the Library's St Pancras site and the Northern Listening Service at Boston Spa have traditionally provided access to the recordings themselves. Although this service will still be available, and is indeed much valued by core users, it is becoming increasingly out-of-date for a global research community and the Library's expanding user base of life-long learners and non-specialist researchers. Until recently online access to actual sound files was limited to 'treasures' of the Sound Archive, such as recordings from *The Voices of the Holocaust* collection. As part of a BL-wide policy to increase access to collections and to reach new constituencies, the Sound Archive has therefore undertaken a programme of web-based digitisation projects, such as the recently completed *Collect Britain* project.

**Collect Britain**

*Collect Britain* is the British Library's largest digitisation project to date. The Library received a £3.2 million grant from the New Opportunities Fund to provide open access for life-long learners to a virtual library of nationally important learning resources relating to places in Britain. Since February 2005 visitors to the site have been able to access a staggering 90,000 samples of images and sounds from the Library's world-renowned collections, without ever needing to visit the prestigious building in London.

The selection has been carefully chosen to include maps, manuscripts, topographical drawings, photographs, rare sound recordings, and even long-forgotten advertisements and music-hall songs that chart the changing face of Britain and its population. Online access has also been provided to two audio collections enabling users to chart the evolution of spoken English over the last fifty years: the Survey of English Dialects (SED) and the Millennium Memory Bank (MMB). Together, they provide a fascinating overview of spoken English during the latter half of the twentieth century. Its rich diversity documents aspects both of continuity and of

change, offering many insights into local history and the fabric of social and working lives.

**The Survey of English Dialects**

The Survey of English Dialects (SED) remains the only systematic survey to have been carried out into the dialects of England. A survey of dialectal English was first considered in the 1930s by Eugen Dieth, Professor of English Language at the University of Zürich, and Harold Orton, then working in the English department at Armstrong College, Newcastle upon Tyne. Little progress was made until the 1940s when Orton, by now installed at the Department of English Language and Medieval English Literature at the University of Leeds, and Dieth believed that the changing social landscape of post-war Britain - particularly the likelihood of increased social and geographical mobility and more widespread access to education and the new broadcast media - would drastically alter the linguistic landscape.

After a series of pilot studies to establish methodology, Orton and Dieth finally settled on a questionnaire as the basis for the survey. The questionnaire comprised over 1300 questions designed to elicit responses from informants which it was felt would best give access to the dialect of a chosen locality. The questions were arranged into nine books, covering such topics as the farm, the human body and social activities, and were chosen so that the responses to them would shed light on the lexical, phonological, morphological and syntactical diversity of spoken English. It was hoped that the results would enable the production of a linguistic atlas of England, and would help to show the historical development of the language. The questionnaire, constantly under revision throughout the survey, was used in conjunction with a number of diagrams and pictures, and the interviews naturally led to passages of spontaneous conversation, which the fieldworkers also duly noted. The questionnaire, however, formed the central part of the interview with an informant in order to ensure the data was comparable between sites and amongst fieldworkers.

With the questionnaire in hand, a team of fieldworkers was assembled and a network of localities identified. The fieldworkers were mainly recruited from undergraduates who had taken their degrees in dialectology under Harold Orton. They all received training from Orton himself and from Dr P.A.D McCarthy of the University of Leeds' Department of Phonetics. This ensured a high degree of internal consistency and continuity in terms of method, practice and ethos, and therefore reliability in terms of results. The selection of the localities was based on a number of criteria: all were in England, and almost all were rural communities, as it was felt that

traditional dialect was best preserved in isolated areas. Preference was given to small communities with a historically stable population. Sites were also chosen to reflect a geographic spread and consideration was given to physical features, such as hills and rivers, which might at one time have formed natural dialect boundaries.

The data was collected over an eleven-year period between 1950 and 1961 in a total of 313 locations. Each fieldworker undertook preliminary research into a chosen locality in order to establish the presence or otherwise of suitable informants. The criteria for selecting informants were crucial to maintain the ultimate goal of comparability. In most cases two or three informants were used and preference was given to older males 'of a restricted social class most likely to have retained the oldest form of local speech' (Ellis 1976: 95). This meant that informants were often farm labourers or workers in allied rural occupations, or in a small number of cases wives of labourers. There is a considerable body of evidence that suggests males are statistically more likely to employ the vernacular, and it was felt that older people with a long-established presence in the community (preferably born of native parents) were the repositories of the traditional dialect of a given speech community. Informants' answers to the questionnaire and any other additional information were written down on recording sheets using a narrow phonetic script. These sheets were then collated and bound into the form of a notebook - one for each location - for analysis by Orton's editorial team during the 1950s and 1960s.

As a result of advances in audio technology during the 1950s, it became increasingly possible to make sound recordings with informants. Initially the cost for such an exercise had been prohibitive, because of the problems associated with electricity and the purchase price of open reel tapes. For the vast majority of locations audio recordings were made on open reel tapes at the same time as the questionnaire; for a small number subsequent visits were made as late as the 1970s, either with the original informant or an informant with a similar profile. Speakers were encouraged to use their most natural form of speech and the interviews were unscripted, unrehearsed and unprepared. 304 recordings were eventually made, which were then edited and dubbed to gramophone discs, and later digitised for public access at The British Library Sound Archive. They vary in length but most last about ten minutes and none are more than thirty minutes in duration. All are spontaneous, informal, and unstructured, covering a myriad of topics but often connected with the speaker's occupation, for example ploughing, harvesting, hedging, pig-killing or bread-making. They are a rich source of social history from a period when very few such recordings with 'ordinary' people were recorded, and still fewer archived.

**The Millennium Memory Bank**

The second collection chosen for the *Collect Britain* audio project is the Millennium Memory Bank (MMB), a joint project between BBC Local Radio and the British Library Sound Archive to create an archival snapshot of 'ordinary' Britons' opinion and experience at the turn of the century. During 1998 and 1999 forty BBC local radio stations across the UK gathered 5429 oral history interviews on mini disc, now catalogued and archived at the British Library. The MMB is one of the largest single oral history collections in Europe, and a unique study of Britain at one particular moment in its history. Sixteen themes were conceived and developed to frame the whole project, including such topics as 'Where We Live', 'Getting Older' and 'Beliefs and Fears'. From the outset, the project sought to focus on local, everyday experiences. Interviewees were encouraged to reflect on events and change at a community level rather than on the wider world stage. Although the primary objective was to record thoughts and attitudes rather than speech patterns, the English spoken has an extremely strong community and place-based resonance. BBC Regional Broadcasting's spread of forty radio stations determined the project's geographical coverage of the United Kingdom, but with thirty-six stations in England, one each in Scotland and Northern Ireland, and two in Wales, no areas were left untouched. A full-time oral history producer and a researcher were appointed for each station on a nine-month contract between September 1998 and June 1999. Amongst the producers some had had oral history experience, most had not. All project staff received some basic background training in oral history methodology, during which an emphasis was placed on the value of longer interviews than many were accustomed to carrying out, and also on the strengths of the life story or biographical approach.

Contributors were either recruited from established groups within the community, such as local history societies, or chosen from respondents to appeals broadcast over the radio. In contrast to the Survey of English Dialects, the Millennium Memory Bank set out to be inclusive: 56% of the contributors were male and 44% female, ranging from 5 to 107 years old and drawn from diverse ethnic and socio-economic backgrounds. The largest occupational groups were school and college students and teachers, followed by farmers (131), nurses, police officers, the unemployed, and secretaries. There were four bishops, five MPs, one lottery winner, a rat catcher, a balloon pilot, a flag maker, a sex therapist, a cricket commentator and one archivist. No geographical area in the UK was unrepresented, although population disparities, for example between the patches served by BBC GLR

(London) and BBC Radio Shropshire, arguably dictate lesser degrees of representation. Most project staff were acutely aware of the need to collect interviews beyond their own audiences and made efforts to target hard-to-reach groups and individuals. A wide range of minority groups was included, among them the aristocracy, the homeless and members of the Traveller community. The result was the 640 half-hour radio programmes in the documentary series, *The Century Speaks* (BBC Radio 4, 1999), broadcast in the final weeks of the Millennium, and an archive of 5429 interviews on mini discs, now catalogued and held in the British Library Sound Archive.

**Collect Britain: The Way We Speak**

*The Way We Speak*, the accents and dialects section of the *Collect Britain* web site, is accessible at http://www.collectbritain.co.uk/collections/dialects/. During the initial phase of the project, the 288 Survey of English Dialects recordings held at the British Library Sound Archive were used as a baseline and cross-matched to an equivalent extract from Millennium Memory Bank interviews. Given the contrasting intentions, scope and scale of the two collections, exact geographical matches were not possible in every case, partly because the MMB is far more representative of Britain's urban population, which was almost entirely ignored by the SED. It was possible, however, to select 267 recordings of individuals recorded for the MMB who were born and brought up either in the same locality or within approximately five miles of the original SED locations. The selection process varied from location to location: in particularly remote areas of the country there might only exist one MMB recording from near an original SED site, and thus this recording had to be used. In other places – most notably near urban areas – it was possible to make a selection from as many as fifty potential MMB interviewees. Dialect speakers from the MMB were favoured, but the SED's gender and occupational bias was not reproduced. Priority was given to speakers who might be said to be representative of their speech community, although an attempt was made to include as broad a cross-section of the population as possible: young and old, female and male, working and middle class.

One of the greatest challenges was in matching speakers from two apparently diverse archives. The SED speakers were, of course, hand-picked dialect speakers with a common socio-economic background. There was a danger in comparing these speakers with a cross-section of speakers from modern Britain that one might present an exaggerated picture of linguistic change. In purely linguistic terms an octogenarian male farm labourer from an isolated 1950s village is not strictly comparable with a young female sixth-form student in a nearby town at the end of the

twentieth century. Nonetheless in attempting to match these speakers over a number of locations, one does get a sense both of linguistic continuity and of change. In addition to the matched pairs of recordings a further 126 MMB audio extracts from key urban centres were added to bring the dialect survey up-to-date and to establish a new baseline for future work.

After an initial audit to select the most suitable recordings, a great deal of attention was given to selecting an appropriate passage for digitisation. The crucial factor here was combining the occasionally competing aspects of content – an interesting, coherent passage that made sense in isolation – and medium – an extract that contained linguistic features, thus enabling a listener to identify a particular speaker with a particular locality. Clearly, in creating a web-based archive of dialect and accent, the latter was given greater importance and every effort was made to select excerpts containing an interesting word, a revealing pronunciation pattern or a noteworthy grammatical construction. Alongside the extracts, new catalogue records were created to provide descriptive content information and comparative data about lexis, phonology and grammar – the words, sounds and structure that define the speech of a given community. Visitors to the site can therefore listen to the recordings with detailed guidance concerning the specific linguistic features that identify a speaker with a particular part of the country. The website is searchable by place-name and subject, and work is still ongoing to provide a map interface to simplify the user's entry point. Most of the original SED recordings are significantly shorter than the MMB recordings, but for each collection five to ten minute audio extracts were selected, edited using Sound Forge digital editing software, stored as .wav files, and then converted to streamed Windows Media for web access.

**Yorkshire**

The following table illustrates the results of the SED/MMB matching process for Yorkshire.

| **SED Localities** | **MMB Sites** |
|---|---|
| Melsonby | Gayles |
| Stokesley | Normanby |
| Skelton | North Skelton and Boosbeck |
| Egton | Whitby |
| Dent | Hawes |
| Muker | Reeth |
| Askrigg | Askrigg |

| | |
|---|---|
| Bedale | Bedale |
| Borrowby | Brawby and Northallerton |
| Rillington | Malton |
| Horton-in-Ribblesdale | No available recording |
| Grassington | Greenhow Hill |
| Pateley Bridge | Ripon |
| Easingwold | Ampleforth |
| Gargrave | Barnoldswick |
| Spofforth | No available recording |
| York | York (x2) |
| Nafferton | Wold Newton |
| Heptonstall | Walsden |
| Wibsey | Bradford (x2) |
| Leeds | Leeds (x2) |
| Cawood | Appleton Roebuck |
| Thornhill | Osset |
| Carleton | South Elmsall |
| Welwick | Withernsea |
| Golcar | Golcar |
| Holmbridge | Holmfirth |
| Skelmanthorpe | Fulstone |
| Ecclesfield | Chapeltown |
| Tickhill | Maltby |
| Sheffield | Sheffield (x3) |

Recordings from the following urban sites have also been added from the MMB archive:

| **Teesside and North Yorkshire** | **West Yorkshire** |
|---|---|
| Billingham | Dewsbury |
| Warrenby | Featherstone |
| Middlesbrough (x4) | Huddersfield |
| Scarborough | **South Yorkshire** |
| **Humberside and East Yorkshire** | Carlton (Barnsley) |
| Old Goole | Cudworth |
| Cherry Burton | Doncaster |
| Hull (x4) | Rotherham |
| | Harthill |

**Outcomes**

The *Way We Speak* web pages enable users both to compare speakers over time, but also to trace the subtle changes to speech patterns as one moves gradually across the country. Above all, the sheer volume of speakers illustrates perfectly that, although spoken English has indeed changed and evolved, in some parts of the country perhaps more rapidly than others, there remains an incredible amount of diversity. Indeed, the website is a celebration of that variety. The project has reconnected the British Library with popular interest in the area of dialects and accents, and the level of international, national, regional, local press, radio and television coverage has been startling. There has been coverage in all the broadsheets, a full page in the *Daily Mirror* (1.9 million readers), *The Daily Mail*, Channel Four News, ITV This Morning, CBBC Exchange, BBC Radios 2, 4 and 5 Live, 27 local radio stations, 22 regional newspapers, and *USA Today* and National Public Radio (22 million listeners every week) in the United States. When web hits are added (41,000 in February 2004 alone and the most visited part of the *Collect Britain* site each month since its launch) this represents a reach of many millions.

Endorsements have been forthcoming from a variety of respected experts in the field, including the chief examiner for English at the Assessments and Qualifications Alliance (AQA), the Head Dialect Coach at the Royal Academy of Dramatic Art (RADA), and senior dialectologists at Leeds University, Lancaster University, University College London, and overseas institutions such as the University of Freiburg in Germany. As a result of press activity after the launch of the updated content in February 2005, the site has been promoted via the Linguistic Association of Great Britain, who in turn informed *The Linguist List* and the EFL (English as a Foreign Language) community, thus giving the BL a presence amongst thousands of specialists across the world.[1] All of this has raised the profile of the dialect holdings at the Sound Archive and re-connected it with traditional partners and users, whilst also opening up the archives to a new audience.

Depending on funding it is hoped that the *Collect Britain* resource can be enhanced in a number of ways. Ideally, the provision of transcripts alongside the sound recordings is envisaged; and additional content from Northern Ireland, Scotland, Wales, ethnic minorities in the UK and even World English could be provided from within the MMB archive. Links have recently been established with the

[1] *The Linguist List* is an online linguistic resource hosted by Eastern Michigan University and Wayne State University, which maintains a mailing list of over 21,000 subscribers worldwide. See *The Linguist List* 2005.

British Library Learning Department's *Texts in Context* project, a web-based initiative sponsored by the Department of Culture, Media and Sport (DCMS), the Department for Education and Skills (DfES), and National and Regional Museums Education Partnerships. This project enables schools and teachers to explore language in context by providing online access to original source material from the BL's collections. These are combined with rich interpretation, case studies of educational projects and suggested learning activities developed in conjunction with AQA. The BL has also recently submitted a proposal to the DfES for funding to create a Sound Portal aimed at secondary schools. If the bid is successful, the Sound Portal will provide online access to a number of additional holdings, such as The Opie Collection of Children's Games and Songs, and will feature interactive resources that enable students to create their own sound recordings and deposit them in a virtual archive or 'Sound Bank'.

In the medium to long-term there are also plans to build on recent successes by re-establishing a programme of acquisition of important recordings in the field of sociolinguistics and dialectology. This is something that has not been possible since the departure of the previous Curator of Accents and Dialects in 1989. Ultimately I would hope that the Sound Archive can engage in actively promoting new fieldwork and even function as a co-ordinating centre for language-related activities within the UK linguistic community. All of this requires the retention of a permanent curatorial post within the Sound Archive. In the current financial climate this looks sadly unlikely, but it is nonetheless worth aiming for given the potential resources available at a national institution such as the British Library.

## References


The British Library (2004). *Collect Britain: Putting History in Place* [Online]. The British Library. Available from the World Wide Web: http://www.collectbritain.co.uk [Accessed 4 March 2005].

_____ (2005). *Texts in Context* [Online]. The British Library. Available from the World Wide Web: http://www.bllearning.co.uk/live/text/dialect/ [Accessed 4 March 2005 via test site].

_____ (2000). *Voices of the Holocaust* [Online]. The British Library. Available from the World Wide Web: http://www.bl.uk/services/learning/curriculum/voices.html [Accessed 4 March 2005].

Ellis, S. (1976). Regional, Social and Economic Influences on Speech: Leeds University Studies. In Viereck, W. (Ed). *Sprachliches Handeln – Soziales Verhalten*. Munich: Fink.

*The Linguist List*. (2005) [Online]. Eastern Michigan University. Available from the World Wide Web: http://www.linguistlist.org [Accessed 4 March 2005].



Edited by Kathryn Jenner, Leeds Archive of Vernacular Culture

University of Leeds

June 2005