10

From Sociolinguistic Research to English Language Teaching

Jenny Cheshire and Susan Fox

Uncorrected Proof

1 Introduction: The Educational Context

Sociolinguistic research on spoken English and language variation has acquired great importance in the UK secondary school curriculum with the introduction in 1981 of the GCE (General Certificate of Education) advanced level examination in English Language. GCE examinations are taken by school leavers at the age of 18 in England, Wales and Northern Ireland; they are also taken by students in Scotland as an alternative to Advanced Higher examinations and as an international qualification around the world. In the UK the grades obtained determine whether students are accepted for university education, with the best universities requiring very high grades. Generally, students take GCE examinations in three or four subjects of their choice.

J. Cheshire
Queen Mary, University of London, London, UK

S. Fox (✉)
University of Bern, Bern, Switzerland

© The Editor(s) (if applicable) and The Author(s) 2016

K.P. Corrigan, A. Mearns (eds.), *Creating and Digitizing Language Corpora*, DOI 10.1057/978-1-137-38645-8_10

GCE English Language has been steadily increasing in popularity since it was introduced and it is now one of the fastest-growing school subjects. During the period 2003–10, for example, the number of entries for all five examining boards rose from 14,751 in 2003 to 23,211 in 2010.[1] Teachers of this subject are required to 'introduce students to the concepts and methods of the disciplines of English language/linguistics in relation to a wide range of spoken and written forms of English, including electronic and multimodal forms' (Department for Education 2014: 1). They must accurately use a range of terminology associated with phonetics, phonology, prosody, lexis and semantics, 'grammar including morphology', 'pragmatics and discourse' and show how the terminology can be applied to a range of contexts for language use, including historical, geographical, social and individual varieties of English, and aspects of language and identity (Department for Education 2014: 2).[2]

While we may applaud the government's drive to educate young people to use linguistic concepts and terminology to accurately analyse language in use, in practice the specifications place heavy demands on teachers. Most English schoolteachers have a background in English literature, not language—two very different disciplines—and so have little or no training in linguistic analysis, especially as applied to spoken rather than written language. In-service training is unavailable in many areas of the UK, so teachers have to learn new subject knowledge largely on their own (Bleiman and Webster 2006: 29). As a result, many teachers are 'seriously lacking in confidence' and 'often feel overwhelmed and uncer-

Uncorrected Proof

[1] See www.phon.ucl.ac.uk/home/dick/ec/stats.htm.

[2] At the time of our project, in 2011, the specifications for this examination incorporated still more sociolinguistics: for example, a section on Language Variation and Change included the study of standard and vernacular dialects and accents, and debates about the role of standard and vernacular varieties in education. For a brief period, between 2010 and 2014, the GCSE (General Certificate of Secondary Education) English Language and English Literature examinations also included the study of spoken English, with the English Language specifications including the study of variation in spoken English and its relation to identity and cultural diversity. This examination is taken at age 16; students choose between 1 and 10 subjects, with English Language or English Literature compulsory. In 2014 the Secretary of State for Education, Michael Gove, changed the GCSE curriculum so that it now focuses almost entirely on written English, with a small and unassessed component dealing with using 'spoken standard English effectively in speeches and presentations' (Department for Education 2013: 6). It is ironic and depressing that, as sociolinguistics and also most teachers well know, this aim is far more likely to be achieved from a starting point of the study of linguistic diversity and the nature of spoken English.

tain about how to teach a course that is so different from their previous experience' (Bleiman and Webster 2006: 29). A further serious problem at the time when we began our project was a lack of classroom materials on spoken English. This meant that hard-pressed teachers had to make time to locate suitable audio or video clips to use in class to prepare their students for the examinations. Once found, transcribing spoken or visual material for classroom use is equally time- consuming, and not easy for teachers with little or no previous experience. The success of the GCE A-level English Language examination was said in 2006 to rely on a small energetic group of teachers who found their own materials on the Internet or who drew on the media and were happy to share their resources with others (Bleiman and Webster 2006: 29). For many teachers, though, this degree of commitment was likely to be daunting and too time-consuming.

Despite the difficulties of teaching GCE English Language, there are good social reasons why it is worth encouraging its take-up in schools. The subject is especially popular in schools in multicultural urban areas, where the focus on sociolinguistic variation and language and identity appeals to students from minority ethnic backgrounds since their own linguistic experiences are relevant, valued and analysed during their studies. It positions speakers of nonstandard varieties to acquire Standard English more effectively through learning about the social diversity of English (Wolfram 1998: 182). The subject allows less able and average students to achieve as well as the most gifted (Bleiman and Webster 2006: 29), so it provides opportunities for social mobility and for improving the skills base of UK society, with significant benefits for both individuals and society. Furthermore, the subject is 'boy-friendly': 6 per cent of all male and 8 per cent of all female A-level entries are for English Language, compared to 14 and 26 per cent for English Literature (Vidal Rodeiro 2006: Table 3). Boys prefer GCE A-level English Language to other language subjects, so it gives them the opportunity to catch up with girls in the acquisition of language skills.

It was against this background that we decided to produce resources for teachers to use in the classroom based on our recent research on the English of young people in London. This was one of several initiatives that were developed at the time in response to the situation just described.

Uncorrected Proof

Others include British Telecom's All Talk: English 14–19; the British Library Sounds Familiar web pages; The Talk of the Toon (Corrigan et al. 2012) web resource, described elsewhere in this volume, and more.

2 The London Projects

The research project Linguistic Innovators: The English of Adolescents in London (Kerswill et al. 2004–2007) was the first large-scale sociolinguistic study of English in London, testing the view that long-standing migration patterns make London the origin of language changes in the UK and beyond. It resulted in a transcribed corpus from audio recordings of 1.4 million words from indigenous Londoners aged 70+ and adolescents aged 16–19 from many different ethnic groups. We found that, contrary to thinking at the time, London is actually not the source of the language changes underway in many UK urban centres. Instead, young people in the multicultural inner-city area used a repertoire of innovative features, in all components of language. We refer to this way of speaking as Multicultural London English (MLE) and argue that MLE has replaced 'Cockney' (see also Fox 2015). Young people from immigrant backgrounds led in the use of MLE features, but white speakers from long-standing 'Cockney' families also used them.

Our second London project, Multicultural London English (Kerswill et al. 2007–2010), aimed to establish how MLE arose. A further 2 million words were recorded and transcribed from speakers aged 4, 8, 12, 16–19, 25 and 40, including parents and caregivers of 12 of the younger children. We found that MLE was well established among the youngest children, suggesting that they acquired it from peers and siblings, not their parents (who were mostly non-native speakers of English). We concluded that children in multilingual areas of London acquire combinations of language features from a rich 'feature pool' of linguistic forms influenced by a wide variety of languages, dialects and learner varieties. The pool serves as a resource and a model for non-native speakers acquiring English where there is no consistent target variety. This is a new dynamic of change affecting a metropolis containing a large minority ethnic and/or immigrant population, with strong implications for

Uncorrected Proof

our understandings of processes of language change; see, for example, Cheshire et al. (2011a, b). [AU1]

Although the research had mainly theoretical objectives, there were also some educational aims. As part of the project work, Susan Fox participated in three Knowledge Exchange workshops with teachers and students of GCE A-level English Language, where sound clips from the project recordings were discussed. Teachers were keen to use these recordings in their classes, because of the difficulties mentioned earlier of obtaining appropriate spoken English material. We learned that they were likewise interested in keeping abreast of relevant current research in linguistics, but that this was not easy because of a lack of time and the difficulties of accessing journals. This encouraged us to successfully seek further funding to develop follow-on resources from the two London projects (Cheshire et al. 2011a, b).

3 The English Language Teaching Resources Archive

We began the follow-on work by setting up an advisory panel of 14 teachers from the north and south of England. Two of the teachers were also experienced examiners for GCE A-level English Language; one of the two also wrote textbooks for both GCE A-level and GCSE English Language courses. The panel advised us at every stage of our work to ensure maximal relevance to the needs and interest of classroom teachers; they also piloted the resources as they were developed. With their guidance, we produced a website containing three types of resource: a Databank of sound clips and accompanying transcripts; a Linguistics Research Digest; and a set of Language Investigations, all interlinked as shown in Fig. 10.1 and as we explain further below.

3.1 A Databank of Spoken English

The Databank contains 12 folders, each with a sound clip taken from the recordings made for the two London projects, a written transcript of the clip, and a set of discussion points. The speakers are from a range

Uncorrected Proof

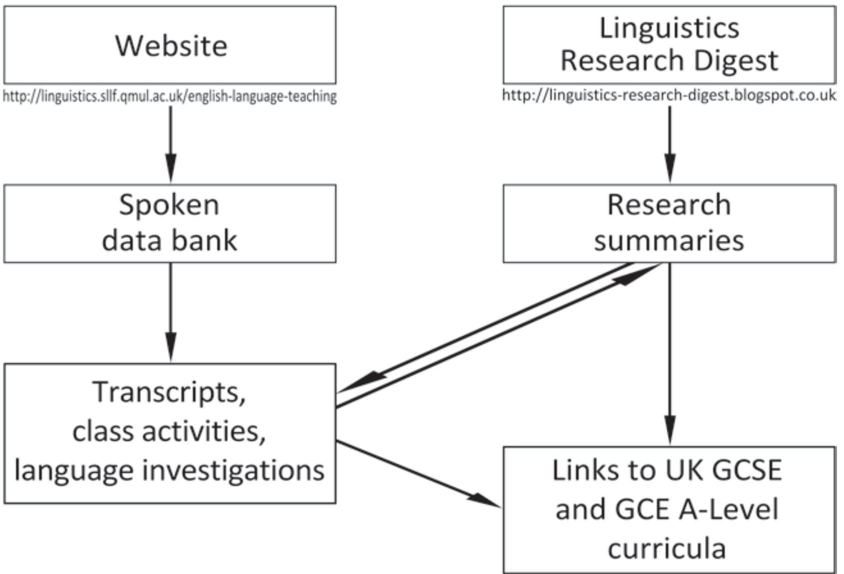


Fig. 10.1 The English language teaching resources archive

of ethnic backgrounds, both male and female, and include 8-year-olds, adolescents and a 77-year-old man. Table 10.1 details the titles we give to the extracts, the speakers (using pseudonyms) and their ages and backgrounds.

The design of the Databank follows the recommendations of Reaser and Adger (2007), which are based on many years' experience of working with teachers. Reaser and Adger point out that materials with descriptions of specific language features are useful resources, and that since teachers are the experts in what is required for the classroom, it is important to provide background information and suggestions so that they can decide for themselves which parts of the materials to use. They caution that teaching materials produced by linguists must be consistent with the goals of the curriculum. We therefore chose sound clips that illustrate specific aspects of spoken English, such as interruption and overlap (Competing for the floor), quotative expressions

Table 10.1 Folders in the Databank of spoken English

Title of extract	Details of speakers
Food stories	Dafne (female, age 8, Nigerian family) and Nandita (age 8, Bangladeshi family)
Competing stories	Derya (female, age 8, Turkish family) and Kareen (female, age 8, Indian family)
The dog story	Howard (male, age 8, white British) and Junior (age 8, Afro-Caribbean)
Competing for the floor	Lydia and Louise (both are female, age 8, white British)
Dressing up	Madeleine (female, age 8, mixed race white British/ Afro-Caribbean)
At the airport	Alex (male, age 16, mixed race white British/ Afro-Caribbean)
Street trouble	Angela (female, age 16, mixed race white British/ Afro-Caribbean)
How Courtney met her boyfriend	Courtney (female, age 17, and Aimee, age 19, both from Jamaican families)
Walking home from cadets	Tina (female, age 18, mixed race white British/Indian)
Problem at college	Laura (female, age 19, white British)
The bike incident	Zack (male, age 16, white British)
Life in the army	Stan (male, age 77, white British)

(The bike incident) or filled and unfilled pauses (Life in the army, and several of the other clips). Some extracts illustrate MLE, as a social and geographic variety of English. The discussion points for each transcript and sound clip draw attention to relevant linguistic features in the clip; and they include most of the linguistic features mentioned in GCE English Language textbooks. We use the linguistic terminology recommended by the Department for Education in the specifications for the English Language GCE examination, plus some additional conventional terminology when needed, and there is a separate glossary in the Databank explaining all the terms.

Example (1) gives, as an illustration of our approach, the introduction to the sound clip and transcript from the Life in the army folder. Examples (2) and (3) show extracts from the discussion points for this folder; line numbers in (2) and (3) refer to the numbered lines in the download version of the transcript.

Uncorrected Proof

Uncorrected Proof

170 1. Stan is 77 years old and lives in Havering, Essex. In this extract he
171 reminisces about his army days. It could be interesting to consider
172 which features of Stan's speech mark him out as a member of an older
173 generation: for example, would a young person use the colloquial
174 words and expressions *chap* (lines 6, 75), *a good half hour* (line 25), *a*
175 *blind bit of notice* (line 31), *blimey* (line 44) or *a great big fat corporal*
176 (line 18)? Stan's hedges, discourse markers and quotative expressions
177 are also more typical of an older speaker. Much of the impact of the
178 story comes from what Stan said to his superior officers and what they
179 said to him, so there is a lot of reported direct speech and reported
180 thought in his story.

181 2. ***er* and *erm* (filled pauses)**
182 These nearly always occur at the beginning of a clause, indicating that
183 Stan wants to keep the floor while planning the grammatical structure
184 of what he is about to say (lines 2, 10, 11, 18, 25, 29, 77). Sometimes
185 there is a silent pause too (lines 11, 25, 29). Stan mainly says *er* (*erm*
186 occurs only once), in line with attested gender differences in the use of
187 *er* and *erm*.

188 3. **Hedges**
189 *Sort of* (lines 21, before the verb *scratched away* and 30, before the verb
190 *walk about*) involves the listener by signalling imprecision—'scratched
191 away' may not be the best way to describe the sound of an old gramo-
192 phone, and the impression Stan gives of the way he walked about (line
193 30) may not be exactly right. *About* (line 75) signals that 30 is an approx-
194 imate number. Note that although young people use *sort of* and *about*
195 too, in these contexts they may be more likely to use *like*.

196 3.2 Ethical Considerations

197 The importance of using research to benefit society has long been recognized
198 by sociolinguists and has been especially pioneered by the two leading US
199 sociolinguists William Labov and Walt Wolfram. Labov's (1982) 'princi-
200 ple of debt incurred' and Wolfram's (1993) 'principle of linguistic gratuity'

both promote the notion that linguists should endeavour to make their research and information about language both available and accessible to the general public as described in some form or other in all contributions to this volume.

The work we describe in this chapter is in line with these principles. It is clearly worthwhile to make corpora such as those compiled from the London projects available for other researchers to use and, equally, to adapt them and make relevant information available to a wider general public. However, there are many ethical issues to take into consideration before allowing access to the data.

First of all, let us consider access to the transcripts. When collecting data during any sociolinguistic study, the participants are guaranteed anonymity and confidentiality of subject matter discussed during the interview. In order to make the transcripts accessible to other researchers or if we want to use examples from the transcripts when explaining concepts to non-linguists, they must therefore be anonymized in order to protect the identity of the speaker, but the question is how far do we need to go in order to ensure anonymity? At the community level, for example, where do we draw the line? In our data, we left in such references as 'Hackney' and 'Havering' (the two London boroughs where the research was conducted). We felt that it was necessary to keep these references in the transcripts because our results cannot be generalized to other areas of London where, perhaps, the demographics differ. The names of London areas such as 'Romford', 'Islington' and 'Wood Green' were also left in the transcripts, as were names of streets if they were used in a general sense as in the examples *I buy my jeans in Mare Street* or *I used to work in a bar down near Liverpool Street.*

Many young people in London refer to the area that they live in by their postcode as in the example *I'm from E8* and these were generally left in the transcripts unless they were mentioned with more specific addresses. However, we anonymized the names of streets when they were used in a more specific sense as in *I live in (name of street).* We also anonymized names of places where they could be used to track down an individual as in the example *if you play football with us yeh over (name of park).* We also left out names of schools if a speaker said the name of the school that they had attended or where a general reference to a school could lead to the

Uncorrected Proof

237 identity of an individual as in the example *some white girl from your area*
238 *…. she goes (name of school) she knows (name of girl).*
239 As far as individuals were concerned, we anonymized the names of
240 the speakers and the names of any individuals that were mentioned—
241 teachers' names for instance. Any obvious private information (such as
242 telephone numbers, addresses or names of specific clubs attended) was
243 also removed. However, is this enough to maintain the level of confi-
244 dentiality and anonymity expected of ethical sociolinguistic researchers?
245 Other issues of confidentiality might revolve around such things as sexual
246 orientation or even sexual activity, highly personal and intimate topics
247 that occur fairly frequently in our data. Should this kind of 'personal'
248 information be removed? In our data (which is currently only available to
249 other researchers) we retained this kind of information, provided we were
250 confident that the speaker was sufficiently anonymous. Similarly, dates of
251 birth were retained, although such decisions would perhaps be different
252 if the transcripts were to be made available to a wider public.
253 Then there is the consideration of other, more 'public' individuals
254 such as celebrities and TV personalities—is it ethical to leave these in the
255 transcripts or do researchers have a responsibility to also protect public
256 figures if disparaging comments or allegations are made about them? The
257 same applies to more 'local' celebrities, for example, the names of locally
258 known music artists (again, particularly when derogatory remarks are
259 made about them)? Where do we draw the line?
260 In a recent volume of *Corpus Linguistics and Linguistic Theory*, Childs
261 et al. (2011) discuss the view that you cannot fully separate the per-
262 son from the conversation because an interview is inherently personal
263 and people are always invoking their personal experiences and personal
264 opinions even if nothing especially 'personal' (or topics that might be
265 considered more private) comes up. With this in mind, is it therefore
266 even possible to achieve complete anonymity in the transcripts? Even if
267 we take the precautions already discussed above (deleting names, private
268 information and so on) we still find examples such as the following:

269 4. Interviewer: where did you go to school?
270 Grant: oh er (name of school). in Archway
271 Interviewer: is that a Catholic school?
272 Grant: yeh

Uncorrected Proof

As we can see in this example, the name of the school has been deleted. However, the fact that the general area of 'Archway' has been left in and the fact that the school is 'Catholic' actually narrows this down to one easily traceable school. Coupled with other more personal information in the transcript about the speaker's age, ethnic background, details about parents and siblings, it becomes possible that the speaker could be traced and (in a worst-case scenario) some type of harm could come to the speaker as a result of taking part in the study. In our London data we often get references to local gangs, both in the sense of in-group and out-group membership. If they have been referred to in a general sense in the two London corpora then they have, on the whole, been left in but we might see a potential for harm if the corpus is more widely available (to journalists, for example).

Then we find more general descriptions such as *that Chinese up the road down erm near the job centre on Mare Street*, referring to a specific Chinese restaurant. Should these references also be removed or does it depend on the context in which the reference occurs? Questions of this nature then feed into whether it is possible to have an 'objective' anonymization protocol or whether these decisions are always going to be subjective. It is clear that different researchers and transcribers will respond differently to these issues depending on the type of community they are researching and the content of the subject matter contained within their recordings. Matsumoto (2015), for example, in her work on Palauan English, has highlighted how different cultures might consider certain issues as being sensitive, of which the researcher or transcribers might be unaware. Quoting one of her participants, Matsumoto (2015) reports:

> When I went to University X in the US and found out how my relatives were quoted in theses, I was really in shock. You know, there're things I swear by God they would never say openly if they'd known their words would be published with their own names. You know, Americans would've thought that we'd never read their theses.

While this quote clearly highlights the need for anonymity, Matsumoto (2015) also notes that researchers need to be aware of 'the existence of locally established beliefs and taboos'. The difficulty that arises here is that the people transcribing the data are often not the researchers them-

Uncorrected Proof

308 selves and these kinds of sensitive issues can easily be overlooked dur-
309 ing the transcribing process. Even with the guarantee of anonymity, one
310 might question whether participants would be willing to discuss certain
311 sensitive topics if they were aware that their words might appear in print
312 somewhere in the public domain.

313 This leads to another important issue: the extent to which the consent
314 form completed by the participants in the projects covers the uses to which
315 the transcripts/recordings will be put. Included here is the situation with
316 regard to children. Children under the age of 16 take part in our projects,
317 with parental consent. However, is this enough? Do researchers then have
318 the right to make the transcripts/recordings of these children available to
319 others? In the consent form for the Linguistic Innovators project we state
320 that the recordings will be used for teaching and 'research' purposes only.
321 The issue here is what is meant by 'research'? We feel sure that most of our
322 participants would accept the use of their recordings and their respective
323 transcripts within the more narrow field of scholarly academic research,
324 but can we be sure that they would be equally accepting of 'research' in a
325 wider sense, such as the type of research carried out by other education-
326 alists or journalists and the media in general if the transcripts/record-
327 ings became more widely available? Furthermore, if the materials become
328 available to the general public, can this still be considered 'research'? In
329 our second project, *Multicultural* London English, the consent form was
330 extended to include permission to use extracts in 'broadcasting' (mainly
331 in response to the many requests from the media to provide sound clips
332 which we were not able to do for the Linguistic Innovators project). We
333 also made it clear to the participants that the anonymized transcripts
334 would be kept in an archive for other researchers to use.

335 Thus far, our discussion has been restricted to the use of the transcripts,
336 but what of the audio recordings? Currently, these are only available to
337 the current research team and to a few members of the academic commu-
338 nity who are working with one or more members of the team. In general,
339 the audio recordings present more ethical concerns than the transcripts.

340 Firstly, we are concerned that the content of the recording, matched
341 with the voice of the speaker, may potentially lead to that speaker's identi-
342 fication, thereby breaching the confidentiality and anonymity guaranteed
343 to the participant. Many of our participants spoke about sensitive topics

Uncorrected Proof

on the understanding that the recording would not be heard by anyone outside of the research team. We have many instances of questions and comments such as *This is not going to be played to anyone, is it? Are you sure no-one will hear this?* or of one friend saying to another during the interview *don't worry, it's confidential* when a speaker might be hesitant about discussing a particularly sensitive or taboo topic or if the nature of the discussion involved gossip about another person. Do we withhold such recordings? Do we assume that others would accept the recordings becoming available just because they *do not* make such explicit remarks?

Our second concern relates to the content of the recordings, some of which is highly confidential, sensitive and, in some cases, incriminating (in fact, some of the recordings in the London projects were withheld by the fieldworker and not used for analysis at all because of the sensitive nature of the content). Again, there is the question of the recognition of the speaker's voice and the extent to which the data can be 'cleaned up'.

Anonymizing the audio files is labour-intensive and therefore time-consuming and costly. Given the time frame and financial constraints of many sociolinguistic projects this is not generally possible for large data sets. The question then arises of how this exercise would be funded and, perhaps more importantly, *who* would carry out this exercise and how much knowledge anonymizers would have about relevant local issues. Assuming that a number of different anonymizers would be working on a data set there is the potential for different levels of anonymity to be applied according to individual ideologies—some may have more relaxed views than others. While we would stress the need for an anonymization protocol there will always be grey areas subject to individual decisions.

Finally, once the corpus is in the public domain it then becomes subject to the ethical decisions of other researchers. We would expect the same stringent ethical considerations to be applied among all academic researchers, but there is still the concern that researchers may not be aware of relevant local issues and we almost certainly cannot guarantee the actions of a wider 'researching' public. Even within the academic community we have already found that researchers in other countries have not always been trained in research ethics and do not feel the same ethical responsibilities towards participants.

Uncorrected Proof

Having highlighted some of the problems that we consider to be of importance in the sharing of data, we nevertheless acknowledge that the benefits of sharing data among researchers and disseminating language information to a wider general public are extremely worthwhile. So far, we have deposited the Linguistic Innovators transcripts with the UK Data Service. The corpus (available as concordance-searchable text files) is available to researchers, teachers and students from any field, organization or country on registration with the UK Data Service. Some data sets have restricted access including, as yet, the MLE transcripts, which are available to other researchers only on request to a member of the research team.

Given the concerns raised in this chapter about the audio files, the sharing of this data remains at the discretion of the research team. We are still considering whether there is any way of making the audio files more widely available but have, so far, rejected the possibility of archiving the recordings with an organization such as the UK Data Service.

4 The Linguistics Research Digest

We designed the Linguistics Research Digest as a way of meeting teachers' desires to keep abreast of research in linguistics. We chose articles published in recent linguistics journals on topics relevant to the specifications for GCE A level and GCSE English Language, and summarized them in a way that aimed to be engaging and jargon-free as well as accurate. The teacher advisory group advised us on the choice of journal articles, the style of the summaries (in terms of their accessibility, interest and their form), and the frequency with which summaries should be posted. We posted the summaries on a blog site—two summaries a week during the life of the Follow-on project, in 2011, and one summary each week from 2012. Figure 10.2 is a screenshot illustrating the Digest.

At the end of the summary we provide the full bibliographic reference to the article, and there are hyperlinks in the summary itself to the web pages of the authors. As Fig. 10.2 shows, the side bars contain links to other language blogs and to sites relevant to the English Language curriculum, and a searchable tool for browsing the Digest by category.

Uncorrected Proof

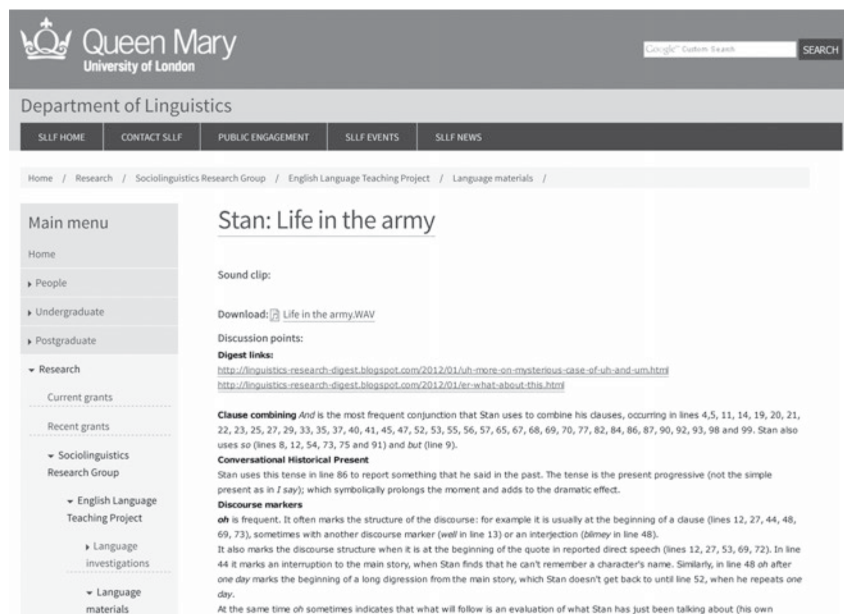
Fig. 10.2 The Linguistics Research Digest: first part of a summary

Uncorre

The transcripts and discussion points in the Databank contain links to summaries that relate to specific linguistic features. For example, the discussion points for the Life in the army folder include Stan's use of filled pauses [see example (1) above] and are linked to the Digest summary illustrated in Fig. 10.2. Figure 10.3 shows how this is done in the Databank folder.

5 Language Investigations

We had intended to provide activity sheets for use in the classroom with the Databank, but the teacher advisory panel found that the sound clips and discussion points provided enough material for their lessons. They



Queen Mary
University of London

SEARCH

Department of Linguistics

SLLF HOME | CONTACT SLLF | PUBLIC ENGAGEMENT | SLLF EVENTS | SLLF NEWS

Home / Research / Sociolinguistics Research Group / English Language Teaching Project / Language materials /

Main menu
Home
People
Undergraduate
Postgraduate
Research
Current grants
Recent grants
Sociolinguistics Research Group
English Language Teaching Project
Language investigations
Language materials

Stan: Life in the army

Sound clip:

Download: Life in the army.WAV

Discussion points:

Digest links:
http://linguistics-research-digest.blogspot.com/2012/01/uh-more-on-mysterious-case-of-uh-and-um.html
http://linguistics-research-digest.blogspot.com/2012/01/er-what-about-this.html

Clause combining *And* is the most frequent conjunction that Stan uses to combine his clauses, occurring in lines 4,5, 11, 14, 19, 20, 21, 22, 23, 25, 27, 29, 33, 35, 37, 40, 41, 45, 47, 52, 53, 55, 56, 57, 65, 67, 68, 69, 70, 77, 82, 84, 86, 87, 90, 92, 93, 98 and 99. Stan also uses *so* (lines 8, 12, 54, 73, 75 and 91) and *but* (line 9).

Conversational Historical Present
Stan uses this tense in line 86 to report something that he said in the past. The tense is the present progressive (not the simple present as in *I say*); which symbolically prolongs the moment and adds to the dramatic effect.

Discourse markers
oh is frequent. It often marks the structure of the discourse: for example it is usually at the beginning of a clause (lines 12, 27, 44, 48, 69, 73), sometimes with another discourse marker (*well* in line 13) or an interjection (*blimey* in line 48).
It also marks the discourse structure when it is at the beginning of the quote in reported direct speech (lines 12, 27, 53, 69, 72). In line 44 it marks an interruption to the main story, when Stan finds that he can't remember a character's name. Similarly, in line 48 *oh* after *one day* marks the beginning of a long digression from the main story, which Stan doesn't get back to until line 52, when he repeats *one day*.
At the same time *oh* sometimes indicates that what will follow is an evaluation of what Stan has just been talking about (his own

Fig. 10.3 First page of one of the Databank documents

Uncorrected

suggested that instead we developed some Language Investigation tasks. In 2011, when the website was set up, students were required to carry out a small-scale piece of research themselves—a Language Investigation. Although for A-level examinations from May/June 2017 the methods of assessment will change to largely formal written examinations, 20 per cent of the final grade will still be obtained from non-examination based assessment of 'Language in Action'. This part of the curriculum aims 'to allow students to explore and analyse language independently and develop and reflect upon their own writing expertise' (AQA AS and A-level specifications 2014: 18). The latter aim requires students to produce a 1500 word piece of original writing and commentary; the former requires them to produce a 2000 word report (excluding data) of a Language Investigation that they have carried out themselves.

So far we have produced seven Language Investigations, all linked both to and from specific summaries in the Linguistics Research Digest. The investigations were piloted by the advisory group and then revised to

take account of feedback from teachers and their students. The Language Investigations give clear directions on data collection and suggestions for how students can analyse the data. The topics are: giving place directions; speech style in call centres; language brokering; compliments; intensifiers; general extenders; and second person plural forms. As an illustration, Fig. 10.4 shows the Language Investigation on the latter.

6 Use of the Resources

Between January 2013 and 31 July 2013 the Archive received 18,000–20,000 visits per month, of which 60 per cent were return visits; in the same period the Digest received 7000–8000 hits per month. Google Analytics showed that approximately 60 per cent of the visitors to the Archive were from the UK, but that others originated in more than 30 different countries, from all continents. Since its inception, the Digest has attracted over 300,000 hits, with the highest number of visitors coming from the USA, closely followed by the UK and then, in descending order among the top ten countries, France, Germany, Russia, Ukraine, India, Australia, Canada and China. Websites set up by schoolteachers recommend the resources: for example, one teacher writes 'There is a fantastic blog produced by Queen Mary University of London's Linguistics Department. They have a real commitment to encouraging A-level students in their study of the English language … even better, they have come up with some possible A2 level investigations and for some, they even suggest a methodology and research question … they have even provided access to a whole data bank of spoken contemporary London English … this is a fantastic opportunity.'[3] We posted an online survey in June 2013 asking for feedback on the Resources Archive. The following quote is typical of the feedback we received from the survey: 'The audio and transcripts have been invaluable in helping me prepare students for the exams and coursework, and the glossary of terms has always been a handy reference point.' We discovered that examiners used the Linguistics

[3] See http://eastnorfolklanguage.blogspot.fr/2013/06/great-leads-for-possible-investigations.html (accessed 8 August 2015).

Uncorrected Proof

Language Investigations in spoken English

"Do you understand who I'm talking to?" Second person plural forms in English

Most languages have separate words for singular and plural pronouns. English used to have separate second person pronouns too, but since *thou* fell out of use the *you* pronoun has had to do double duty. So, in the scene depicted here, is the speaker accusing one of his friends, or all of them? How do we deal with this problem?

you took my biscuit!

You could investigate how English speakers make clear who they're talking to when there is more than one person around. How do they show that they are speaking to just one person? Or to two people? Or to the whole group?

How to investigate?

Listen and note

One way to find out is to listen to what your teachers say when they are addressing one person, the whole class, or a small group of students. Note this down during the course of a day, so that you end up with a collection of phrases. Perhaps the teacher uses the student's name as well as *you* (but this obviously wouldn't be possible for a group

Fig. 10.4 Example of a Language Investigation

of students!) You'll probably find that the phrases include *you all* (for example, *will you all now think about this?*) and *the two of you* or *both of you* (for example, *would the two of you do this?*) What other phrases does your teacher use to show who he or she is addressing? Does the teacher sometimes make it clear through eye contact, or pointing?

Some linguists* have claimed that we have an unconscious rule about how to address two or more people: if it isn't clear from the context, the speaker must make it perfectly clear whether they are referring to one person, to everyone who is there or to a subset of the people who are there. They usually do this by using people's names, or a phrase like *all of you*, *you fellows*, *both of you*, or by gesturing (usually pointing). Once this has been made clear, it is OK to use *you* from then on, but only until the next ambiguous moment in the conversation. If someone joins the group or if someone leaves, the speaker has to make it clear all over again just who they are talking to.

[* Andrew Pawley and Frances Hodgetts Syder (1983) Natural selection in syntax: notes on adaptive variation and change in vernacular and literary grammar. *Journal of Pragmatics* 7: 551-579.]

Watch TV

Researchers have found that in the *Friends* series, the speakers often use *you guys* when they are addressing more than one person. You could watch an episode of *Friends* and note down all the words and phrases used when people address more than one person. How often do speakers say *you guys*? Are there any other words or phrases that they use to make it clear whether they are talking to one person or more than one person? Is it always clear what *you* means?

Perhaps more interestingly, watch a British TV sitcom where people sometimes address more than one person (such as *Big Brother*). Do people use *you guys* here? If not, how do they make it clear who they are addressing?

Do some dialect research

Fig. 10.4 (continued)

Many varieties of English have a separate second person plural pronoun, unlike standard English. There are many different forms, including *youse*, *you all*, *yinz* or *you uns* (and more). You could browse the internet or look at some Linguistics textbooks to gather examples. Find as many second plural pronoun forms as you can, and note down in which parts of the world they are heard.

TIP:

Try googling "second person pronouns" and national varieties of English around the world, such as "Irish English", "America", "Australia", "Jamaica" or "South Africa". You could also search for second person pronouns in regional varieties (dialects) of British and American English.

You could also try to find out what has happened to the old singular pronoun *thou*. Is it still used? If so, where?

In conclusion

Once you've done your investigation, consider whether *you* in English is really ambiguous. Do people really not know who is being addressed when they hear *you*? Or do they find other ways of showing that they are addressing more than one person?

Suggested Reading:

Theresa Heyd (2010) How you guys doin'? Staged orality and emerging plural address in the television series *Friends*. *American Speech* 85 (1): 33-66. (Click here for a summary of this paper).

Fig. 10.4 (continued)

Research Digest with teachers at training sessions 'to encourage them to find new material to point students towards and it's always gone down very well'.

We learned, too, that the resources were useful for teachers and students of EFL/ESL and university-level English Language in the UK and beyond. A typical response from the survey is the following: 'I teach linguistics, working with prospective teachers—and practicing teachers—in the USA. This is a terrific site for them to look at … and to talk about the phenomena here in the context of analogous topics in varieties of American English they know and are likely to encounter.'

7 Workshops for English Language Teachers: Analysing Spoken English

A further spin-off of our project for producing resources for teachers was the organization, in collaboration with Dr Heike Pichler (Newcastle University), of workshops for teachers of English Language. The workshops, held in April 2012 at the University of Salford, July 2012 at Queen Mary, University of London and December 2012 at Newcastle University, aimed to disseminate insights from scholarly research about language variation and change and to provide teachers with an overview of databanks and resources available online for use in the classroom. These one-day workshops consisted of talks from the organizers and other invited researchers and consisted of two parts.

In the first part of the workshop we aimed to break down persisting prejudices against the use and users of discourse-pragmatic features such as *innit* (as in *It's only an hour from Edinburgh and Newcastle, innit?—Oh, I've answered this one before, innit?*), quotative forms such as *be like* (for example, *And they were like, 'we divn't want you here'. And we were like, 'why?'*) or intensifiers such as *dead* (as in *It was dead funny*). The aim was to demonstrate that it is wrong to dismiss these features as mere fillers which contribute nothing to the content or communicative force of an utterance or, even worse, to perceive them as a sign of inarticulateness, laziness or lack of intelligence. We aimed to break down existing preju-

Uncorrected Proof

dices against the use of such features and to demonstrate how these features develop, what communicative function they perform in interaction (such as to signal tentativeness or assertiveness, or to facilitate speaker change), and how they change over time. We thereby hoped to raise participants' awareness of discourse-pragmatic features and to demonstrate that they play a vital role in interaction.

In the second part of the workshop, we provided teachers with an overview of currently available resources for working with spoken data in the classroom, focusing in particular on a demonstration of two projects specifically aimed at providing teachers with relevant resources. The first, the Diachronic Electronic Corpus of Tyneside English (DECTE) is a corpus of spoken language from north-eastern England spanning five decades. The linked Talk of the Toon website developed by the same project team which is more fully described elsewhere in this volume is a multi-media publicly-available resource containing audio recordings and transcriptions as well as still and moving images relating to themes relevant to subject areas in the National Curriculum. The second project is the one discussed in this chapter, The English Language Teaching Resources Archive, which focuses on London English and had the aim of developing accessible classroom materials arising from sociolinguistic research on spoken language.

We also compiled a Resource Booklet for the teachers, which contained factsheets summarizing relevant insights from current research on the selected topics as well as relevant scholarly articles that teachers would find useful. The materials included suggestions for classroom activities that would enhance students' theoretical knowledge about spoken language and language variation and change. We provided answers and commentaries to these activities. We also included photocopiable worksheets for practical investigations into spoken language, similar in nature to the Language Investigations described earlier.

The workshops were very well attended, with around 30 participants at each one. At the end of each workshop we asked participants to complete an evaluation questionnaire. The value of such workshops for teachers is captured in some of the comments they made on the questionnaires, examples of which are given below:

Uncorrected Proof

5. The presentations had a very clear sense of the needs and level of understanding of the audience and was [*sic*] consequently very accessible and engaging. Also—a very coherent focus to the whole event.
6. Very interesting to hear about recent research. Made relevant to our A-level teaching context. Excellent resource pack.
7. Up-to-date information/details about recent research. The website resources look really useful especially the recordings/transcripts/ condensed research data.

Overwhelmingly, many of the teachers asked for further sessions to be organized on the same and other topics. They asked for the workshops to be provided in different locations around the UK and there were many requests for similar workshops to be made available to students of A-level English.

8 The Future

As we mentioned earlier, the English Language Teaching Resources Archive was developed with the help of external research funding for one year. Further funding is now needed to develop the resources further. For example, we would like to extend the Databank so that it contains sound clips and accompanying transcripts from other regional or indeed national varieties of English, and we would like to post additional language investigations based on articles summarized in the Research Digest. Even without this, however, the existing Databank remains as an online resource, and it continues to be used.

Of course, lack of funding not only hinders development of the resources; it could also mean that it is difficult to sustain the website and to deal with any technological problems that may arise. However, UK higher education institutions are subject to a national assessment of the quality of the research carried out by their academic staff, the results of which determine part of the government funding given to each institution. The most recent assessment system, the Research Excellence Framework (REF), includes

Uncorrected Proof

567 assessment of the impact of research.[4] As a result, many universities have
568 now set aside a budget to encourage public engagement with the research of
569 their staff, and to ensure that relevant 'users' are able to access the research.
570 The Databank of spoken English has benefited from this as the university
571 where the resources were developed (Queen Mary, University of London)
572 was willing to provide the financial resources and the person power for the
573 website to be made part of the web page of the Department of Linguistics.
574 This means that it can be maintained along with the department's web
575 pages, and its future is assured.

576 We have been fortunate in obtaining financial sponsorship for the
577 Linguistics Research Digest from the Linguistics Association for Great
578 Britain and the British Association for Applied Linguistics, so far every
579 year since 2012; and in 2012 the Archive won further financial support
580 from the British Association for Applied Linguistics' 'Applying Linguistics'
581 competition. We use the funds to recruit able graduate students, who try
582 their hand at writing summaries of relevant journal articles and thereby
583 gain experience of writing for a lay audience. We edit these summaries
584 and also write some ourselves but, as is the case for the teachers for whom
585 the Digest was developed, time and resources are in short supply. Our
586 digest was modelled on the British Psychological Society's very successful
587 Research Digest,[5] which is maintained with the support of a permanent
588 part-time post. We have not so far been able to secure a post of this kind
589 for work on the Digest, but we continue to explore possibilities.

[AU3] 590
References

591
Books and Articles

592 AQA Education. 2014. *AQA AS and A-level English Language Specifications*.
593 Version 1.0, 14 October 2014. Manchester: AQA Education.
594 Bleiman, Barbara, and Lucy Webster. 2006. *English at A level: A Guide for*
595 *Lecturers in Higher Education. A Report to the Higher Education Academy*

[4] For further details, see the REF website at http://www.ref.ac.uk.

[5] See http://digest.bps.org.uk.

Uncorrected Proof

English Subject Centre. London: English and Media Centre Report Series no. 12.

Cheshire, Jenny, Susan Fox, and Paul Kerswill. Jan–Dec 2011a. *From Sociolinguistic Research to English Language Teaching.* UK Economic and Social Research Council project RES 189-25-0181.

Cheshire, Jenny, Susan Fox, Paul Kerswill, and Eivind Torgersen. 2011b. Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics* 15(2): 151–196.

Childs, Becky, Gerard Van Herk, and Jennifer Thorburn. 2011. Safe harbour: Ethics and accessibility in sociolinguistic corpus building. *Corpus Linguistics and Linguistic Theory* 7(1): 163–180.

Department for Education. 2013. *English Language GCSE Subject Content and Assessment Objectives*. London: Department for Education DFE-002320-2013. Available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254497/GCSE_English_language.pdf (accessed 9 June 2015).

———. 2014. *GCE AS and A level Subject Content for English Language*. London: Department for Education, DFE-00362-2014. Available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/302109/A_level_English_language_subject_content.pdf (accessed 9 June 2015).

Fox, Susan. 2015. *The New Cockney: New Ethnicities and Adolescent Speech in the Traditional East End of London*. Basingstoke: Palgrave Macmillan.

Kerswill, Paul, Jenny Cheshire, Susan Fox, and Eivind Torgersen. 2004–2007. *Linguistic Innovators: The English of Adolescents in London*. UK Economic and Social Research Council project RES 000 23 0680.

———. 2007–2010. *Multicultural London English.* UK Economic and Social Research Council project RES 062 23 0814.

Labov, William. 1982. Objectivity and commitment in linguistic science. *Language in Society* 11(2): 165–201.

Matsumoto, Kazuko. 2015. Beware that the community changes: Issues of the returns for linguistic favours. *Doing Research within Communities: Stories and Lessons from Language and Education Fieldwork*. Kerry Taylor-Leech and Donna Starks (eds). London: Routledge.

Reaser, Jeffrey, and Caroline Temple Adger. 2007. Developing language awareness materials for non-linguists: Lessons learned from the *Do You Speak American?* Project. *Language and Linguistic Compass* 1(3): 155–167.

Vidal Rodeiro, and Carmen L. 2006. *Uptake of GCE A-level Subjects in England 2001–2005.* Statistics Report Series no. 3. Cambridge: Cambridge

Uncorrected Proof

635 Assessment, Local Examinations Syndicate, University of Cambridge.
636 Available at http://www.cambridgeassessment.org.uk/Images/109897-
637 uptake-of-gce-a-level-subjects-in-england-2001-2005.pdf (accessed 8 August
638 2015).
639 Wolfram, Walt. 1993. Ethical considerations in language awareness programs.
640 *Issues in Applied Linguistics* 4(2): 225–255.
641 ———. 1998. Dialect awareness and the study of language. In *Students as*
642 *Researchers of Culture and Language*, eds. Ann Egan-Robertson, and David
643 Bloome, 167–190. Cresskill, NJ: Hampton Press.

644 Websites and Online Resources

645 British Library. *Sounds Familiar*. http://www.bl.uk/learning/langlit/sounds.
646 (accessed 8 August 2015).
647 British Telecom. *All Talk: English 14–19*. http://www.btplc.com/BetterFuture/
648 ConnectedSociety/LearningAndSkillsFreeResources/AllTalk (accessed 8
649 August 2015).
650 Cheshire, Jenny, and Susan Fox. 2011a. *English Language Teaching Resources*
651 *Archive*. http://linguistics.sllf.qmul.ac.uk/english-language-teaching
652 (accessed 8 August 2015).
653 ———. 2011b. *Linguistics Research Digest*.http://linguistics-research-digest.
654 blogspot.co.uk (accessed 8 August 2015).
655 Corrigan, Karen P., Isabelle Buchstaller, Adam Mearns and Hermann Moisl.
656 2012. *The Talk of the Toon*. Newcastle University. http://research.ncl.ac.uk/
657 decte/toon (accessed 8 August 2015).
658 REF. 2014. Research Excellence Framework. http://www.ref.ac.uk (accessed 8
659 August 2015).
660 *The British Psychological Society Research Digest*. http://digest.bps.org.uk (accessed
661 8 August 2015).

Uncorrected Proof

Author Queries

Chapter No.: 10 0002695185

Queries	Details Required	Author's Response
AU1	The reference "Cheshire et al. (2011)" has been changed to "Cheshire et al. (2011a, b)" here and in subsequent occurrence as per the reference list. Please check.	
AU2	The references "Cheshire and Fox (2011a, b)", were not cited anywhere in the text. Please provide a citation. Alternatively, delete the items from the list.	

Uncorrected Proof