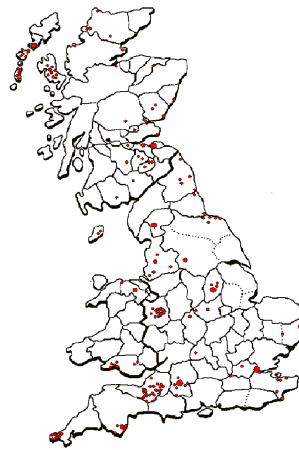


USER'S GUIDE

to

F R E D

Freiburg Corpus of English Dialects



Nuria Hernández

English Dialects Research Group
Albert-Ludwigs-Universität Freiburg
2006

Contents	page
Tables and figures	ii
<i>Acknowledgements</i>	ii
1. General introduction to FRED	1
2. Sources and sampling techniques	2
3. Corpus design	2
3.1 The recordings	3
3.2 The transcripts	3
3.3 Dialect areas	3
3.4 Counties and locations	3
3.5 File names	6
3.6 Speakers	6
4. Text markup and orthography	7
4.1 Format and bracket types	7
4.2 Text headers	8
4.3 Indexing of semi-standardised tags in the running text	9
4.4 Orthography	9
4.5 Rendering of non-standard pronunciations	10
5. Concordance tools	11
6. Storage and documentation	11
6.1 Availability	12
6.2 Copyright restrictions and confidentiality	12
References	13
Notes	14
Appendix	16

Tables and figures

Table 1: Speaker distribution by recording date	3
Table 2: FRED area and county coverage	4
Table 3: Speaker distribution and text production by speaker sex	6
Table 4: Speaker distribution and text production by age group	6
Table 5: Speaker distribution and text production by birth decade	7
Table 6: FRED text list	16
Table 7: FRED speaker list	25
Table 8: Semi-standardised tags	35
Map : FRED locations	5
Figure 1: Header and first lines of a FRED interview	8

Acknowledgements

The Freiburg Corpus of English Dialects FRED was compiled by the project group ‘English Dialect Syntax from a Typological Perspective’ at the English Department of Freiburg University under the supervision of Professor Bernd Kortmann. It was funded by the Deutsche Forschungsgemeinschaft under grant No. Ko-1181/1,2 from 2000 to 2005.

Over the years, the following researchers were involved in the compilation and completion of the corpus:

Prof. Bernd Kortmann (project supervisor), Lieselotte Anderwald (assistant supervisor), Tanja Herrmann, Lukas Pietsch and Susanne Wagner (research assistants, text collection and compilation, transcription, original markup), Nuria Hernández and Benedikt Szmrecsanyi (research assistants, markup, transcription, digitisation, anonymisation, homepage).

We could not have tackled the time-consuming tasks of transcribing and editing the large amount of data without the continuous help of our student assistants; a big thank-you to all of them!

We would also like to thank our research partners: the English departments of the Universities of Leeds, Helsinki and Basel, and the National Centre for English Cultural Tradition at the University of Sheffield, for sharing with us their valuable experience on dialect research.

And most important: thank you to all the dedicated fieldworkers and employees of archives, museums and libraries who have entrusted their data to us, making it possible to build a corpus which will greatly profit the linguistic research on varieties of English spoken in the British Isles.

1. General introduction to FRED

What type of corpus is FRED, and how can linguistic research benefit from it?

FRED is a monolingual spoken-language dialect corpus¹ which contains full-length interviews² with native speakers from England, Wales, Scotland, the Hebrides and the Isle of Man. The texts reflect the ‘traditional’ varieties of British English spoken in these areas during the second half of the 20th century. The corpus consists of sound recordings and orthographic transcripts. Sound files and transcripts are not aligned, but alignment, as well as part-of-speech (POS) tagging, may be carried out in the future.

The data originate from so-called oral history projects where informants were interviewed to record their life memories.³ Oral history interviews provide the quantity of spontaneous speech data that is needed to investigate both frequent and rare syntactic phenomena.⁴ All texts are spontaneous face-to-face conversations between an interviewer and an informant⁵ in a private environment (in most cases the speaker’s home). The interviewers were native speakers themselves and, although the interviewees knew they were being recorded, the whole setting and the interest expressed in their life stories were designed to distract them from their own linguistic behaviour.

The present version of FRED contains approximately 2.5 mill. words⁶ and 300 hours of recorded speech. It consists of 372 interviews with male and female speakers (see Table 3 for gender distribution) from 163 different locations in 43 different counties in 9 major dialect areas. It is an extensive database for both research on single dialects and cross-dialectal studies.

The primary incentive for compiling FRED was the research group’s interest in morphosyntactic variation in British English dialects and the lack of a geographically well-balanced, easy-to-access, machine-readable database.⁷ The wider theoretical framework is based on functional typology, as language-specific variation (in English, in this case) may be integrated into global variation patterns.⁸

FRED can ideally be used to investigate morphological and syntactic phenomena. As far as discourse is concerned, researchers should act with caution, since interviewers usually tried to keep in the background. Their comments and questions clearly aim at making the speakers talk, which can result in longer stretches of monologue.⁹

FRED was not designed as a sociolinguistic corpus, but a limited set of sociolinguistic variables is specified for each text (region, age and gender of the speaker, see 4.2).¹⁰

As the primary research purpose was to investigate grammatical phenomena, the transcripts are orthographic. Nevertheless, sound files are available for most interviews, and researchers interested in phonetic or phonological aspects will find this great collection of recordings useful.

Information on ongoing research as well as sample texts and audio files for the different dialect areas can be found on the FRED homepage (<http://www.anglistik.uni-freiburg.de/institut/lkortmann/FRED>).

2. Sources and sampling techniques

The compilation process of recordings and transcripts for FRED started in 2000, according to the following criteria:

1. the project group's main aim was to investigate morphosyntactic dialect phenomena (predominantly)
2. the goal was to collect traditional dialect data, one reason being the comparability to other, already existing sources on morphosyntactic variation (e.g. the Survey of English Dialect (SED))¹¹
3. tape recordings had to be in acceptable quality (ideally, but not necessarily, including word by word transcripts)
4. the material had to be freely available to the research group and researchers of future generations (for copyright restrictions see 6.2)

Since morphosyntactic phenomena are usually rarer than phonetic phenomena, a large quantity of text was needed, and due to restricted time and resources it was decided to use already recorded material. All of the above criteria suggested one source which is hardly ever used for linguistic or dialectological purposes: oral history interviews.

Tape and mini-disc copies were made of pre-selected original tape recordings made available by various fieldworkers, historians, local museums, libraries and archives from different locations in England, Scotland, Wales, the Hebrides and the Isle of Man. Back at Freiburg University, the tapes were digitised for protection (see 6. below) and stored on DVD. The interviews deemed most suitable for our purposes¹² were then transcribed (either from scratch or revised) by English native speakers and linguistically trained staff.¹³

For a number of texts, transcripts were already available but had to be checked and adapted to the FRED markup. In these cases, there was the obvious advantage of having preliminary transcripts which occasionally included unknown place names and specialist vocabulary (e.g. specialist tools or types of apples used for cider making). Still, the available transcripts had been made for oral history projects. While the normalisation of linguistic utterances and the summarising of interview contents is a perfectly acceptable method for other purposes, it often results in omitting those features which are especially interesting for linguistic research.¹⁴ Pre-existing transcripts were therefore carefully checked against the corresponding sound recordings. Various morphological, syntactic and discourse features (e.g. zero relatives, double negation, repetitions and truncations) were reinserted; paralinguistic features (e.g. laughter) were marked. Irrelevant phonetic and phonological features and instances of eye-dialect, on the other hand, were regularised, except for those which might be relevant from a morphosyntactic perspective (e.g. contracted forms like *gonna*) or those which facilitate word searches (see special case *mi* in 4.4). For a full description of the transcription guidelines see chapter 4.

3. Corpus design

FRED consists of 372 texts, which total approx. 2.5 million words (excluding interviewer utterances) or 300 hours of recorded speech.

3.1 The recordings

All conversations were recorded between 1968 and 2000, the majority during the 1970s and 1980s. The recording date and/or recording decade of each text is specified in the text header (<RECDAT *year*>, <RECDEC *3 digits decade reference*>; see Figure 1 for an example). Most recordings are full-length interviews¹⁵ between 30 and 90 minutes.

recording date	number of speakers	% of all speakers
1961 – 1969	2	0.5%
1970 – 1979	125	29.0%
1980 – 1989	176	40.8%
1990 – 1999	57	13.2%
2000	2	0.5%
unknown	69	16.0%
Total	431	100%

Table 1: Speaker distribution by recording date

3.2 The transcripts

For each recording there is a transcript with the same text identification number (e.g. YKS_001.wav and YKS_001.txt). A number of texts have no source recording but only a source transcript. These source transcripts were only included in FRED if they were word by word transcripts by experienced linguists or fieldworkers with a linguistic background. They were then adapted to the general FRED markup and orthographic guidelines described in chapter 4. Table 6 (Appendix) lists all FRED texts, including word counts. It also indicates for which texts recordings are available (column 5).

3.3 Dialect areas

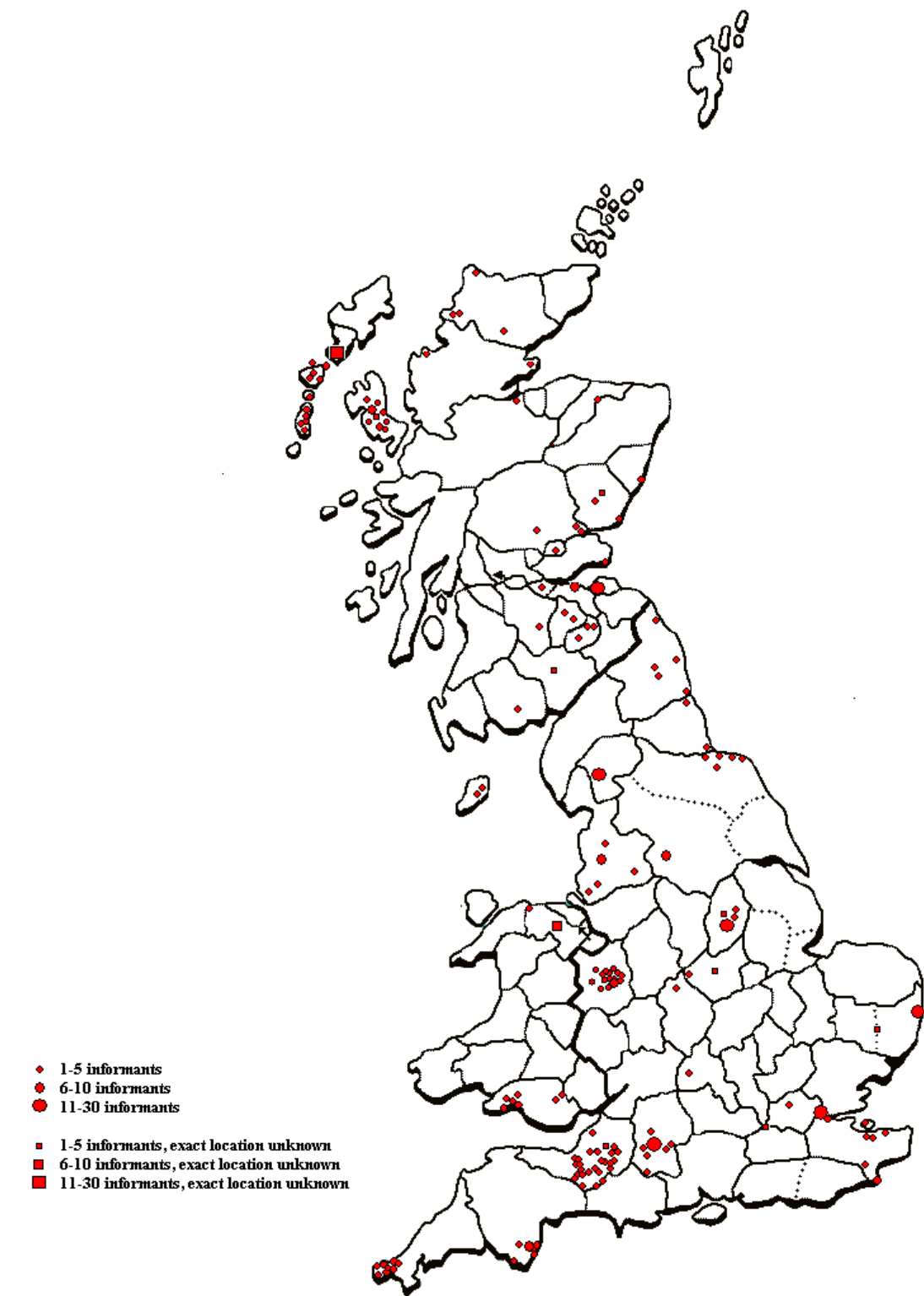
FRED is designed as a dialect corpus. It covers 9 dialect areas,¹⁶ the major 4 of which - England's Southwest, Southeast, Midlands and North - are represented by 350,000 to 625,000 words each. Scotland is divided into 2 areas, Lowlands and Highlands, which amount to a total of almost 200,000 words. The Hebridean Isles are represented by more than 142,000 words, Wales by over 88,000, and, last but not least, the Isle of Man, by almost 10,500 words. The dialect area of each text is specified in the text header (<area *value*>).

3.4 Counties and locations

For a more fine-grained analysis of language use, researchers may want to look at differences between counties. The county of each text is specified in the text header (<county *value*>) and can be easily told from the file name (*LAN_001*, for example, is an interview from Lancashire; see the county codes in Table 2 below). The location of each text is specified in the text header as <location *value*>. The dialect area, county and location of each text are also specified in the FRED text list (Table 6, Appendix), which can be combined with the FRED speaker list (Table 7) to check for speaker details or vice versa.

dialect area/ county	words	% of dialect area	% of corpus	dialect area/ county	words	% of dialect area	% of corpus
SE				ScL			
Kent (KEN)	174,420	27.93 %	6.99 %	Angus (ANS)	17,759	10.10 %	0.71 %
London (LND)	106,549	17.06 %	4.27 %	Banffshire (BAN)	5,468	3.11 %	0.22 %
Middlesex (MDX)	31,048	4.97 %	1.25 %	Dumfriesshire (DFS)	9,282	5.28 %	0.37 %
Suffolk (SFK)	312,414	50.03 %	12.53 %	East Lothian (ELN)	36,032	20.49 %	1.45 %
SE TOTAL	624,431	100.0 %	25.04 %	Fife (FIF)	3,465	1.97 %	0.14 %
SW				Kincardineshire (KCD)	7,092	4.03 %	0.28 %
Cornwall (CON)	102,447	17.40 %	4.11 %	Kinrosshire (KRS)	2,218	1.26 %	0.09 %
Devon (DEV)	92,174	15.65.0 %	3.70 %	Lanarkshire (LKS)	3,599	2.05 %	0.14 %
Oxfordshire (OXF)	14,043	2.38 %	0.56 %	Midlothian (MLN)	30,123	17.13 %	1.21 %
Somerset (SOM)	204,239	34.68 %	8.19 %	Peebleshire (PEE)	14,584	8.29 %	0.58 %
Wiltshire (WIL)	176,028	29.89 %	7.06 %	Pertshire (PER)	20,545	11.69 %	0.82 %
SW TOTAL	588,931	100.0 %	23.62 %	Selkirkshire (SEL)	9,032	5.14 %	0.36 %
Mid				West Lothian (WLN)	16,620	9.45 %	0.67 %
Leicestershire (LEI)	8,311	2.37 %	0.33 %	ScL TOTAL	175,819	100.0 %	7.05 %
Nottinghamshire (NTT)	160617	45.72 %	6.44 %	ScH			
Shropshire (SAL)	174,180	49.58 %	6.99 %	Inverness-shire (INV)	3,126	13.09 %	0.13 %
Warwickshire (WAR)	8,176	2.33 %	0.33 %	Ross and Cromarty (ROC)	10,131	42.44 %	0.41 %
Mid TOTAL	351,284	100.0 %	14.09 %	Sutherland (SUT)	10,615	44.47 %	0.4 3%
N				ScH TOTAL	23,872	100.0 %	0.96 %
Durham (DUR)	26,602	5.46 %	1.07 %	Heb TOTAL	142,682	100.0 %	5.72 %
Lancashire (LAN)	195,111	40.04 %	7.82 %	Wal			
Northumberland (NBL)	28,432	5.83 %	1.14 %	Denbighshire (DEN)	37,284	42.01 %	1.50 %
Westmorland (WES)	149,747	30.73 %	6.01 %	Glamorgan (GLA)	51,471	57.99 %	2.06 %
Yorkshire (YKS)	87,389	17.93 %	3.50 %	Wal TOTAL	88,755	100.0 %	3.56 %
N TOTAL	487281	100.0 %	19.54 %	Man TOTAL	10,461	100.0 %	0.42 %

Table 2: FRED area and county coverage¹⁷



Map: FRED locations¹⁸

3.5 File names

One look at a file name will tell you which county the text is from. FRED file names are composed of a three letter Chapman county code, followed by underscore and a running number (001, 002, ...).¹⁹ Corresponding text and sound files have the same identification, e.g. DEV_004.txt and DEV_004.wav. Note that, for longer interviews, there may be more than one wav-file (e.g. NTT_006A.wav and NTT_006B.wav).

3.6 Speakers

FRED is a traditional-dialects corpus which aims at regional, not social, representativeness.

About two thirds of the informants are so-called NORMs - non-mobile old rural males - who typically left school at age fourteen or younger.²⁰ The ratio of male to female speakers is roughly 2:1 (with 277 male and 132 female informants). 77.8% of the overall textual material in FRED was demonstrably produced by male speakers, and 21.3% by female speakers.²¹

	number of speakers	% of speakers	% text production
Male	277	64.3 %	77.8 %
Female	132	30.6 %	21.3 %
Unknown	22	5.1 %	0.9 %
Total	431	100 %	100 %

Table 3: Speaker distribution and text production by speaker sex

One feature of oral history interviews which is a crucial prerequisite for investigating dialects in their 'original' form is that speakers spent their life in one specific geographic area (region/county/location) and did not leave it for any considerable period of time. This is true for most FRED speakers.

The corpus contains speech data from 431²² speakers (interviewers excluded) who all grew up in Britain. Most of them were born before World War I/II and were aged 60 or over when interviewed. About three quarters of the textual material in FRED was produced by the 60+ age group. The younger speakers included in the corpus produced only a small percentage of the overall textual material. The mean age is 75.2 years at recording date. A breakdown of the amount of text produced by the different age groups is presented in Table 4. For a breakdown according to date of birth see Table 5 below.

age group	number of speakers	words	% text production
0 – 14	9	7,501	0.30 %
15 – 24	14	30,336	1.22 %
25 – 34	2	5,312	0.21 %
35 – 44	4	10,633	0.43 %
45 – 59	17	108,472	4.34 %
60+	239	1,866,202	74.74 %
unknown	146	468,307	18.76 %

Table 4: Speaker distribution and text production by age group

The oldest FRED informant was born in 1877 (aged 102 at the time of the interview); 89% of all informants were born before 1920. Information on age, birth date/decade, sex and interview date/decade for all speakers can be obtained from the FRED speaker list (Table 7, Appendix).

decade of birth	number of speakers	% of speakers	words	% text production
1870-1879	1	0.23 %	7,003	0.28 %
1880-1889	13	3.02 %	131,252	5.26 %
1890-1899	62	14.4 %	468,874	18.78 %
1900-1909	101	23.4 %	823,686	32.99 %
1910-1919	64	14.8 %	464,741	18.61 %
1920-1929	24	5.57 %	174,839	7.00 %
1930-1939	5	1.16 %	31,748	1.27 %
1940-1949	3	0.70 %	8,878	0.36 %
1950-1959	2	0.47 %	5,311	0.21 %
unknown	156	36.20 %	380,431	15.24 %
TOTAL	432	100.0 %	2,496,763	100.0 %

Table 5: Speaker distribution and text production by birth decade

All the criteria affecting the choice of informants, i.e. their age, strong affiliation with their native region, relatively low education level and little mobility, have resulted in a largely homogeneous speaker profile which, although restricting the use of FRED for sociolinguistic studies, facilitates the comparison with former dialect surveys (e.g. the SED).

4. Text markup and orthography

The basic idea was to produce a dialect corpus approximating the format of already existing language corpora (e.g. the British National Corpus (BNC)) and compatible with the text retrieval system TACT.²³ A consistent markup was applied to both the texts transcribed at Freiburg University and pre-existing transcripts from other sources. FRED is an orthographically transcribed corpus.²⁴ The decision in favour of orthographic transcripts was made for a number of reasons, the most important being the project group's predominant interest in morphosyntactic features.²⁵

4.1 Format and bracket types

All transcriptions are pure ASCII texts with no text-processor formatting. A number sign (#) marks the start of each sentence ("sentence" in an intuitive sense with no particular analytic meaning) – a simple means of dividing speech into bits for reference purposes. All markup is separated from the running text by brackets.

- [] square brackets for non-standardised markup such as header information, page references of source transcripts, tape position references and editorial notes
- < > angular brackets for text and speaker information
- { } curly brackets for interviewer utterances
- () round brackets for semi-standardised tags (see 4.3 below)

All speaker utterances start with <u *speaker ID*>. Interviewer utterances - which are part of the recorded dialogue but might be less relevant for some studies - read {<u *interviewer ID*> *text*>. The curly brackets are an additional markup feature that allows researchers to exclude interviewer utterances from word searches in concordance programmes like WordSmith. The utterance tags clearly signal turn-taking between the speakers involved in the conversation.

4.2 Text headers

Each text is preceded by a header. Figure 1 shows the beginning of an interview with two informants in Swarland, Northumberland, in 1974.

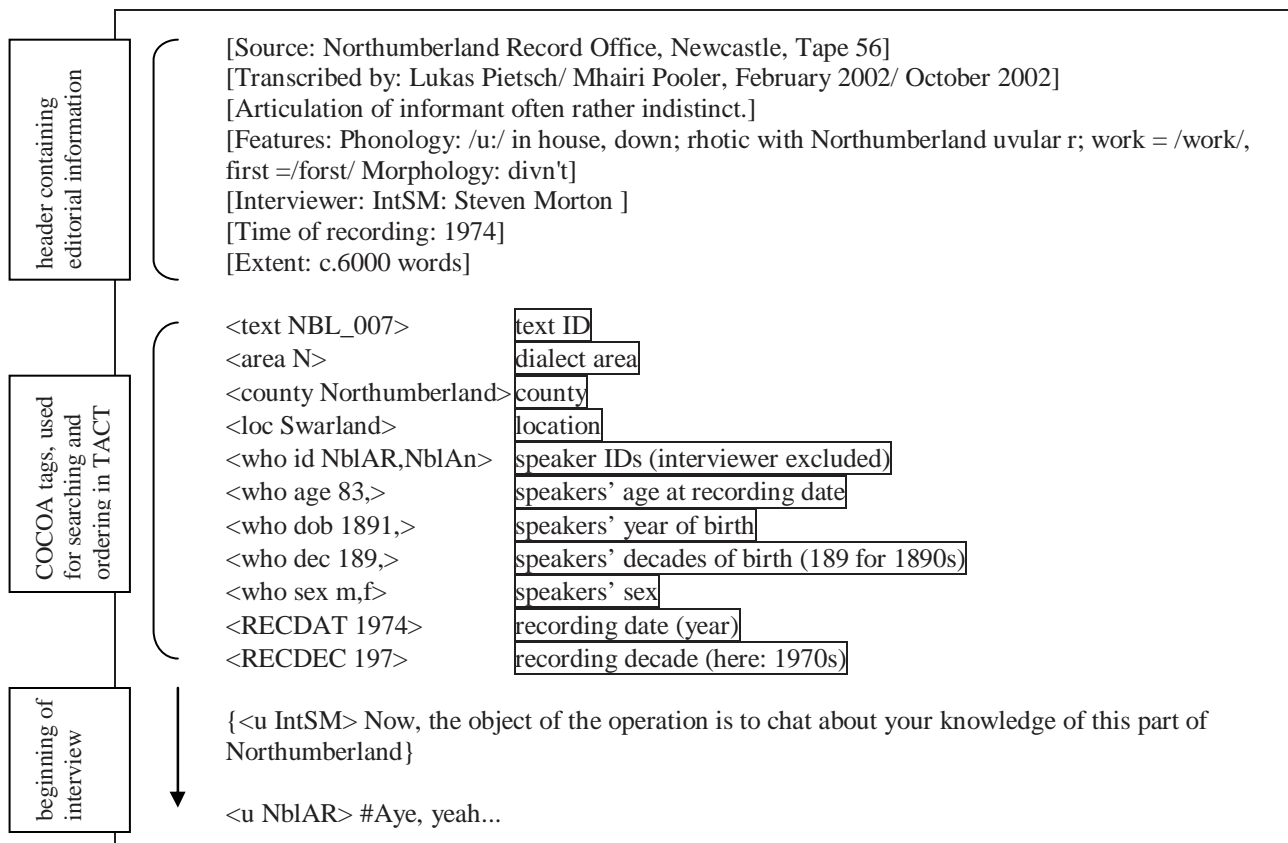


Figure 1: Header and first lines of a FRED interview

The first part of this header, in square brackets, contains the most important editorial information about the text, such as a full descriptive title of the text file, copyright information, source, place and date of the interview, and in some cases a short description of the speakers' accent and dialect features.²⁶ In most cases, some information on the interviewer is also included.²⁷

The second part of the header consists of a set of standardised COCOA-type tags in angular brackets (<tagname value>) which specify the text ID, dialect area, county and location, the recording date and decade, as well as the speaker ID, age (at recording date), year of birth, decade of birth (e.g. 189 for 1890s) and sex (*f* or *m*). The simple reason for adding a separate tag for recording decade and decade of birth is that, in various cases, the exact recording date of the interview and/or year of birth of the speaker could not be deduced from the text.

Some interviews were made with more than one informant, in which case the tags contain information on all speakers, separated by single commas. Information is given where available; otherwise, the value is empty.

4.3 Indexing of semi-standardised tags in the running text

Round brackets are used for a number of semi-standardised tags; they cannot be used for automatic searches in TACT but can be easily transformed into standardised SGML tags. These tags come as either start-end pairs (xxx) ... (/xxx) with a space after the start-tag and a space before the end-tag, single tags containing a tag name and a value (xxx 'yyy') or single tags containing a tag name, an attribute name and an attribute value (xxx yyy=zzz). They indicate the occurrence of non-verbal sounds (such as laughs and coughs), pauses and truncations, and were used for editorial corrections, orthographic regularisations and uncertain or missing words. They were also used in the anonymisation process (see 6.2). A complete list of the semi-standardised tags used in FRED is given in Table 8.

4.4 Orthography

FRED transcripts almost read like normal orthographic writing - with a few exceptions:

Upper case: as in standard orthography, upper case is used for proper names and at the beginning of sentences. It is also used to indicate indirect speech (see below).

Acronyms: spoken acronyms (such as *USA* or *TV*) maintain their usual form.

Abbreviations: written abbreviations (such as *etc.* or *vs.*) are spelled out (*etcetera*, *versus*), with the exception of titles (*Mr*, *Mrs*, *Dr* – all without dots).

Numbers: numbers, including measurements and monetary units, are spelled out in words (e.g. *nineteen hundred and sixty-four* or *nineteen oh seven*), with the exception of calendar years which were rendered in numerals in pre-existing source transcripts without recordings (e.g. *1899* or *1927*).

Compound words: one-word spellings were usually preferred over hyphenated spellings

(*teacup* rather than *tea-cup*), and hyphenated spellings were usually preferred over two-word spellings (*tea-time* rather than *tea time*).

Punctuation marks: every sentence starts with a number sign (#) which serves as a sentence counter. Full stops, commas, exclamation marks, question marks, semicolons and dashes were used intuitively according to their normal prosodic and syntactic meaning. Dashes are typeset as double hyphens surrounded by spaces (--) to distinguish them from hyphenated spellings.

Indirect speech: no quotation marks are used in FRED. Indirect speech is indicated by a comma followed by a capital letter (as in *he said, No, he said, Never seen it*).

Contractions: contrary to standard orthography, contractions with *be*, *have*, *do* and *modal verbs* are divided by space: *he 's* (not *he's*), *there 's* (not *there's*), *they 'll* (not *they'll*). In the case of 's, this serves to differentiate contracted verbal 's from possessive 's. Non-standard contractions or other reduced forms (for example *t'* for *to*) may be reflected in two different ways: they are either preserved in the transcript together with their standard equivalent ((*reg sic=t'*) *to (/reg)*) or have been regularised, in which case the header of the corresponding text will contain a comment on the procedure. Contracted verbal negations, whether standard or non-standard, are written as one word (*hasn't*, *can't*, *canna*, *dinna*).

special case *mi*: possessive /mi/ is spelled *mi* (e.g. *mi brother*) in order to distinguish it from object-case *me*. As Anderwald and Wagner (2005) mention, using the orthographic form *me* “not only suggests a certain etymology for this form (at worst, a “substitution” of the object form of the personal pronoun for the possessive function), but also complicates computer-based searches considerably, as all instances of the object case of the personal pronoun (*He saw me*) have to be manually excluded, at least as long as the corpus is not tagged for word class”.

special cases 'course and 'cause: 'course (abbreviation of 'of course') is distinguished from 'cause (abbreviation of 'because') wherever possible, and both expressions are distinguished from the verbs/nouns *course* and *cause* respectively. Where both items are distinguishable neither in pronunciation (/z/ vs. /s/) nor function (e.g. when used as a kind of universal sentence connector), 'cos is used.

pause fillers: expressions of hesitation and fillers read *eh*, *ehm*, *ehr* or similar, with an -h- to distinguish them from reduced pronoun forms like 'em ('them') or 'er ('her').

4.5 Rendering of non-standard pronunciations

As mentioned before, FRED was compiled for studies on morphosyntactic variation. The transcripts do not offer any consistent renderings of phonological features, but transcribers were encouraged to include orthographical hints to such features if they were characteristic of the informant's speech. Instances of regular phonologically conditioned non-standard pronunciations (such as "h-dropping" with 3rd person pronouns or -g dropping at the end of words ending in <ng>) are either marked in the text or mentioned in the text header. Other common non-standard pronunciations, such as /hu:s/ (house) or /mek/ (make), are usually regularised and mentioned in the header of the respective texts.

Non-standard realisations are rendered most consistently with grammatical items (pronouns, auxiliaries, etc.), especially in those areas of the English grammar known to display interesting regional variation. Just to mention a few examples: the reduced definite article *t*' (distinguished from neuter pronoun 't), contracted negations (*-nae*, *-na*, *ain't*, *inn't*), neutralised weak forms of verbs (e.g. *was/were* → *wa'*), prepositions (*o'*, *i'*) and modals (*mun* 'must', *gonna*).

So-called "pseudo-dialectisms" (which may be part of the colloquial standard) are generally avoided (e.g. *fellow*, not *feller*; *should 've done*, not *should of done*; *maybe*, not *mebbe*).

Non-standard forms of individual words that are not phonologically conditioned in a straightforward way (i.e. that cannot be accounted for by "accent") are usually given the (reg) ... (/reg) tag: non-standard content words were rendered when they were felt to be indicative of a "broad", markedly local form of speech, and if a written form that clearly indicated the actual pronunciation was available. In most cases, the (reg) tag is used to enhance comprehensibility (*to fetch it (reg sic=yame) home (/reg)*). The same tag is used for non-standard morphological features (*they (reg sic=gien) gave (/reg) us*).

5. Concordance tools

FRED transcripts can be analysed with different concordance programmes. The markup norms were originally set up with regard to the requirements of the DOS text retrieval system TACT but are compatible with the now widely used WordSmith tools. Both of these programmes can be used to retrieve data from one or several texts and do simple kinds of analyses (such as sorted frequencies of words or word combinations and type-token ratios). For analysis with TACT, the text files first have to be converted into textual databases (*.tdb) by using TACT Makebase. For a full description of the different text analysis computing tools in TACT (Usebase, Collgen, TACTstat, etc.) go to <http://khnt.hit.uib.no/icame/manuals/tact.htm>.

Researchers should be aware that TACT is limited as to the quantity of text it can process in one go. For high frequency words (articles, pronouns, etc.) it is therefore advisable to use a different programme, e.g. WordSmith.²⁸ Of the three main WordSmith tools (Concord/ Wordlist/ Keywords), the first two are especially useful for syntactic analyses: the concordancer can be used to search and display words or phrases in context, while the wordlist tool lists words or word-clusters of a text in either alphabetical or frequency order.

6. Storage and documentation

The tape and MD copies of the original recordings are kept at Freiburg University together with the documentation of the collection and compilation process.²⁹ The transcripts are stored as text files. The sound recordings were digitised³⁰ and stored on DVD (slightly denoised, but otherwise unaltered, sound files). Text and audio samples can be found on the FRED homepage.

6.1 Availability

Linguists interested in using FRED for research purposes may check out the project website at www.anglistik.uni-freiburg.de/institut/lskortmann/FRED or contact us directly:

Chair Prof. Bernd Kortmann
English Department
University of Freiburg
Rempartstr. 15
79098 Freiburg
Germany
tel. +49-761-203-3320
fax +49-761-203-3330
melitta.cocan@anglistik.uni-freiburg.de

6.2 Copyright restrictions and confidentiality

Please note that some material included in the corpus is subject to copyright restrictions; check the homepage for more details.

Prior to publication, all texts were anonymised. The stories told in FRED are private in nature, and for the protection of the speakers as well as other people mentioned in the interviews (relatives, friends, etc.), all person and family names were replaced by names of corresponding length and syllable structure. Different tags were used for female and male first names and nicknames, initials and surnames (see Table 8 for a complete list of name tags and examples). Names of ships, trains, larger companies, military regiments, brands, etc. remain unchanged.

References

The first linguistic studies based on FRED have recently appeared in print and online. These include investigations of single dialects as well as comparative studies which take into account the different dialect areas represented in the corpus; an up-to-date account of completed and ongoing research, including master's and doctoral theses can be found on the homepage. Also see Anderwald & Wagner (2005) for a more detailed account of the very first studies conducted by members of the research group.

- Anderwald, Lieselotte. and Bernd Kortmann (2002): "Typology and dialectology: a programmatic sketch." In: J. van Marle and J. Berns (eds.): *Present Day Dialectology, Vol. I: Problems and Discussions*. Berlin: Mouton de Gruyter. 159–71.
- Anderwald, Lieselotte and Susanne Wagner (2005, to appear): "FRED – The Freiburg English Dialect corpus." In: J. Beal, K. Corrigan and H. Moisl (eds.): *Papers from the 20th Sociolinguistics Symposium at Newcastle, April 2004*. London: Macmillan.
- Chambers, J. K. and P. Trudgill (1998²): *Dialectology*. Cambridge: CUP.
- Kortmann, Bernd (1999): "Typology and Dialectology." In: B. Caron (ed.): *Proceedings of the 16th International Congress of Linguists, Paris 1997*. Amsterdam: Elsevier Science.
- Kortmann, Bernd (2002a): "New approaches to dialectology: Introduction." In: D. Kastovsky, G. Kaltenböck and S. Reichl (eds.): *Anglistentag 2001 Wien – Proceedings*. Trier: Wissenschaftlicher Verlag. 3–6.
- Kortmann, Bernd (2002b): "New prospects for the study of English dialect syntax: impetus from syntactic theory and language typology." In: S. Barbiers, L. Cornips and S. van der Kleij (eds.): *Syntactic Microvariation*. Amsterdam: *Online*. 185–213.
<http://www.meertens.nl/books/synmic/>
- Kortmann, Bernd (2003). "Comparative English dialect grammar: a typological approach." In: I.M. Palacios, M. J. López Couso, P. Fra and E. Seoane (eds.), *XXVI AEDEAN Conference Proceedings*. Santiago de Compostela: University of Santiago. 63-81.
- Kortmann, Bernd (2004a). "Why Dialect Grammar Matters." *The European English Messenger* XIII: 24-29.
- Kortmann, Bernd, ed. (2004b). *Dialectology meets Typology: Dialect Grammar from a Cross-Linguistic Perspective*. Berlin: Mouton de Gruyter.
- Kortmann B., T. Herrmann, L. Pietsch, S. Wagner (2005): *A Comparative Grammar of British English Dialects: Agreement, Gender, Relative Clauses*. Berlin: Mouton de Gruyter.
- Orton, Harold (1962). *Survey of English Dialects – An Introduction*. Leeds: Arnold.
- Orton, Harold and W.J.Halliday (eds.) (1962). *Survey of English Dialects. The Basic Material*. Leeds: Arnold.
- Trudgill, Peter (1999²): *The Dialects of England*. Oxford: Blackwell.

Notes

¹ FRED is a specialised corpus, as opposed to general corpora like the BNC which contain different registers and spoken and written language data. For other European dialect research projects see, e.g., www.ncl.ac.uk/necte (Newcastle Electronic Corpus of Tyneside English), asis-cnr.unipd.it (Syntactic Atlas of Northern Italy), meertens.nl/projecten/sand/sandeng.html (Syntactic Atlas of the Dutch Dialects), www.research-projects.unizh.ch/phil/unit64100/area477/p1794.htm (Dialect Syntax of Swiss German) and www.ling.man.ac.uk/Research/RMS/proj.html (The Morphosyntactic Typology of Romani Dialects).

² As opposed to elicited single lexical items/sentences or speech samples.

³ Unlike most dialectological research in the past, this corpus is not based on questionnaires.

⁴ The fact that usually large quantities of text are needed to investigate syntactic phenomena, led to the decision to use already recorded and/or transcribed data. For more information on oral history consider www.oralhistory.org.uk. For the suitability and value of oral history recordings for dialectological studies and contemporary linguistics see Anderwald “Oral history material as a source for dialectological studies” on the FRED homepage. For linguistic consequences of using oral history material, e.g. the predominance of past-tense paradigms, see Anderwald & Wagner 2005.

⁵ In very few cases, there is more than one interviewer. Some interviews were done with more than one speaker at a time or were ‘interrupted’ by a family member or friend coming into the room. The speakers of each interview are shown in Table 6, column 6.

⁶ 2,493,516 words, interviewer utterances, tags and transcribers’ comments not included. Word counts may of course vary, according to the definition of ‘word’ (e.g. hyphenation, truncations, contractions, ...).

⁷ “The morphology and syntax of regional varieties are a much neglected topic in dialect research in general – not only in English dialectology. One reason for this neglect is certainly the fact that for this type of research a much greater amount of data is needed than for phonological studies; another is that the relevant material is difficult to elicit.” (FRED homepage)

For the importance of dialect research see Kortmann (2004a) “Why Dialect Grammar Matters”.

⁸ As reflected in the research project’s name ‘English Dialect Syntax from a Typological Perspective’.

“[...] a typological framework is indispensable, both in terms of questions that need to be addressed, methods that should be employed, and explanations that can be put forward. Thus, the project opens new perspectives for dialectology, typology, corpus linguistics and for sociolinguistics, since we also want to explore how regional dialect features develop into general features of informal spoken Standard English.” (FRED homepage)

For the original motivation of linking dialectology and typology, see, for example, Anderwald & Kortmann (2002) and Kortmann (1999, 2002a/b, 2003, 2004b).

⁹ “However, this limitation is probably not specific to FRED, but applies to dialectological and sociolinguistic interviews alike, as the main objective is always to record the speakers’ speech, rather than one’s own.” (Anderwald & Wagner 2005)

¹⁰ “Although this restricts the range of investigations that can be conducted with the help of FRED in sociolinguistic terms, it represents exactly the same bias as in earlier dialectological work, where we find a preponderance of NORM speakers as well, so that results from work on FRED will be comparable to earlier studies or to material from earlier investigations.” (Anderwald & Wagner 2005)

¹¹ Note that large-scale linguistic changes such as dialect levelling can be stimulated by increased mobility and a higher education level among speakers as well as the influence of mass media. Most FRED informants are “speakers who grew up *before* the Second World War [or even WWI], as this date seems to be the major cataclysmic event after which wide-ranging social and economic changes (with concomitant linguistic changes) came into effect.” (Anderwald & Wagner 2005)

¹² More material is stored in Freiburg. It was not included in the current version of FRED but may be used in the future.

¹³ Dialect transcripts are usually more error-prone than standard-language transcripts, and, in spite of strict transcription guidelines, there may be differences between texts by different transcribers.

¹⁴ Consider the difference between the following two renderings of the same text passage mentioned in Anderwald & Wagner (forthcoming):

(1) That pot? Oh, I, I don’t know, I don’t remember what I made he for. I don’t collect no pots now. (word-for-word transcription)

(2) I don't remember what I made that pot for. I don't collect pots now. (possible wording in pre-existing oral history transcript)

¹⁵ Except for a few texts where the tape was interrupted or the recording stops before the actual conversation ends (which is noted in the corresponding transcripts). These texts were included nevertheless, as they still contain long passages of free speech. The longest interview is KEN_003, with almost 4.5 hours (roughly 47,000 words, 244 KB).

¹⁶ For a geographical division of modern dialect areas see Trudgill (1999: 65).

¹⁷ The numbers, again, may vary slightly depending on the definition of 'word'.

¹⁸ Special thanks to Benedikt Szmrecsanyi for providing this map.

¹⁹ Administrative borders changed a few times during the 20th century, and county borders mentioned by different sources do not always tally. The Chapman county codes, which are the standard format for genealogical purposes, present a clear-cut solution to this problem. As most FRED interviews were recorded during the 1970s and 80s and speakers are more likely to identify with county borders of their own time, it was decided to use the Chapman codes before 1974 (1975 for Scotland) when the local government system in England, Scotland and Wales was reorganised; see <http://www.genuki.org.uk> for historic counties and successor counties after 1986.

All interviews from the Hebrides, outer and inner, have dialect area HEB (Hebrides), even if some of the isles form part of administrative counties on the Scottish "mainland" (e.g. the islands North Uist, South Uist and Skye would belong to Invernessshire according to the Chapman codes).

The present version of FRED contains a maximum of 41 text per county (Hebrides), but since, in the future, FRED might possibly grow to include more texts, three digit numbers were used.

²⁰ There are rare exceptions, in which case the text header gives additional information on the speaker's educational and/or professional background.

²¹ For the stronger tendency of male speakers to use dialect features see Chambers and Trudgill (1998: 30).

²² Some speakers were interviewed more than once, and some interviews have more than one speaker - hence the discrepancy between the total number of speakers and the total number of texts.

²³ <http://www.chass.utoronto.ca/tact/index.html>

²⁴ Researchers can of course align the text and sound files and transcribe the recordings phonetically if they wish to do so.

²⁵ Other factors were of course time, funding and a number of technical issues, as well as the fact that our research cooperation partners also worked with orthographic transcripts (see Anderwald/Wagner 2005).

²⁶ During the editorial process, headers contained additional information on the stage of editing (work in progress, for further revision, completed, etc.). This proved very useful, since the whole process involved various editorial stages that were often carried out by different members of the project team.

²⁷ Interviewer utterances in FRED usually consist of single sentences and questions aimed at eliciting longer passages of free speech from the speaker. They might not be instructive for dialectological studies but might be worth considering for other effects, e.g. those of syntactic priming.

²⁸ To order WordSmith go to www.oup.co.uk.

²⁹ This includes correspondence with the different individuals and institutions who made the material available.

³⁰ Special thanks to the New Media Centre of our University Library for their help.

Appendix

FRED text list						
Dialect Area	County	Location	Text	Recording	Speaker(s)	Words
SE	Kent	Whitstable	KEN_001	yes	JC	9,493
SE	Kent	Faversham	KEN_002	yes	FK	13,881
SE	Kent	Faversham	KEN_003	yes	HM	42,253
SE	Kent	Tenterden	KEN_004	yes	KentPB	13,721
SE	Kent	Sheerness	KEN_005	yes	KentTAD	10,487
SE	Kent	Sittingbourne	KEN_006	yes	WS	11,946
SE	Kent	Lydd	KEN_007	yes	KentEDG, WG	14,383
SE	Kent	Lydd	KEN_008	yes	KentSPR	17,588
SE	Kent	Faversham	KEN_009	yes	KentAW1	12,089
SE	Kent	Tenterden	KEN_010	yes	KentPB	13,981
SE	Kent	Tenterden	KEN_011	yes	KentPB	14,598
SE	London	London North	LND_001	yes	CA	17,114
SE	London	London North	LND_002	yes	EB,Bhu	15,799
SE	London	London North	LND_003	yes	FM	13,954
SE	London	London North	LND_004	yes	JG,DG,DL,CG	16,296
SE	London	LondonNorth	LND_005	yes	LA,DA	12,552
SE	London	Poplar, London Port	LND_006	yes	GA	15,074
SE	London	Poplar, London Port	LND_007	yes	GA	15,760
SE	Middlesex	Pinner	MDX_001	yes	MdxCG	13,686
SE	Middlesex	Pinner	MDX_002	yes	CP	17,362
SE	Suffolk	Lowestoft	SFK_001	no	AnnieShort	9,906
SE	Suffolk	Lowestoft	SFK_002	no	EAA1	7,690
SE	Suffolk	Lowestoft	SFK_003	no	EF1	8,058
SE	Suffolk	Lowestoft	SFK_004	no	ELF	12,222
SE	Suffolk	Lowestoft	SFK_005	no	HED	7,252
SE	Suffolk	Lowestoft	SFK_006	no	EJM	14,939
SE	Suffolk	Lowestoft	SFK_007	no	RR1	8,367
SE	Suffolk	Lowestoft	SFK_008	no	REK	2,078
SE	Suffolk	Lowestoft	SFK_009	no	JR1	8,476
SE	Suffolk	Lowestoft	SFK_010	no	GCS	7,774
SE	Suffolk	Lowestoft	SFK_011	no	JRS	10,072
SE	Suffolk	Lowestoft	SFK_012	no	WT	7,381
SE	Suffolk	Lowestoft	SFK_013	no	HJT1	8,865
SE	Suffolk	Lowestoft	SFK_014	no	SufEAA2	8,851
SE	Suffolk	Lowestoft	SFK_015	no	SufEAA3	7,930
SE	Suffolk	Lowestoft	SFK_016	no	SufEAA4	8,424
SE	Suffolk	Yarmouth	SFK_017	no	SufHFJE	8,054
SE	Suffolk	Lowestoft	SFK_018	no	SufEF2	7,636
SE	Suffolk	Lowestoft	SFK_019	no	EJF	4,033
SE	Suffolk	Lowestoft	SFK_020	no	SufGFK	7,192

SE	Suffolk	Lowestoft	SFK_021	no	SufHED2	6,678
SE	Suffolk	Lowestoft	SFK_022	no	SufHED3	6,676
SE	Suffolk	Lowestoft	SFK_023	no	SufHED4	6,518
SE	Suffolk	Lowestoft	SFK_024	no	SufWEJ1	7,321
SE	Suffolk	Lowestoft	SFK_025	no	SufWEJ1	7,756
SE	Suffolk	Lowestoft	SFK_026	no	SufWEJ1	7,550
SE	Suffolk	Lowestoft	SFK_027	no	JR1	3,547
SE	Suffolk	Lowestoft	SFK_028	no	SufHJT2	8,977
SE	Suffolk	Lowestoft	SFK_029	no	SufHJT2	8,268
SE	Suffolk	Lowestoft	SFK_030	no	SufGCS2	7,239
SE	Suffolk	Lowestoft	SFK_031	no	SufHED5	7,177
SE	Suffolk	Lowestoft	SFK_032	no	SufWCG1	10,052
SE	Suffolk	Lowestoft	SFK_033	no	SufGCS2	7,291
SE	Suffolk	Lowestoft	SFK_034	no	SufSMW	8,459
SE	Suffolk	Lowestoft	SFK_035	no	SufEAA5	7,478
SE	Suffolk	Lowestoft	SFK_036	no	WT	9,968
SE	Suffolk	Lowestoft	SFK_037	no	HJWC	9,928
SE	Suffolk	Lowestoft	SFK_038	no	SufEAA5	8,247
SE	Suffolk	Lowestoft	SFK_039	no	SufEQU	8,084
SW	Cornwall	Churchtown	CON_001	yes	CAVA_DB	3,007
SW	Cornwall	Carnelloe	CON_002	yes	CAVA_HJ	4,035
SW	Cornwall	St.Ives	CON_003	yes	CAVA_PV, Inf	5,610
SW	Cornwall	St.Ives	CON_004	yes	CAVA_TC	1,819
SW	Cornwall	Gurnards Head	CON_005	yes	CAVA_WJB	7,560
SW	Cornwall	Pendeen	CON_006	yes	CAVA_WW, CW	4,756
SW	Cornwall	St.Just	CON_007	no	ICS_JE	10,857
SW	Cornwall	Heamoor	CON_008	no	ICS_JP	9,298
SW	Cornwall	Ludgvan	CON_009	no	ICS_EW, ICS_MW	9,401
SW	Cornwall	Gulval	CON_010	no	ICS_RM	8,630
SW	Cornwall	Sennen	CON_011	yes	CorJLN	37,078
SW	Cornwall	Nancredra	CON_012	no	CorPO	396
SW	Devon	Blackawton	DEV_001	yes	TCA_RA, unknown	10,204
SW	Devon	Galmpton	DEV_002	yes	TCA_GB	5,571
SW	Devon	Brixham	DEV_003	yes	TCA_EC, TCA_C	6,577
SW	Devon	Totnes	DEV_004	yes	TCA_SC	4,764
SW	Devon	Totnes	DEV_005	yes	TCA_FP	10,949
SW	Devon	Dartmouth	DEV_006	yes	TCA_GR	7,521
SW	Devon	Totnes	DEV_007	yes	TCA_EA	8,869
SW	Devon	Totnes	DEV_008	yes	TCA_FK, TCA_K	12,499
SW	Devon	Totnes	DEV_009	yes	TCA_WH	11,345
SW	Devon	Buckfast	DEV_010	yes	TCA_WC	9,652
SW	Devon	Blackawton	DEV_011	no	TCA_PR, TCA_DR	4,223
SW	Oxfordshire	Leafield	OXF_001	yes	FP,BF	6,464

SW	Oxfordshire	Filkins	OXF_002	yes	GW,XW	7,579
SW	Somerset	West Stoughton	SOM_001	yes	SRLM_HR	5,459
SW	Somerset	unknown	SOM_002	yes	SRLM_PG	3,561
SW	Somerset	Babcary	SOM_003	no	SRLM_B	3,662
SW	Somerset	N.Burrowbridge	SOM_004	yes	SRLM_CK	4,968
SW	Somerset	Street	SOM_005	yes	SRLM_RF	7,506
SW	Somerset	Sunnyside	SOM_006	yes	SRLM_AW	4,270
SW	Somerset	Evercreech	SOM_007	no	SRLM_PT, wifeofPT	6,408
SW	Somerset	Petherton	SOM_008	yes	SRLM_CH, SRLM_ES	5,573
SW	Somerset	Horton	SOM_009	yes	SRLM_CA	10,523
SW	Somerset	Compton Dundon	SOM_010	no	SRLM_WB	7,586
SW	Somerset	Compton Dundon	SOM_011	no	SRLM_WB	17,286
SW	Somerset	Buckleigh	SOM_012	yes	SRLM_SH	6,215
SW	Somerset	Yeovil	SOM_013	yes	SRLM_GG	4,573
SW	Somerset	Barton St.David	SOM_014	yes	SRLM_RM,M	10,480
SW	Somerset	Hatch Beauchamp	SOM_015	no	SRLM_HH	2,697
SW	Somerset	Stoke St.Gregory	SOM_016	no	SRLM_WP	3,135
SW	Somerset	Baltonsborough	SOM_017	no	SRLM_LD	3,277
SW	Somerset	Fivehead	SOM_018	no	SRLM_AB	2,852
SW	Somerset	Coley	SOM_019	no	SRLM_W	11,318
SW	Somerset	Moorwood Oakhill	SOM_020	no	SRLM_AE	3,402
SW	Somerset	Ashwick	SOM_021	no	SRLM_R	6,964
SW	Somerset	Churchingfold Beaminster	SOM_022	no	SRLM_T, SRLM_S	2,609
SW	Somerset	Wedmore	SOM_023	no	SRLM_M, SRLM_CM	1,685
SW	Somerset	North Bridgwater	SOM_024	no	SRLM_HM	4,132
SW	Somerset	unknown	SOM_025	no	SRLM_CJ	1,066
SW	Somerset	North Curry	SOM_026	no	SRLM_HC	1,578
SW	Somerset	Kingston Seymour	SOM_027	no	SRLM_TP, SRLM_GP	3,074
SW	Somerset	Bridgwater	SOM_028	no	SRLM_WT	14,014
SW	Somerset	Evercreech	SOM_029	no	SRLM_JA	6,716
SW	Somerset	Henley	SOM_030	no	SRLM_HW,W,D	5,191
SW	Somerset	Galhampton	SOM_031	no	SRLM_Y	5,632
SW	Somerset	Butleigh	SOM_032	no	SRLM_HG	8,869
SW	Somerset	Glastonbury	SOM_033	no	SRLM_EC	8,016
SW	Somerset	North Petherton	SOM_034	yes	SRLM_EW	2,967
SW	Somerset	Glastonbury	SOM_035	yes	SRLM_SC	3,485
SW	Somerset	Watchet	SOM_036	no	SRLM_ES	3,490
SW	Wiltshire	Trowbridge	WIL_001	yes	TrbrGR	12,009

SW	Wiltshire	Trowbridge	WIL_002	yes	TrbrJH	5,953
SW	Wiltshire	Westbury	WIL_003	yes	TrbrJS	9,118
SW	Wiltshire	Trowbridge	WIL_004	yes	TrbrDP	7,780
SW	Wiltshire	Trowbridge	WIL_005	yes	TrbrCS, TrbrViv, TrbrD1, TrbrD2, TrbrCN, TrbrLW, TrbrDH	10,151
SW	Wiltshire	Trowbridge	WIL_006	yes	TrbrGSR	7,640
SW	Wiltshire	Trowbridge	WIL_007	yes	TrbrBJ, TrbrKJ	5,591
SW	Wiltshire	Westbury	WIL_008	yes	TrbrRCC	12,924
SW	Wiltshire	Enford	WIL_009	no	WflsFP,Y	4,267
SW	Wiltshire	Urchfont	WIL_010	yes	WflsWGP	4,419
SW	Wiltshire	Edington	WIL_011	no	WflsWH, WflsGW	4,570
SW	Wiltshire	Donhead St.Andrew	WIL_012	no	WflsEF	2,503
SW	Wiltshire	Trowbridge	WIL_013	no	TrbrVL	1,093
SW	Wiltshire	Trowbridge	WIL_014	no	TrbrGHG	4,269
SW	Wiltshire	Trowbridge	WIL_015	no	TrbrEGC	10,513
SW	Wiltshire	Trowbridge	WIL_016	no	TrbrEGC	4,552
SW	Wiltshire	Trowbridge	WIL_017	no	TrbrAH, TrbrQD, TrbrBC	10,251
SW	Wiltshire	Melksham	WIL_018	no	TrbrDC,ML, Jack	12,126
SW	Wiltshire	Trowbridge	WIL_019	no	TrbrKM	5,208
SW	Wiltshire	Trowbridge	WIL_020	no	TrbrRS	2,679
SW	Wiltshire	Trowbridge	WIL_021	no	TrbrS,TrbrRS	1,233
SW	Wiltshire	Trowbridge	WIL_022	no	TrbrFC, TrbrJC	24,135
SW	Wiltshire	Trowbridge	WIL_023	no	TrbrAS	8,061
SW	Wiltshire	Trowbridge	WIL_024	no	TrbrIS	4,983
Mid	Leicestershire	unknown	LEI_001	yes	LeiA,LeiB	2,423
Mid	Leicestershire	Caldwell	LEI_002	yes	Lei2	5,888
Mid	Nottinghamshire	Nottingham	NTT_001	yes	NotA103	5,664
Mid	Nottinghamshire	Nottingham	NTT_002	yes	NotA25	11,247
Mid	Nottinghamshire	Nottingham	NTT_003	yes	NotA30	9,221
Mid	Nottinghamshire	Nottingham	NTT_004	yes	NotA40	4,959
Mid	Nottinghamshire	Southwell	NTT_005	yes	NotA52	11,836
Mid	Nottinghamshire	Nottingham	NTT_006	yes	NotA58	14,344
Mid	Nottinghamshire	Nottingham	NTT_007	yes	NotA65	6,992
Mid	Nottinghamshire	Nottingham	NTT_008	yes	NotA71	5,508
Mid	Nottinghamshire	Nottingham	NTT_009	yes	NotA80	8,610
Mid	Nottinghamshire	Nottingham	NTT_010	yes	NotA85	9,418
Mid	Nottinghamshire	Nottingham	NTT_011	yes	NotA99	9,135
Mid	Nottinghamshire	Nottingham	NTT_012	yes	ALI	8,376
Mid	Nottinghamshire	Nottingham	NTT_013	yes	NotA8	13,259
Mid	Nottinghamshire	unknown	NTT_014	yes	A15	15,686
Mid	Nottinghamshire	Lambley	NTT_015	yes	A109,A109w	11,832
Mid	Nottinghamshire	Nottingham	NTT_016	yes	NotA48	14,530
Mid	Shropshire	Coalport	SAL_001	no	IBMrsA	764

Mid	Shropshire	Coalport	SAL_002	no	IBBB	4,848
Mid	Shropshire	Coalport	SAL_003	no	IBBB	1,739
Mid	Shropshire	Coalbrookdale	SAL_004	no	IBWB	4,790
Mid	Shropshire	Dawley	SAL_005	no	IBCB	2,605
Mid	Shropshire	Dawley	SAL_006	no	IBNB	4,982
Mid	Shropshire	Broseley	SAL_007	no	IBIB,IBJ, NBall	1,926
Mid	Shropshire	Dawley	SAL_008	no	IBAB	4,971
Mid	Shropshire	Madeley	SAL_009	no	IBGB	2,941
Mid	Shropshire	Madeley	SAL_010	no	IBLB	1,501
Mid	Shropshire	Wellington	SAL_011	no	IBBeBo	7,416
Mid	Shropshire	Shifnal	SAL_012	no	IBPB	1,768
Mid	Shropshire	Dawley	SAL_013	no	IBFB	5,851
Mid	Shropshire	Telford	SAL_014	no	IBWiBr	6,411
Mid	Shropshire	Telford	SAL_015	no	IBWiBr	2,019
Mid	Shropshire	Coalport	SAL_016	no	IBED	2,831
Mid	Shropshire	unknown	SAL_017	no	IBAD	5,336
Mid	Shropshire	unknown	SAL_018	no	IBAD	5,196
Mid	Shropshire	Ketley	SAL_019	no	IBJD	4,677
Mid	Shropshire	Madeley	SAL_020	no	IBErDi	6,927
Mid	Shropshire	Ironbridge	SAL_021	no	IBCD	3,931
Mid	Shropshire	Ironbridge	SAL_022	no	IBMrsD	1,262
Mid	Shropshire	Oakengates	SAL_023	no	IBME	15,201
Mid	Shropshire	Craigside	SAL_024	no	IBLE	6,003
Mid	Shropshire	Craigside	SAL_025	no	IBLE	4,474
Mid	Shropshire	Ironbridge	SAL_026	no	IBCE	6,301
Mid	Shropshire	Ironbridge	SAL_027	no	IBFF	4,196
Mid	Shropshire	Ironbridge	SAL_028	no	IBMrsF	1,769
Mid	Shropshire	Ironbridge	SAL_029	no	IBSH, IBAnBa	5,380
Mid	Shropshire	Coalport	SAL_030	no	IBAL	4,366
Mid	Shropshire	Dawley	SAL_031	no	IBGL	5,665
Mid	Shropshire	Madeley	SAL_032	no	IBGJ	4,706
Mid	Shropshire	Telford	SAL_033	no	IBTJ	4,289
Mid	Shropshire	Wellington	SAL_034	no	IBCH	3,332
Mid	Shropshire	Ironbridge	SAL_035	no	IBRH	1,925
Mid	Shropshire	Broseley	SAL_036	no	IBBM	6,775
Mid	Shropshire	Madeley	SAL_037	no	IBHP	3,436
Mid	Shropshire	Lawley	SAL_038	no	IBJP	6,142
Mid	Shropshire	Farley	SAL_039	no	IBEP	5,528
Mid	Warwickshire	Walsall	WAR_001	yes	Wal128LG, WAR001unknown	8,176
N	Durham	Birtley	DUR_001	yes	DurML	9,094
N	Durham	Hartlepool	DUR_002	yes	DurNB	7,932
N	Durham	Hartlepool	DUR_003	yes	DurEL	9,576
N	Lancashire	Barrow	LAN_001	yes	LanD1B	11,392
N	Lancashire	Barrow	LAN_002	yes	LAN002, thirdperson	9,287

N	Lancashire	Preston	LAN_003	yes	Langlp	9,389
N	Lancashire	Preston	LAN_004	yes	LanWIP	8,882
N	Lancashire	Preston	LAN_005	yes	LanC5P	10,720
N	Lancashire	Preston	LAN_006	yes	LanD1P	10,080
N	Lancashire	Preston	LAN_007	yes	LanF1P	8,585
N	Lancashire	Preston	LAN_008	yes	LanT3P	10,193
N	Lancashire	Preston	LAN_009	yes	LanHIP	11,110
N	Lancashire	Preston	LAN_010	yes	H3L,MH3L	9,671
N	Lancashire	Crompton	LAN_011	yes	LanMJ	13,677
N	Lancashire	Prescott	LAN_012	yes	LanEG	27,977
N	Lancashire	Prescott	LAN_013	yes	LanPK	727
N	Lancashire	Huyton	LAN_014	yes	LanCD	53
N	Lancashire	Wigan	LAN_015	no	WigMWH	6,634
N	Lancashire	Wigan	LAN_016	no	WigMS	4,283
N	Lancashire	Wigan	LAN_017	no	WigMB	3,055
N	Lancashire	Wigan	LAN_018	no	WigPS	926
N	Lancashire	Wigan	LAN_019	no	WigHT	6,326
N	Lancashire	Wigan	LAN_020	no	WigWHJ	14,958
N	Lancashire	Wigan	LAN_021	no	WigJH	615
N	Lancashire	Wigan	LAN_022	no	WigGRB	7,764
N	Lancashire	Wigan	LAN_023	no	WigWO	8,807
N	Northumberland	Cammerse	NBL_001	yes	NblWL	395
N	Northumberland	Choppington	NBL_003	yes	NblMN	7,886
N	Northumberland	Fenwick Steads	NBL_006	yes	NblJB	7,294
N	Northumberland	Swarland	NBL_007	yes	NblAR,NblAn	8,540
N	Northumberland	Fenwick Steads	NBL_008	yes	NblGS	4,317
N	Westmorland	Ambleside	WES_001	no	WesAD	10,252
N	Westmorland	Ambleside	WES_002	no	WesAE	4,650
N	Westmorland	Ambleside	WES_003	yes	WesAQ	4,538
N	Westmorland	Ambleside	WES_004	yes	WesAY	7,514
N	Westmorland	Ambleside	WES_005	no	WesBE	11,022
N	Westmorland	Ambleside	WES_006	yes	WesBL	3,051
N	Westmorland	Ambleside	WES_007	no	WesBP	4,749
N	Westmorland	Ambleside	WES_008	yes	WesBR	10,527
N	Westmorland	Ambleside	WES_009	no	WesBX	5,618
N	Westmorland	Ambleside	WES_010	no	WesCB	7,862
N	Westmorland	Ambleside	WES_011	yes	WesCE,WesCEa	3,190
N	Westmorland	Ambleside	WES_012	no	WesCU	6,211
N	Westmorland	Ambleside	WES_013	no	WesCY	3,191
N	Westmorland	Ambleside	WES_014	no	WesDA	5,473
N	Westmorland	Ambleside	WES_015	no	WesDB	5,478
N	Westmorland	Ambleside	WES_016	no	WesDT	13,473
N	Westmorland	Ambleside	WES_017	no	WesDX,WesAnon	22,820
N	Westmorland	Ambleside	WES_018	no	WesHN,WesJN	5,169
N	Westmorland	Ambleside	WES_019	no	WesZb,WesZs	14,959
N	Yorkshire	Middlesbrough	YKS_001	yes	YksWF	10,009

N	Yorkshire	Middlesbrough	YKS_002	yes	YksJV	9,439
N	Yorkshire	Guisborough	YKS_003	yes	YksSL	11,891
N	Yorkshire	Guisborough	YKS_004	yes	YksEH	1,907
N	Yorkshire	Guisborough	YKS_005	no	YksGR	2,954
N	Yorkshire	Hinderwell	YKS_006	yes	YksHS	7,191
N	Yorkshire	Middlesbrough	YKS_007	yes	YksML	11,321
N	Yorkshire	Redcar	YKS_008	yes	YksMW	9,689
N	Yorkshire	Loftus	YKS_009	yes	YksWG,YksMG	5,638
N	Yorkshire	Hebden Bridge	YKS_010	yes	Yks_10	9,442
N	Yorkshire	Hebden Bridge	YKS_011	yes	YksBR,YksJS, YksMR,YksAnon	7,908
Wal	Denbighshire (Gwynedd)	Conwy	DEN_001	no	AL	5,562
Wal	Denbighshire (Gwynedd)	unknown	DEN_002	no	GwyBV	7,218
Wal	Denbighshire (Gwynedd)	unknown	DEN_003	no	GwyHHJ,GwyME, GwyWL	10,842
Wal	Denbighshire (Gwynedd)	unknown	DEN_004	no	GwyR	13,662
Wal	Glamorgan (Maesteg)	Cwmfelin	GLA_001	no	CwmWTC	6,755
Wal	Glamorgan (Bedlinog)	Bedlinog	GLA_002	no	BedHT	8,463
Wal	Glamorgan (Glydach)	Swansea	GLA_003	no	SwanBE	7,896
Wal	Glamorgan (Swansea)	Bishopston	GLA_004	no	WalWTA	5,505
Wal	Glamorgan (Swansea)	Plasmarl	GLA_005	no	WalDJN	8,712
Wal	Glamorgan (Swansea)	Townhill	GLA_006	no	WalDTC	5,469
Wal	Glamorgan (Swansea)	Pontardulais	GLA_007	no	WalJSD	8,671
Man	Isle of Man	Laxey	IOM_001	no	IomBK	5,822
Man	Isle of Man	Ramsey	IOM_002	no	IomRC	4,639
Heb	Skye	Balmaquien	HEB_001	no	HebMMcl,Heb18	3,027
Heb	Skye	Staffin	HEB_002	no	Heb21,Heb22	608
Heb	Skye	Staffin	HEB_003	no	Heb23,Heb24	583
Heb	Skye	Portree	HEB_004	no	Heb42,Heb43	305
Heb	unknown	unknown	HEB_005	no	HebNN	541
Heb	unknown	unknown	HEB_006	no	Heb53,Heb54	7,057
Heb	North Uist	unknown	HEB_007	no	Heb57	4,098
Heb	North Uist	unknown	HEB_008	no	Heb57	3,054
Heb	Skye	Portree	HEB_009	no	Heb2	1,111
Heb	Skye	Struan	HEB_010	no	Heb3	3,249
Heb	Skye	North Duntulm	HEB_011	no	Heb7	4,631
Heb	Skye	Portree	HEB_012	no	Heb8	1,568
Heb	Skye	Portree	HEB_013	no	Heb9	1,656
Heb	unknown	unknown	HEB_014	no	Heb12, Heb014unknown	2,494
Heb	Skye	unknown	HEB_015	no	Heb14	8,906
Heb	Skye	Portree Kingsburgh	HEB_016	no	Heb16, Heb016unknown, Heb016husband	2,896
Heb	Skye	Bornaskitaig	HEB_017	no	Heb19	3,971
Heb	South Uist	Staffin Clachan	HEB_018	no	Heb25	14,330

Heb	Skye	Balmaquien	HEB_019	no	Heb26	1,980
Heb	Skye	Portree Staffin	HEB_020	no	Heb37	1,070
Heb	unknown	Maliger	HEB_021	no	Heb38	5,098
Heb	unknown	unknown	HEB_022	no	Heb39, Heb022unknown, Heb022wife	3,375
Heb	Skye	Portree Staffin	HEB_023	no	Heb41	2,137
Heb	Skye	Stenscholl	HEB_024	no	Heb44	4,118
Heb	Skye	unknown	HEB_025	no	Heb51	7,300
Heb	unknown	unknown	HEB_026	no	Heb52	554
Heb	North Uist	Clachan	HEB_027	no	Heb56	2,516
Heb	North Uist	unknown	HEB_028	no	Heb58	3,127
Heb	North Uist	Locheport	HEB_029	no	Heb59	1,460
Heb	unknown	unknown	HEB_030	no	Heb60	4,911
Heb	unknown	unknown	HEB_031	no	Heb62	5,335
Heb	South Uist	unknown	HEB_032	no	Heb63	8,829
Heb	unknown	unknown	HEB_033	no	Heb64, Heb033unknown	3,907
Heb	South Uist	Howbeg	HEB_034	no	Heb67	4,289
Heb	South Uist	Howbeg	HEB_035	no	Heb68	2,653
Heb	South Uist	Iochdar	HEB_036	no	Heb76	2,655
Heb	North Uist	unknown	HEB_037	no	Heb77	4,008
Heb	Benbecula	Hacklett	HEB_038	no	Heb72	2,123
Heb	Berneray	Innes	HEB_039	no	Heb73	927
Heb	South Uist	Loch Carnan	HEB_040	no	Heb74	4,426
Heb	Carinish	Claddach	HEB_041	no	Heb78,Heb79, Heb80	1,799
SCL	Angus	Kirkton	ANS_001	yes	AnsJT,AnsAN	5,677
SCL	Angus	Forfar	ANS_002	yes	AnsBS	1,707
SCL	Angus	Forfar	ANS_003	no	AnsWB	6,315
SCL	Angus	Arbroath	ANS_004	no	AnsLC,AnsJL	4,060
SCL	Banffshire	Keith	BAN_001	yes	BanJS	5,468
SCL	Dumfriesshire	Mickle Kirkland	DFS_001	yes	DfsWH,DfsAN	9,282
SCL	East Lothian	Tranent	ELN_001	no	ElnCD,ElnKR	1,214
SCL	East Lothian	Tranent	ELN_002	no	ElnAS,ElnAC	2,454
SCL	East Lothian	Tranent	ELN_003	no	ElnAS,ElnGB	1,015
SCL	East Lothian	Tranent	ELN_004	yes	ElnLS,ElnAB	2,238
SCL	East Lothian	Tranent	ELN_005	yes	ElnAG,ElnML	1,162
SCL	East Lothian	Tranent	ELN_006	yes	ElnEL	1,198
SCL	East Lothian	Tranent	ELN_007	yes	ElnLS,ElnAB	2,070
SCL	East Lothian	Tranent	ELN_008	yes	ElnAB,ElnMS	3,738
SCL	East Lothian	Tranent	ELN_009	yes	ElnLS,ElnAB	7,412
SCL	East Lothian	Tranent	ELN_010	yes	ElnIB	2,089
SCL	East Lothian	Tranent	ELN_011	yes	ElnEL,ElnIB	2,356
SCL	East Lothian	Tranent	ELN_012	yes	ElnAG,ElnML	2,808
SCL	East Lothian	Tranent	ELN_013	yes	ElnSH,ElnTR	787

SCL	East Lothian	Tranent	ELN_014	yes	ElnAS,ElnAC	3,234
SCL	East Lothian	Tranent	ELN_015	no	EdbBK,EdbIL	2,257
SCL	Fife	Anstruther	FIF_001	yes	FifPS	3,465
SCL	Kincardineshire	Gourdon	KCD_001	yes	KcdRS,KcdES	7,092
SCL	Kinross-shire	Crook-o-Devon	KRS_001	yes	KrsWS	2,218
SCL	Lanarkshire	Symington	LKS_001	yes	LksGH	3,599
SCL	Midlothian	Edinburgh	MLN_001	no	MlnRS	2,049
SCL	Midlothian	Edinburgh	MLN_002	no	MlnCM,MlnAnon	1,975
SCL	Midlothian	Edinburgh	MLN_003	no	MlnAA,Mln2Anon	1,846
SCL	Midlothian	Edinburgh	MLN_004	yes	MlnSM,MlnDM	2,110
SCL	Midlothian	Edinburgh	MLN_005	yes	MlnJM	3,299
SCL	Midlothian	Edinburgh	MLN_006	yes	MlnRH	6,045
SCL	Midlothian	Edinburgh	MLN_007	yes	MlnJH	12,799
SCL	Peeblesshire	West Linton	PEE_001	yes	PeeJP	2,185
SCL	Peeblesshire	Nether Urd	PEE_002	yes	PeeMT	12,399
SCL	Perthshire	Perth	PER_001	yes	PerSS	3,713
SCL	Perthshire	Abernethy	PER_002	yes	PerTJ	2,219
SCL	Perthshire	Perth	PER_003	yes	PerAB	8,362
SCL	Perthshire	Perth	PER_004	no	PerHK	4,721
SCL	Perthshire	Crieff	PER_005	no	PerWM,PerWG	1,530
SCL	Selkirkshire	Selkirk	SEL_001	yes	SelJB	4,357
SCL	Selkirkshire	Ettrick	SEL_002	yes	SelJM	2,883
SCL	Selkirkshire	Yarrow	SEL_003	yes	SelAN	1,792
SCL	West Lothian	Falkirk	WLN_001	yes	WlnTW	1,790
SCL	West Lothian	Falkirk	WLN_002	yes	WlnMO	1,968
SCL	West Lothian	Falkirk	WLN_003	yes	WlnMD	3,764
SCL	West Lothian	Falkirk	WLN_004	yes	WlnMD	5,217
SCL	West Lothian	Falkirk	WLN_005	yes	WlnES	936
SCL	West Lothian	Falkirk	WLN_006	yes	WlnES	2,945
ScH	Inverness-shire	Inverness	INV_001	no	InvWR	3,126
SCH	Ross and Cromarty	Aultbea	ROC_001	yes	RocAM	5,685
SCH	Ross and Cromarty	Hill of Fearn	ROC_002	yes	RocPN	4,446
SCH	Sutherland	Terryside	SUT_001	yes	SutDS	163
SCH	Sutherland	Boulstoer	SUT_002	yes	SutGM	3,809
SCH	Sutherland	Durness	SUT_003	yes	SutGMcK	3,194
SCH	Sutherland	Assynt	SUT_004	yes	SutTR	3,449

Table 6: FRED text list

FRED speaker list								
Speaker ID	Text(s)	Age	Birth Date	Birth Decade	Sex	Interview Date	Interview Decade	Words
A109	NTT_015	unknown	1912	191	m	unknown	198	10449
A109w	NTT_015	unknown	1912	191	f	unknown	198	1383
A15	NTT_014	71	1911	191	m	1982	198	15686
AL	DEN_001	unknown	unknown	189	f	unknown	unknown	5562
ALI	NTT_012	78	1906	190	f	1984	198	8376
AnnieShort	SFK_001	unknown	unknown	unknown	f	unknown	unknown	9906
AnsAN	ANS_001	unknown	unknown	unknown	unknown	1993	199	10
AnsBS	ANS_002	81	1919	191	m	2000	200	1686
AnsJL	ANS_004	unknown	unknown	192	m	1993	199	1973
AnsJT	ANS_001	75	1918	191	m	1993	199	5698
AnsLC	ANS_004	65	1928	192	m	1993	199	2103
AnsWB	ANS_003	100	1896	189	m	1996	199	6315
BanJS	BAN_001	76	1914	191	f	1990	199	5452
BedHT	GLA_002	78	1896	189	m	1974	197	8463
BF	OXF_001	73	1899	189	m	1972	197	2971
BHu	LND_002	unknown	unknown	unknown	m	1982	198	512
CA	LND_001	61	1917	191	m	1978	197	17114
CAVA_DB	CON_001	unknown	1895	189	m	unknown	198	3007
CAVA_HJ	CON_002	unknown	unknown	unknown	f	unknown	198	4035
CAVA_PV	CON_003	86	1892	189	m	1978	197	5610
CAVA_TC	CON_004	unknown	unknown	unknown	m	unknown	198	1819
CAVA_WJB	CON_005	74	1904	190	m	1978	197	7560
CAVA_WW	CON_006	unknown	unknown	190	m	1979	197	4756
CG	LND_004	unknown	unknown	unknown	unknown	1982	198	515
CorJLN	CON_011	69	1910	191	m	1979	197	37078
CorPO	CON_012	64	1915	191	m	1979	197	396
CP	MDX_002	70	1905	190	m	1975	197	17362
CW	CON_006	unknown	unknown	unknown	f	1979	197	321
CwmWTC	GLA_001	89	1884	188	m	1973	197	6755
D	SOM_030	unknown	unknown	unknown	unknown	1985	198	369
DA	LND_005	unknown	unknown	unknown	f	1981	198	38
DfsAN	DFS_001	unknown	unknown	unknown	f	1982	198	13
DfsWH	DFS_001	72	1910	191	m	1982	198	9269
DG	LND_004	unknown	unknown	unknown	f	1982	198	40
DL	LND_004	unknown	unknown	unknown	f	1982	198	380
DurEL	DUR_003	unknown	1897	189	f	unknown	198	9576
DurML	DUR_001	unknown	unknown	191	m	unknown	198	9094
DurNB	DUR_002	77	1908	190	m	1985	198	7932
EAA1	SFK_002	74	1902	190	m	1976	197	7690
EB	LND_002	56	1926	192	f	1982	198	15287
EdbBK	ELN_015	unknown	unknown	unknown	m	unknown	unknown	1425
EdbIL	ELN_015	unknown	unknown	unknown	m	unknown	unknown	832
EF1	SFK_003	79	1898	189	m	1977	197	8058
EJF	SFK_019	64	1912	191	m	1976	197	4033

EJM	SFK_006	80	1896	189	m	1976	197	7383
ELF	SFK_004	71	1905	190	m	1976	197	12222
ElnAB	ELN_004 ELN_007 ELN_008 ELN_009	unknown	unknown	unknown	f	unknown	197	7564
ElnAC	ELN_002 ELN_014	unknown	unknown	unknown	m	unknown	197	4056
ElnAG	ELN_005 ELN_012	unknown	unknown	unknown	m	unknown	197	2241
ElnAS	ELN_002,ELN_003,ELN_014	unknown	unknown	unknown	f	unknown	197	1990
ElnCD	ELN_001	15	unknown	unknown	f	unknown	197	603
ElnEL	ELN_006,ELN_011	unknown	unknown	unknown	f	unknown	197	2934
ElnGB	ELN_003	15	unknown	unknown	m	unknown	197	657
ElnIB	ELN_010,ELN_011	17	unknown	unknown	m	unknown	197	2709
ElnKR	ELN_001	15	unknown	unknown	m	unknown	197	611
ElnLS	ELN_004 ELN_007 ELN_009	18	unknown	unknown	f	unknown	197	5561
ElnML	ELN_005 ELN_012	18	unknown	unknown	f	unknown	197	1729
ElnMS	ELN_008	18	unknown	unknown	f	unknown	197	2333
ElnSH	ELN_013	unknown	unknown	unknown	f	unknown	197	447
ElnTR	ELN_013	unknown	unknown	unknown	m	unknown	197	340
FifPS	FIF_001	69	1914	191	m	1983	198	3465
FK	KEN_002	88	1887	188	m	1975	197	13881
FM	LND_003	67	1914	191	f	1981	198	13954
FP	OXF_001	81	1891	189	m	1972	197	3493
GA	LND_006 LND_007	61	1924	192	m	1985	198	30834
GCS	SFK_010	73	1903	190	m	1976	197	7774
GW	OXF_002	102	1877	187	m	1979	197	7003
GwyBV	DEN_002	unknown	1904	190	m	unknown	unknown	7218
GwyHHJ	DEN_003	unknown	unknown	unknown	m	unknown	unknown	8057
GwyME	DEN_003	unknown	unknown	unknown	m	unknown	unknown	513
GwyR	DEN_004	87	unknown	unknown	f	unknown	unknown	13662
GwyWL	DEN_003	unknown	1912	191	m	unknown	unknown	2272
H3L	LAN_010	71	1904	190	m	1975	197	9644
Heb014unknown	HEB_014	unknown	unknown	unknown	unknown	unknown	unknown	104
Heb016husband	HEB_016	unknown	unknown	unknown	m	unknown	unknown	15
Heb016unknown	HEB_016	unknown	unknown	unknown	f	unknown	unknown	7
Heb022unknown	HEB_022	unknown	unknown	unknown	f	unknown	unknown	13
Heb022wife	HEB_022	unknown	unknown	unknown	f	unknown	unknown	3
Heb033unknown	HEB_033	unknown	unknown	unknown	unknown	unknown	unknown	455

Heb12	HEB_014	75	unknown	unknown	f	unknown	unknown	2390
Heb14	HEB_015	50	unknown	unknown	m	unknown	unknown	8906
Heb16	HEB_016	90	unknown	unknown	f	unknown	unknown	2874
Heb18	HEB_001	67	unknown	unknown	m	unknown	unknown	1900
Heb19	HEB_017	90	unknown	unknown	f	unknown	unknown	3971
Heb2	HEB_009	15	unknown	unknown	f	unknown	unknown	1111
Heb21	HEB_002	11	unknown	unknown	f	unknown	unknown	119
Heb22	HEB_002	11	unknown	unknown	m	unknown	unknown	489
Heb23	HEB_003	11	unknown	unknown	f	unknown	unknown	55
Heb24	HEB_003	11	unknown	unknown	m	unknown	unknown	528
Heb25	HEB_018	81	unknown	unknown	m	unknown	unknown	14330
Heb26	HEB_019	40	unknown	unknown	unknown	unknown	unknown	1980
Heb3	HEB_010	17	unknown	unknown	f	unknown	unknown	3249
Heb37	HEB_020	10	unknown	unknown	f	unknown	unknown	1070
Heb38	HEB_021	45	unknown	unknown	f	unknown	unknown	5098
Heb39	HEB_022	unknown	unknown	unknown	m	unknown	unknown	3359
Heb41	HEB_023	16	unknown	unknown	m	unknown	unknown	2137
Heb42	HEB_004	8	unknown	unknown	f	unknown	unknown	210
Heb43	HEB_004	8	unknown	unknown	m	unknown	unknown	95
Heb44	HEB_024	45	unknown	unknown	m	unknown	unknown	4118
Heb51	HEB_025	unknown	unknown	unknown	m	unknown	unknown	7300
Heb52	HEB_026	unknown	unknown	unknown	unknown	unknown	unknown	554
Heb53	HEB_006	unknown	unknown	unknown	m	unknown	unknown	4097
Heb54	HEB_006	unknown	unknown	unknown	m	unknown	unknown	2960
Heb56	HEB_027	47	unknown	unknown	f	unknown	unknown	2516
Heb57	HEB_007 HEB_008	77	unknown	unknown	unknown	unknown	unknown	7152
Heb58	HEB_028	27	unknown	unknown	unknown	unknown	unknown	3127
Heb59	HEB_029	37	unknown	unknown	m	unknown	unknown	1460
Heb60	HEB_030	unknown	unknown	unknown	m	unknown	unknown	4911
Heb62	HEB_031	unknown	unknown	unknown	unknown	unknown	unknown	5335
Heb63	HEB_032	77	unknown	unknown	f	unknown	unknown	8829
Heb64	HEB_033	unknown	unknown	unknown	f	unknown	unknown	3452
Heb67	HEB_034	19	unknown	unknown	f	unknown	unknown	4289
Heb68	HEB_035	68	unknown	unknown	f	unknown	unknown	2653
Heb7	HEB_011	65	unknown	unknown	m	unknown	unknown	4631
Heb72	HEB_038	15	unknown	unknown	m	unknown	unknown	2123
Heb73	HEB_039	6	unknown	unknown	m	unknown	unknown	927
Heb74	HEB_040	66	unknown	unknown	f	unknown	unknown	4426
Heb76	HEB_036	68	unknown	unknown	f	unknown	unknown	2655
Heb77	HEB_037	11	unknown	unknown	f	unknown	unknown	4008
Heb78	HEB_041	62	unknown	unknown	m	unknown	unknown	1522
Heb79	HEB_041	unknown	unknown	unknown	unknown	unknown	unknown	268
Heb8	HEB_012	15	unknown	unknown	f	unknown	unknown	1568
Heb80	HEB_041	unknown	unknown	unknown	unknown	unknown	unknown	9
Heb9	HEB_013	17	unknown	unknown	f	unknown	unknown	1656
HebMMcl	HEB_001	64	unknown	unknown	f	unknown	unknown	3322

HebNN	HEB_005	unknown	unknown	unknown	m	unknown	unknown	541
HED	SFK_005	77	1900	190	m	1977	197	14808
HJT1	SFK_013	72	1904	190	m	1976	197	8865
HJWC	SFK_037	74	1902	190	m	1976	197	9928
HM	KEN_003	85	1890	189	m	1975	197	42253
IBAB	SAL_008	unknown	unknown	unknown	m	1988	198	4971
IBAD	SAL_017 SAL_018	85	1901	190	m	1986	198	10532
IBAL	SAL_030	77	1902	190	m	1979	197	4366
IBAnBa	SAL_029	76	1911	191	f	1987	198	3396
IBBB	SAL_002 SAL_003	86	1889	188	m	1975	197	6587
IBBeBo	SAL_011	91	1903	190	m	1994	199	7416
IBBM	SAL_036	unknown	unknown	unknown	m	1979	197	6775
IBCB	SAL_005	88	1893	189	m	1981	198	2605
IBCD	SAL_021	77	1907	190	m	1983	198	3931
IBCE	SAL_026	78	1914	191	m	1992	199	6301
IBCH	SAL_034	89	1893	189	m	1982	198	3332
IBED	SAL_016	72	1903	190	m	1975	197	2831
IBEP	SAL_039	81	1905	190	m	1986	198	5528
IBErDi	SAL_020	82	1901	190	m	1983	198	6927
IBFB	SAL_013	74	1914	191	m	1988	198	5851
IBFF	SAL_027	unknown	unknown	unknown	m	1982	198	4196
IBGB	SAL_009	91	1902	190	m	1993	199	2941
IBGJ	SAL_032	83	1899	189	m	1982	198	4706
IBGL	SAL_031	85	1906	190	f	1991	199	5665
IBHP	SAL_037	69	1915	191	m	1984	198	3436
IBIB	SAL_007	unknown	unknown	unknown	f	1991	199	1897
IBJ	SAL_007	unknown	unknown	unknown	unknown	1991	199	24
IBJD	SAL_019	69	1916	191	m	1985	198	4677
IBJP	SAL_038	73	1906	190	m	1979	197	6142
IBLB	SAL_010	81	1913	191	f	1994	199	1501
IBLE	SAL_024 SAL_025	89	1895	189	m	1984	198	10477
IBME	SAL_023	76	1906	190	m	1982	198	15201
IBMrsA	SAL_001	76	1902	190	f	1978	197	764
IBMrsD	SAL_022	81	1894	189	f	1975	197	1262
IBMrsF	SAL_028	89	1890	189	f	1979	197	1769
IBNB	SAL_006	83	1901	190	m	1984	198	4982
IBPB	SAL_012	88	1907	190	f	1995	199	1768
IBRH	SAL_035	79	1901	190	m	1980	198	1925
IBSH	SAL_029	79	1908	190	m	1987	198	1984
IBTJ	SAL_033	82	1901	190	m	1983	198	4289
IBWB	SAL_004	78	1906	190	m	1984	198	4790
IBWiBr	SAL_014 SAL_015	99	1886	188	m	1985	198	8430
ICS_EW	CON_009	66	1914	191	m	1980	198	6739

ICS_JE	CON_007	59	1920	192	m	1979	197	10857
ICS_JP	CON_008	unknown	unknown	unknown	m	1979	197	9298
ICS_MW	CON_009	68	1912	191	f	1980	198	2662
ICS_RM	CON_010	70	1910	191	m	1980	198	8630
Inf	CON_003	unknown	unknown	unknown	unknown	1978	197	406
InvWR	INV_001	45	1952	195	m	unknown	unknown	3126
IomBK	IOM_001	85	1895	189	m	1980	198	5822
IomRC	IOM_002	77	1894	189	m	1971	197	4639
Jack	WIL_018	unknown	unknown	unknown	m	1982	198	183
JC	KEN_001	89	1886	188	m	1975	197	9493
JG	LND_004	66	1916	191	m	1982	198	15361
JR1	SFK_009 SFK_027	50	1926	192	m	1976	197	12023
JRS	SFK_011	85	1891	189	m	1976	197	10072
KcdES	KCD_001	unknown	unknown	unknown	f	1995	199	1666
KcdRS	KCD_001	71	1924	192	m	1995	199	5426
KentAW1	KEN_009	81	1894	189	m	1975	197	12089
KentEDG	KEN_007	81	1895	189	m	1976	197	13797
KentPB	KEN_004 KEN_010 KEN_011	87	1889	188	m	1976	197	42300
KentSPR	KEN_008	83	1893	189	m	1976	197	17588
KentTAD	KEN_005	86	1890	189	m	1976	197	10487
KrsWS	KRS_001	69	1917	191	m	1986	198	2218
LA	LND_005	80	1901	190	f	1981	198	12552
LAN002	LAN_002	83	1892	189	f	1975	197	9287
LanC5P	LAN_005	61	1919	191	f	1980	198	10720
LanCD	LAN_014	63	1913	191	m	1976	197	53
LanD1B	LAN_001	76	1899	189	f	1975	197	11392
LanD1P	LAN_006	71	1908	190	f	1979	197	10080
LanEG	LAN_012	71	1906	190	m	1977	197	27977
LanF1P	LAN_007	73	1906	190	m	1979	197	8585
Lang1p	LAN_003	76	1903	190	m	1979	197	9389
LanH1P	LAN_009	67	1911	191	f	1978	197	11110
LanMJ	LAN_011	47	1930	193	f	1977	197	13677
LanPK	LAN_013	62	1914	191	m	1976	197	727
LanT3P	LAN_008	93	1886	188	m	1979	197	10193
LanW1P	LAN_004	80	1899	189	f	1979	197	8882
Lei2	LEI_002	unknown	1903	190	m	unknown	197	5888
LeiA	LEI_001	unknown	unknown	unknown	m	unknown	197	1034
LeiB	LEI_001	unknown	unknown	unknown	m	unknown	197	1389
LksGH	LKS_001	75	1910	191	m	1985	198	3599
M	SOM_014	78	1902	190	f	1980	198	19
MdxCG	MDX_001	79	1897	189	m	1976	197	13686
MH3L	LAN_010	unknown	unknown	unknown	f	1975	197	27
ML	WIL_018	unknown	unknown	unknown	unknown	1982	198	62
Mln2Anon	MLN_003	unknown	unknown	unknown	f	unknown	197	50

MlnAA	MLN_003	unknown	1907	190	m	unknown	197	1796
MlnAnon	MLN_002	unknown	unknown	unknown	unknown	unknown	197	118
MlnCM	MLN_002	unknown	unknown	unknown	f	unknown	197	1857
MlnDM	MLN_004	unknown	unknown	unknown	m	unknown	197	1722
MlnJH	MLN_007	unknown	1919	191	m	unknown	197	12799
MlnJM	MLN_005	unknown	unknown	unknown	f	unknown	197	3299
MlnRH	MLN_006	unknown	unknown	unknown	m	unknown	197	6045
MlnRS	MLN_001	unknown	unknown	unknown	m	unknown	197	2049
MlnSM	MLN_004	unknown	unknown	unknown	f	unknown	197	388
NBall	SAL_007	unknown	unknown	unknown	unknown	1991	199	5
NblAn	NBL_007	unknown	unknown	unknown	f	1974	197	36
NblAR	NBL_007	83	1891	189	m	1974	197	8504
NblGS	NBL_008	unknown	unknown	189	f	1975	197	3171
NblJB	NBL_006	unknown	unknown	189	f	1975	197	8440
NblMN	NBL_003	83	1890	189	f	1973	197	7886
NblWL	NBL_001	80	1891	189	m	1971	197	395
NotA103	NTT_001	85	1899	189	m	1984	197	5664
NotA25	NTT_002	85	1897	189	m	1982	198	11247
NotA30	NTT_003	81	1902	190	m	1983	198	9221
NotA40	NTT_004	66	1917	191	m	1983	198	4959
NotA48	NTT_016	unknown	1884	188	m	unknown	unknown	14530
NotA52	NTT_005	85	1898	189	m	1983	198	11839
NotA58	NTT_006	81	1902	190	f	1983	198	14341
NotA65	NTT_007	86	1897	189	f	1983	198	6992
NotA71	NTT_008	81	1902	190	m	1983	198	5508
NotA8	NTT_013	unknown	1908	190	m	unknown	198	13259
NotA80	NTT_009	78	1906	190	f	1984	198	8610
NotA85	NTT_010	72	1912	191	f	1984	198	9418
NotA99	NTT_011	91	1893	189	m	1984	198	9135
PeeJP	PEE_001	33	1952	195	m	1985	198	2185
PeeMT	PEE_002	79	1905	190	f	1984	198	12399
PerAB	PER_003	80	1913	191	m	1993	199	8362
PerHK	PER_004	73	1913	191	m	1986	198	4721
PerSS	PER_001	84	1902	190	m	1986	198	3713
PerTJ	PER_002	unknown	unknown	unknown	m	1986	198	2219
PerWG	PER_005	unknown	unknown	unknown	m	unknown	198	1463
PerWM	PER_005	unknown	1893	189	m	unknown	198	67
REK	SFK_008	92	1884	188	m	1976	197	2078
RocAM	ROC_001	81	1901	190	f	1982	198	5685
RocPN	ROC_002	79	1921	192	m	2000	200	4446
RR1	SFK_007	70	1906	190	m	1976	197	8367
SelAN	SEL_003	unknown	1931	193	m	unknown	198	1792
SelJB	SEL_001	82	1902	190	m	1984	198	4357
SelJM	SEL_002	59	1926	192	m	1985	198	2883
SRLM_AB	SOM_018	72	1901	190	m	1973	197	2852
SRLM_AE	SOM_020	86	1892	189	m	1978	197	3402
SRLM_AW	SOM_006	89	1894	189	m	1982	198	4270

SRLM_B	SOM_003	86	1897	189	m	1983	198	3662
SRLM_CA	SOM_009	75	1916	191	m	1991	199	10523
SRLM_CH	SOM_008	86	1904	190	m	1990	199	5497
SRLM_CJ	SOM_025	unknown	unknown	unknown	m	1983	198	1066
SRLM_CK	SOM_004	71	1916	191	m	1987	198	4968
SRLM_CM	SOM_023	unknown	unknown	unknown	m	1983	198	754
SRLM_EC	SOM_033	72	1923	192	m	1995	199	8016
SRLM_ES	SOM_008 SOM_036	85	1904	190	m	1989/1990	198/199	3565
SRLM_EW	SOM_034	83	1910	191	f	1993	199	2967
SRLM_GG	SOM_013	78	1918	191	m	1996	199	4573
SRLM_GP	SOM_027	unknown	unknown	unknown	f	1985	198	1549
SRLM_HC	SOM_026	71	1912	191	m	1983	198	1578
SRLM_HG	SOM_032	82	1904	190	m	1986	198	8869
SRLM_HH	SOM_015	84	1889	188	m	1973	197	2697
SRLM_HM	SOM_024	81	1902	190	m	1983	198	4132
SRLM_HR	SOM_001	unknown	unknown	unknown	m	1983	198	5459
SRLM_HW	SOM_030	88	1897	189	m	1985	198	4455
SRLM_JA	SOM_029	70	1915	191	m	1985	198	6716
SRLM_LD	SOM_017	85	1888	188	m	1973	197	3277
SRLM_M	SOM_023	unknown	unknown	unknown	f	1983	198	931
SRLM_PG	SOM_002	unknown	unknown	unknown	m	1983	198	3561
SRLM_PT	SOM_007	82	1898	189	m	1980	198	6321
SRLM_R	SOM_021	unknown	unknown	unknown	m	1978	197	6976
SRLM_RF	SOM_005	76	1905	190	m	1981	198	7506
SRLM_RM	SOM_014	78	1902	190	m	1980	198	10480
SRLM_S	SOM_022	unknown	unknown	unknown	m	1968	196	1252
SRLM_SC	SOM_035	80	1914	191	m	1994	199	3485
SRLM_SH	SOM_012	68	1927	192	m	1995	199	6215
SRLM_T	SOM_022	unknown	unknown	unknown	m	1968	196	1345
SRLM_TP	SOM_027	81	1904	190	m	1985	198	1525
SRLM_W	SOM_019	unknown	unknown	unknown	m	1973	197	11318
SRLM_WB	SOM_010 SOM_011	76	1904	190	m	1980	198	24872
SRLM_WP	SOM_016	89	1884	188	m	1973	197	3135
SRLM_WT	SOM_028	87	1897	189	m	1984	198	14014
SRLM_Y	SOM_031	83	1894	189	m	1981	198	5632
SufEAA2	SFK_014	74	1902	190	m	1978	197	8851
SufEAA3	SFK_015	75	1902	190	m	1977	197	7930
SufEAA4	SFK_016	75	1902	190	m	1977	197	8424
SufEAA5	SFK_035 SFK_038	78	1902	190	m	1980	198	15725
SufEF2	SFK_018	79	1898	189	m	1977	197	7636
SufEQU	SFK_039	72	1911	191	m	1983	198	8084
SufGCS2	SFK_030 SFK_033	75	1903	190	m	1978	197	14530
SufGFK	SFK_020	74	1907	190	m	1981	198	7192

SufHED2	SFK_021	78	1900	190	m	1978	197	6678
SufHED3	SFK_022	77	1900	190	m	1978	197	6676
SufHED4	SFK_023	78	1900	190	m	1978	197	6518
SufHED5	SFK_031	80	1900	190	m	1980	198	7177
SufHFJE	SFK_017	67	1912	191	m	1979	197	8054
SufHJT2	SFK_028 SFK_029	73	1904	190	m	1977	197	17245
SufSMW	SFK_034	79	1900	190	m	1979	197	8459
SufWCG1	SFK_032	67	1909	190	m	1976	197	10052
SufWEJ1	SFK_024 SFK_025 SFK_026	73	1906	190	m	1978	197	22627
SutDS	SUT_001	69	1914	191	m	1983	198	163
SutGM	SUT_002	61	1922	192	m	1983	198	3809
SutGMcK	SUT_003	82	1901	190	m	1983	198	3194
SutTR	SUT_004	54	1929	192	m	1983	198	3449
SwanBE	GLA_003	87	1886	188	m	1973	197	7896
TCA_C	DEV_003	unknown	unknown	unknown	m	unknown	198	1687
TCA_DR	DEV_011	unknown	unknown	unknown	f	1985	198	501
TCA_EA	DEV_007	93	1892	189	f	1985	198	8869
TCA_EC	DEV_003	unknown	unknown	unknown	f	unknown	198	4890
TCA_FK	DEV_008	unknown	1910	191	m	unknown	unknown	8345
TCA_FP	DEV_005	82	1902	190	m	1984	198	10949
TCA_GB	DEV_002	unknown	1900	190	m	unknown	198	5571
TCA_GR	DEV_006	83	1901	190	m	1984	198	7521
TCA_K	DEV_008	unknown	unknown	unknown	f	unknown	unknown	4154
TCA_PR	DEV_011	87	1898	189	m	1985	198	3722
TCA_RA	DEV_001	76	1909	190	m	1985	198	10204
TCA_SC	DEV_004	unknown	unknown	unknown	m	1985	198	4708
TCA_WC	DEV_010	unknown	1913	191	f	unknown	unknown	9652
TCA_WH	DEV_009	79	1906	190	m	1985	198	11345
thirdperson	LAN_002	unknown	unknown	unknown	f	1975	197	265
TrbrAH	WIL_017	unknown	unknown	unknown	f	1992	199	4952
TrbrAS	WIL_023	81	1913	191	f	1994	199	8061
TrbrBC	WIL_017	unknown	unknown	unknown	f	1992	199	867
TrbrBJ	WIL_007	unknown	unknown	unknown	m	1992	199	3906
TrbrCN	WIL_005	unknown	unknown	unknown	unknown	1992	199	5
TrbrCS	WIL_005	91	1901	190	m	1992	199	7966
TrbrD1	WIL_005	unknown	unknown	unknown	f	1992	199	1908
TrbrD2	WIL_005	unknown	unknown	unknown	f	1992	199	109
TrbrDC	WIL_018	unknown	1909	190	f	1982	198	11881
TrbrDH	WIL_005	unknown	unknown	unknown	unknown	1992	199	11
TrbrDP	WIL_004	80	1912	191	m	1992	199	7780
TrbrEGC	WIL_015 WIL_016	83	1909	190	m	1992	199	15065
TrbrFC	WIL_022	71	1923	192	m	1994	199	17690
TrbrGHG	WIL_014	79	1913	191	m	1992	199	4269

TrbrGR	WIL_001	85	1907	190	m	1992	199	12009
TrbrGSR	WIL_006	66	1926	192	m	1992	199	7640
TrbrIS	WIL_024	74	1920	192	m	1994	199	4983
TrbrJC	WIL_022	61	1933	193	f	1994	199	6445
TrbrJH	WIL_002	58	1934	193	f	1992	199	5953
TrbrJS	WIL_003	unknown	unknown	unknown	m	1987	198	9169
TrbrKJ	WIL_007	52	1940	194	f	1992	199	1685
TrbrKM	WIL_019	92	1901	190	f	1993	199	5208
TrbrLW	WIL_005	unknown	unknown	unknown	f	1992	199	57
TrbrQD	WIL_017	unknown	unknown	unknown	f	1992	199	4432
TrbrRCC	WIL_008	71	1922	192	m	1993	199	12924
TrbrRS	WIL_020 WIL_021	79	1914	191	m	1993	199	2710
TrbrS	WIL_021	73	1920	192	f	1993	199	1151
TrbrViv	WIL_005	unknown	unknown	unknown	f	1992	199	95
TrbrVL	WIL_013	78	1914	191	f	1992	199	1093
'unknown'	DEV_001	unknown	unknown	unknown	f	1985	198	14
W	SOM_030	unknown	unknown	unknown	unknown	1985	198	396
Wal128LG	WAR_001	unknown	unknown	unknown	m	unknown	198	8176
WalDJN	GLA_005	81	1895	189	m	1976	197	8712
WalDTC	GLA_006	77	1899	189	m	1976	197	5469
WalJSD	GLA_007	78	1898	189	m	1976	197	8671
WalWTA	GLA_004	unknown	unknown	unknown	m	1975	197	5505
WAR001unknown	WAR_001	unknown	unknown	unknown	unknown	unknown	198	6
WesAD	WES_001	84	1896	189	f	1980	198	10252
WesAE	WES_002	76	1905	190	m	1981	198	4650
WesAnon	WES_017	unknown	unknown	unknown	f	1991	199	20
WesAQ	WES_003	84	1898	189	m	1982	198	4538
WesAY	WES_004	79	1905	190	m	1984	198	7514
WesBE	WES_005	74	1910	191	m	1984	198	11022
WesBL	WES_006	84	1901	190	f	1985	198	3051
WesBP	WES_007	73	1912	191	f	1985	198	4749
WesBR	WES_008	78	1908	190	m	1986	198	10527
WesBX	WES_009	79	1907	190	m	1986	198	5618
WesCB	WES_010	73	1914	191	m	1987	198	7862
WesCE	WES_011	85	1902	190	m	1987	198	1927
WesCEa	WES_011	unknown	unknown	192	m	1987	198	1263
WesCU	WES_012	90	1899	189	f	1989	198	6211
WesCY	WES_013	85	1904	190	f	1989	198	3191
WesDA	WES_014	unknown	unknown	unknown	m	1989	198	5473
WesDB	WES_015	83	1906	190	f	1989	198	5478
WesDT	WES_016	80	1911	191	f	1991	199	13473
WesDX	WES_017	81	1910	191	m	1991	199	22800
WesHN	WES_018	77	1922	192	m	1999	199	5064
WesJN	WES_018	unknown	unknown	unknown	f	1999	199	105
WesZb	WES_019	70	1910	191	m	1980	198	11876
WesZs	WES_019	82	1898	189	f	1980	198	3083

WflsEF	WIL_012	93	1890	189	f	1983	198	2503
WflsFP	WIL_009	unknown	1914	191	m	unknown	unknown	4048
WflsGW	WIL_011	unknown	unknown	unknown	m	unknown	198	1387
WflsWGP	WIL_010	unknown	unknown	189	m	1982	198	4419
WflsWH	WIL_011	86	1900	190	m	unknown	198	3183
WG	KEN_007	unknown	unknown	unknown	f	1976	197	586
wife	SOM_007	unknown	unknown	unknown	f	1980	198	88
WigGRB	LAN_022	56	1925	192	f	1981	198	7764
WigHT	LAN_019	37	1945	194	m	1982	198	6267
WigJH	LAN_021	53	1929	192	m	1982	198	615
WigM	LAN_019	unknown	unknown	unknown	f	1982	198	59
WigMB	LAN_017	78	1904	190	f	1982	198	3055
WigMS	LAN_016	82	1899	189	f	1981	198	4283
WigMWH	LAN_015	56	1926	192	m	1982	198	6634
WigPS	LAN_018	35	1946	194	f	1981	198	926
WigWHJ	LAN_020	76	1906	190	m	1982	198	14958
WigWO	LAN_023	73	1909	190	m	1982	198	8807
WlnES	WLN_005 WLN_006	57	1931	193	m	1988	198	3881
WlnMD	WLN_003 WLN_004	83	1908	190	f	1991	198	8981
WlnMO	WLN_002	unknown	unknown	unknown	f	1984	198	1968
WlnTW	WLN_001	60	1927	192	m	1987	198	1790
WS	KEN_006	84	1892	189	m	1976	197	11946
WT	SFK_012 SFK_036	68	1908	190	m	1976	197	17349
XW	OXF_002	unknown	unknown	unknown	f	1979	197	576
Y	WIL_009	unknown	unknown	unknown	f	unknown	unknown	219
Yks_10	YKS_010	85	1899	189	m	1984	198	9442
YksAnon	YKS_011	unknown	unknown	unknown	unknown	unknown	198	282
YksBR	YKS_011	unknown	unknown	unknown	m	unknown	198	3688
YksEH	YKS_004	unknown	unknown	unknown	f	unknown	198	1907
YksGR	YKS_005	unknown	unknown	190	m	unknown	198	2954
YksHS	YKS_006	unknown	1910	191	m	unknown	198	7191
YksJS	YKS_011	unknown	unknown	unknown	m	unknown	198	1408
YksJV	YKS_002	unknown	1899	189	m	unknown	198	9439
YksMG	YKS_009	unknown	unknown	unknown	f	unknown	198	363
YksML	YKS_007	unknown	1900	190	m	unknown	198	11321
YksMR	YKS_011	unknown	1912	191	f	unknown	198	2530
YksMW	YKS_008	unknown	1900	190	f	unknown	198	9689
YksSL	YKS_003	unknown	1912	191	m	unknown	198	11891
YksWF	YKS_001	unknown	1910	191	m	unknown	198	10009
YksWG	YKS_009	80	unknown	190	m	unknown	198	5275

Table 7: FRED speaker list

Tag	Indicates	Description/examples
(pause)	pause in speech	also used with values, e.g. (pause '6 seconds') or (pause 'very long')
(v '...')	non-verbal element, vocal	e.g. (v 'laughs') or (v 'coughs')
(k '...')	non-verbal element, kinesic	e.g. (k 'points to the picture')
(e '...')	non-verbal element, event	e.g. (e 'cat jumps on the table') or (e 'visitor enters room')
(trunc) ...(/trunc)	truncation	e.g. He (trunc) ha (/trunc) he has two dogs.
(corr sic=...) ... (/corr)	editorial correction	e.g. They took (corr sic=births) berths (/corr) on those larger boats. Predominantly used for editing source transcripts (also as temporary mark during the editing process). Corrections of obvious misspellings are not indicated.
(reg sic=...) ...(/reg)	regularisation disambiguation enhances comprehensibility	e.g. (reg sic=so's) so as (/reg) or (reg sic=summat) somewhat (/reg) Used where source transcripts rendered non-standard pronunciations, especially uncommon ones. This method allows for word searches using standard orthography, without losing the information of the source transcript. Also used to disambiguate reduced word forms, e.g. they (reg sic='re) were (/reg) (not 'are') and to enhance comprehensibility of uncommon dialectal realisations of words, e.g. to fetch it (reg sic=yame) home (/reg) The regularisation of constant dialect features, e.g. h-dropping, is usually not mentioned in the text but in the respective text header.
(sic) ... (/sic)	dubious item	Used to indicate that the item in question was found in the source transcript; e.g. that didn't belong (sic) toll (/sic) him
(unclear) ...(/unclear)	unclear utterances	Also used for source transcripts without recordings when original handwriting was difficult to read.
(gap '...')	gap in transcript	Used whenever a part of the recorded dialogue could not be understood at all; also used for unreadable source transcripts without recordings; e.g. there was a (gap'indistinct') you know; also possible (gap 'tape interrupted'), (gap 'three words inaudible'), (gap 'placename unclear'), etc. (note the difference between the (unclear) and the (gap) tag)
(name) ...(/name)	anonymisation of person and family names	(iname) ... (/iname) for initials (sname) ... (/sname) for surnames (mname) ... (/mname) for male first names and male nicknames (fname) ... (/fname) for female first names and female nicknames Titles (Mr/Mrs/Dr/Lord, etc.) precede the tags. Examples: (iname) J. (/iname) (iname) B. (/iname) (sname) Tugster (/sname) Dr (sname) Thomas (/sname) (mname) John (/mname) (sname) Towers (/sname) Mr (mname) John Gary (/mname) (sname) Hearty (/sname) (mname) Bob's (/mname) Mr (iname) M. (/iname) (sname) Guillian (/sname) Mr and Mrs (sname) Portman (/sname) Reverend (sname) Gardner (/sname) the (sname) Tates (/sname) Lord (sname) Henry (/sname) Mrs (mname) John (/mname) (iname) G. (/iname) (sname) Green (/sname)

Table 8: Semi-standardised tag