Research Unit for Variation, Contacts and Change in English (VARIENG)

# Manual for the Cambridgeshire sampler

Helsinki Archive of Regional English Speech

Simo Ahava

# Contents

# 1 ACKNOWLEDGEMENTS

The Cambridgeshire sampler, HARES and the audio recordings themselves have all been compiled and collected by a vast group of people. Thus the entire dialect project should be considered a collaborative effort between students and scholars, laypersons and experts, informants and fieldworkers, and many others who have provided help and support throughout the over 40-year-long project.

These acknowledgements list some of the key figures who have helped sculpt the Cambridgeshire sampler from the original hand-written transcriptions by the fieldworker into a fully functional multimedia corpus.

The Research Unit for Variation, Contacts and Change in English (VARIENG) has been the host of the project and has also provided most of the funding. The project has also benefited from the scholarly expertise of Prof. Terttu Nevalainen, Prof. Irma Taavitsainen and Dr. Anneli Meurman-Solin, who have each provided invaluable assistance to the project and to some of the preliminary research papers, which have, in turn, helped create a foundation for most of the decisions that were made during corpus compilation.

Another important source of funding was the City Centre Campus Online Services, which made it possible for the HARES team to hire two research assistants for the years 2008 and 2009.

We also thankfully acknowledge the hard work put in by the previous project coordinators: Dr. Kirsti Peitsara and Prof. Ossi Ihalainen. Even though they worked with the project before HARES was on the drawing board, their contributions to preliminary transcription work and to research and their efforts to keep the dialect project active are some of the things that the Cambridgeshire sampler owes its existence to.

The project benefited from the assistance of Ms. Alice Beal, who helped the team finish transcription work far earlier than was thought possible.

Research assistant Mr. Simo Ahava has been with HARES since its beginning. He oversaw the digitisation of the audio, post-processed the audio files, created the concept of HARES, worked on the transcriptions, did all of the XML work, converted the sampler into two corpora (the XML and the TXT corpora), helped promote the corpus in symposia and conferences and wrote this manual.

Research assistant Mr. Joseph McVeigh was with the project from early 2009 to mid-2010. His main contribution was to the transcription work, and he actually did most of it. His native language proficiency helped decipher a lot of the ‘brogue’ that the Finns working on

the project just could not grasp. He was also present at ICAME 31 (with Simo Ahava) presenting the sampler for the first time to the academic public. He now resides in the United States, where, to his great relief, he no longer has to transcribe anything.

Dr. Anna-Liisa Vasko is the original fieldworker for the Cambridgeshire interviews and the coordinator for the HARES project. She has persistently helped gather funding for the project, and she did most of the preliminary transcription work. Her main focus over the last few years has been to finish work on the enormous *Cambridgeshire Dialect Grammar*, which was published in May 2010. The *Grammar* is a companion to the Cambridgeshire sampler (or vice versa) and should help flesh out most of the interesting dialect features that the informants so often use.

## 2 INTRODUCTION

The Helsinki Archive of Regional English Speech (HARES) is a collection of audio-recorded interviews that were gathered in England in the 1970s and 1980s. The fieldworkers were Finnish graduate and post-graduate students from the University of Helsinki, who shared a common interest in the study of dialect syntax. The informants were elderly persons who had lived in the region all their lives and who had left school at an early age. HARES combines the digital audio files with orthographic transcriptions and XML-annotated metadata.

Since its conception in 2008, the primary goal of HARES has been to harness the valuable and unique cultural content within the recorded interviews for use in academic research and for the general public as well. This goal, coupled with the main methodological concern of keeping the audio as primary data, has defined all the efforts in transcribing and annotating the audio data, because keeping in mind all the different applications of the archive during the compilation stages requires that the approach be as transparent and non-exclusive as possible.

Due to the enormity of the project, HARES compilation has evolved in stages, with the first product being the Cambridgeshire sampler (HARES-CAM). This sampler contains approximately 18 hours of interview data from the Cambridgeshire county. The fieldworker, Anna-Liisa Vasko, recorded the interviews in the years 1974-1977, and the sampler contains 20 interviews in 15 villages. The sampler was compiled by Simo Ahava (research assistant), Joseph McVeigh (research assistant) and Anna-Liisa Vasko (project co-ordinator), with casual assistance provided by Alice Beal (summer 2009). The project was funded by the Research Unit for Variation, Contacts and Change in English (VARIENG) and the City Centre Campus Online Services, both at the University of Helsinki.

HARES contains spoken testimonies from a diverse group of people from diverse backgrounds, who all comment enthusiastically on a wide variety of topics, ranging from farming to war and from neighbourhood gossip to local history. The Cambridgeshire sampler reflects this quality exceptionally well, as the informants are eager to share their expertise on a lot of different matters having to do with 'the good old days'. The resources found within can be used for (but are certainly not limited to) research on language, local history, anthropology, ethnography of communication and cultural discourse.

This manual is meant to be used as a reference tool for extra information about the recordings and the project itself. Any and all corrections, suggestions and additions should be directed to the HARES team.

## 3 OVERVIEW OF THE CAMBRIDGESHIRE SAMPLER

The Cambridgeshire sampler contains 20 interviews recorded in 15 villages. The interviewees are 'typical' HARES informants, in that they are elderly (youngest is in his 60s and the oldest in his late 90s), non-mobile (the informants have been born and bred in the area, with little or no time spent away), minimally educated (having left school at an early age) and rural (with professions such as horsekeeper, farmer and housewife). The sampler represents a 'modern' trend in dialectological surveys, because some of the interviews have women as primary informants[1].

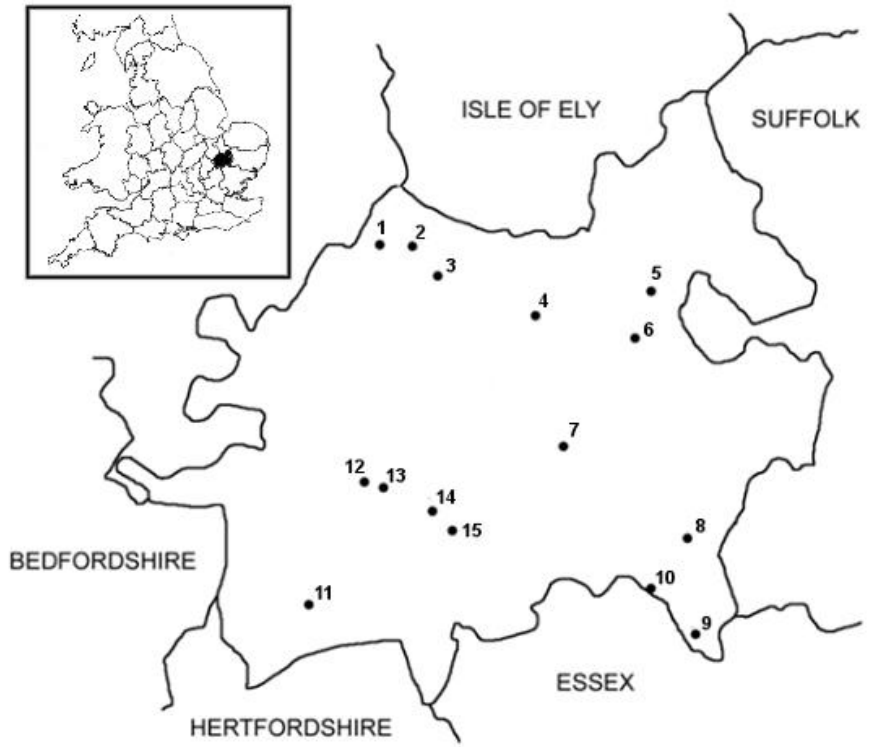
See MAP 1 below for a complete list of the villages included in the sampler. Note that the map and all subsequent references to Cambridgeshire county refer to pre-1974 borders, due to the fact that the project began before the Local Government Act of 1972, which effectively reorganised the county division around Cambridgeshire in 1974.

The sampler is primarily intended for browsing with XML-capable (eXtensive Markup Language) software. The transcriptions have been annotated with TEI (Text Encoding Initiative) P5 guidelines, and the metadata provide a large amount of useful information not accessible via plain text corpus browsers. However, the sampler is distributed in plain text format as well, in which case it can be accessed with the appropriate tools (e.g. WordSmith or AntConc).

[1] 'Primary' informants are speakers who were chosen as interview subjects and who take part actively in the interviews. 'Secondary' speakers would be, for example, friends, spouses or relatives also present in the interview situation, but with less dialogue.

The audio files should be listened to together with the transcriptions. This is because the best and most reliable results in terms of research and interpreting transcriber judgment are arrived at by listening to the audio while reading the transcription of the interview. Indeed, we have compiled the transcriptions with the mentality that they are a 'necessary evil' and serve as subtitles to the audio. We urge for the end user not to trust blindly the transcriptions but to make use of the audio whenever and wherever possible.

**MAP 1**. **Villages in the Cambridgeshire sampler**[2]




| | | |
|---|---|---|
| 1 Over | 6 Swaffham Prior | 11 Bassingbourn |
| 2 Willingham | 7 Fulbourn | 12 Little Eversden |
| 3 Rampton | 8 West Wickham | 13 Harlton |
| 4 Waterbeach | 9 Castle Camps | 14 Harston |
| 5 Burwell | 10 Bartlow | 15 Newton |

[2] This map and table are modified from the ones found in Ahava (2010: 114).

## 3.1 PROJECT HISTORY

HARES began as an effort to make use of the valuable dialect recordings collected in the 1970s and 1980s. The outline of the archive was that it should be considered a cultural repository of spoken testimonials by people who wanted to talk about some aspects of their everyday life. This means that compilation principles have been designed to highlight this transparency and not to steer the archive in any one direction (e.g. linguistic database).

In 2008, the original reel-to-reel tapes were digitised, and Anna-Liisa Vasko (Cambridgeshire fieldworker and project co-ordinator) and Simo Ahava (research assistant) began work on HARES. Having the digital audio meant that transcriptions were not considered as central in the archive anymore. All previous work on the dialect recordings had focused on transcriptions, which is understandable considering the difficulties with having to consult the original reel-to-reel tapes to verify the research. To make use of the audio files, it was necessary to adopt a transcription method that would include the audio. XML annotation was chosen to realise this goal because with XML it was possible to utilise existing and universal specifications for encoding spoken language data.

In May 2009, first glimpses of the new project were revealed in the form of a poster presentation by Ahava and Vasko at the International Computer Archive of Modern and Medieval English (ICAME 30) conference. After the conference, work began on a sampler of Cambridgeshire speech. A sampler was chosen as the first product because of the enormity of transcribing, annotating and publishing such a multi-faceted spoken language archive. An 18-hour sampler would also help refine the transcription protocol and XML schema so that future work on HARES would not be hindered by the most basic problems of e.g. devising principles for orthographic transcription.

In 2009, Joseph McVeigh joined the HARES team. His native speaker intuition was invaluable in transcribing and annotating the data. Also, in July 2009, Alice Beal joined the group for two months to help on the transcription work. Thanks to her help, the final audio files of the archive were transcribed in time for Christmas holidays.

At ICAME 30, the HARES team promised that the sampler would be completed by ICAME 31. This goal was partially met, as Ahava and McVeigh presented the finished corpus as a work-in-progress presentation at the conference. However, the sampler was not completed yet, as still a number of questions regarding e.g. anonymity and censoring data persisted.

## 3.2 DATA COLLECTION

The Cambridgeshire interviews were done with the informal interview method. This meant that the interviewer was optimally a silent participant in the conversation, allowing the

informant to steer the talk freely and to choose his or her preferred topic. The interviewer's main strategy was to ask leading questions and then let the informant answer at length in a manner that did not seem premeditated.

Anna-Liisa Vasko, the fieldworker, considered it paramount to create a relaxed atmosphere during the conversation. She would visit the informants a day or so before without the microphone, so that the informants would become accustomed to her presence and thus feel more at ease during the actual recorded conversation. During the interviews it was important to play to the informant's strengths and focus on questions that would highlight his or her expertise in any matter. Thus a farmer would be repeatedly asked to tell about old farming practices, and an extrovert person would be asked to tell about friends, feasts and other social events.

Because the fieldworkers were interested in dialect syntax, this informal interview method was considered better than the other fieldwork methods of the day, such as using a wordlist or a questionnaire. Allowing the speakers to talk without interruption resulted in longer stretches of speech – something that syntactic studies rely on. A downside was, of course, that the conversations were relatively unconstrained, and any rare syntactic feature that could have been uncovered with a simple elicitation strategy might not occur once during the entire interview. However, Vasko felt that revealing her research ambitions to the informant could have resulted in the informant speaking in a way that did not reflect his or her everyday speech. Vasko's decision to hide her research goals helped reduce the effect of interviewer intervention and the Observer's Paradox (how to observe the speech of people without them feeling like they are being observed).

The more one listens to the recordings the more one becomes ascertained that the speakers are, truly, relaxed and comfortable, recounting their experiences with enthusiasm that could not be feigned.

## 3.3 INFORMANT SELECTION

Vasko found her informants with the help of local assistants and via post offices, pubs and village shops. When available, the local assistants were most helpful because they usually knew the older people and could set up meetings and interviews for the fieldworkers. However, as soon as Vasko had established herself as a person interested in interviewing older 'experts' on local matters, she didn't need to make an effort in finding the informants – they practically lined up for her.

Vasko chose her informants on the basis of the prototypical dialect speaker: an elderly person with minimal mobility to or from the area, with minimal education and a profession

in farming or work on the land. Focusing on these aspects instead of choosing the informant bluntly on the basis of their 'dialect' speaking skills resulted in the colourful mixture of rural stories and rural speech that this sampler is abundant with.

Contrary to previous dialect projects, Vasko interviewed a number of women as primary informants. Two of these interviews are included in the sampler (cam04 and cam16), and they clearly reveal that any preconceptions of women being 'less dialectal' and thus unfit for dialectological surveys are misguided.

Vasko had help during her fieldwork. A local man, Mike Hopkins, proved invaluable to her work among the informants. He was interested in local history, family lineages and the language. He helped Vasko find informants for her survey and did much of the interviewing himself. He had a lot of knowledge on local matters and could thus get the conversation going without effort. Vasko was also assisted by MG[3], a non-local female who joined in a number of interviews as the second interviewer, but her role was significantly smaller then Hopkins'.

For the Cambridgeshire sampler, informant and locality selection criteria were dependent on Vasko's decisions during the writing of the *Cambridgeshire Dialect Grammar*. The sampler is intended to complete the written *Grammar*, as most of the grammatical examples within can be listened to via links to the interviews in the sampler. This meant that the informants chosen for the sampler had to be the same people who speak in the examples in the *Grammar*.

Villages were chosen so that enough geographical diversity would be represented in the corpus. Even though the county is not a large one, previous research has indicated that dialectal variants differ depending on, for example, what the nearest neighbouring county is. Only one speaker per village was the original plan, but as Willingham speakers populated the *Grammar* (probably because Vasko spent a great deal of time in Willingham and thus became familiar with the residents), we chose five speakers from the village.

## 3.4 ORTHOGRAPHIC TRANSCRIPTION PROTOCOL (OTP)

The OTP chosen for HARES leans towards Standard English. What this means in practice is that we have avoided the use of 'eye-dialect' (using orthography like *me an' you* to denote dropping the /d/-sound in running speech) because it just leads to confusion and places a significant burden on the corpus browser to identify all the different spellings for a single

[3] For the sake of anonymity, only the initials are provided for MG.

utterance. In HARES, we have minimised the number of alternate spellings in favour of the Standard English spelling (e.g. *anything* for *anythin', anythink, anythan'*).

Words that have no Standard English equivalent (we have mainly consulted the Oxford English Dictionary and English Dialect Dictionary) receive a new spelling devised by the HARES team. Such examples are *een't* (related to *isn't* and *ain't*), *war* (a positive past BE verb related to *was* and *were*), *cos* (truncated from either *(of) course* or *'cause*) and *hapenny* (shortened version of *halfpenny*). Whenever we devise a new spelling for a word that we hear, it is because its use is somehow special or different from the Standard English alternatives. In the case of *een't*, for example, it is phonetically clearly distinguishable from the diphthongised *ain't* and the Standard English *isn't*, and some speakers use it so much that it deserves a spelling that highlights its special quality.

We have retained vocalisations such as *mm*, *er, aha* and *oh*. We have transcribed these to the best of our efforts, but in some cases, for example when the interviewer says *mm* in the background, we have left them out in the transcriptions because they carry no obvious meaning nor do they interrupt the speaker's flow.

Stuttering and false starts are also preserved and marked with the **<seg/>** tag in XML (see below). These are the only cases where we use hyphens (e.g. *w- we*). We have left the hyphens out in compounds and other similar places because they are not features of spoken language. Similarly, all punctuation is removed from the transcriptions (no full stops, commas, question marks), and capitalisation exists only in proper nouns (no sentence-initial capitalisation).

## 3.5 COMPILATION

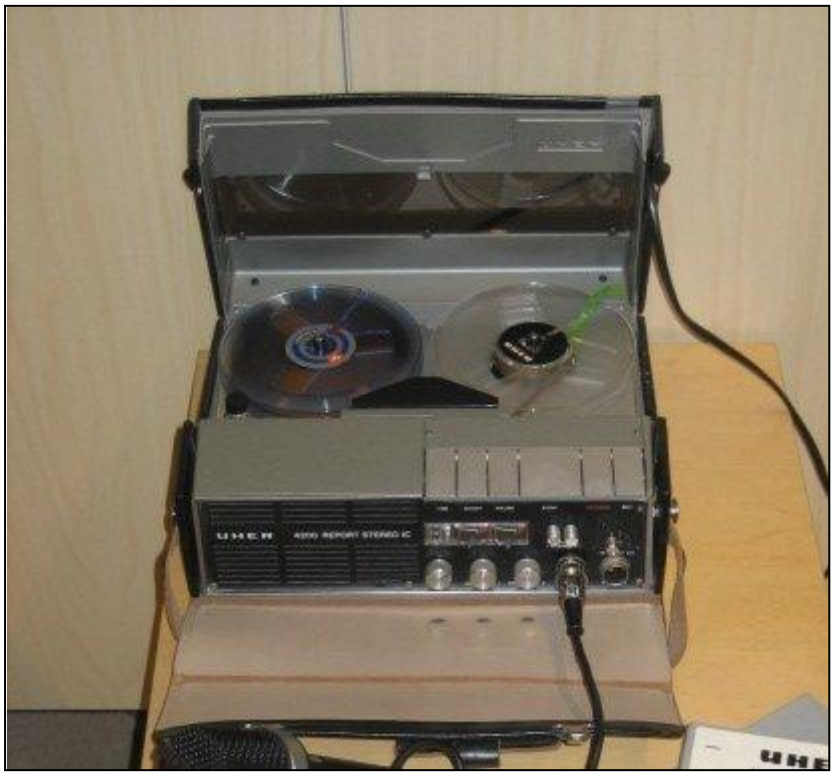
In this chapter, we will explain the compilation process of a single, finished HARES interview. A finished interview consists of the audio file (see TABLE 1), an XML file (see Chapter 4.3.1) and a plain text file (see Chapter 4.3.2). Because spoken language corpora with a focus on rural varieties of English are few and far between, some of the decisions we made in the early stages of corpus compilation could have been made differently. We refined our methods as we progressed in the compilation. These 'poor decisions' had ramifications especially during the transcription stage, as explained below.

### 3.5.1 Audio: archiving and digitisation

The audio was originally recorded on reel-to-reel tapes (see FIGURE 1). A typical tape in the HARES recordings could store up to 45 minutes of audio per track. Even though the tapes contain four tracks (two per side), usually just one track per side was used. The tapes were

labelled with information about the interview: who are the informants, what is the locality and what date the interview was recorded.

**FIGURE 1. A reel-to-reel tape and recorder**

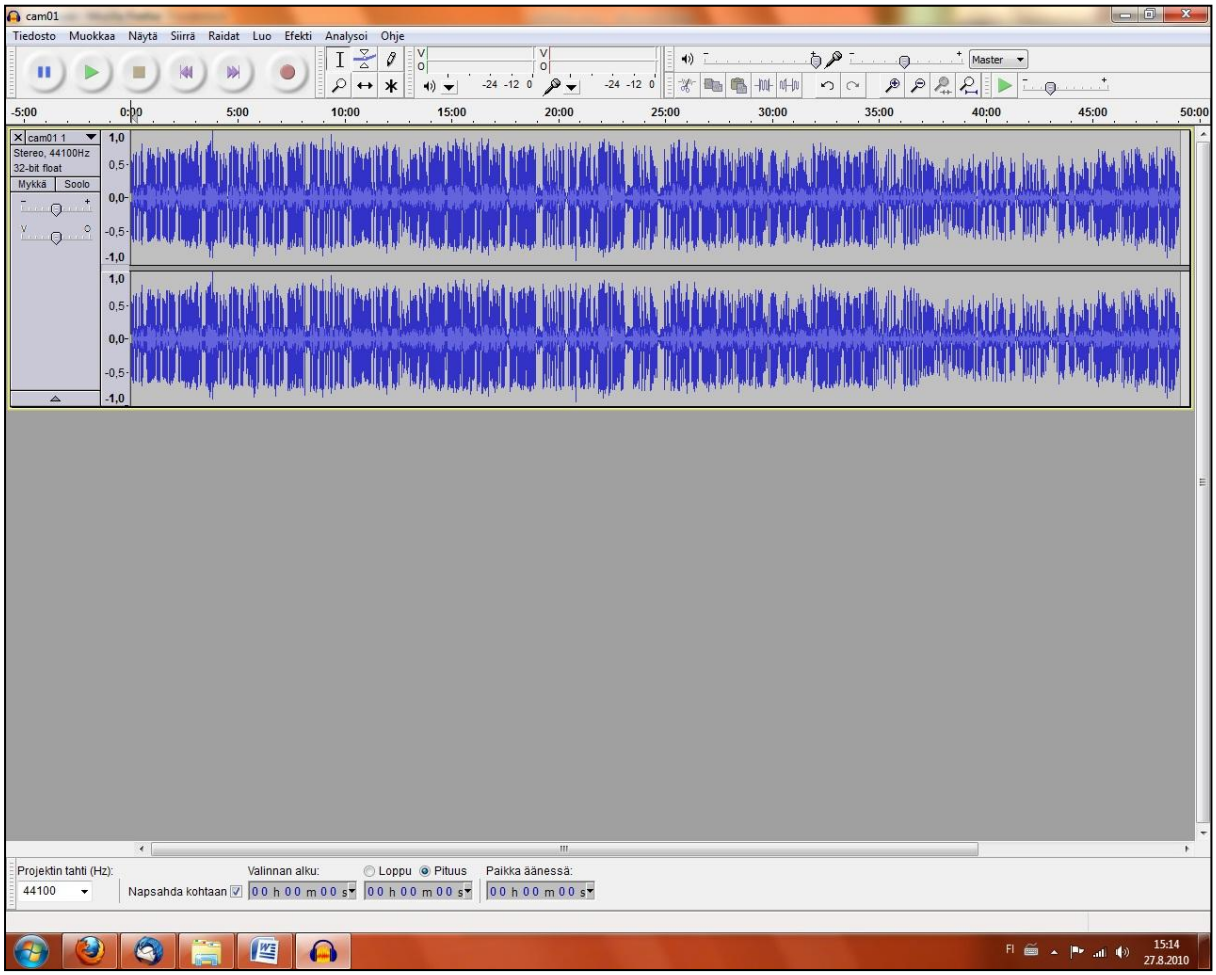


Prior to sending the tapes to be digitised, each tape was identified and archived so that we ended up with a list containing details of each track on each tape. This list was then sent along with the tapes to Diginord (http://www.diginord.fi/). The company made digital copies of the recordings as high-quality WAV files stored on DAT tapes. Diginord also provided us with MP3 copies on CD-ROM disks.

### 3.5.2 Audio: post-processing the files

We used Audacity to post-process the audio files (http://audacity.sourceforge.net/; see FIGURE 2). Because a single side of an original reel-to-reel tape could contain audio from as many as three interviews, the first order of action was to join together all the various parts of a single interview. This was a lengthy task because some of the original tapes were insufficiently labelled or archived, and a successful identification of the interview on an audio file was sometimes a combination of comparing the original transcriptions with what one heard, detective work and sheer good luck. Some of the audio files still contain interview data from separate dates: sometimes it was impossible to distinguish when one interview ends and when the next begins.

**FIGURE 2. Audacity**




After the interviews were combined into audio files, we converted the WAV files to MP3. Due to lack of foresight, we did not think to save WAV copies of these repaired files, as we thought that MP3 quality is more than good enough for any corpus work.

Next, we stripped the audio files of unnaturally long silences. We wanted to ensure that the audio played smoothly and fluently, without long gaps or background noise. To remove the latter, we used equalisation filters and noise removal tools. We also manually removed most of the unnatural 'peaks' in the audio. Not all the clicks and pops were erased, partly because we didn't have time to go over the audio with a fine-toothed comb and partly because using noise removal tools with a lower threshold would have resulted in degraded audio quality.

As briefly mentioned above, one problem with the audio processing was that we only saved the WAV files as compressed MP3 files. The MP3 files are in perfectly good quality to be listened to by casual corpus users, but if one wishes to do more demanding acoustic

analyses on the audio files, some of the files are simply not up to par. However, we still have the original, untouched WAV files, which we can distribute to those who want better quality audio. These files are raw, meaning that they have not been compiled into single interview files nor have they been cleaned up in any way. It might thus be difficult to locate the desired segment from the original WAV file. This is a problem we will likewise address once we are at a stage of HARES compilation that allows for such an increase in workload.

### 3.5.3 Audio: anonymisation

The interviews were recorded without written and signed consent forms. When the data were recorded in the 1970s and 1980s, there was no need to explicitly ask for consent because sharing the recordings to a larger public was not an option in that day. However, the interviewees were asked for permission to use the data for the fieldworkers' own research.

Because we lack explicitly stated consent for distributing the audio to a larger public, we gave anonymisation measures a lot of thought. However, in the end we decided not to anonymise the content. We only give the initials of each speaker in the headers of the interviews, so a minimal level of anonymity is achieved. Of course, if a relative or close friend of the interviewee would ever happen across the recordings, it would be a simple matter to identify the speaker regardless of how much data we provided on him or her.

We also decided against bleeping or muting any personal information from the data. Our first reason was that cutting out these utterances would discard a plethora of data useful for later analysis. Also, we found that using bleeps and muted passages is obtrusive and irritating to the listener. As the data are intended for use solely within academia, we do not expect any consequences for our decision to leave the data untouched. However, if and when we do decide to publish it to the general public too, the question of anonymity will need to be brought up and dealt with to the satisfaction of the compilers, the end users, authorities and all living relatives of the informants.

### 3.5.4 Transcription: turning speech to text

We benefited from the original hand-written and computerised transcriptions for most of the interviews in the Cambridgeshire sampler. However, each transcription we wrote was disputed, corrected, re-corrected and modified countless times as the entire HARES team was involved in reviewing them. The transcriptions still contain passages that we did not agree upon as well as segments that we have labelled unclear. These segments are most often due to either poor quality of the audio, overlapping speech or because someone is speaking in the far background. Attempts to decipher such unclear speech are futile, and it

is a safe bet to say that even if the original speaker would listen to the audio recording, he or she would find it equally difficult to understand what is being spoken.

For transcription work, we used Transcriber (http://trans.sf.net/). We chose it because it is freeware and because it produces XML output. It is relatively simple to use. First you indicate which audio file you wish to transcribe. Then you segment the audio by hitting the ENTER key at a place of your choosing. The segment is then attributed with a speaker ID and textual content.

After one of the transcribers finished work on a transcription, it was sent over to the others for reviewing. These second, third and fourth passes (and sometimes even more) ensured that each audio was listened to carefully and each person working on the audio could influence the end result. It is surprising how much of the original transcription changes during these reviews. This is not because the original transcriber is incompetent in his or her work but is rather a fitting tribute to the, at times, incredibly difficult task of deciphering rural speech that is more often than not uttered through a pipe, through false teeth or as a whisper during a segment where two or three people speak over each other[4].

During the last stages of reviewing the interview files, we encountered a problem with Transcriber software that proved a rather large nuisance. For some reason, the software introduces a lag the further you transcribe a single, large audio file. By the time the audio passed the 45 minute mark, the lag was approximately 10 seconds. What this meant in practice was that the tags we had used for anchoring the audio to the text (see Chapters 4.3.1.4 and 4.3.2.4) were off the mark by an increasing amount of time. So we had to go over each interview manually and move the anchors to their correct places in the transcription. We did not foresee this problem, as Transcriber had worked flawlessly throughout transcription work, and we simply did not think to double-check the anchor tags until at the very end. This unfortunate setback cost us over 20 hours of work and should serve as a cautionary note to the developers and users of Transcriber software.

### 3.5.5 Transcription: conversion to XML

At one point, we thought of using Transcriber's native XML output as the encoding schema for HARES. However, since we have strived to make HARES as transparent as possible, we decided to choose a set of specifications that have been more or less universally accepted as an up-and-coming standard for encoding corpora: TEI (Text Encoding Initiative; see http://www.tei-c.org/).

[4] Correction suggestions to the transcriptions can be directed to the HARES team (see Chapter 8).

Transcriber's XML output was structured so that conversion to TEI was relatively simple. We compared Transcriber's XML file with the TEI specifications and came up with a combination of modules and tags that best suit a corpus of spoken language data. After this, we progressed in the following steps:

1) We removed the header from the Transcriber XML file because the TEI specifications have very specific guidelines for header content
2) We compiled a set of regular expression searches in Microsoft Word to perform the conversion of the Transcriber tags to TEI
3) We coded these searches into Word macros
4) We ran these macros over all the Transcriber XML files
5) We manually corrected and appended the output of the macros

Step (1) was simple enough. We created a new document that contained the bare structure of a TEI header. Then we simply copy-pasted the Transcriber output to the beginning of the body in the new TEI XML file. We removed Transcriber's own header at this point. Because it differed significantly from the TEI specifications, we opted against any conversion method for the header.

The regular expression searches we created for step (2) were of the type: 'look for tag <x> and convert it to <y>'. In practice, this meant that we were able to retain all important content in the conversion, such as speaker IDs. However, when we compiled these searches into macros (3), we did not code any new information into the operation, that is to say we left a lot of work for manual correction.

After running the macros over the Transcriber files (4), we started manually correcting the output (5). This was, of course, the most laborious stage of work because it involved, among other things, the following:

- Attributing a unique ID to each utterance
- Attributing a unique ID to each **<seg>**
- Coding overlapping speech segments

Working with overlapping segments was especially time-consuming because we changed our protocol of transcription in the middle of the work. Initially, we had decided to transcribe every single overlapping utterance, but soon we noticed this was an enormous task and, to be truthful, quite unnecessary. Most of the overlapping segments had the interviewer saying things like "mm" and "oh" in the background. We decided to discard all the instances where the overlapping backchannel does not audibly influence the informant's flow of speech.

We are sensible of the fact that we could have decreased our workload by using XML stylesheet transformations (XSLT) or a more detailed macro scheme. However, since we were, in many ways, sailing through uncharted waters and learning by doing, it is understandable that not all of the decisions we made were the best for any given situation. All the same, this is one of the reasons that pioneering projects exist: they devise and refine protocols for future projects to come.

### 3.5.6 Final touches

After work on the audio and the transcriptions was finished, we moved onto a seemingly endless loop of rechecking and redoing. First of all, as mentioned above, we came across a serious glitch in the time anchors in the transcriptions, which was the result of lag in Transcriber software. We also sent the finished interviews to Anna-Liisa Vasko, who, as the fieldworker and collector of the Cambridgeshire data, had final say on what to retain and what to dismiss. She requested that we remove some of the interviewer utterances that had little to do with the interview situation. During the final stage of work we also compiled the XML corpus master file (see Chapter 4.3).

A couple of lingering questions remain: what should we do with copyright and anonymity issues and where should the files be stored? Both questions are burdened with solutions that either only work within a certain time frame or provide hypothetical scenarios that may or may not come to be. The question of storage is of course bound to which ever institution hosts the project itself. At the time of writing, the corpus is available to VARIENG members, since it is HARES' host institution. However, after VARIENG, whether or not the archive will find itself in the repositories of the Department of Modern Languages in the University of Helsinki or some other institution entirely remains to be seen.

As for the copyright and anonymity problems, our method of publishing the archive for academic use only should solve these questions. After all, if the archive is released for research use only, the original research motivation of the fieldworkers, for which (oral) consent *was* obtained, would be respected. Nevertheless, it has been one of the objectives of the HARES project from its conception to publish the archive for the general public, too. It contains such a wealth of information about life from mid-19$^{th}$ century onwards that it would be a shame not to make it available to all.

# 4 ARCHIVE CONTENT

## 4.1 RECORDINGS

HARES recordings were digitised by Diginord, a company that specialises in preserving archive material in digital format. The original reel-to-reel tapes were digitised so that each

side of the tape was first converted to digital format and then spliced into separate files, depending on how many interview parts were on one side. The files were stored as high quality WAV files on DAT tapes and as compressed MP3 files (128kbps 44.1kHz) on audio CDs.

**TABLE 1**. **Audio files in the Cambridgeshire sampler**[5]

| **AUDIO FILE** | **LENGTH** (hh:mm:ss) | **SIZE** (KB) |
|---|---|---|
| cam01.mp3 | 00:49:08 | 46 067 |
| cam02.mp3 | 00:40:03 | 37 550 |
| cam03.mp3 | 00:54:07 | 50 741 |
| cam04.mp3 | 00:36:33 | 34 266 |
| cam05.mp3 | 01:01:26 | 57 600 |
| cam06.mp3 | 01:32:22 | 86 609 |
| cam07.mp3 | 00:54:35 | 51 188 |
| cam08.mp3 | 01:34:59 | 89 054 |
| cam09.mp3 | 00:47:41 | 44 715 |
| cam10.mp3 | 00:51:50 | 48 598 |
| cam11.mp3 | 00:46:53 | 43 963 |
| cam12.mp3 | 00:47:29 | 44 521 |
| cam13.mp3 | 00:46:52 | 43 947 |
| cam14.mp3 | 00:22:17 | 20 892 |
| cam15.mp3 | 01:16:32 | 71 753 |
| cam16.mp3 | 01:14:00 | 69 384 |
| cam17.mp3 | 00:35:12 | 33 003 |
| cam18.mp3 | 00:47:23 | 44 438 |
| cam19.mp3 | 01:34:35 | 88 685 |
| cam20.mp3 | 00:46:43 | 43 805 |
| **TOTAL** | **18:40:40** | **1 050 824** |

Even though the audio files were spliced, a single interview could be spread out onto several audio files. This is because interviews that lasted more than 45 minutes (length of one side of the reel-to-reel tape) had to be continued on the other side or on a completely different tape. So the first course of action in editing the audio files had to be to glue the interview snippets together and thus form complete recordings. This was done only to the MP3 files to save time.

[5] Modified from Ahava (2010: 113)

After this, the MP3 files were normalised with equalisation filters. Finally, most of the audible clicks were removed manually. The places where clicks resulted from stopping the recorder are annotated with the **<gap/>** tag in the XML corpus and the **<GAP>** tag in the TXT corpus.

Some of the interviews files contain recorded data from more than one day. This is because Vasko sometimes returned in a few days' time to continue talking with the informant. In some of these cases it is quite difficult to distinguish where an interview ends and the next begins, and thus these audio files contain data from different days. However, because the speakers are the same and the difference in time is only a few days, this was not considered a big issue.

## 4.2 SAMPLER FILE STRUCTURE

The corpus contains the following files. The files should be stored in their original directories.

- /HARES
- /HARES/**hares.rng** – The schema for the XML files (RELAX NG)
- /HARES/**hares.xml** – The corpus master file (XML)
- /HARES/**manual.pdf** – This manual (PDF)
- /HARES/**quickstart.pdf** – Quick Start Guide & Reference Sheet (PDF)
- /HARES/AUDIO/**cam01…cam20.mp3** – The audio files (MP3)
- /HARES/TXT/**cam01…cam20.txt** – The interview files (plain text)
- /HARES/TXT/**ex-ir.tag** – Tag settings for WordSmith Tools (see Chapter 4.3.3)
- /HARES/TXT/**tags.tag** – Tag settings for WordSmith Tools (see Chapter 4.3.3)
- /HARES/XML/**cam01…cam20.xml** – The interview files (XML)

## 4.3 CORPUS CONTENT

This section introduces the two corpora that make up the Cambridgeshire sampler of HARES: the XML corpus and the TXT corpus. It is good to remember that the annotation schemata we have chosen for both corpora are, in the end, arbitrary and only provide extra information about the corpus content. This means that the end user can discard all annotation from corpus browsing and queries (by using filters, for example) if he or she so chooses.

### 4.3.1 The XML corpus

The XML-annotated corpus contains the following files:

**hares.rng** – RELAX NG schema for HARES XML files
**hares.xml** – The corpus master file
**cam01…cam20.xml** – The interviews.

The RELAX NG (REgular LAnguage for XML Next Generation) schema contains specifications for how the HARES documents should be structured and what kind of syntax should be used to describe their content. The schema file contains the TEI specifications against which the corpus files are validated. If the corpus contains syntax or structure that is not native to the TEI specifications described within the schema file, the validator reports an error.

**FIGURE 3. XML corpus master file**

| | |
|---|---|
| <?xml version="1.0" encoding="UTF-8"?><br><?oxygen RNGSchema="hares.rng" type="xml"?><br><br><teiCorpus xmlns="http://www.tei-c.org/ns/1.0"<br>xmlns:xi="http://www.w3.org/2001/XInclude"><br><teiHeader type="corpus"><br><fileDesc><br><titleStmt><br><title>…</title><br><author>…</author><br><funder>…</funder><br><principal>…</principal><br><respStmt><br><persName>…</persName><br><resp>…</resp><br></respStmt><br><respStmt><br><persName>…</persName><br><resp>…</resp><br></respStmt><br><respStmt><br><persName>…</persName><br><resp>…</resp><br></respStmt><br><respStmt><br><persName>…</persName><br><resp>…</resp><br></respStmt><br></titleStmt><br><publicationStmt><br><publisher>…</publisher><br><availability><br> … <br></availability><br></publicationStmt><br><sourceDesc><br><recordingStmt><br> … <br></recordingStmt><br></sourceDesc><br></fileDesc><br><encodingDesc><br><projectDesc><br> … | XML declaration and reference to an external RNG file.<br><br>Corpus declaration (with reference to TEI specifications) and the corpus header.<br><br>The various <…Desc> and <…Stmt> tags contain relevant information about the recordings and the project in general. However, most of this information is supplied in this manual, which is why the header in the **hares.xml** file contains mostly references to this document.<br><br>Note that some of the <…Desc> and <…Stmt> elements are left empty. Even though they are not relevant to the HARES project, TEI specifications require them even if left empty.<br><br>Headers in the individual interview files (**cam01-cam20.xml**) contain further information about the interviews themselves. |

| | |
|---|---|
| </projectDesc><br><editorialDecl><br> … <br></editorialDecl><br></encodingDesc><br><profileDesc><br><particDesc><br> … <br></particDesc><br><settingDesc><br> … <br></settingDesc><br></profileDesc><br></teiHeader><br><br><xi:include href="cam01.xml"/><br><xi:include href="cam02.xml"/><br><xi:include href="cam03.xml"/><br><xi:include href="cam04.xml"/><br><xi:include href="cam05.xml"/><br><xi:include href="cam06.xml"/><br><xi:include href="cam07.xml"/><br><xi:include href="cam08.xml"/><br><xi:include href="cam09.xml"/><br><xi:include href="cam10.xml"/><br><xi:include href="cam11.xml"/><br><xi:include href="cam12.xml"/><br><xi:include href="cam13.xml"/><br><xi:include href="cam14.xml"/><br><xi:include href="cam15.xml"/><br><xi:include href="cam16.xml"/><br><xi:include href="cam17.xml"/><br><xi:include href="cam18.xml"/><br><xi:include href="cam19.xml"/><br><xi:include href="cam20.xml"/><br><br></teiCorpus> | The interview files are embedded in the master file via external references. This is simply to keep the corpus more compact and more accessible. Also, it was desirable to have the individual interviews as separate corpus files, so that focusing on a single interview would be more convenient. |

**FIGURE 4. XML interview header**

| | |
|---|---|
| <TEI xml:id="cam04" xmlns="http://www.tei-c.org/ns/1.0"><br><br><teiHeader type="text"><br><fileDesc><br><titleStmt><br><title>HARES interview cam04</title><br><respStmt><br><resp>Transcribed from original audio by</resp><br><orgName>HARES team</orgName> | The header contains various <…Desc> and <…Stmt> tags again, and these provide additional information about the file.<br><br>Audio duration is provided in the **<recording type="audio" dur="PT…">** tag. Here, **PT** prefixes the following time format, which is provided in minutes |

| | |
|---|---|
| </respStmt><br></titleStmt><br><publicationStmt><br><publisher><br></publisher><br></publicationStmt><br><sourceDesc><br><recordingStmt><br><recording type="audio" dur="PT36M33S"><br><respStmt><br><resp>Interview recorded by</resp><br><persName>Anna-Liisa Vasko</persName><br></respStmt><br></recording><br></recordingStmt><br></sourceDesc><br></fileDesc><br><profileDesc><br><particDesc> | (e.g. **36M**) and seconds (e.g. **33S**). |
| <person xml:id="ier1cam04" role="interviewer" sex="1"><br><persName>Mike Hopkins</persName><br></person><br><br><person xml:id="iee1cam04" role="interviewee" sex="2"><br><persName><br><forename full="init">E</forename><br><surname full="init">T</surname><br></persName><br><age><br>70-80<br></age><br><residence><br>Willingham<br></residence><br><occupation><br>Housewife<br></occupation><br><education><br>Left school at 13<br></education><br></person><br><br><person xml:id="iee2cam04" role="interviewee" sex="1"><br><persName><br><forename full="init">A</forename><br><surname full="init">T</surname><br></persName><br><age><br>70-80<br></age><br><residence> | The **<person>** tags give information about the various speakers in the interview (sex "1" is male and "2" is female). Each subsequent utterance in the body text refers back to the **xml:id** code provided in these header tags. Note that only the initials of the informants' names (and one interviewer name) are provided in the tag to ensure anonymity. |

| | |
|---|---|
| Willingham<br></residence><br><occupation><br>Farmer<br></occupation><br><education><br>Left school at 12<br></education><br></person><br><br></particDesc><br><br><settingDesc><br><setting><br><name type="locality">Willingham</name><br><date>16 June 1974</date><br></setting><br></settingDesc><br></profileDesc><br></teiHeader> | The **<setting>** at the end of the file describes the setting of the actual interview. |

#### 4.3.1.1 XML schema

HARES users are expected to know basic XML, which is why this manual does not include information on XML document structure and syntax. The Text Encoding Initiative (TEI) website (http://www.tei-c.org/Support/Learn/) includes comprehensive guides on XML. The website also contains TEI P5 guidelines, which have been used for HARES annotation. In the following chapters, each XML tag used in the interview body is explained.

#### 4.3.1.2 XML tags: <u/>

**Example 1. <u/>**

```
<u xml:id="cam04q91" who="#ier1cam04">
 how about <anchor xml:id="cam04hares0500"/> blackberrying and things down the Fen did you used to go as well
</u>
<u xml:id="cam04s96" who="#iee1cam04">
 yes
</u>
<u xml:id="cam04s97" who="#iee2cam04">
 yeah
</u>
```

Each individual utterance is contained within the **<u>** and **</u>** tags. The utterance tag has two attributes: **xml:id=””**and **who=””**. The first gives a unique, sequential ID for each utterance. The ID is always prefixed with the interview ID and then either **q** or **s** for

interviewer and interviewee, respectively. The number after the prefix increases by one for **q** whenever a new interviewer utterance is in question and by one for **s** in the case of interviewee utterances.

The **who=""** attribute refers back to the respective **<person>** tag in the header (see below). The number sign (**#**) associates the following ID with the more detailed description of the participant in the header. The next four symbols, **ierX** or **ieeX**, denote interviewer or interviewee, respectively, and the number refers to the order in which the person is introduced in the header.

Because each utterance is an uninterrupted stretch of speech by a single speaker, nested utterances are not possible. Overlapping speech is described with the **<seg>** tag (see below).

#### 4.3.1.3 XML tags: <seg/>

**Example 2. <seg/>**

```
<u xml:id="cam20s32" who="#iee3cam20">
<anchor xml:id="cam20hares0100"/> Burt is your nephew <pause/> <seg
synch="#cam20s33-1"> Burt </seg> is your <seg synch="#cam20s33-2"> nephew
</seg>
</u>
<u xml:id="cam20s33" who="#iee1cam20">
<seg xml:id="cam20s33-1"> yeah </seg>
<seg xml:id="cam20s33-2"> yes </seg>
yes <pause/> <seg synch="#cam20s34-1"> Burt is my nephew </seg>
</u>
```

The **<seg/>** tag has two purposes in the XML corpus. Firstly, it is used to describe overlapping speech. We have included overlapping segments to the best of our ability, but in some cases the overlap is insignificant (an utterance such as "**mm**" by another speaker, which doesn't interrupt the flow of the one whose speech is overlapped) or completely indecipherable (someone talking in the far background). In these cases the overlap is not included in the transcription.

The segment which is overlapped is annotated with the **<seg synch="#..."> </seg>** tag. The segment is synched with the respective segment in a following utterance. In the example above, the first overlapped segment ("**Burt**") is linked to the first overlapping segment in the following utterance ("**yeah**"). The ID refers to the utterance ID of the overlapping segment (**<u xml:id="cam20s33"**) and the number after the hyphen is **1** for the first overlapping segment, **2** for the second and so on.

The second purpose for the **<seg>** tag is to denote hesitations and false starts which are stuttered by the speaker. For example: **<seg type="truncation"> w- </seg> what**. This is the only case in the transcription where the dash is used.

#### 4.3.1.4 XML tags: <anchor/>

**Example 3. <anchor/>**

```
<u xml:id="cam20s12" who="#iee2cam20">
 <anchor xml:id="cam20hares0040"/> five in the morning till six at night
<pause/> sometimes seven
</u>
```

The **<anchor/>** tag is used to anchor the text transcription with the respective audio segment. This is done in 10 second intervals. The interval time was chosen arbitrarily, but a 10 second audio-text synchronised segment is good enough for verifying concordance hits and a fair compromise between required detailed and efficiency on the transcriber's behalf.

The ID is prefixed by the interview ID and the text "**hares**". The last four digits are time elapsed from the beginning in seconds. In the example above, 40 seconds have passed from the beginning of the interview audio file.

#### 4.3.1.5 XML tags: <unclear/>, <vocal/>, <pause/>, <gap/>

**Example 4. <pause/>, <vocal/>, <unclear/>**

```
<u xml:id="cam08s281" who="#iee1cam08">
 I said <pause/> I said <vocal> <desc> laugh </desc> </vocal> <pause/>
<anchor xml:id="cam08hares1510"/> <unclear/> I'm brought a parcel from
David Cole's
</u>
```

The **<unclear/>** tag is used for segments in the audio that the transcribers have not understood or heard properly. Due to the quality of the audio, the broad dialect speech of some of the informants and other disturbances (such as speaking through a pipe or pacing around the room while talking) there are a fair number of unclear utterances throughout the audio.

The **<vocal/>** tag is used for any features that have no lexical method of transcribing. Laughter and coughing are two most common vocalisations in the interviews.

The **<pause/>** tag is used to denote a pause in speech. The pauses vary in length, and they are introduced when it is clear that the speaker interrupts his flow of speech for any amount of time.

The **<gap/>** tag is used when the audio file itself contains a gap in the recording. This is a result of the interviewer pressing the STOP or PAUSE button on the recorder. The **<gap/>** tag contains the attribute **reason=""** which provides a reason for the gap (usually just **<gap reason="break in recording"/>**).

#### 4.3.1.6 XML tags: combining empty elements

**Example 5. Combining empty elements**

```
<u xml:id="cam20s15" who="#iee3cam20">
 yes <pause/> <unclear synch="#cam20s16-1"/> gave way
</u>
<u xml:id="cam20s16" who="#iee1cam20">
 <unclear xml:id="cam20s16-1"/>
 eh
</u>
```

In some cases with the **<seg>** tag, the element contains only the **<unclear/>** tag. In these cases, the tags are combined within the **<unclear/>** tag in order to keep the syntax more compact. In the example above, **<unclear synch="#cam20s16-1"/>** is the same as **<seg synch="#cam20s16-1"> <unclear/> </seg>**, and in the following utterance **<unclear xml:id="cam20s16-1"/>** is the same as **<seg xml:id="cam20s16-1"> <unclear/> </seg>**.

Another place where this type of compacting occurs is when a **<seg/>** tag is nested within a **<seg/>** tag. For example:

**Example 6. Combining empty elements**

```
<u xml:id="cam20s37" who="#iee1cam20">
 <seg xml:id="cam20s37-1" type="truncation"> it's ne- it's nearly </seg>
time somebody kicked isn't it
</u>
```

Here, the truncated **"ne-"** is included in the higher level **<seg xml:id="…"/>** tag. So **<seg xml:id="cam20s37-1" type="truncation"> it's ne- … </seg>** would be the same as **<seg xml:id="cam20s37-1"> it's <seg type="truncation"> ne- </seg> … </seg>**. Even though the compacted syntax doesn't provide the same detail of information about the location of the truncated element (in the longer syntax, the truncation marker is immediately adjacent to the truncated element: **<seg type="truncation"> ne- </seg>**), this is not a problem, as the dash in **"ne-"** reveals the truncation.

### 4.3.2 The TXT corpus

In addition to the XML corpus, HARES is distributed as a minimally annotated plain text corpus (TXT corpus). This is because we want HARES to reach a larger audience, especially because XML browsers and query interfaces applicable for spoken language corpora that combine audio and transcriptions are scarce. The TXT corpus is distributed as .txt-files, and it is intended to be used with corpus software such as WordSmith or AntConc with which the user can manipulate the tag set included in or excluded from the corpus use.

The TXT corpus is identical to the XML corpus content-wise, and all but one of the tags in the XML interview body have their equivalents in the TXT corpus (the **<seg/>** tag is omitted). The corpus is distributed as 20 individual interview files labelled **cam01-cam20.txt**. A single interview consists of a header and a body. The header looks like this:

**FIGURE 5. TXT interview header**

```
<CAM04>

<R Anna-Liisa Vasko>
<L Willingham>
<D 16/06/1974>
<T 00:36:33>
<IR1 N=Mike Hopkins X=M>
<IE1 N=ET X=F A=70-80 H=Willingham O=Housewife E=13>
<IE2 N=AT X=M A=70-80 H=Willingham O=Farmer E=12>
```

```
<File ID>

R Recorded by
L Location
D Date
T Length
IRn Interviewer(s)
IEn Interviewee(s)

N name
X sex
A age group
H residence
O occupation
E school until
```

#### 4.3.2.1 TXT corpus annotation schema

All the TXT corpus tags are arbitrarily chosen and are unique to HARES. The tags exist solely to provide descriptive metadata about the content, with the exception of the **<sound>** tag, which can be used in WordSmith for media file playback while using the corpus. The TXT body tags are explained below with relevant examples. Note that with one exception (see the first tag entry below), a TXT tag is always a single tag (e.g. **<GAP>**) unlike the XML tags which, in order to meet well-formedness criteria for XML documents, require an opening and closing tag.

#### 4.3.2.2 TXT tags: <IEn/>, <IRn/>

**Example 7. <IEn/>, <IRn/>**

```
<IR1> did you say there were two horses <sound=../AUDIO/cam19.mp3 240 10>
pulling </IR1>
<IE1> two horse pulling <...> yeah two horses in front <...> yes two
horses in front </IE1>
<IR2> how many horses were there </IR2>
```

The **<IRn/>** and **<IEn/>** tags are similar to the **<u/>** tags in the XML corpus in that they contain each individual utterance within the opening and closing tags. The unique speaker ID is given in the tag name itself (e.g. **<IR1>**) and refers back to the header description. These utterance tags are the only ones in the TXT corpus with an opening and closing tag. This is because some corpus software permits the users to choose the tags they wish to exclude from corpus searches. This way, if the user would like to only analyse interviewee speech, he or she can exclude all **<IR*/>** tags from searches.

#### 4.3.2.3 TXT tags: (#n ), (n )

**Example 8. (#n ), (n )**

```
<IE1> (#1 postman come in ) give father <sound=../AUDIO/cam20.mp3 400 10>
a postcard there it is your boy hasn't (#1 been to school this week )
<...> (#1 his <UNCLEAR> was up ) <...> he had to there then <...> (#1 but
he's been washing ) </IE1>
<IR1> (1 is he ) </IR1>
<IE2> (1 <UNCLEAR> ten acres was about ) </IE2>
<IR1> (1 yes <...> yeah ) </IR1>
<IE2> (1 oh oh was his <UNCLEAR> ) </IE2>
```

The overlap markers in the TXT corpus are not tags, strictly speaking. We have devised a unique syntax for marking overlapping segments in the TXT corpus. The overlapped segment is anchored to the overlapping segment with the number sign **"#"**. The first overlapping segment from the same speaker is marked with **"#1"**, the second with **"#2"** and so on.

In the example above, all the segments are marked with **"#1"** because a single speaker has only one overlapping segment a time. The first overlap occurs with the pair **"postman come in"** and "**is he"**. Then the person who overlaps changes, and **IE2** overlaps **"been to school this week"** with **"<UNCLEAR> ten acres was about"**, etc. Had the third overlapped segment (**"his <UNCLEAR> was up"**) been from **IE2** as well, it would have been marked with **#2**.

#### 4.3.2.4 TXT tags: <sound>

**Example 9. <sound>**

```
<IR1> <sound=../AUDIO/cam14.mp3 110 10> mm </IR1>
<IE1> w- when I started at work <...> on Moat Farm </IE1>
```

```
<IR1> mm </IR1>
<IE1> that's the name of the farm where mister Harlot live now Moat Farm
</IE1>
<IR1> <sound=../AUDIO/cam14.mp3 120 10> mm </IR1>
```

The **<sound>** tag provides the location and name of the audio file first (**../AUDIO/cam14.mp3**), then the amount of seconds elapsed from the beginning and then the default length of audio playback (**10** seconds).

#### 4.3.2.5 TXT tags: <UNCLEAR>, <LAUGH> etc., <...>, <GAP>

**Example 10. <...>, <UNCLEAR>, <COUGH>**

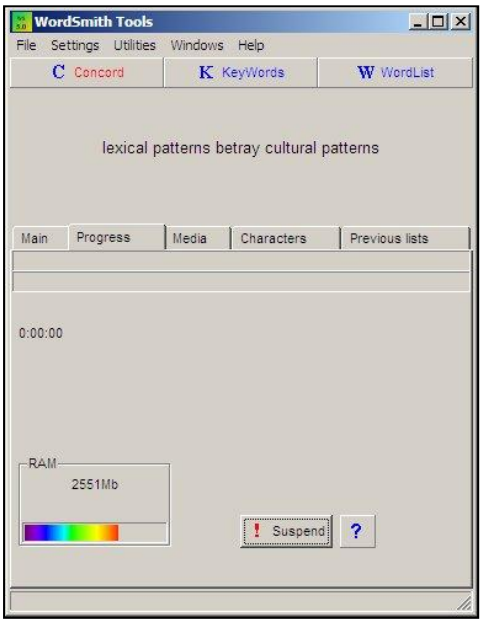
```
<IE1> and cut it up for chaff (#1 for the ) stock and so on and <...> some
</IE1>
<IR1> (1 yes ) so (#1 nothing ) was wasted </IR1>
<IE1> (1 <UNCLEAR> ) <COUGH> no there was nothing wasted </IE1>
```

Unclear segments, pauses, vocalisations and gaps are tagged in the TXT corpus in a similar manner to the XML corpus.

### 4.3.3 Wordsmith-specific instructions

The TXT corpus has been designed for use with WordSmith Tools (tested on version 5.0). This becomes apparent in two cases: first, all the tags are included between the **<** and **>** symbols, so that WordSmith can ignore them in corpus searches, if the user so wishes. Second, the **<sound>** tag has been designed so that the corpus user can listen to the audio while doing corpus searches.

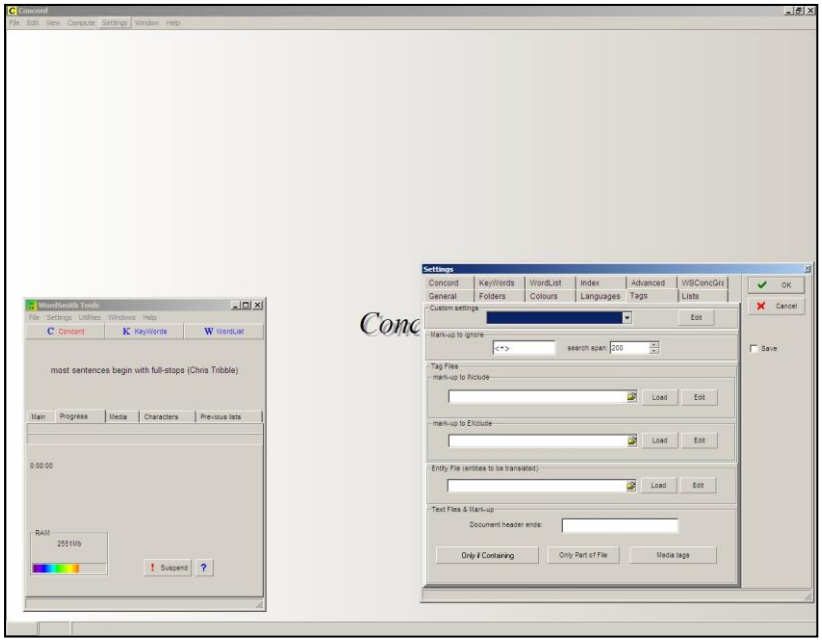
**FIGURE 6. WordSmith Tools main window**




### 4.3.3.1 Set up media tags for concordance searches

1. In the WordSmith Tools main window (see FIGURE 6), choose **Concord.**
2. In the **Concord** window, choose **Settings -> Tags** (see FIGURE 7).

**FIGURE 7. Concord tag settings**



3. In the area "**mark-up to INclude**", click the open folder image in the end of the text area and browse to the TXT corpus directory.
4. Choose the file **tags.tag** and click **Open**.
5. Click **Load** next to the "**mark-up to INclude**" text area (which should now contain the location of the file **tags.tag**).
6. Click **OK** on the screen that pops up to inform you that 1 tag file entry has been found: the media tag **<sound>** (see FIGURE 8).

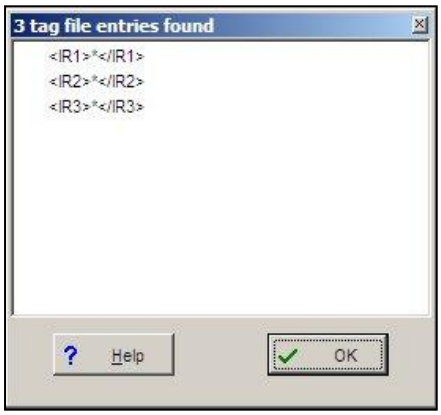
**FIGURE 8. Media tag found**


7. Click **OK** on the settings screen to save the changes you have made.
8. Choose texts and perform a regular concordance search.
9. If you want to hear the 10 second audio segment which contains the concordance hit, choose the concordance line and hit **Ctrl+M** (or choose **File -> Play media file…**).

#### 4.3.3.2 Excluding interviewer questions from searches

1. Repeat steps **1** and **2** above.
2. In the area "**mark-up to EXclude**", click the open folder image in the end of the text area and browse to the TXT corpus directory.
3. Choose the file **ex-ir.tag** and click **Open**.
4. Click **Load** next to the "**mark-up to EXclude**" text area (which should now contain the location of the file **ex-ir.tag).**
5. Click **OK** on the screen that pops up to inform you that 3 tag file entries have been found (see FIGURE 9).

**FIGURE 9. Interviewer tags found**




6. Click **OK** on the settings screen to save the changes you have made.
7. Choose texts and perform a regular concordance search. With these settings, your searches will now only return results from the interviewee utterances.

# 5 INTERVIEW PROFILES

In this section, each interview in the corpus is introduced. The first table of each subsection (subsections are named after each interview ID) contains general information about the interviews: location, date recorded, who recorded the interview, interview length, name(s) of the interviewer(s) and details of the informant(s). Participant information is prefixed by the identifier **ierX** for interviewer and **ieeX** for informant (see Chapter 4.3.1.2).

The second table contains a list of the topics of conversation in the interview, and the third table all non-standard expressions uttered by the informant(s). In the topic index, topics in capital letters represent a general category and topics in small letters represent more particular category labels. Each entry in these two tables is followed by time stamps in the interview. A simple Find operation with the time stamp in any corpus software or word processor will locate the relevant utterance.

## 5.1 CAM01

**TABLE 2. Cam01 details**

| | |
|---|---|
| **INTERVIEW** | |
| Location | Willingham |
| Date recorded | 22 July 1977 |
| Recorded by | Anna-Liisa Vasko |
| Length | 49:08 |
| **IER1CAM01** | |
| Name / initials | Mike Hopkins |

| **IEE1CAM01** | |
|---|---|
| Initials | AA |
| Sex | m |
| Age | n/a |
| Residence | Willingham |
| Left school at | 12 |
| Occupation | Farmer |

**TABLE 3. Cam01 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|
| ACQUIRING FOOD | birding | 2710 |
| | eel gleaving | 2700 |
| DAILY LIFE | Sunday school | 1750 |
| FOOD | | 2340 |
| | flitters | 1520 |
| IMPORTANT EVENTS | flood | 0560 |
| TOOLS AND ITEMS | flatten basket | 0170 |
| WORK ON THE LAND | | 1600 |
| | cow-keeping | 0030 0360 |
| | digging drains | 0840 |
| | filling ponds | 0270 |
| | milking cows | 2410 |
| | ploughing | 0240 |

**TABLE 4. Cam01 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| AGIN (prep) | 0620 0730 0930 1760 2250 |
| BOR (n) | 0680 0880 0900 1130 1640 1900 1950 2300 2450 2480 2590 2830 |
| DRAWED (v) | 0790 |
| EEN'T (v) | 0630 0770 0780 0800 1160 1190 1210 1450 2210 2490 2700 |
| FLITTER (n) | 1520 |
| HARKY (n) | 1610 |
| LALLYGAG (n) | 1040 2060 |
| MESELF (pron) | 0770 |
| OFFWARD (adj) | 2430 |
| PONDERSTICK (n) | 1690 1710 |
| RICKAGE (n) | 1080 |
| WAR (v) | 0180 0200 0280 0400 0520 0810 1030 1050 1090 1110 1500 1740 2530 2680 2850 |
| WARN'T (v) | 0560 0810 0900 1380 1410 1420 1810 1830 1900 1920 2560 2640 2680 2770 2780 2840 |

## 5.2 CAM02

**TABLE 5. Cam02 details**

| | |
|---|---|
| **INTERVIEW** | |
| Location | Willingham |
| Date recorded | 21 June 1974 |
| Recorded by | Anna-Liisa Vasko |
| Length | 40:03 |
| **IER1CAM02** | |
| Name / initials | Mike Hopkins |
| **IEE1CAM02** | |
| Initials | SS |
| Sex | m |
| Age | 60-70 |
| Residence | Willingham |
| Left school at | n/a |
| Occupation | Work on the land |

**TABLE 6. Cam02 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|
| ACQUIRING FOOD | birding | 0220 |
| FOOD | | 1440 |
| GAMES | dominoes | 1550 |
| MEASURES | weights | 0750 |
| TOOLS AND ITEMS | dilly sheaver | 1120 |
| | reaping hook | 0910 |
| WORK ON THE LAND | | 0320 |
| | digging fires | 1880 |
| | draining | 0980 |
| | milking cows | 0540 |
| | mowing | 0950 |
| | ploughing | 2050 |

**TABLE 7. Cam02 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| AGIN (prep) | 0060 0470 1250 1850 2040 |
| BOR (n) | 1520 |
| DILLY SHEAVER (n) | 1120 |
| EEN'T (v) | 0400 0710 0730 0830 1010 1020 1040 1160 1170 1200 1410 1820 1830 1910 1990 2200 |
| HAPENCE (n) | 0970 |
| HAPENNY (n) | 0090 0970 |
| HAUSEN (n) | 0710 |
| HISN (pron) | 1510 |
| HISSELF (pron) | 0500 |

| | |
|---|---|
| OFFWARD (adj) | 0540 0560 0570 0580 0590 0600 0610 0620 0640 |
| THRAIL (n) | 0890 |
| THRUPPENCE (n) | 0970 |
| WAR (v) | 0340 0780 1090 1320 1480 |
| WARN'T (v) | 0060 0340 0970 1280 1320 1350 1400 1410 1430 1480 1630 1720 1760 2350 2400 |

## 5.3 CAM03

**TABLE 8. Cam03 details**

| | |
|---|---|
| **INTERVIEW** | |
| Location | Rampton |
| Date recorded | 19 June 1974 |
| Recorded by | Anna-Liisa Vasko |
| Length | 54:07 |
| **IER1CAM03** | |
| Name / initials | Mike Hopkins |
| **IEE1CAM03** | |
| Initials | HP |
| Sex | m |
| Age | 70-80 |
| Residence | Rampton |
| Left school at | 14 |
| Occupation | Work on the land |
| **IEE2CAM03** | |
| Initials | n/a |
| Sex | f |
| Age | n/a |
| Residence | n/a |
| Left school at | n/a |
| Occupation | n/a |

**TABLE 9. Cam03 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|
| ACQUIRING FOOD | birding | 0470 |
| DAILY LIFE | holidays | 0040 |
| | school | 1990 |
| FOOD | | 2190 |
| GAMES | quoits | 1040 |
| GEOGRAPHY | field names | 1530 |
| IMPORTANT EVENTS | war | 0890 |
| TOOLS AND ITEMS | pritcher | 2860 |
| | shoes | 2380 |
| WORK ON THE LAND | fruit picking | 2630 |

| | | |
|---|---|---|
| | mowing | 0520 |
| | threshing | 2810 |

**TABLE 10 Cam03 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| AGIN (prep) | 1330 2160 |
| BLOWED (v) | 1630 |
| DOCKEY (n) | 2230 |
| EEN'T (v) | 0500 1800 2100 2590 3160 |
| KNOWED (v) | 2500 |
| LUCERNE (n) | 3040 |
| MATER (n) | 1390 |
| PRITCHER (n) | 2860 2890 |
| WAR (v) | 0280 1380 1400 1410 1620 |
| WARN'T (v) | 0140 0150 1530 2480 |

## 5.4 CAM04

**TABLE 11. Cam04 details**

| | |
|---|---|
| **INTERVIEW** | |
| Location | Willingham |
| Date recorded | 16 June 1974 |
| Recorded by | Anna-Liisa Vasko |
| Length | 36:33 |
| **IER1CAM04** | |
| Name / initials | Mike Hopkins |
| **IEE1CAM04** | |
| Initials | ET |
| Sex | f |
| Age | 70-80 |
| Residence | Willingham |
| Left school at | 13 |
| Occupation | Housewife |
| **IEE2CAM04** | |
| Initials | AT |
| Sex | m |
| Age | 70-80 |
| Residence | Willingham |
| Left school at | 12 |
| Occupation | Farmer, road work, cemetery work |

**TABLE 12. Cam04 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|

| | | |
|---|---|---|
| FOOD | | 1770 |
| DAILY LIFE | school | 1290 |
| WORK ON THE LAND | | 0390 |

**TABLE 13. Cam04 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| DOCKEY (n) | 0400 |
| EEN'T (v) | 0360 0390 0570 0640 0810 0870 1480 1740 |
| KNOWED (v) | 0570 |
| WAR (v) | 0770 1130 1140 |
| WARN'T (v) | 0230 0260 0330 0370 0600 0730 0950 1170 1290 1320 2070 |

## 5.5 CAM05

**TABLE 14. Cam05 details**

| | |
|---|---|
| **INTERVIEW** | |
| Location | Rampton |
| Date recorded | 22 June 1974 |
| Recorded by | Anna-Liisa Vasko |
| Length | 61:26 |
| **IER1CAM05** | |
| Name / initials | Mike Hopkins |
| **IEE1CAM05** | |
| Initials | TR |
| Sex | m |
| Age | 90+ |
| Residence | Rampton |
| Left school at | 9 |
| Occupation | Horsekeeper |

**TABLE 15. Cam05 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|
| DAILY LIFE | post | 3570 |
| FOOD | | 2340 |
| GAMES | four corners | 0200 |
| GEOGRAPHY | railway | 2580 |
| TOOLS AND ITEMS | binder | 2650 |
| WORK ON THE LAND | gleaning | 1970 |
| | horsekeeping | 0700 1150 |
| | mowing | 3040 |
| | ploughing | 1530 |
| | shoeing horses | 1190 |
| | thatching | 0320 |

**TABLE 16. Cam05 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| AGIN (prep) | 1400 2180 2920 3600 3680 |
| DOCKEY (n) | 1060 1170 |
| DRAWED (v) | 2760 |
| EEN'T (v) | 0260 0640 0760 1450 1540 2240 2350 2700 2880 |
| KNOWED (v) | 0300 0550 0930 1410 2110 2640 3000 3440 3490 3500 |
| WAKED (v) | 1000 |
| WAR (v) | 0030 0140 0260 0420 0530 1330 1360 1370 1870 3140 3190 3310 3340 3360 3520 3650 |
| WARN'T (v) | 0040 0430 0720 0960 0990 1050 1480 1590 1600 1660 1870 2400 2750 2860 3190 3260 3300 3310 3470 3520 |

## 5.6 CAM06

**TABLE 17. Cam06 details**

| | |
|---|---|
| **INTERVIEW** | |
| Location | Waterbeach |
| Date recorded | 20 June 1974, 21 June 1974 |
| Recorded by | Anna-Liisa Vasko |
| Length | 91:22 |
| **IER1CAM06** | |
| Name / initials | Anna-Liisa Vasko |
| **IEE1CAM06** | |
| Initials | BB |
| Sex | m |
| Age | 80-90 |
| Residence | Waterbeach |
| Left school at | 14 |
| Occupation | Farmer, delivery man |

**TABLE 18. Cam06 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|
| ACQUIRING FOOD | birding | 4990 |
| DAILY LIFE | post | 1080 |
| | ice skating | 1790 |
| | town crier | 3760 |
| | water pumps | 2740 |
| FOOD | | 1450 3310 |
| | baking bread | 2970 |
| IMPORTANT EVENTS | Queen Victoria's Jubilee | 2500 |
| | war | 2210 4680 |

| | | |
|---|---|---|
| LANGUAGE | Suffolk dialect | 2690 |
| STORIES | ghosts | 0600 5280 |
| TOOLS AND ITEMS | lamps | 0940 |
| | penny-farthing | 0160 |
| | vehicles | 1120 |
| WORK | shop-keeping | 2850 |
| WORK ON THE LAND | | 3910 |
| | eel-gleaving | 4880 |
| | fishing | 4840 |
| | horsekeeping | 4340 |
| | milking cows | 4160 |
| | ploughing | 1970 |

**TABLE 19. Cam06 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| DOCKEY (n) | 1590 1600 1620 |
| BURSTED (v) | 4720 |
| DRAWED (v) | 1970 |
| EEN'T (v) | 0390 0500 0570 0600 0840 1020 1180 1210 1280 1290 1520 1650 2030 2110 2480 3360 3390 3710 4180 4300 4400 4420 4540 4620 4810 4830 4860 4910 5260 |
| FRUMETY (n) | 3300 3330 3350 |
| GRET (n) | 4010 4030 |
| HAPENNY (n) | 3070 |
| KNOWED (v) | 0090 0210 1360 3620 5180 |
| WAR (v) | 1100 1490 2270 2370 2430 2950 3020 3330 4640 |
| WARN'T (v) | 0040 0140 0150 0300 0630 0760 0870 1510 1700 1750 2230 2270 2550 3690 4200 |

## 5.7 CAM07

**TABLE 20. Cam07 details**

| | |
|---|---|
| **INTERVIEW** | |
| Location | Swaffham Prior |
| Date recorded | 6 August 1975 |
| Recorded by | Anna-Liisa Vasko |
| Length | 54:36 |
| **IER1CAM07** | |
| Name / initials | Anna-Liisa Vasko |
| **IER2CAM07** | |
| Name / initials | MG |
| **IEE1CAM07** | |
| Initials | EW |
| Sex | m |

| | |
|---|---|
| Age | 70-80 |
| Residence | Swaffham Prior |
| Left school at | 11 |
| Occupation | Farmer, turf-cutter |
| **IEE2CAM07** | |
| Initials | GW |
| Sex | f |
| Age | 70-80 |
| Residence | Swaffham Prior |
| Left school at | n/a |
| Occupation | Housewife |

**TABLE 21. Cam07 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|
| DAILY LIFE | gardening | 1930 |
| FOOD | making butter | 0920 |
| TOOLS AND ITEMS | spade | 0530 |
| WORK | butchery | 1520 |
| WORK ON THE LAND | | 2160 |
| | cow-keeping | 1060 |
| | milking cows | 0830 |
| | turf-cutting | 0230 |

**TABLE 22. Cam07 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| AGIN (prep) | 0410 1040 2030 2330 |
| DOCKEY (n) | 1470 1480 |
| DRAWED (v) | 0490 |
| EEN'T (v) | 1370 1870 1900 1910 2020 2570 2610 3100 |
| GROWED (v) | 1970 |
| STALCH (n) | 0360 |
| THRUPPENCE (n) | 2520 |
| WARN'T (v) | 1560 2280 2310 2510 |

## 5.8 CAM08

**TABLE 23. Cam08 details**

| | |
|---|---|
| **INTERVIEW** | |
| Location | Willingham |
| Date recorded | 23 June 1974 |
| Recorded by | Anna-Liisa Vasko |
| Length | 94:59 |
| **IER1CAM08** | |

| Name / initials | Mike Hopkins |
|---|---|
| **IEE1CAM08** | |
| Initials | ES |
| Sex | m |
| Age | 80-90 |
| Residence | Willingham |
| Left school at | 14 |
| Occupation | Farmer |
| **IEE2CAM08** | |
| Initials | n/a |
| Sex | f |
| Age | 80-90 |
| Residence | Willingham |
| Left school at | n/a |
| Occupation | Housewife |

**TABLE 24. Cam08 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|
| DAILY LIFE | fishing | 0770 |
| FOOD | | 0360 |
| IMPORTANT EVENTS | war | 5430 |
| TOOLS AND ITEMS | lamps | 0970 |
| | shoes | 2850 |

**TABLE 25. Cam08 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| AGIN (prep) | 0170 0410 0690 0750 1050 1520 1940 2140 2280 2340 2360 2370 2650 3610 4360 4630 4810 4960 5210 5440 |
| COMED (v) | 4030 |
| DRAWED (v) | 0980 |
| EEN'T (v) | 1340 2590 4610 5090 5200 5500 |
| KNOWED (v) | 2280 2380 3180 3450 4090 4930 |
| WAR (v) | 0110 0230 0330 1230 1750 1840 1900 2000 2120 2340 2860 3150 3550 3560 4320 4890 5030 5240 5500 |
| WARN'T (v) | 0030 1010 1030 1060 1360 2280 2920 3050 3110 3310 3680 3940 4010 4210 4220 4450 4700 4900 5240 5250 5260 |

## 5.9 CAM09

**TABLE 26. Cam09 details**

| INTERVIEW | |
|---|---|
| Location | Burwell |
| Date recorded | 7 August 1975 |

| | |
|---|---|
| Recorded by | Anna-Liisa Vasko |
| Length | 47:41 |
| **IER1CAM09** | |
| Name / initials | Anna-Liisa Vasko |
| **IER2CAM09** | |
| Name / initials | MG |
| **IEE1CAM09** | |
| Initials | GW |
| Sex | m |
| Age | 90+ |
| Residence | Burwell |
| Left school at | 10 |
| Occupation | Farmer |

**TABLE 27. Cam09 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|
| FOOD | brewing beer | 1790 |
| | making butter | 1680 |
| MEASURES | weights | 0510 |
| WORK ON THE LAND | breaking horses | 0950 |
| | milking cows | 0030 |
| | mowing | 1980 |
| | ploughing | 0260 |

**TABLE 28. Cam09 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| AGIN (prep) | 1330 2550 2820 |
| EEN'T (v) | 0570 0590 0670 1460 1470 1640 1680 1830 1890 1920 2410 |
| WAR (v) | 0450 2590 |
| WARN'T (v) | 0040 0240 0260 0820 1250 1570 2170 2280 2540 |

## 5.10 CAM10

**TABLE 29. Cam10 details**

| | |
|---|---|
| **INTERVIEW** | |
| Location | Harlton |
| Date recorded | 15 July 1975 |
| Recorded by | Anna-Liisa Vasko |
| Length | 51:50 |
| **IER1CAM10** | |
| Name / initials | Anna-Liisa Vasko |
| **IEE1CAM10** | |
| Initials | MP |

| | |
|---|---|
| Sex | m |
| Age | 90+ |
| Residence | Harlton |
| Left school at | 12 |
| Occupation | Farmer |
| **IEE2CAM10** | |
| Initials | PP |
| Sex | m |
| Age | 70-80 |
| Residence | Harlton |
| Left school at | 13 |
| Occupation | Gas worker |

**TABLE 30. Cam10 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|
| DAILY LIFE | post | 2620 |
| MEASURES | weights | 2260 |
| WORK | blacksmithing | 0590 |
| WORK ON THE LAND | | 1840 |
| | mowing | 0190 |

**TABLE 31. Cam10 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| AGIN (prep) | 0610 |
| DOCKEY (n) | 0130 0140 0170 |
| DRAWED (v) | 0290 |
| EEN'T (v) | 1310 1480 1510 2840 2970 3080 3100 |
| HAUSEN (n) | 0620 1740 2700 |
| WAR (v) | 1050 1260 1400 3020 |
| WARN'T (v) | 1690 2670 |

## 5.11 CAM11

**TABLE 32. Cam11 details**

| | |
|---|---|
| **INTERVIEW** | |
| Location | Newton |
| Date recorded | 30 July 1975 |
| Recorded by | Anna-Liisa Vasko |
| Length | 46:53 |
| **IER1CAM11** | |
| Name / initials | Anna-Liisa Vasko |
| **IEE1CAM11** | |
| Initials | JF |

| | |
|---|---|
| Sex | m |
| Age | 70-80 |
| Residence | Newton |
| Left school at | 12 |
| Occupation | Farmer, orchard worker |

**TABLE 33. Cam11 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|
| MEASURES | weights | 1190 |
| TOOLS AND ITEMS | lamps | 1970 |
| | machines | 1020 |
| WORK ON THE LAND | | 0050 |
| | breaking horses | 0780 |
| | ploughing | 0890 |

**TABLE 34. Cam11 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| DOCKEY (n) | 0020 |
| EEN'T (v) | 0320 0510 0670 0760 0940 0970 1060 1110 1220 1430 1500 2000 2060 2100 2440 |
| HAPENNY (n) | 0710 0720 |
| MESELF (pron) | 2040 |

## 5.12 CAM12

**TABLE 35. Cam12 details**

| | |
|---|---|
| **INTERVIEW** | |
| Location | Harston |
| Date recorded | 16 July 1975 |
| Recorded by | Anna-Liisa Vasko |
| Length | 47:29 |
| **IER1CAM12** | |
| Name / initials | Anna-Liisa Vasko |
| **IER2CAM12** | |
| Name / initials | MG |
| **IEE1CAM12** | |
| Initials | AS |
| Sex | m |
| Age | 80-90 |
| Residence | Harston |
| Left school at | 13 |
| Occupation | Farmer |

**TABLE 36. Cam12 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|
| DAILY LIFE | post | 2700 |
| IMPORTANT EVENTS | war | 0160 |
| MEASURES | weights | 1970 |
| WORK ON THE LAND | milking cows | 1650 |
| | ploughing | 0610 1740 |

**TABLE 37. Cam12 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| DOCKEY (n) | 2250 2260 2270 |
| EEN'T (v) | 0410 0850 |
| GROWED (v) | 0490 0600 |
| KNOWED (v) | 0940 1000 1400 2810 |
| THEIRSELF (pron) | 2170 |
| WARN'T (v) | 0400 0930 1100 2210 2340 |

## 5.13 CAM13

**TABLE 38. Cam13 details**

| | |
|---|---|
| **INTERVIEW** | |
| Location | Bassingbourn |
| Date recorded | 18 July 1975 |
| Recorded by | Anna-Liisa Vasko |
| Length | 46:52 |
| **IER1CAM13** | |
| Name / initials | Anna-Liisa Vasko |
| **IER2CAM13** | |
| Name / initials | MG |
| **IEE1CAM13** | |
| Initials | BR |
| Sex | m |
| Age | 80-90 |
| Residence | Bassingbourn |
| Left school at | 9 |
| Occupation | Farmer |
| **IEE2CAM13** | |
| Initials | ER |
| Sex | m |
| Age | 70-80 |
| Residence | Bassingbourn |
| Left school at | 13 |
| Occupation | Farmer |

**TABLE 39. Cam13 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|
| DAILY LIFE | post | 2260 |
| FOOD | | 0670 |
| MEASURES | weights | 0960 |
| TOOLS AND ITEMS | lamps | 2390 |
| WORK ON THE LAND | growing food | 0310 |
| | ploughing | 0000 |

**TABLE 40. Cam13 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| AGIN (prep) | 1700 2330 2400 |
| DOCKEY (n) | 0580 0590 |
| GROWED (v) | 0400 |
| HAPENCE (n) | 0470 0890 |
| HAPENNY (n) | 0480 0500 |
| PENNORTH (n) | 0470 |
| POSTMANS (n) | 2270 |
| SOWED (v) | 0300 |
| THINGBOB(BY) (n) | 0690 1080 |
| THRUPPENCE (n) | 1170 |
| WARN'T (v) | 1410 |

## 5.14 CAM14

**TABLE 41. Cam14 details**

| | |
|---|---|
| **INTERVIEW** | |
| Location | Castle Camps |
| Date recorded | 5 July 1974 |
| Recorded by | Anna-Liisa Vasko |
| Length | 22:17 |
| **IER1CAM14** | |
| Name / initials | Anna-Liisa Vasko |
| **IEE1CAM14** | |
| Initials | JH |
| Sex | m |
| Age | 80-90 |
| Residence | Castle Camps |
| Left school at | 11 |
| Occupation | Farmer |
| **IEE2CAM14** | |
| Initials | BH |

| | |
|---|---|
| Sex | f |
| Age | 50-60 |
| Residence | Castle Camps |
| Left school at | n/a |
| Occupation | Takes care of her father (JH) |
| **IEE3CAM14** | |
| Initials | FH |
| Sex | m |
| Age | 50-60 |
| Residence | Castle Camps |
| Left school at | n/a |
| Occupation | n/a |

**TABLE 42. Cam14 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|
| WORK | butchery | 1150 |
| WORK ON THE LAND | ditching | 0610 |
| | milking cows | 0820 |
| | ploughing | 0450 |

**TABLE 43. Cam14 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| AGIN (prep) | 0320 |
| THEIRSELF (pron) | 0340 |
| WARN'T (v) | 0410 |

## 5.15 CAM15

**TABLE 44. Cam15 details**

| | |
|---|---|
| **INTERVIEW** | |
| Location | Bartlow |
| Date recorded | 5 July 1974, 10 July 1974, 13 July 1974 |
| Recorded by | Anna-Liisa Vasko |
| Length | 76:32 |
| **IER1CAM15** | |
| Name / initials | Anna-Liisa Vasko |
| **IEE1CAM15** | |
| Initials | CP |
| Sex | m |
| Age | 70-80 |
| Residence | Bartlow |
| Left school at | 13 |
| Occupation | Farmer |

| IEE2CAM15 | |
|---|---|
| Initials | MP |
| Sex | f |
| Age | 60-70 |
| Residence | Bartlow |
| Left school at | 14 |
| Occupation | Housewife |

**TABLE 45. Cam15 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|
| DAILY LIFE | post | 1930 |
| | school | 3780 |
| GAMES | | 3630 |
| | dominoes | 4320 |
| IMPORTANT EVENTS | cold winter | 2890 |
| | war | 0870 |
| TOOLS AND ITEMS | bicycle | 2030 |
| | lamps | 1810 |
| WORK | butchery | 0390 |
| WORK ON THE LAND | ditching | 0610 |
| | feeding animals | 0330 |
| | growing food | 1460 |
| | milking cows | 0480 |
| | ploughing | 0090 |
| | shoeing horses | 2150 |
| | thatching | 2720 |
| | threshing | 3360 |

**TABLE 46. Cam15 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| DOCKEY (n) | 0990 1000 1020 1030 1040 1050 |
| HAPENNY (n) | 4350 |
| RUNNED (v) | 3720 3730 3770 |
| THEIRSELF (pron) | 0330 2620 |
| WARN'T (v) | 1720 1740 3690 |

## 5.16 CAM16

**TABLE 47. Cam16 details**

| INTERVIEW | |
|---|---|
| Location | Waterbeach |
| Date recorded | 19 June 1974, 27 June 1974 |
| Recorded by | Anna-Liisa Vasko |

| | |
|---|---|
| Length | 74:00 |
| **IER1CAM16** | |
| Name / initials | Anna-Liisa Vasko |
| **IEE1CAM16** | |
| Initials | SB |
| Sex | f |
| Age | 90+ |
| Residence | Waterbeach |
| Left school at | 11 |
| Occupation | Housewife |

**TABLE 48. Cam16 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| EEN'T (v) | 3440 3590 3780 3790 3800 3810 3880 4010 4280 4330 |
| GROWED (v) | 4270 |
| HAPENNY (n) | 0520 |

## 5.17 CAM17

**TABLE 49. Cam17 details**

| | |
|---|---|
| **INTERVIEW** | |
| Location | Fulbourn |
| Date recorded | 18 July 1975 |
| Recorded by | Anna-Liisa Vasko |
| Length | 35:12 |
| **IER1CAM17** | |
| Name / initials | Anna-Liisa Vasko |
| **IER2CAM17** | |
| Name / initials | MG |
| **IEE1CAM17** | |
| Initials | CM |
| Sex | m |
| Age | 70-80 |
| Residence | Fulbourn |
| Left school at | 14 |
| Occupation | Land worker, coal delivery |
| **IEE2CAM17** | |
| Initials | EM |
| Sex | f |
| Age | 70-80 |
| Residence | Fulbourn |
| Left school at | n/a |
| Occupation | Housewife |

**TABLE 50. Cam17 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|
| FOOD | | 0800 |
| IMPORTANT EVENTS | war | 1050 |
| WORK ON THE LAND | feeding animals | 0590 |
| | growing food | 0340 |
| | keeping sheep | 0960 |
| | milking cows | 0460 |
| | ploughing | 0070 |
| | threshing | 0260 |

**TABLE 51. Cam17 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| DOCKEY (n) | 0800 0810 |
| DRAWED (v) | 0280 |
| EEN'T (v) | 1310 |
| OFFWARD (adj) | 0480 |
| WAR (v) | 0530 |

## 5.18 CAM18

**TABLE 52. Cam18 details**

| | |
|---|---|
| **INTERVIEW** | |
| Location | Little Eversden |
| Date recorded | 23 July 1975 |
| Recorded by | Anna-Liisa Vasko |
| Length | 47:23 |
| **IER1CAM18** | |
| Name / initials | Anna-Liisa Vasko |
| **IER2CAM18** | |
| Name / initials | MG |
| **IEE1CAM18** | |
| Initials | SC |
| Sex | m |
| Age | 80-90 |
| Residence | Little Eversden |
| Left school at | 12 |
| Occupation | Land worker |

**TABLE 53. Cam18 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|
| DAILY LIFE | school | 0300 |

| | | |
|---|---|---|
| WORK ON THE LAND | growing food | 0700 |
| | keeping sheep | 1270 |
| | mowing | 0580 |
| | thatching | 1110 |

**TABLE 54. Cam18 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| DOCKEY (n) | 0800 0830 |
| EEN'T (v) | 0270 0520 0690 2400 2420 2760 2820 |
| KNOWED (v) | 0100 0910 0920 2290 |
| WAR (v) | 0430 0440 2320 |
| WARN'T (v) | 0400 1730 1990 2060 |

## 5.19 CAM19

**TABLE 55. Cam19 details**

| | |
|---|---|
| **INTERVIEW** | |
| Location | West Wickham |
| Date recorded | 17 July 1975, 22 July 1975 |
| Recorded by | Anna-Liisa Vasko |
| Length | 94:35 |
| **IER1CAM19** | |
| Name / initials | Anna-Liisa Vasko |
| **IER2CAM19** | |
| Name / initials | MG |
| **IEE1CAM19** | |
| Initials | CC |
| Sex | m |
| Age | 70-80 |
| Residence | West Wickham |
| Left school at | 13 |
| Occupation | Farmer |

**TABLE 56. Cam19 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|
| ACQUIRING FOOD | wild rabbit | 2110 |
| FOOD | | 0550 |
| GEOGRAPHY | Roman road | 4130 |
| MEASURES | weights | 1030 |
| TOOLS AND ITEMS | farm cart | 3240 |
| WORK ON THE LAND | fruit picking | 1860 |
| | growing food | 0900 2420 |
| | horsekeeping | 3350 |

| | |
|---|---|
| ploughing | 0150 |
| threshing | 0300 |

**TABLE 57. Cam19 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| DRAWED (v) | 3950 |
| EEN'T (v) | 0380 1090 1640 2160 2780 3340 3510 3530 3790 3800 3820 3890 4040 4520 4640 4910 4930 4970 5190 5470 5500 5650 |
| HAUSEN (n) | 0000 0010 1790 1800 |
| WARN'T (v) | 1570 1580 1600 1760 3840 3850 4510 4710 5170 5270 5280 5590 |

## 5.20 CAM20

**TABLE 58. Cam20 details**

| | |
|---|---|
| **INTERVIEW** | |
| Location | Over |
| Date recorded | 26 September 1974 |
| Recorded by | Anna-Liisa Vasko |
| Length | 46:43 |
| **IER1CAM20** | |
| Name / initials | Mike Hopkins |
| **IER2CAM20** | |
| Name / initials | Anna-Liisa Vasko |
| **IEE1CAM20** | |
| Initials | TR |
| Sex | m |
| Age | 90+ |
| Residence | Rampton |
| Left school at | 9 |
| Occupation | Farmer |
| **IEE2CAM20** | |
| Initials | EF |
| Sex | m |
| Age | 90+ |
| Residence | Over |
| Left school at | 11 |
| Occupation | Horsekeeper |
| **IEE3CAM20** | |
| Initials | HH |
| Sex | f |
| Age | 70-80 |
| Residence | Over |
| Left school at | n/a |
| Occupation | Housewife |

**TABLE 59. Cam20 topic index**

| GENERAL CATEGORY | DETAILED CATEGORY | TIME STAMP(S) |
|---|---|---|
| FOOD | | 0710 |

**TABLE 60. Cam20 non-standard expressions**

| EXPRESSION | TIME STAMP(S) |
|---|---|
| AGIN (prep) | 0500 0810 1040 1050 1290 1340 1510 1980 2350 2640 2650 2770 |
| DOCKEY (n) | 0700 0710 0720 |
| EEN'T (v) | 0000 0150 0190 0290 0340 1030 1040 1160 1480 1910 2020 2270 2640 2720 |
| GOODUNS (n) | 2230 |
| HISN (pron) | 1590 |
| KNOWED (v) | 0040 0300 0820 1560 1830 2670 |
| THESELF (pron) | 0120 |
| WAR (v) | 0320 0450 0460 1560 1580 2350 2380 2500 2600 2770 |
| WARN'T (v) | 0340 0410 0540 0590 0860 1170 1550 1590 1720 1930 1940 1950 1990 2160 2180 2260 |

# 6 LIST OF NON-STANDARD EXPRESSIONS

**TABLE 61. Glossary of non-standard expressions with definitions[6]**

| EXPRESSION | WORD CLASS | DEFINITION | EXAMPLE | FOUND IN interview |
|---|---|---|---|---|
| AGIN | Preposition | Next to or near something (cf. *against*) | somebody stood **agin** the gate (cam03) | cam01<br>cam02<br>cam03<br>cam05<br>cam07<br>cam08<br>cam09<br>cam10<br>cam13<br>cam14<br>cam16<br>cam20 |
| BLOWED | Verb | Past tense *blow* (cf. *blew*) | bought the bought the place what what was <anchor xml:id="cam03hares163 | cam03 |

[6] Thank you to Michael McCarthy for helping with some of the definitions

| | | | | |
|---|---|---|---|---|
| | | | 0"/> **blowed** up (cam03) | |
| BOR | Noun | A calling name (cf. *boy*) | no <pause/> I remember that **bor** <pause/> yeah (cam01) | cam01<br>cam02 |
| BURSTED | Verb | Past tense *burst* (cf. *burst*) | the river bank <pause/> **bursted** down here just (cam06) | cam06 |
| COMED | Verb | Past tense *come* (cf. *came*) | what's his name what <pause/> er <pause/> **comed** over here (cam08) | cam08 |
| DILLY SHEAVER | Noun | A type of scythe used for clearing out drains | and that was a **dilly sheaver** (cam02) | cam02 |
| DOCKEY | Noun | Lunch | eleven o'clock eleven o'clock generally <pause/> that's **dockey** time (cam12) | cam03<br>cam04<br>cam05<br>cam06<br>cam07<br>cam10<br>cam11<br>cam12<br>cam13<br>cam15<br>cam17<br>cam18<br>cam20 |
| DRAWED | Verb | Past tense *draw* (cf. *drew*) | if you when you **drawed** a <seg type="truncation"> m-</seg> mark (cam10) | cam01<br>cam05<br>cam06<br>cam07<br>cam08<br>cam10<br>cam17<br>cam19 |
| EEN'T | Verb | Negative present tense BE (cf. *isn't, ain't*) | we know we **een't** they **een't** got so hard work to do now as what they used to (cam19) | cam01<br>cam02<br>cam03<br>cam04<br>cam05<br>cam06<br>cam07<br>cam08<br>cam09<br>cam10<br>cam11 |

| | | | | |
|---|---|---|---|---|
| | | | | cam12<br>cam16<br>cam17<br>cam18<br>cam19<br>cam20 |
| FLITTER | Noun | Dried and crispy bits of meat | we used to have **flitters** for Saturday dinner (cam01) | cam01 |
| FRUMETY | Noun | A wheat dish (cf. *frumenty*) | there was some stuff they call **frumety** (cam06) | cam06 |
| GOODUNS | Noun | Cf. *good ones* | some tiny **gooduns** <pause/> they ain't all been bad (cam20) | cam20 |
| GRET | Noun | Work for which compensation is according to the amount of work done (cf. *piecework*) | I took good many by what they called by the **gret** (cam06) | cam06 |
| GROWED | Verb | Past tense *grow* (cf. *grew*) | and off we used to had to go <pause/> as we **growed** up (cam07) | cam07<br>cam12<br>cam13<br>cam16 |
| HAPENCE | Noun | Coin half the value of a penny (cf. *halfpenny*), see also HAPENNY | our beer and that <pause/> three **hapence** a pint (cam13) | cam02<br>cam13 |
| HAPENNY | Noun | Coin half the value of a penny (cf. *halfpenny*), see also HAPENCE | **hapenny** a lot they used to get what a pint whatever it was they used to fill your can (cam11) | cam02<br>cam06<br>cam11<br>cam13<br>cam15<br>cam16 |
| HARKY | Noun | Unknown | he had a bit of wurzels in <pause/> in the **harky** home (cam01) | cam01 |
| HAUSEN | Noun | Plural of *house* (cf. *houses*) | I mean <pause/> one word which <anchor xml:id="cam02hares0710"/> always sticks in my mind **hausen** (cam02) | cam02<br>cam10<br>cam19 |
| HISN | Pronoun | Third person | cos it warn't so big | cam02 |

| | | | | |
|---|---|---|---|---|
| | | possessive pronoun (cf. *his / his one*) | as **hisn** **(cam20)** | cam20 |
| HISSELF | Pronoun | Third person reflexive pronoun (cf. *himself*) | and all he ever had <anchor xml:id="cam02hares0500"/> for **hisself** was sixpence a week (cam02) | cam02 |
| KNOWED | Verb | Past tense *know* (cf. *knew*) | he **knowed** in a minute <pause/> as I went on the road I was on here (cam18) | cam03<br>cam04<br>cam05<br>cam06<br>cam08<br>cam12<br>cam18<br>cam20 |
| LALLYGAG | Noun | Straps that keep the trousers rolled up below the knees | the little old short old man <pause/> he always used to have **lallygags** <anchor xml:id="cam01hares2070"/> on (cam01) | cam01 |
| LUCERNE | Noun | A crop variety of alfalfa used in green manure and in fodder | <seg xml:id="cam03s600-1"> I eh </seg> used to cut **lucerne** (cam03) | cam03 |
| MATER | Noun | Mother | and she said <pause/> Arthur <pause/> yes **mater** | cam03 |
| MESELF | Pronoun | First person reflexive pronoun (cf. *myself*) | I throw <anchor xml:id="cam01hares0770"/> **meself** out of me boat that's too much I een't good at that (cam01) | cam01<br>cam11 |
| OFFWARD | Adjective | The right-hand (or off) side of a cow | you you milk your cows on the **offward** side <seg synch="#cam01s551-1"> didn't you (cam01) | cam01<br>cam02<br>cam17 |
| PENNORTH | Noun | Worth a penny (cf. *pennyworth*) | and get **pennorth** of milk hapence of milk <pause/> have a canful (cam13) | cam13 |
| PONDERSTICK | Noun | Unknown | and we were on about **pondersticks** and one thing or another (cam01) | cam01 |
| POSTMANS | Noun | Plural of *postman* (cf. | used to be <anchor xml:id="cam13hares227 | cam13 |

| | | | | |
|---|---|---|---|---|
| | | *postmen*) | 0"/> no end of **postmans** | |
| PRITCHER | Noun | A tool used for clearing drains of bushes | I found a **pritcher** the other day do you know what that is (cam03) | cam03 |
| RICKAGE | Noun | Unknown | went in there they were threshing in that **rickage** you know (cam01) | cam01 |
| RUNNED | Verb | Past tense *run* (cf. *ran*) | some of them got some young turkeys you see and they **runned** in cos they it it was going dark (cam15) | cam15 |
| SOWED | Verb | Past participle of *sow* (cf. *sown*) | not far off of the <pause/> where the corn was sowed (cam13) | cam13 |
| STALCH | Noun | Long bundles of straw used in thatching | wheel them up into great <pause/> <anchor xml:id="cam07hares0360"/> **stalches** (cam07) | cam07 |
| THESELF | Pronoun | Third person reflexive pronoun (cf. *themselves*) | they got a bungalow down the bottom they built it **theself** (cam20) | cam20 |
| THEIRSELF | Pronoun | Third person reflexive pronoun (cf. *themselves*) | they enjoy <anchor xml:id="cam12hares2170"/> **theirself** there's more together (cam12) | cam12<br>cam14<br>cam15 |
| THINGBOB(BY) | Noun | Colloquial word to represent a thing | the milk skimmed milk out the <seg type="truncation">th- </seg> **thingbobby** out (cam13) | cam13 |
| THRAIL | Noun | Variant pronunciation of *flail* | they always used to thresh beans with a **thrail** (cam02) | cam02 |
| THRUPPENCE | Noun | Sum of money equal to three pennies | cos I mean you could go in a pub and you could get a pint of beer then for <pause/> for **thruppence** (cam07) | cam02<br>cam07<br>cam13 |
| WAKED | Verb | Past tense *wake* (cf. *woke*) | out he said Peter come over night afore the last he said and when I **waked** up yesterday morning I was dressed (cam05) | cam05 |
| WAR | Verb | Past tense | and I was born | cam01 |

| | | | | |
|---|---|---|---|---|
| | | positive BE (cf. *was*, *were*) | <pause/> the same day as he **war** (cam08) | cam02<br>cam03<br>cam04<br>cam05<br>cam06<br>cam08<br>cam09<br>cam10<br>cam17<br>cam18<br>cam20 |
| WARN'T | Verb | Past tense negative BE (cf. *wasn't*, *weren't*) | and three on the double plough but there **warn't** very many (cam12) | cam01<br>cam02<br>cam03<br>cam04<br>cam05<br>cam06<br>cam07<br>cam08<br>cam09<br>cam10<br>cam12<br>cam13<br>cam14<br>cam15<br>cam18<br>cam19 |

# 7 TOPIC INDEX

**TABLE 62. Index of conversation topics**

| CATEGORY | | FOUND IN |
|---|---|---|
| **GENERAL** | **particular** | interview |
| ACQUIRING FOOD | birding | cam01<br>cam02<br>cam03<br>cam06 |
| ACQUIRING FOOD | eel gleaving | cam01<br>cam06 |
| ACQUIRING FOOD | wild rabbit | cam19 |
| DAILY LIFE | gardening | cam07 |
| DAILY LIFE | fishing | cam06<br>cam08 |
| DAILY LIFE | holidays | cam03 |
| DAILY LIFE | ice skating | cam06 |

| | | |
|---|---|---|
| DAILY LIFE | post | cam05 |
| | | cam06 |
| | | cam10 |
| | | cam12 |
| | | cam13 |
| | | cam15 |
| DAILY LIFE | school | cam03 |
| | | cam04 |
| | | cam15 |
| | | cam18 |
| DAILY LIFE | Sunday school | cam01 |
| DAILY LIFE | town crier | cam06 |
| DAILY LIFE | water pumps | cam06 |
| FOOD | | cam01 |
| | | cam02 |
| | | cam03 |
| | | cam04 |
| | | cam05 |
| | | cam06 |
| | | cam08 |
| | | cam09 |
| | | cam13 |
| | | cam17 |
| | | cam19 |
| | | cam20 |
| FOOD | baking bread | cam06 |
| FOOD | brewing beer | cam09 |
| FOOD | flitters | cam01 |
| FOOD | making butter | cam07 |
| | | cam09 |
| GAMES | | cam15 |
| GAMES | dominoes | cam02 |
| | | cam15 |
| GAMES | four corners | cam05 |
| GAMES | quoits | cam03 |
| GEOGRAPHY | field names | cam03 |
| GEOGRAPHY | railway | cam05 |
| GEOGRAPHY | Roman road | cam19 |
| IMPORTANT EVENTS | cold winter | cam15 |
| IMPORTANT EVENTS | flood | cam01 |
| IMPORTANT EVENTS | Queen Victoria's Jubilee | cam06 |
| IMPORTANT EVENTS | war | cam03 |
| | | cam06 |
| | | cam08 |
| | | cam12 |
| | | cam15 |

| | | |
|---|---|---|
| | | cam17 |
| LANGUAGE | Suffolk dialect | cam06 |
| MEASURES | weights | cam09 |
| | | cam10 |
| | | cam11 |
| | | cam12 |
| | | cam13 |
| | | cam19 |
| STORIES | ghosts | cam06 |
| TOOLS AND ITEMS | bicycle | cam15 |
| TOOLS AND ITEMS | binder | cam05 |
| TOOLS AND ITEMS | dilly sheaver | cam02 |
| TOOLS AND ITEMS | farm cart | cam19 |
| TOOLS AND ITEMS | flatten basket | cam01 |
| TOOLS AND ITEMS | lamps | cam06 |
| | | cam08 |
| | | cam11 |
| | | cam13 |
| | | cam15 |
| TOOLS AND ITEMS | machines | cam11 |
| TOOLS AND ITEMS | penny-farthing | cam06 |
| TOOLS AND ITEMS | pritcher | cam03 |
| TOOLS AND ITEMS | reaping hook | cam02 |
| TOOLS AND ITEMS | shoes | cam03 |
| | | cam08 |
| TOOLS AND ITEMS | spade | cam07 |
| TOOLS AND ITEMS | vehicles | cam06 |
| WORK | butchery | cam07 |
| | | cam14 |
| | | cam15 |
| WORK | blacksmithing | cam10 |
| WORK | shop-keeping | cam06 |
| WORK ON THE LAND | | cam01 |
| | | cam02 |
| | | cam04 |
| | | cam06 |
| | | cam07 |
| | | cam10 |
| | | cam11 |
| WORK ON THE LAND | breaking horses | cam09 |
| | | cam11 |
| WORK ON THE LAND | cow keeping | cam01 |
| | | cam07 |
| WORK ON THE LAND | digging drains | cam01 |
| WORK ON THE LAND | digging fires | cam02 |
| WORK ON THE LAND | ditching | cam14 |

| | | |
|---|---|---|
| | | cam15 |
| WORK ON THE LAND | draining | cam02 |
| WORK ON THE LAND | eel-gleaving | cam06 |
| WORK ON THE LAND | feeding animals | cam15 |
| | | cam17 |
| WORK ON THE LAND | filling ponds | cam01 |
| WORK ON THE LAND | fishing | cam01 |
| WORK ON THE LAND | fruit picking | cam03 |
| | | cam19 |
| WORK ON THE LAND | gleaning | cam05 |
| WORK ON THE LAND | growing food | cam13 |
| | | cam15 |
| | | cam17 |
| | | cam18 |
| | | cam19 |
| WORK ON THE LAND | horsekeeping | cam05 |
| | | cam06 |
| | | cam19 |
| WORK ON THE LAND | keeping sheep | cam17 |
| | | cam18 |
| WORK ON THE LAND | milking cows | cam01 |
| | | cam02 |
| | | cam06 |
| | | cam07 |
| | | cam09 |
| | | cam12 |
| | | cam14 |
| | | cam15 |
| | | cam17 |
| WORK ON THE LAND | mowing | cam02 |
| | | cam03 |
| | | cam05 |
| | | cam09 |
| | | cam10 |
| | | cam18 |
| WORK ON THE LAND | ploughing | cam01 |
| | | cam02 |
| | | cam05 |
| | | cam06 |
| | | cam09 |
| | | cam11 |
| | | cam12 |
| | | cam13 |
| | | cam14 |
| | | cam15 |
| | | cam17 |

| | | |
|---|---|---|
| | | cam19 |
| WORK ON THE LAND | shoeing horses | cam05<br>cam15 |
| WORK ON THE LAND | thatching | cam05<br>cam15<br>cam18 |
| WORK ON THE LAND | threshing | cam03<br>cam15<br>cam17<br>cam19 |
| WORK ON THE LAND | turf-cutting | cam07 |

# 8 REFERENCE LINE AND COPYRIGHT



## 8.1 REFERENCE LINE

*HARES-CAM = Helsinki Archive of Regional English Speech - Cambridgeshire sampler*. 2010. Compiled by Ahava, Simo, Joseph McVeigh and Anna-Liisa Vasko at the Department of English, University of Helsinki.

## 8.2 CITATION

To refer to a HARES interview in running text, add the following after the quote: (interviewID-hares). For example:

(1) from nineteen seventeen <pause/> nineteen nineteen (cam12-hares)

# 9 CONTACT INFORMATION

**Helsinki Archive of Regional English Speech**
VARIENG
P.O. Box 24
FIN-00014 University of Helsinki
Finland

**Simo Ahava – HARES contact person**
simo.ahava@alumni.helsinki.fi

# 10 REFERENCES

## 10.1 PRINTED TEXTS

Vasko, Anna-Liisa. 2005. *Up Cambridge. Prepositional locative expressions in dialect speech: a corpus-based study of the Cambridgeshire dialect*. Helsinki: Société Néophilologique.

## 10.2 ONLINE SOURCES

Ahava, Simo. 2010. *Intermediate Past BE: A Paradigm Reshaped With Data Drawn From HARES*. Unpublished MA thesis. University of Helsinki. Visited 9 September 2010. http://urn.fi/URN:NBN:fi-fe201006212088

*Cambridgeshire Dialect Grammar*. Anna-Liisa Vasko. Visited 6 September 2010. http://www.helsinki.fi/varieng/journal/volumes/04/