Manual of Information to accompany the *Freiburg Corpus of English Dialects Sampler* ("FRED-S")

Benedikt Szmrecsanyi
benedikt.szmrecsanyi@uni-freiburg.de

Nuria Hernández
nuria.hernandez@pluto.uni-freiburg.de

English Department
University of Freiburg
2007

Contents

List of Tables

List of Figures

Acknowledgements

The *Freiburg Corpus of English Dialects* (FRED) was compiled by the research group 'English Dialect Syntax from a Typological Perspective', based at the English Department of the University of Freiburg and supervised by Bernd Kortmann. Generous support by the *Deutsche Forschungsgemeinschaft* (grant no Ko 1181/1-3) in the period between 2000 and 2005 is gratefully acknowledged.

Over the years, the following researchers were substantially involved in the compilation and completion of the corpus: Bernd Kortmann (supervisor), Lieselotte Anderwald (assistant supervisor), Tanja Herrmann, Lukas Pietsch and Susanne Wagner (text collection and compilation, transcription, original markup), Nuria Hernández and Benedikt Szmrecsanyi (markup, transcription, digitisation).

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.

Lastly, we are also indebted to all the dedicated fieldworkers, archives, museums and libraries who have entrusted their data to us, making it possible to build a corpus which will continue to profit linguistic research on varieties of English spoken in the British Isles.

1. General introduction to the *Freiburg Corpus of English Dialects*

1.1. The project

What type of corpus is the *Freiburg Corpus of English Dialects* (FRED), and how can linguistic research benefit from it? FRED is a monolingual spoken-language dialect corpus[1] which contains full-length interviews with native speakers from England, Scotland, as well as (in its full version) Wales, 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, nor POS-tagged or syntactically parsed. The data originate from so-called oral history projects where informants were interviewed to record their life memories. Owing to the fact that large quantities of text are needed to investigate morphosyntactic phenomena, oral history interviews provide just the appropriate quantity of spontaneous speech data needed to investigate both frequent and medium-range frequent syntactic phenomena.[2] All texts are face-to-face conversations between (usually) one interviewer and an informant 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 setting and the interest expressed in their life stories helped to sufficiently distract them from their own linguistic behaviour.

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. FRED is not designed as a sociolinguistic corpus, but a limited set of sociolinguistic variables is specified for each text (geographic data, and – often – age and sex of the speaker; see section 3.5 [p. 8]). Transcripts are orthographic; nevertheless, sound files are available for most interviews (in the case of FRED-S, for all interviews). Even researchers primarily interested in phonetic or phonological phenomena might find this sizable collection of audio recordings useful.

1 FRED and FRED-S are specialised corpora, as opposed to general corpora such as the BNC which contain different registers and spoken and written language data.

2 For linguistic consequences of using oral history material – for example, the predominance of past-tense paradigms – see Anderwald & Wagner (2007).

1.2. Previous research based on FRED

The primary aim of compiling FRED was the research group's interest in morphosyntactic variation in British English dialects and the lack of geographically well-balanced, easy-to-access, machine-readable databases. The wider theoretical framework is based on functional Greenbergian typology, as language-internal variation (in English, in this case) may be integrated into global variation patterns (see Kortmann 2004 for a collection of papers in this spirit). Anderwald & Wagner (2007) present a detailed account of the very first studies conducted by members of the research group. Morphosyntactic phenomena subject to analysis have included the following: relativization (Herrmann, 2003, 2005), pronoun usage (Wagner, 2004a,b, 2005), verbal agreement (Pietsch, 2005a,b), morphosyntactic persistence (Szmrecsanyi, 2005, 2006), genitive variation (Szmrecsanyi & Hinrichs to appear), and non-standard verbal morphology (Anderwald to appear).

For an up-to-date account of published and ongoing research, including master's and doctoral theses, consult the project web site at `www.anglistik.uni-freiburg.de/institut/lskortmann/FRED/`.

1.3. FRED versus FRED-S

The full version of FRED contains approximately 2.5 mio. words and 300 hours of recorded speech. It consists of 372 interviews with male and female speakers from 163 different locations in 43 different counties in 9 major dialect areas (cf. Hernández 2006 for a manual). FRED, in its full version, is available to researchers and visiting scholars at the University of Freiburg.

FRED-S spans a subset of FRED texts not subject to copyright restrictions. More specifically, FRED-S covers

- 1,011,396 running words,[3]
- c. 123 hours of recorded speech,
- 121 interviews,
- 144 dialect speakers,
- 57 different locations,
- 18 different counties,
- 5 major dialect areas (the Southwest of England, the Southeast of England, the Midlands, the North of England, and the Scottish Lowlands).

3 Word counts may of course vary, according to the definition of 'word'. The numbers given in this manual include truncations and false starts; hyphenated words are counted as one word, and so are *not*-contractions. All other contractions (e.g. *he 's going to school*), however, are transcribed and counted as two words.

2. Sources and sampling techniques

The process of compiling recordings and transcripts for the corpus started in 2000. The following criteria guided the selection of material:

1. The main aim was to compile material for investigating morphosyntactic dialect phenomena.
2. There was a preference for traditional dialect data, one reason being the comparability to other, already existing sources mapping morphosyntactic variation in British English dialects (for instance, the *Survey of English Dialects* [SED])
3. Tape recordings had to be in acceptable quality (ideally, but not necessarily, accompanied by transcripts)

Since more textual material is usually needed to investigate morphosyntactic phenomena than, e.g., phonological phenomena, the compilers aimed at a comparatively large database of dialect material. Due to restricted time and resources, there was a preference for 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 Chapter 5 [p. 16]) and stored electronically. Those interviews deemed most suitable for our purposes[4] were then transcribed (either from scratch or revised) by English native speakers and linguistically trained staff.

For a number of texts, transcripts were available but had to be checked against and adapted to corpus markup conventions. In these cases, there was still the obvious advantage of having preliminary transcripts which occasionally included unknown place names and specialist vocabulary (e.g. specialist tools or, e.g., types of apples used for cider making). Still, the available transcripts had been intended for oral history projects. This means that 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 very features which are of particular interest for linguistic research. Consider the difference between the following two renderings of the same text passage (cf. Anderwald & Wagner 2007):

4 More material is stored in Freiburg. It is not included in the current version of either FRED or FRED-S but may be used in the future.

(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)

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 re-inserted; 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 which might facilitate word searches (cf. the special case *mi* in section 4.4 [p. 12]). For a full description of the transcription guidelines see Chapter 4 (p. 10).

3. Corpus design

3.1. The recordings

All conversations in FRED-S were recorded between 1970 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 4.1 [p. 11] for an example). Except for a few texts where the tape was interrupted or the recording stops before the actual conversation ends[5], most recordings are full-length interviews which last between 30 and 90 minutes.[6]

recording date	number of texts	% of textual material
1970–1979	47	42.2%
1980–1989	56	43.9%
1990–1999	15	10.3%
unknown	3	3.6%
Total	121	100%

Table 3.1.: Text distribution by recording date in FRED-S

3.2. Transcripts and file names

Each recording yields a transcript with the same text identification number (e.g. DUR_001.txt). Table B.1 (p. 19) lists all FRED-S texts, along with word counts.

FRED-S file names are composed of a three letter Chapman county code, followed by a underscore and a number (001, 002, ...).[7] Corresponding text and sound files have the same iden-

5 These texts were included nevertheless, as they still contain long passages of free speech.

6 The longest interview is KEN_003, with almost 4.5 hours (roughly 47,000 words, or 241 KB).

7 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 a standard format for genealogical purposes, present a clear-cut solution to this problem. As most interviews were recorded during the 1970s and 1980s 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.

tification, e.g. `DEV_004.txt` and `DEV_004.wav`. Note that, for longer interviews, there may be more than one audio file (e.g., `NTT_006A.wav` and `NTT_006B.wav`).

3.3. Dialect areas

FRED-S covers 5 dialect areas (or a geographical division of modern dialect areas, see Trudgill 1999: 65). Table 3.2 displays a breakdown by number of texts and number of running words (excluding interviewer utterances). The dialect area of each text is specified in the text header (`<area value>`).

dialect area	number of texts	running words	% of textual material in corpus
Southwest (SW)	38	264,863	26.2%
Southeast (SE)	17	260,643	25.8%
Midlands (Mid)	16	152,535	15.1%
North (N)	30	266,955	26.4%
Scottish Lowlands (ScL)	20	66,400	6.6%
Total	121	1,011,396	100%

Table 3.2.: Number of texts and running words (excluding interviewer utterances) by dialect area in FRED-S

3.4. Counties and locations

The county of each text is specified in the text header (`<county value>`) and can be easily derived from the Chapman county code on which file names are based (`LAN_001`, for example, is an interview from Lancashire). Table 3.3 [p. 8] below lists counties represented in FRED-S, along with word counts

The location of each text – if known – is also specified in the text header, as `<location value>`. Table A.1 (p. 17) provides an overview of locations sampled in the corpus; Figure 3.1 [p. 7] visualizes the areal coverage of FRED-S. For convenience, the dialect area, county and location of each text are also specified in the FRED-S text list (B.1 [p. 19]), which can be combined with the FRED-S speaker list (C.1 [p. 22]) to obtain speaker details or vice versa.

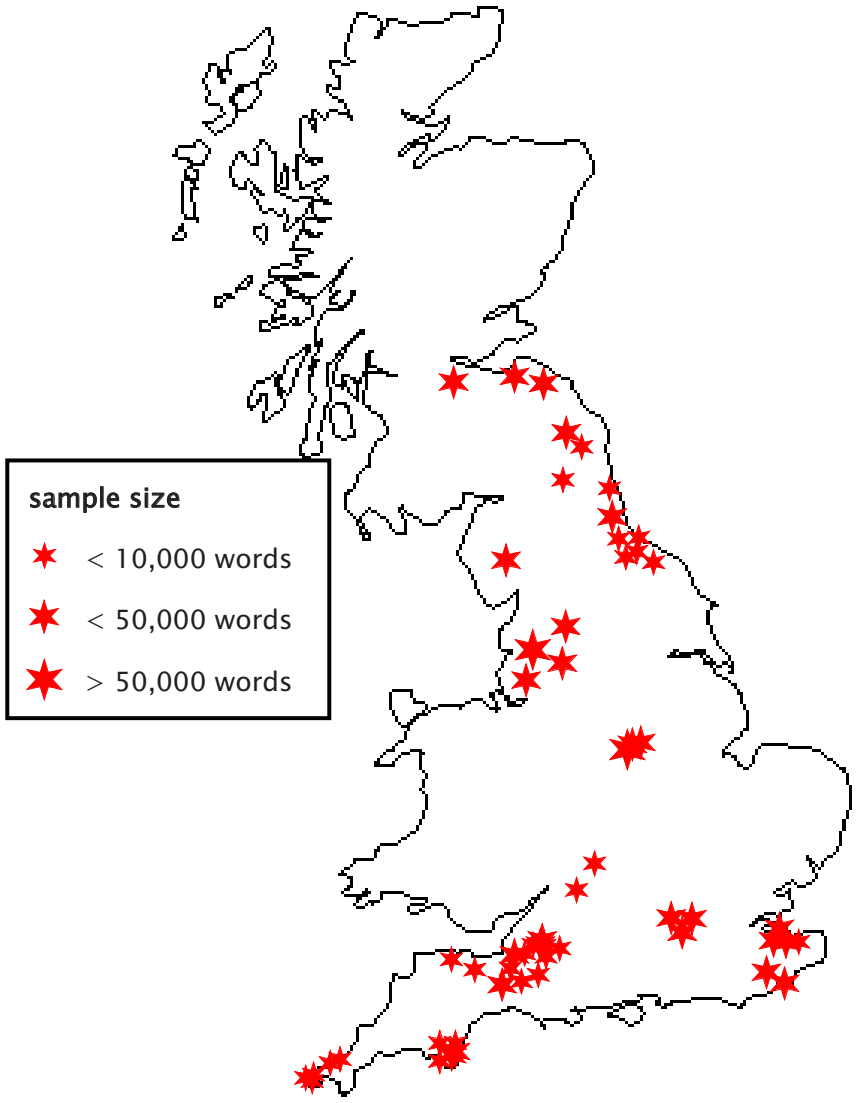

Figure 3.1.: FRED-S areal coverage: locations sampled in the corpus

county	Chapman county code	dialect area	running words	% of textual material within corpus
Cornwall	CON	SW	26,535	2.6%
Devon	DEV	SW	79,870	7.9%
Oxfordshire	OXF	SW	13,801	1.4%
Somerset	SOM	SW	69,321	6.9%
Wiltshire	WIL	SW	75,336	7.4%
Kent	KEN	SE	155,192	15.3%
London	LND	SE	74,856	7.4%
Middlesex	MDX	SE	30,595	3.0%
Leicestershire	LEI	Mid	2,341	0.2%
Nottinghamshire	NTT	Mid	150,194	14.9%
Durham	DUR	N	26,507	2.6%
Lancashire	LAN	N	139,845	13.8%
Northumberland	NBL	N	27,777	2.7%
Westmorland	WES	N	21,304	2.1%
Yorkshire	YKS	N	51,522	5.1%
East Lothian	ELN	SCL	28,985	2.9%
Midlothian	MLN	SCL	21,068	2.1%
West Lothian	WLN	SCL	16,347	1.6%

Table 3.3.: Running words (excluding interviewer utterances) by county in FRED-S

3.5. Speakers

Most informants are so-called NORMs – non-mobile old rural males – who typically left school at age fourteen or younger. The ratio of running text produced by male speakers to running text produced by female speakers is roughly 74:26 (with 87 male and 52 female informants in total; cf. Table 3.4 below for a breakdown). About three quarters of the overall textual material in FRED-S is produced by male speakers.[8]

	number of speakers	% of speakers	running words	% of textual material within corpus
male	87	60.4%	742,873	73.5%
female	52	36.1%	267,966	26.5%
sex unknown	5	3.5%	398	0.0%

Table 3.4.: Speaker distribution and text production by speaker sex in FRED-S

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 and did not

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

leave it for any considerable period of time. This is true for most FRED-S speakers.

FRED-S samples 144 speakers[9] (interviewers excluded) who all grew up in Britain. Most of them were born before World War I (mean date of birth is 1905) and were aged 60 or over when interviewed (mean age is 74.5 years at recording date). About 70% of the textual material in FRED-S was produced by the 60+ age group. The younger speakers included in the corpus produced only a small percentage of the overall textual material. A breakdown of the amount of text produced by the different age groups is presented in Table 3.5. For a breakdown according to date of birth see Table 3.6.

age group	number of speakers	running words	% of textual material within corpus
0–44 years	4	12,287	1.2%
45–59 years	5	40,258	4.0%
60+ years	71	726,134	71.8%
age unknown	64	232,558	23.0%

Table 3.5.: Speaker distribution and text production by age group in FRED-S

decade of birth	number of speakers	running words	% of textual material within corpus
1870–1879	1	6,899	0.7%
1880–1889	5	89,615	8.9%
1890–1899	28	260,909	25.8%
1900–1909	30	281,068	27.8%
1910–1919	21	176,507	17.5%
1920–1929	7	74,365	7.4%
1930–1939	3	23,494	2.3%
1940–1949	1	1,684	0.2%
date of birth unknown	48	96,696	9.5%

Table 3.6.: Speaker distribution and text production by birth decade in FRED-S

The oldest FRED-S informant was born in 1877; 88.5% of all informants were born before 1920. Exhaustive sociological information on individual speakers can be gleaned from the FRED-S speaker list (Table C.1 [p. 22]).

9 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.

4. Text markup and orthography

A consistent markup was applied to both the texts transcribed at Freiburg University and pre-existing transcripts from other sources. FRED-S is an orthographically transcribed corpus. The decision in favour of orthographic transcripts was made for a number of reasons, the most important being the research group's predominant interest in morphosyntactic features.

4.1. Format and bracket types

All transcriptions are pure ASCII texts with no text-processor formatting. All markup is separated from the running text by brackets.

- `< >` angular brackets enclose text and speaker information
- `{ }` curly brackets enclose interviewer utterances
- `( )` round brackets enclose semi-standardised tags (see Table 4.1 [p. 13] below for an exhaustive list)

All speaker utterances start in `<u speakerID>`. Interviewer utterances – which are part of the recorded dialogue but might be less relevant for some studies – have the following format: `{<u interviewerID> text}`. Curly brackets are an additional markup feature that allows researchers to exclude interviewer utterances from word searches in concordance programmes such as *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 4.1 shows the beginning of an interview with two informants in Swarland, Northumberland, in 1974.

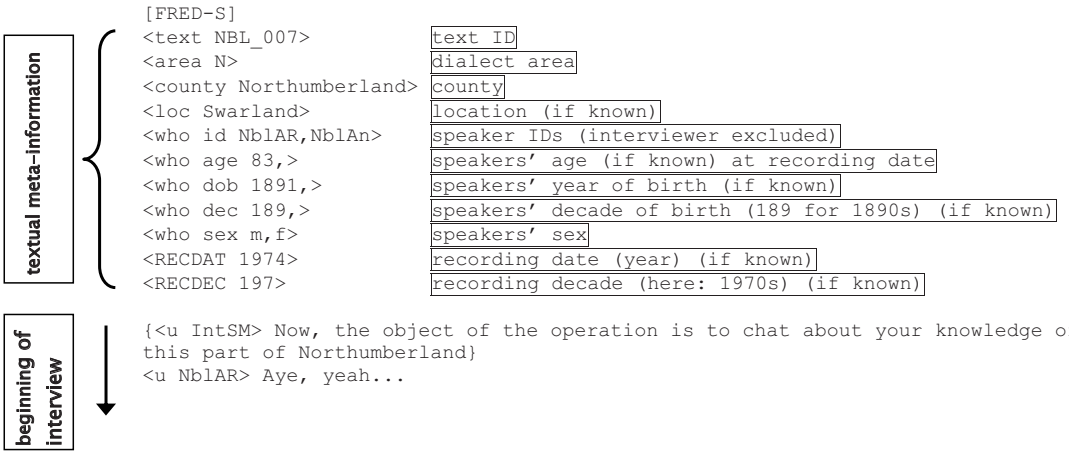

Figure 4.1.: Header and first lines of a FRED-S interview

The header consists of a set of tags in angular brackets (<*tagname value*>) which specifies the following information:

text ID	<text *value*>
dialect area	<area *value*>
county	<county *value*>
location	<loc *value*>
speaker IDs	<who id *value*>
speaker age at recording date	<who age *value*>
speaker date of birth	<who dob *value*>
speaker decade of birth, e.g. 189 for 1890s	<who dec *value*>
speaker sex, i.e. f (female) or m (male)	<who sex *value*>
recording date	<RECDAT *value*>
recording decade, e.g. 197 for 1970s	<RECDEC *value*>

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 is not known. Information is given where available; otherwise, the value is empty.

Notice here that some interviews were made with more than one informant, in which case the tags contain information on all speakers, separated by single commas. To illustrate: in text NBL_007 (cf. Figure 4.1), two speakers are interviewed: speaker NblAR and speaker NblAn (<who

`id NblAR,NblAn>`). Speaker `NblAR` is 83 years of age while the age of speaker `NblAn` is unknown (`<who age 83,>`); speaker `NblAR` was born in 1891 while the date of birth of speaker `NblAn` is unknown (`<who dob 1891,>`); and so on.

4.3. Indexing of semi-standardised tags in the running text

Round brackets are used for a number of semi-standardised tags. These tags come in three varieties:

1. as start-end pairs `(`*`value`*`)` ... `(/`*`value`*`)` with a space after the start-tag and a space before the end-tag
2. as single tags containing a tag name and a value `(`*`tagname 'value'`*`)`
3. or as single tags containing a tag name, an attribute name and an attribute value `(`*`tag name attribute=value`*`)`.

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 section 4.6 [p. 15]). A complete list of the semi-standardised tags used in FRED-S is given in Table 4.1 (p. 13).

4.4. Orthography

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 direct speech (see below).

Alphabetisms. Spoken alphabetisms (such as *USA* or *TV*) are rendered as alphabetisms.

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 typically spelled out in words (e.g. *nineteen hundred and sixty-four* or *nineteen oh seven*).

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

Punctuation marks. Full stops, commas, exclamation marks, question marks, semicolons, and dashes were used impressionistically according to their ordinary prosodic and syntactic meaning. Dashes are typeset as double hyphens surrounded by spaces (`- -`) to distinguish them from hyphenated spellings.

Direct speech. Direct speech is indicated not by quotation marks, but by a comma followed by a capital letter (as in `he said, No, he said, Never seen it`).

tag	meaning	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. Corrections of obvious misspellings are not indicated.
`(reg sic=...)` ... `(/reg)`	regularisation; disambiguation	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 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)` and to enhance comprehensibility of uncommon dialectal realisations of words, e.g. `to fetch it (reg sic=yame) home (/reg).`
`(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	e.g. `Oh he was uh supposed to be (unclear) a (/unclear) groom gardener`
`(gap '...')`	gap in transcript	Used whenever a part of the recorded dialogue could not be understood at all; e.g. `there was a (gap 'indistinct ') you know`; also possible: `(gap 'tape interrupted')`, `(gap 'three words inaudible')`, `(gap 'place name 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)`

Table 4.1.: Semi-standardised tags in FRED-S

Contractions. Contrary to standard orthography, auxiliary contraction (i.e. contractions with *be*, *have*, *do* and modal verbs) is rendered as two orthographic words: `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*) are often preserved in the transcript along with their standard equivalent (e.g. `(reg sic=t') to (/reg)`). Contracted verbal negation, whether standard or non-standard, is rendered as one word (e.g. `hasn't, can't, cannae, dinna`).

Special case *mi*. possessive /mɪ/ is spelled *mi* (e.g. `mi brother`) in order to distinguish it from object-case *me*.

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 in 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 such as `'em` (*them*) or `'er` (*her*).

4.5. Rendering of non-standard pronunciations

Due to the corpus compilers' primary interest in morphosyntax, FRED-S transcripts do not offer consistent renderings of phonological features. However, 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 g-dropping at the end of words ending in <ng>) are often marked in the text. Other common non-standard pronunciations, such as /huːs/ (*house*) or /mek/ (*make*), are usually regularised.

Non-standard realisations are rendered most consistently with grammatical items (pronouns, auxiliaries, etc.), especially in those areas of English morphosyntax 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 tokens that are not phonologically conditioned in a straightforward way 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 (e.g., `to fetch it (reg sic=yame) home (/reg)`). The same tag is used for non-standard morphological features (e.g., `they (reg sic=gien) gave (/reg) us`).

4.6. Anonymisation

All FRED-S transcripts are anonymised. The stories told in the corpus are private in nature, which is why 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 4.1 [p. 13] for a complete list of name tags and examples). Names of ships, trains, larger companies, military regiments, brands, etc. remain unchanged.

5. Storage and documentation

The tape and MD copies of the original recordings are kept at Freiburg University, along with the documentation of the collection and compilation process. The transcripts are stored as text files. The sound recordings were digitised and stored electronically (slightly denoised, but otherwise unaltered) in .wav and compressed .mp3 format. Text and audio samples can be found on the project website: `www.anglistik.uni-freiburg.de/institut/lskortmann/FRED/`.

Researchers interested in obtaining FRED-S sound files can contact the corpus compilers at:

Chair Prof. Bernd Kortmann
English Department
University of Freiburg
Rempartstr. 15
79098 Freiburg
Germany

phone +49-(0)761-203-3320
fax +49-(0)761-203-3330
`melitta.cocan@anglistik.uni-freiburg.de`

Appendix A. FRED-S locations

location	county	dialect area	running words
–	Somerset (SOM)	SW	3,525
Barton St. David	Somerset (SOM)	SW	10,448
Blackawton	Devon (DEV)	SW	10,195
Brixham	Devon (DEV)	SW	6,534
Buckfast	Devon (DEV)	SW	9,645
Buckleigh	Somerset (SOM)	SW	6,207
Carnelloe	Cornwall (CON)	SW	3,999
Churchtown	Cornwall (CON)	SW	2,975
Filkins	Oxfordshire (OXF)	SW	7,461
Galmpton	Devon (DEV)	SW	5,244
Glastonbury	Somerset (SOM)	SW	3,459
Gurnards Head	Cornwall (CON)	SW	7,483
Horton	Somerset (SOM)	SW	10,503
Leafield	Oxfordshire (OXF)	SW	6,340
North Burrowbridge	Somerset (SOM)	SW	4,966
North Petherton	Somerset (SOM)	SW	2,938
Pendeen	Cornwall (CON)	SW	4,751
Petherton	Somerset (SOM)	SW	5,566
St. Ives	Cornwall (CON)	SW	7,327
Street	Somerset (SOM)	SW	7,482
Sunnyside	Somerset (SOM)	SW	4,237
Totnes	Devon (DEV)	SW	48,252
Trowbridge	Wiltshire (WIL)	SW	49,064
Urchfont	Wiltshire (WIL)	SW	4,389
West Stoughton	Somerset (SOM)	SW	5,422
Westbury	Wiltshire (WIL)	SW	21,883
Yeovil	Somerset (SOM)	SW	4,568
Faversham	Kent (KEN)	SE	67,298
London North	London (LND)	SE	45,212
Lydd	Kent (KEN)	SE	14,051
Pinner	Middlesex (MDX)	SE	30,595
Poplar London Port	London (LND)	SE	29,644
Sheerness	Kent (KEN)	SE	10,453
Sittingbourne	Kent (KEN)	SE	11,874
Tenterden	Kent (KEN)	SE	42,177
Whitstable	Kent (KEN)	SE	9,339
–	Leicestershire (LEI)	Mid	2,341
–	Nottinghamshire (NTT)	Mid	15,651
Lambley	Nottinghamshire (NTT)	Mid	11,773
Nottingham	Nottinghamshire (NTT)	Mid	110,936
Southwell	Nottinghamshire (NTT)	Mid	11,834

location	county	dialect area	running words
Ambleside	Westmorland (WES)	N	21,304
Barrow	Lancashire (LAN)	N	20,559
Birtley	Durham (DUR)	N	9,085
Choppington	Northumberland (NBL)	N	7,859
Crompton	Lancashire (LAN)	N	13,656
Fenwick Steads	Northumberland (NBL)	N	11,558
Guisborough	Yorkshire (YKS)	N	1,856
Hartlepool	Durham (DUR)	N	17,422
Hebden Bridge	Yorkshire (YKS)	N	17,180
Hinderwell	Yorkshire (YKS)	N	7,197
Loftus	Yorkshire (YKS)	N	5,637
Middlesbrough	Yorkshire (YKS)	N	9,962
Prescott	Lancashire (LAN)	N	27,709
Preston	Lancashire (LAN)	N	77,921
Redcar	Yorkshire (YKS)	N	9,690
Swarland	Northumberland (NBL)	N	8,360
Edinburgh	Midlothian (MLN)	ScL	21,068
Falkirk	West Lothian (WLN)	ScL	16,347
Tranent	East Lothian (ELN)	ScL	28,985

Table A.1.: Locations in FRED-S (unknown values are indicated by '–')

Appendix B. FRED-S texts

text	location	county	area	speaker(s)	running words
CON_001	Churchtown	Cornwall (CON)	SW	CAVA_DB	2,975
CON_002	Carnelloe	Cornwall (CON)	SW	CAVA_HJ	3,999
CON_003	St. Ives	Cornwall (CON)	SW	CAVA_PV	5,542
CON_004	St. Ives	Cornwall (CON)	SW	CAVA_TC	1,785
CON_005	Gurnards Head	Cornwall (CON)	SW	CAVA_WJB	7,483
CON_006	Pendeen	Cornwall (CON)	SW	CAVA_WW	4,751
DEV_001	Blackawton	Devon (DEV)	SW	TCA_RA	10,195
DEV_002	Galmpton	Devon (DEV)	SW	TCA_GB	5,244
DEV_003	Brixham	Devon (DEV)	SW	TCA_EC, TCA_C	6,534
DEV_004	Totnes	Devon (DEV)	SW	TCA_SC, DEV4ano	4,735
DEV_005	Totnes	Devon (DEV)	SW	TCA_FP	10,941
DEV_007	Totnes	Devon (DEV)	SW	TCA_EA	8,848
DEV_008	Totnes	Devon (DEV)	SW	TCA_FK, TCA_K	12,410
DEV_009	Totnes	Devon (DEV)	SW	TCA_WH	11,318
DEV_010	Buckfast	Devon (DEV)	SW	TCA_WC	9,645
OXF_001	Leafield	Oxfordshire (OXF)	SW	FP, BF	6,340
OXF_002	Filkins	Oxfordshire (OXF)	SW	GW, XW	7,461
SOM_001	West Stoughton	Somerset (SOM)	SW	SRLM_HR	5,422
SOM_002	–	Somerset (SOM)	SW	SRLM_PG	3,525
SOM_004	North Burrowbridge	Somerset (SOM)	SW	SRLM_CK	4,966
SOM_005	Street	Somerset (SOM)	SW	SRLM_RF	7,482
SOM_006	Sunnyside	Somerset (SOM)	SW	SRLM_AW	4,237
SOM_008	Petherton	Somerset (SOM)	SW	SRLM_CH, SRLM_ES	5,566
SOM_009	Horton	Somerset (SOM)	SW	SRLM_CA	10,503
SOM_012	Buckleigh	Somerset (SOM)	SW	SRLM_SH	6,207
SOM_013	Yeovil	Somerset (SOM)	SW	SRLM_GG	4,568
SOM_014	Barton St. David	Somerset (SOM)	SW	SRLM_RM	10,448
SOM_034	North Petherton	Somerset (SOM)	SW	SRLM_EW	2,938
SOM_035	Glastonbury	Somerset (SOM)	SW	SRLM_SC	3,459
WIL_001	Trowbridge	Wiltshire (WIL)	SW	TrbrGR	12,010
WIL_002	Trowbridge	Wiltshire (WIL)	SW	TrbrJH	5,951
WIL_003	Westbury	Wiltshire (WIL)	SW	TrbrJS	9,116
WIL_004	Trowbridge	Wiltshire (WIL)	SW	TrbrDP	7,784
WIL_005	Trowbridge	Wiltshire (WIL)	SW	TrbrCS, TrbrViv, TrbrD1, TrbrD2, TrbrCN, TrbrLW, TrbrDH	10,098
WIL_006	Trowbridge	Wiltshire (WIL)	SW	TrbrGSR	7,629
WIL_007	Trowbridge	Wiltshire (WIL)	SW	TrbrBJ, TrbrKJ	5,592

text	location	county	area	speaker(s)	running words
WIL_008	Westbury	Wiltshire (WIL)	SW	TrbrRCC	12,767
WIL_010	Urchfont	Wiltshire (WIL)	SW	WflsWGP	4,389
KEN_001	Whitstable	Kent (KEN)	SE	JC	9,339
KEN_002	Faversham	Kent (KEN)	SE	FK	13,612
KEN_003	Faversham	Kent (KEN)	SE	HM	41,721
KEN_004	Tenterden	Kent (KEN)	SE	KentPB	13,665
KEN_005	Sheerness	Kent (KEN)	SE	KentTAD	10,453
KEN_006	Sittingbourne	Kent (KEN)	SE	WS	11,874
KEN_007	Lydd	Kent (KEN)	SE	KentEDG, WG	14,051
KEN_009	Faversham	Kent (KEN)	SE	KentAW1	11,965
KEN_010	Tenterden	Kent (KEN)	SE	KentPB	13,945
KEN_011	Tenterden	Kent (KEN)	SE	KentPB	14,567
LND_002	London North	London (LND)	SE	EB, BHu	15,503
LND_003	London North	London (LND)	SE	FM	13,853
LND_004	London North	London (LND)	SE	JG, DG, DL, CG	15,856
LND_006	Poplar London Port	London (LND)	SE	GA	14,037
LND_007	Poplar London Port	London (LND)	SE	GA	15,607
MDX_001	Pinner	Middlesex (MDX)	SE	MdxCG	13,657
MDX_002	Pinner	Middlesex (MDX)	SE	CP	16,938
LEI_001	–	Leicestershire (LEI)	Mid	LeiA, LeiB	2,341
NTT_001	Nottingham	Nottinghamshire (NTT)	Mid	NotA103	5,633
NTT_002	Nottingham	Nottinghamshire (NTT)	Mid	NotA25	11,181
NTT_003	Nottingham	Nottinghamshire (NTT)	Mid	NotA30	9,171
NTT_004	Nottingham	Nottinghamshire (NTT)	Mid	NotA40	4,934
NTT_005	Southwell	Nottinghamshire (NTT)	Mid	NotA52	11,834
NTT_006	Nottingham	Nottinghamshire (NTT)	Mid	NotA58	14,208
NTT_007	Nottingham	Nottinghamshire (NTT)	Mid	NotA65	6,962
NTT_008	Nottingham	Nottinghamshire (NTT)	Mid	NotA71	5,477
NTT_009	Nottingham	Nottinghamshire (NTT)	Mid	NotA80	8,568
NTT_011	Nottingham	Nottinghamshire (NTT)	Mid	NotA99	9,046
NTT_012	Nottingham	Nottinghamshire (NTT)	Mid	ALI	8,250
NTT_013	Nottingham	Nottinghamshire (NTT)	Mid	NotA8	13,122
NTT_014	–	Nottinghamshire (NTT)	Mid	A15	15,651
NTT_015	Lambley	Nottinghamshire (NTT)	Mid	A109, A109w	11,773
NTT_016	Nottingham	Nottinghamshire (NTT)	Mid	NotA48	14,384
DUR_001	Birtley	Durham (DUR)	N	DurML	9,085
DUR_002	Hartlepool	Durham (DUR)	N	DurNB	7,880
DUR_003	Hartlepool	Durham (DUR)	N	DurEL	9,542
LAN_001	Barrow	Lancashire (LAN)	N	LanD1B	11,274
LAN_002	Barrow	Lancashire (LAN)	N	LAN002	9,285
LAN_003	Preston	Lancashire (LAN)	N	Lang1p	9,376
LAN_004	Preston	Lancashire (LAN)	N	LanW1P	8,876
LAN_005	Preston	Lancashire (LAN)	N	LanC5P	10,577
LAN_006	Preston	Lancashire (LAN)	N	LanD1P	9,904
LAN_007	Preston	Lancashire (LAN)	N	LanF1P	8,469
LAN_008	Preston	Lancashire (LAN)	N	LanT3P	10,110
LAN_009	Preston	Lancashire (LAN)	N	LanH1P	11,031
LAN_010	Preston	Lancashire (LAN)	N	H3L, MH3L	9,578
LAN_011	Crompton	Lancashire (LAN)	N	LanMJ	13,656
LAN_012	Prescott	Lancashire (LAN)	N	LanEG	27,709
NBL_003	Choppington	Northumberland (NBL)	N	NblMN	7,859
NBL_006	Fenwick Steads	Northumberland (NBL)	N	NblJB	7,235

text	location	county	area	speaker(s)	running words
NBL_007	Swarland	Northumberland (NBL)	N	NblAR, NblAn	8,360
NBL_008	Fenwick Steads	Northumberland (NBL)	N	NblGS	4,323
WES_003	Ambleside	Westmorland (WES)	N	WesAQ	4,539
WES_006	Ambleside	Westmorland (WES)	N	WesBL	3,051
WES_008	Ambleside	Westmorland (WES)	N	WesBR	10,525
WES_011	Ambleside	Westmorland (WES)	N	WesCE, WesCEa	3,189
YKS_001	Middlesbrough	Yorkshire (YKS)	N	YksWF	9,962
YKS_004	Guisborough	Yorkshire (YKS)	N	YksEH	1,856
YKS_006	Hinderwell	Yorkshire (YKS)	N	YksHS	7,197
YKS_008	Redcar	Yorkshire (YKS)	N	YksMW	9,690
YKS_009	Loftus	Yorkshire (YKS)	N	YksWG, YksMG	5,637
YKS_010	Hebden Bridge	Yorkshire (YKS)	N	Yks10	9,443
YKS_011	Hebden Bridge	Yorkshire (YKS)	N	YksBR, YksJS, YksMR, YksAnon	7,737
ELN_004	Tranent	East_Lothian (ELN)	ScL	ElnLS, ElnAB	2,237
ELN_005	Tranent	East_Lothian (ELN)	ScL	ElnAG, ElnML	1,139
ELN_006	Tranent	East_Lothian (ELN)	ScL	ElnEL	1,198
ELN_007	Tranent	East_Lothian (ELN)	ScL	ElnLS, ElnAB	2,069
ELN_008	Tranent	East_Lothian (ELN)	ScL	ElnAB, ElnMS	3,730
ELN_009	Tranent	East_Lothian (ELN)	ScL	ElnLS, ElnAB	7,393
ELN_010	Tranent	East_Lothian (ELN)	ScL	ElnIB	2,077
ELN_011	Tranent	East_Lothian (ELN)	ScL	ElnEL, ElnIB	2,343
ELN_012	Tranent	East_Lothian (ELN)	ScL	ElnAG, ElnML	2,800
ELN_013	Tranent	East_Lothian (ELN)	ScL	ElnSH, ElnTR	783
ELN_014	Tranent	East_Lothian (ELN)	ScL	ElnAS, ElnAC	3,216
MLN_004	Edinburgh	Midlothian (MLN)	ScL	MlnSM, MlnDM	2,105
MLN_006	Edinburgh	Midlothian (MLN)	ScL	MlnRH	6,176
MLN_007	Edinburgh	Midlothian (MLN)	ScL	MlnJH	12,787
WLN_001	Falkirk	West_Lothian (WLN)	ScL	WlnTW	1,782
WLN_002	Falkirk	West_Lothian (WLN)	ScL	WlnMO	1,952
WLN_003	Falkirk	West_Lothian (WLN)	ScL	WlnMD	3,657
WLN_004	Falkirk	West_Lothian (WLN)	ScL	WlnMD	5,065
WLN_005	Falkirk	West_Lothian (WLN)	ScL	WlnES	938
WLN_006	Falkirk	West_Lothian (WLN)	ScL	WlnES	2,953

Table B.1.: Texts (interviews) in FRED-S (unknown values are indicated by '–')

Appendix C. FRED-S speakers

speaker	text(s)	location	county	area	age	date of birth	record. decade	sex	running words
BF	OXF_001	Leafield	OXF	SW	73	1899	189	m	2,891
CAVA_DB	CON_001	Churchtown	CON	SW	–	1895	189	m	2,974
CAVA_HJ	CON_002	Carnelloe	CON	SW	–	–	–	f	3,998
CAVA_PV	CON_003	St. Ives	CON	SW	86	1892	189	m	5,541
CAVA_TC	CON_004	St. Ives	CON	SW	–	–	–	m	1,784
CAVA_WJB	CON_005	Gurnards Head	CON	SW	74	1904	190	m	7,482
CAVA_WW	CON_006	Pendeen	CON	SW	–	–	190	m	4,750
DEV4ano	DEV_004	Totnes	DEV	SW	–	–	–	–	51
FP	OXF_001	Leafield	OXF	SW	81	1891	189	m	3,447
GW	OXF_002	Filkins	OXF	SW	102	1877	187	m	6,899
SRLM_AW	SOM_006	Sunnyside	SOM	SW	89	1894	189	m	4,236
SRLM_CA	SOM_009	Horton	SOM	SW	75	1916	191	m	10,502
SRLM_CH	SOM_008	Petherton	SOM	SW	86	1904	190	m	5,490
SRLM_CK	SOM_004	North Burrowbridge	SOM	SW	71	1916	191	m	4,965
SRLM_ES	SOM_008	Petherton	SOM	SW	–	–	–	–	74
SRLM_EW	SOM_034	North Petherton	SOM	SW	83	1910	191	f	2,937
SRLM_GG	SOM_013	Yeovil	SOM	SW	78	1918	191	m	4,567
SRLM_HR	SOM_001	West Stoughton	SOM	SW	–	–	–	m	5,421
SRLM_PG	SOM_002	–	SOM	SW	–	–	–	m	3,524
SRLM_RF	SOM_005	Street	SOM	SW	76	1905	190	m	7,481
SRLM_RM	SOM_014	Barton St. David	SOM	SW	78	1902	190	m	10,447
SRLM_SC	SOM_035	GlaStonbury	SOM	SW	80	1914	191	m	3,458
SRLM_SH	SOM_012	Buckleigh	SOM	SW	68	1927	192	m	6,206
TCA_C	DEV_003	Brixham	DEV	SW	–	–	–	m	1,679
TCA_EA	DEV_007	Totnes	DEV	SW	93	1892	189	f	8,847
TCA_EC	DEV_003	Brixham	DEV	SW	–	–	–	f	4,853
TCA_FK	DEV_008	Totnes	DEV	SW	–	1910	191	m	8,296
TCA_FP	DEV_005	Totnes	DEV	SW	82	1902	190	m	10,940
TCA_GB	DEV_002	Galmpton	DEV	SW	–	1900	190	m	5,243
TCA_K	DEV_008	Totnes	DEV	SW	–	–	–	f	4,112
TCA_RA	DEV_001	Blackawton	DEV	SW	76	1909	190	m	10,194
TCA_SC	DEV_004	Totnes	DEV	SW	–	–	–	m	4,682
TCA_WC	DEV_010	BuckfaSt	DEV	SW	–	1913	191	f	9,644
TCA_WH	DEV_009	Totnes	DEV	SW	79	1906	190	m	11,317
TrbrBJ	WIL_007	Trowbridge	WIL	SW	–	–	–	m	3,906
TrbrCN	WIL_005	Trowbridge	WIL	SW	–	–	–	–	4
TrbrCS	WIL_005	Trowbridge	WIL	SW	91	1901	190	m	7,922

speaker	text(s)	location	county	area	age	date of birth	record. decade	sex	running words
TrbrD1	WIL_005	Trowbridge	WIL	SW	–	–	–	f	1,897
TrbrD2	WIL_005	Trowbridge	WIL	SW	–	–	–	f	108
TrbrDH	WIL_005	Trowbridge	WIL	SW	–	–	–	–	10
TrbrDP	WIL_004	Trowbridge	WIL	SW	80	1912	191	m	7,783
TrbrGR	WIL_001	Trowbridge	WIL	SW	85	1907	190	m	12,009
TrbrGSR	WIL_006	Trowbridge	WIL	SW	66	1926	192	m	7,628
TrbrJH	WIL_002	Trowbridge	WIL	SW	58	1934	193	f	5,950
TrbrJS	WIL_003	WeStbury	WIL	SW	–	–	–	m	9,115
TrbrKJ	WIL_007	Trowbridge	WIL	SW	52	1940	194	f	1,684
TrbrLW	WIL_005	Trowbridge	WIL	SW	–	–	–	f	56
TrbrRCC	WIL_008	WeStbury	WIL	SW	71	1922	192	m	12,766
TrbrViv	WIL_005	Trowbridge	WIL	SW	–	–	–	f	94
WflsWGP	WIL_010	Urchfont	WIL	SW	–	–	189	m	4,388
XW	OXF_002	Filkins	OXF	SW	–	–	–	f	560
BHu	LND_002	London North	LND	SE	–	–	–	m	421
CG	LND_004	London North	LND	SE	–	–	–	f	451
CP	MDX_002	Pinner	MDX	SE	70	1905	190	m	16,937
DG	LND_004	London North	LND	SE	–	–	–	f	37
DL	LND_004	London North	LND	SE	–	–	–	f	353
EB	LND_002	London North	LND	SE	56	1926	192	f	15,080
FK	KEN_002	Faversham	KEN	SE	88	1887	188	m	13,611
FM	LND_003	London North	LND	SE	67	1914	191	f	13,852
GA	LND_006, LND_007	Poplar London Port	LND	SE	61	1924	192	m	29,642
HM	KEN_003	Faversham	KEN	SE	85	1890	189	m	41,720
JC	KEN_001	WhitStable	KEN	SE	89	1886	188	m	9,338
JG	LND_004	London North	LND	SE	66	1916	191	m	15,011
KentAW1	KEN_009	Faversham	KEN	SE	81	1894	189	m	11,964
KentEDG	KEN_007	Lydd	KEN	SE	81	1895	189	m	13,580
KentPB	KEN_004, KEN_010, KEN_011	Tenterden	KEN	SE	87	1889	188	m	42,174
KentTAD	KEN_005	Sheerness	KEN	SE	86	1890	189	m	10,452
MdxCG	MDX_001	Pinner	MDX	SE	79	1897	189	m	13,656
WG	KEN_007	Lydd	KEN	SE	–	–	–	f	469
WS	KEN_006	Sittingbourne	KEN	SE	84	1892	189	m	11,873
A109	NTT_015	Lambley	NTT	Mid	–	1912	191	m	10,394
A109w	NTT_015	Lambley	NTT	Mid	–	1912	191	f	1,377
A15	NTT_014	–	NTT	Mid	71	1911	191	m	15,650
ALI	NTT_012	Nottingham	NTT	Mid	78	1906	190	f	8,249
LeiA	LEI_001	–	LEI	Mid	–	–	–	m	989
LeiB	LEI_001	–	LEI	Mid	–	–	–	m	1,350
NotA103	NTT_001	Nottingham	NTT	Mid	85	1899	189	m	5,632
NotA25	NTT_002	Nottingham	NTT	Mid	85	1897	189	m	11,180
NotA30	NTT_003	Nottingham	NTT	Mid	81	1902	190	m	9,170
NotA40	NTT_004	Nottingham	NTT	Mid	66	1917	191	m	4,933
NotA48	NTT_016	Nottingham	NTT	Mid	–	1884	188	m	14,383
NotA52	NTT_005	Southwell	NTT	Mid	85	1898	189	m	11,833
NotA58	NTT_006	Nottingham	NTT	Mid	81	1902	190	f	14,207
NotA65	NTT_007	Nottingham	NTT	Mid	86	1897	189	f	6,961
NotA71	NTT_008	Nottingham	NTT	Mid	81	1902	190	m	5,476

speaker	text(s)	location	county	area	age	date of birth	record. decade	sex	running words
NotA8	NTT_013	Nottingham	NTT	Mid	–	1908	190	m	13,121
NotA80	NTT_009	Nottingham	NTT	Mid	78	1906	190	f	8,567
NotA99	NTT_011	Nottingham	NTT	Mid	91	1893	189	m	9,045
DurEL	DUR_003	Hartlepool	DUR	N	–	1897	189	f	9,541
DurML	DUR_001	Birtley	DUR	N	–	–	191	m	9,084
DurNB	DUR_002	Hartlepool	DUR	N	77	1908	190	m	7,879
H3L	LAN_010	PreSton	LAN	N	71	1904	190	m	9,552
LAN002	LAN_002	Barrow	LAN	N	83	1892	189	f	9,284
LanC5P	LAN_005	PreSton	LAN	N	61	1919	191	f	10,576
LanD1B	LAN_001	Barrow	LAN	N	76	1899	189	f	11,273
LanD1P	LAN_006	PreSton	LAN	N	71	1908	190	f	9,903
LanEG	LAN_012	Prescott	LAN	N	71	1906	190	m	27,708
LanF1P	LAN_007	PreSton	LAN	N	73	1906	190	m	8,468
Lang1p	LAN_003	PreSton	LAN	N	76	1903	190	m	9,375
LanH1P	LAN_009	PreSton	LAN	N	67	1911	191	f	11,030
LanMJ	LAN_011	Crompton	LAN	N	47	1930	193	f	13,655
LanT3P	LAN_008	PreSton	LAN	N	93	1886	188	m	10,109
LanW1P	LAN_004	PreSton	LAN	N	80	1899	189	f	8,875
MH3L	LAN_010	PreSton	LAN	N	–	–	–	f	24
NblAn	NBL_007	Swarland	NBL	N	–	–	–	f	35
NblAR	NBL_007	Swarland	NBL	N	83	1891	189	m	8,323
NblGS	NBL_008	Fenwick Steads	NBL	N	–	–	189	f	3,176
NblJB	NBL_006	Fenwick Steads	NBL	N	–	–	189	f	8,379
NblMN	NBL_003	Choppington	NBL	N	83	1890	189	f	7,858
WesAQ	WES_003	Ambleside	WES	N	84	1898	189	m	4,538
WesBL	WES_006	Ambleside	WES	N	84	1901	190	f	3,050
WesBR	WES_008	Ambleside	WES	N	78	1908	190	m	10,524
WesCE	WES_011	Ambleside	WES	N	85	1902	190	m	1,925
WesCEa	WES_011	Ambleside	WES	N	–	–	192	m	1,262
Yks10	YKS_010	Hebden Bridge	YKS	N	85	1899	189	m	9,442
YksAnon	YKS_011	Hebden Bridge	YKS	N	–	–	–	–	259
YksBR	YKS_011	Hebden Bridge	YKS	N	–	–	–	m	3,588
YksEH	YKS_004	Guisborough	YKS	N	–	–	–	f	1,855
YksHS	YKS_006	Hinderwell	YKS	N	–	1910	191	m	7,196
YksJS	YKS_011	Hebden Bridge	YKS	N	–	–	–	m	1,381
YksMG	YKS_009	Loftus	YKS	N	–	–	–	f	362
YksMR	YKS_011	Hebden Bridge	YKS	N	–	1912	191	f	2,505
YksMW	YKS_008	Redcar	YKS	N	–	1900	190	f	9,689
YksWF	YKS_001	Middlesbrough	YKS	N	–	1910	191	m	9,961
YksWG	YKS_009	Loftus	YKS	N	80	–	190	m	5,273
ElnAB	ELN_004, ELN_007, ELN_008, ELN_009	Tranent	ELN	ScL	–	–	–	f	7,549
ElnAC	ELN_014	Tranent	ELN	ScL	–	–	–	m	2,228
ElnAG	ELN_005, ELN_012	Tranent	ELN	ScL	–	–	–	m	2,212
ElnAS	ELN_014	Tranent	ELN	ScL	–	–	–	f	986
ElnEL	ELN_006, ELN_011	Tranent	ELN	ScL	–	–	–	f	2,922

speaker	text(s)	location	county	area	age	date of birth	record. decade	sex	running words
ElnIB	ELN_010, ELN_011	Tranent	ELN	ScL	17	–	–	m	2,692
ElnLS	ELN_004, ELN_007, ELN_009	Tranent	ELN	ScL	18	–	–	f	5,546
ElnML	ELN_005, ELN_012	Tranent	ELN	ScL	18	–	–	f	1,723
ElnMS	ELN_008	Tranent	ELN	ScL	18	–	–	f	2,326
ElnSH	ELN_013	Tranent	ELN	ScL	–	–	–	f	443
ElnTR	ELN_013	Tranent	ELN	ScL	–	–	–	m	338
MlnDM	MLN_004	Edinburgh	MLN	ScL	–	–	–	m	1,716
MlnJH	MLN_007	Edinburgh	MLN	ScL	–	1919	191	m	12,786
MlnRH	MLN_006	Edinburgh	MLN	ScL	–	–	–	m	6,175
MlnSM	MLN_004	Edinburgh	MLN	ScL	–	–	–	f	387
WlnES	WLN_005, WLN_006	Falkirk	WLN	ScL	57	1931	193	m	3,889
WlnMD	WLN_003, WLN_004	Falkirk	WLN	ScL	83	1908	190	f	8,720
WlnMO	WLN_002	Falkirk	WLN	ScL	–	–	–	f	1,951
WlnTW	WLN_001	Falkirk	WLN	ScL	60	1927	192	m	1,781

Table C.1.: Speakers in FRED-S (unknown values are indicated by '–')

Bibliography

ANDERWALD, LIESELOTTE. t.a.. *The Morphology of English Dialects: Verb-Formation in Non-standard English.* Cambridge: Cambridge University Press.

ANDERWALD, LIESELOTTE, & WAGNER, SUSANNE. 2007. FRED – The Freiburg English Dialect corpus. *In:* BEAL, JOAN C., CORRIGAN, KAREN P., & MOISL, H. (eds), *Creating and Digitizing Language Corpora,* Vol. 1: Synchronic Databases. London: Macmillan.

CHAMBERS, JACK, & TRUDGILL, PETER. 1998. *Dialectology*. 2nd edn. Cambridge: Cambridge University Press.

HERNÁNDEZ, NURIA. 2006. *User's Guide to* FRED. Freiburg: English Department, University of Freiburg (available online at `http://www.freidok.uni-freiburg.de/volltexte/2489`).

HERRMANN, TANJA. 2003. *Relative clauses in dialects of English. A typological approach*. PhD-Thesis, University of Freiburg (available online at `http://www.freidok.uni-freiburg.de/volltexte/830/`).

HERRMANN, TANJA. 2005. Relative clauses in English dialects of the British Isles. *Pages 21–124 of:* KORTMANN, BERND, HERRMANN, TANJA, PIETSCH, LUKAS, & WAGNER, SUSANNE (eds), *A Comparative Grammar of British English Dialects: Agreement, Gender, Relative Clauses*. Berlin/New York: Mouton de Gruyter.

KORTMANN, BERND (ed). 2004. *Dialectology Meets Typology: Dialect Grammar from a Cross-Linguistic Perspective*. Berlin/New York: Mouton de Gruyter.

PIETSCH, LUKAS. 2005a. "Some do and some doesn't": Verbal concord variation in the north of the British Isles. *Pages 125–210 of:* KORTMANN, BERND, HERRMANN, TANJA, PIETSCH, LUKAS, & WAGNER, SUSANNE (eds), *A Comparative Grammar of British English Dialects: Agreement, Gender, Relative Clauses*. Berlin/New York: Mouton de Gruyter.

PIETSCH, LUKAS. 2005b. *Variable Grammars: Verbal Agreement in Northern Dialects of English*. Tübingen: Niemeyer.

SZMRECSANYI, BENEDIKT. 2005. Language users as creatures of habit: a corpus-linguistic analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory*, **1**(1), 113–149.

SZMRECSANYI, BENEDIKT. 2006. *Morphosyntactic Persistence in Spoken English: A Corpus Study at the Intersection of Variationist Sociolinguistics, Psycholinguistics, and Discourse Analysis*. Berlin/New York: Mouton de Gruyter.

SZMRECSANYI, BENEDIKT, & HINRICHS, LARS. t.a.. *Probabilistic determinants of genitive variation in spoken and written English: a multivariate comparison across time, space, and genres*.

TRUDGILL, PETER. 1999. *The Dialects of England.* 2nd edn. Oxford: Blackwell.

WAGNER, SUSANNE. 2004a. *Gender in English Pronouns*. PhD-Thesis, University of Freiburg (available online at `http://freidok.ub.uni-freiburg.de/volltexte/1412/`).

WAGNER, SUSANNE. 2004b. 'Gendered' pronouns in English dialects - a typological perspective. *In:* KORTMANN, BERND (ed), *Dialectology Meets Typology*. Berlin/New York: Mouton de Gruyter.

WAGNER, SUSANNE. 2005. Gender in English pronouns: Southwest England. *Pages 211–367 of:* KORTMANN, BERND, HERRMANN, TANJA, PIETSCH, LUKAS, & WAGNER, SUSANNE (eds), *A Comparative Grammar of British English Dialects: Agreement, Gender, Relative Clauses*. Berlin/New York: Mouton de Gruyter.