A METHOD OF DATA COLLECTION FOR REGIONAL AND SOCIAL LANGUAGE VARIATION STUDIES

CARMEN LLAMAS

Abstract

This paper presents an innovative method of data elicitation for use in both large-scale regional language variation studies, and for use in studies of socially-stratified variation in a particular variety. The methodology combines the elicitation of data for various levels of linguistic analysis simultaneously as well as allowing for the collection of metalinguistic comment and attitudinal data. It was originally devised and designed to fit the requirements of a proposed national collaborative venture, the Survey of Regional English (SuRE). It was then expanded for use in a sociolinguistic study of Middleborough English undertaken by the author. It has since been used in adapted form for various studies of language variation and change in British English of both the in-depth single variety type and the large-scale survey type.

1. Introduction

Although the field methods involved in a sociolinguistic study are often not reported in great detail, many and varied methods of eliciting data for analysis of language variation exist (see for example Crawford (2002), Milroy and Gordon (2003), Llamas (2007a); see further discussion). The direct comparison of findings of different studies is therefore potentially problematic as individual projects have different aims and employ differing methodologies. Studies which utilize the same field methods and techniques of data elicitation not only share a transparent, replicable method but also produce data which are directly comparable. Such comparative potential is advantageous to obtaining an overview of, for example, processes of leveling and diffusion across a large geographical area.

In terms of British English, although a considerable amount is known about current phonological and grammatical variation and change, knowledge of contemporary regional and social lexical variation is sparse, with few studies being or having been undertaken in recent years. The studies which have been made generally utilize a similar method of data collection, namely the questionnaire. However, these studies, again, are not necessarily comparable, since different notion words are used or sought.

Few individual studies of a given variety attempt to combine investigation of social variation in spreading and localized features found in phonology, grammar and lexicon. This paper presents a methodology designed to allow the simultaneous elicitation of data for analysis on three levels. The core methodology was originally created for use in a proposed new survey of variation in the spoken English of the British Isles, the Survey of Regional English (SuRE) (Kerswill, Llamas and Upton 1999) and Upton and Llamas 1999). However, it can be, and has been used in individual studies of socially-stratified variation in a given variety, either in its core form or in an expanded or adapt form. This was demonstrated by the methodology used in the Middleborough study (Llamas 2001, 2006, 2007b). Furthermore, it is currently being used in various doctoral studies of variation in British English (see, for example, Arnaud forthcoming, Flumina forthcoming, Pringle forthcoming, Wallace forthcoming). It has been used in adapted form for the popular BBC Voices series on language variation throughout the British Isles.

2. Background

The larger picture of the concept of a planned large-scale survey of British English dictated the design of the data elicitation techniques presented in this paper. Some understanding, therefore, of the requirements of such a survey is necessary for an appreciation of the potential of the methodology.

2.1. A large-scale survey of British English: aims and difficulties

As the Survey of English Dialects (SED) (Orton and Ditch 1650-71), which was carried out in the 1950s, represents the only consistently-collected nation-wide survey of diatlectal variation in English, a deficiency exists in the knowledge and awareness of current variation in British English on a national scale. The basic intention of a large-scale survey project would be to create a computer-held database of consistently-collected material from a planned network of British localities which would record and document the facts of linguistic variation throughout Britain, permitting detailed analyses of issues concerning the diffusion of language change and the spread of current vernacular changes in British English. The form of the survey would be guided by the necessity for the primary data to be the object of analytical work addressing current research questions such as the extent and processes of accent and dialect levelling, for example. At the same time, its form must be sufficiently broad as not to prejudice the potential for analysis which addresses research questions arising in the future.

In order for such a project to obtain as complete a picture of regional language variation as possible, data must be obtained which can be analyzed on three levels of possible variation: phonological, grammatical and lexical. To discount any levels would be to obtain an incomplete picture of regional variation in spoken English found throughout Britain. These multi-levelled data must be comparable across the localities to be studied, permitting quantitative analyses of the different levels of regional and social variation where possible.

The primary aim of a methodology for such a project would be to obtain samples of informal speech from which analyses can be made at the phonological and the grammatical levels. As this is a fundamental requirement of any methodology for such a project, a problem lies in combining the level of comparable lexical variation with the necessity of obtaining casual speech, as to control the lexical items used in a conversation is to make the interaction less than spontaneous. This control can have the effect of formalizing the speech style, thus hindering the possibility of gaining access to the 'vernacular' or 'the style in which the minimum strain is given to the monitoring of speech' (Labov 1972: 208).
2.2. Previous methods and their applicability

As a means of eliciting data, the questionnaire has been employed in traditional dialectology since the nineteenth century and was the 'fundamental instrument' of the SED (Orton 1962: 15). Although it proves successful in eliciting lexical and some grammatical data, it is inappropriate for a survey whose intention is to access and collect samples of informal speech which are sizeable enough to allow phonological and grammatical analyses which permit quantification.

Additionally, the methods employed by the SED, and by other studies undertaken within the traditional dialectological paradigm, give scant information on language variation associated with social factors within a given variety, this not being the focus of interest of such research. Social variables, however, are central to current studies of language variation. As such, many more informants are required from each location than the two or three used in the SED. As a practical consideration, therefore, the methodology for a large-scale survey of current usage must be reasonably rapid and efficient, demanding considerably less of the informant's and the fieldworker's time than the lengthy SED questionnaire which contained nine books of questions each one taking at least two hours to complete (Orton 1962: 17).

Thus, methods which are associated with traditional dialectological studies of language variation are quite inappropriate to a large-scale survey of current usage.

However, methods used to obtain data for research undertaken within a quantitative paradigm are also unsuitable. Various attempts have been made to access the vernacular, or the informant's least overtly careful speech style, for example, the interview situation in which the fieldworker asks questions to elicit personal narratives (cf. Labov 1972, Trudgill 1974), allowing informants to converse in pairs on topics of their own choosing with minimal fieldworker involvement (cf. Docherty, Foulkes, Milroy, Milroy and Waisbrow 1997, Llamas 1998, Stuart-Smith 1999). Although successful in obtaining informal speech, these methods almost completely remove the possibility of obtaining comparable information on lexical variation. The ethnographic technique of participant observation, as used by, for example, Cheshire (1982) in Reading and Milroy (1987a) in Belfast, although successful in gaining quantitative and qualitative data, is also far too time-consuming for a collaborative large-scale projection.

Thus, because data are to be elicited rapidly and with ease, and because lexical data are to be included, which eliminates the option of 'free' conversation, an interview of some sort appears the only option. However, an innovative approach to the elicitation of lexical data than that of the traditional questionnaire is required as the interview must simultaneously elicit data which are analysable phonologically and also grammatically. As the data are to be quantifiable (where possible), comparable, analysable on at least three levels of variation and admissible to a relatively large number of informants, an innovative mode of data elicitation and collection is necessary, as no existing data elicitation technique is entirely suitable or applicable to the needs of a multi-layered, large-scale survey.

3. The data elicitation method: the core

3.1 Overall aims

The primary aims of the method are to obtain informal speech from the informant (from which multi-layered analyses of both regionally and socially comparable data are possible), and to elicit the data as efficiently as is possible. A methodology which is perceived to be too complicated or lengthy to administer may result in the unwillingness of potential fieldworkers to use it for their research purposes.

Although the sociolinguistic interview is regarded as the basic tool for recording conversation in sociolinguistic variation, according to Tagliamonte (2005: 37) it should be 'anything but an interview'. The interview as a speech event is not the ideal means through which to elicit casual conversation due to the 'asymmetrical distribution of power suggested by the roles of questioner and respondent', as Milroy (1987b: 49) notes, but it proves to be the only practical way of obtaining the necessary data. It is vital therefore to lessen the formality of the interview situation as much as possible, and to make the interview 'anything but an interview': rather an unintrusive and, if possible, enjoyable experience for the informant.

In order to obtain the required informal speech style combined with data on lexical variation in the interview, the fieldworker 'leads' a conversation around semantic fields. To lessen the formality of the interview context, the interview is undertaken with socially paired informants, permitting interaction to be more like a conversation than an interview. Discussion on local lexical items is prompted by the fieldworker, with informants encouraged to discuss 'local' words, how they are used and what connotations they have. As well as producing informal conversations from which phonological and grammatical analyses can be undertaken, the ensuing conversation produces a mass of information on the lexical data produced. This can include age and gender differences in usage, computational and collocational information, perceived social variation in usage, perceived etymologies, perceived geographical distribution of usage, knowledge and use differentiation and attitudinal information on dialect communities.

Although the method of discussing lexical items in pairs produces the sample of informal speech for analysis, control must still be exercised over the specific lexical items elicited in order for direct comparisons of variants to be made.

3.2 Sense Relation Network sheet

The principal tool devised and designed to allow the information on lexical items to be comparable regionally and socially, and to give a somewhat flexible structure to the interview, is the Sense Relation Network sheet (SRN). The three SRNs which form the core of the interview are shown in Figures 1, 2 and 3 below.

3.2.1 SRNs: visual design and content design

Both the visual design and the content design of the SRNs are inspired by the idea that there exists a 'web of words' (Altindol 1997: 61), or a series of interconnected networks which define, delimit and store linguistic expressions in the mind. The visual design of the SRNs is also inspired by materials and aids used in language teaching, such as word trees and word field diagrams (see Galton and Redman 1986), in which visual impact is crucial.

As can be seen in Figures 1, 2 and 3, visually, networks are designed in which the standard notion words are connected to subdivisions. The subdivisions, in turn, are connected to the semantic field of the SRN, symbolizing, in a way, the interconnected network or 'web of words'. Spaces are then provided under the standard notion word for the insertion of a distinctive partial synonym. Each SRN is printed in a different colour (presented here in black and white), the aim being for the visual impact of the SRNs to be positive, unobtrusive and inviting, and for the SRNs to engage the interest of the informant to a level at which the desire is to complete them.
Figure 1: Sense Relation Network sheet (first of three)

Figure 2: Sense Relation Network sheet (second of three)
In terms of content design, the SRNs are built around semantic fields (Lehman 1974) and, as such, are akin to the grouping of questions by subject matter in the SED questionnaire. According to Johnston (1995: 83), the grouping of questions by subject matter, as opposed to alphabetically or randomly, allows for a level of spontaneity in the responses. On the SRNs, standard notion words are offered as prompts for the elicitation of dialectal variants, as interviews which use indirect elicitation techniques are much more time-consuming than those which use direct ones. Additionally, indirect questioning may make the interaction feel more like an interview or a test than a conversation, so skewing speech style towards the more formal.

The selection of semantic fields and standard notion words in the three SRNs is the result of trialling and revision of the method during which eight original SRNs have been submitted under the present three. The assimilation was made in the interests of reducing the time needed by informants to complete the SRNs, as well as the time necessary to conduct the interview. None of the initial semantic fields have been discarded entirely, but the fields have become broader to encompass a greater area of notion words. Standard notion words producing little or no variation in trialling have been removed. However, each sub-division carries space for dialectal variants of notion words not included on the SRN which the informant wishes to include. When selecting standard notion words, the wish to include the same standard notion word as the SED where possible and appropriate was borne in mind, as a direct comparison could reveal potential real time change. Due to the urban bias of the proposed survey and the study of Middlesbrough English, however, this proved inappropriate in most cases, with few SED notion words remaining.

The SRNs then, as well as being a visual network, rather than a list of questions, represent the interrelated network of paradigmatic and syntagmatic sense relations in which linguistic expressions from similar semantic fields define and delimit each other's meaning. They also represent the sense relation of partial synonymy, which the dialectal variant holds with the standard notion word. Additionally, in time they will represent a geographical sense relation network of dialectal variation of partial synonyms found across a number of locations throughout the British Isles.

3.2.2 SRNs: technique of administration

Coupled with their concept and design, the technique of administering the SRNs is an essential part of their success as a method of eliciting lexical data. Informants are given the SRNs some days before the interview, with both verbal instructions from the fieldworker and written instructions as part of the interview pack (see Appendix 2 for instruction sheet). The innovatory step of allowing informants to know the content of the interview prior to the event has implications for both the content of the interview and for the interview as a speech event.

Giving informants the SRNs prior to the interview allows them time to consider the lexical items they use. This has a considerable effect upon the amount of lexical data yielded from the interview. If asked to produce a dialectal variant as an immediate response to a prompt, there is a danger of the informant's mind going blank. This results in minimal data being yielded. This may also necessitate an undesirable level of prompting from the fieldworker. More importantly, however, there could be a harmful effect on the required speech style and the willingness of informants to speak at length, due to a feeling of unease in the interview situation. Thus, the technique of administering the materials prior to the interview maximizes the amount of data yielded.

Additionally, feelings of unease in the interview situation may be heightened if the informant perceives the interview as a test of some sort. By having prior knowledge of the content of the interview however, it is thought that suspicion on the part of the informant is
diminished appreciably. Thus, combined with the fact of experiencing the interview in a social dyad, allows informants to settle into a relatively casual speech style as soon as possible. To ensure the ready recruiting of informants and to maximise the possibility of gaining access to their least overtly careful or monitored speech style, it was crucial that informants feel at ease and enjoy the interview as much as possible. Discussing the use of experiments in mainstream linguistic research, Meyer and Nelson (2006: 103) note it is possible to minimize the effects of the observer's paradox by giving subjects explicit instructions outlining precisely which kinds of judgments the experimenter wishes them to give. The same can be said to apply to the use of questions and data elicitation techniques in the sociolinguistic interview, and having prior knowledge of the context of the interview minimizes any sense of unease that may be felt.

When the informants had some days in which to complete the SRNs at their convenience, eliciting responses with others who should wish (differentiating between their own and others' responses on the SRNs), the paired interview is undertaken and recorded. The interview consists of the written response on the SRNs being read out by the informants with responses being discussed in terms of whether informants use the variants or only know them, situations in which they would be used, connotations and collocations associated with the variant, as well as anything else which informants might initiate. The fieldworker can use an interviewer's guide to ensure that all the notion words are covered (the informants keep their own SRNs until the end of the interview) and to offer suitable probe questions, e.g. the use of intensifiers, gender differences in use, age differences in use, varying degrees of a state, additional notion words or senses of the notion words given, all of which can provide vital additional information and extend the discussion. During the interview other known or used variants which crop up are noted on the SRNs in different coloured ink by the informant. Thus the written record of the informant's responses on the SRNs (which the fieldworker collects after the interview), a recording of the informant's spoken responses in real speech style and a range of attitudinal information on the lexical items elicited in an informal speech style are all secured by means of the use of the SRNs in the recorded interview.

3.3 SRNs: data yielded

In terms of lexical items elicited through the SRNs, the richness of the data yielded can be seen in the three completed SRNs from a Middlesbrough informant which appear as Appendix 3. The potential for the study of the differences and problematic distinctions between dialectal variants, regional slang, national slang and standard colloquialisms are clear. The study of nonstandard orthography is also promoted by the method. Additionally, the difference between items produced before and items produced during the interview may be of interest.

From the recorded discussion about the responses, more lexical data are produced. For example, the three SRNs shown as Appendix 3 present 215 variants for 80 standard notion words; however, by including all the variants the informant mentioned but did not write on the SRNs and those she claimed knowledge of during the recorded interview, a total of 272 variants were counted. Informants can use dialectal variants without necessarily being aware they are doing so and informants' self-reports can be inaccurate; for example, one informant, when discussing the notion word 'man', claimed that she would never use *lute* before already having done so during the interview. Additionally, informants may become aware only when they hear someone else use a particular word that they themselves use it. Also informants' insights into which variants are considered to be local, as opposed to those which are more widely used can be revealed. For example, one informant claimed not to have inserted a variant for soft shoes worn by children for P.E., because she 'couldn't think of another word for sandshoes', indicating that she believed sandals to be a widely used or standard variant. Therefore, discussion about responses can increase the amount of lexical data produced to a considerable extent.

Once read in isolation lexical items are immediately put into context by the informant. Thus, the individual lexical item is recorded and can then be disregarded for the purposes of a phonological analysis of informal speech (the nature of the written response on the SRN being read aloud perhaps constituting a more formal reading style of speech). The context of the interaction makes it clear which particular lexical items are read aloud and which are not. Alternatively, the use of different coloured ink on the SRNs is an indication of which variants were written before the interview (and thus read aloud), and which were noted down during the interview (written after having been known).

After having been read aloud, the lexical items are generally elaborated upon and discussed in the context of casual conversation, giving the sample of informal speech which can be analysed phonologically and grammatically. For example, after having given the responses *wel, tex, nick, shank, and snipe* for the notion word 'ideal', two informants from Middlesbrough proceeded to discuss at length precisely what each term referred to and their perceptions of the words' etymologies. Similarly, gender and age differences in responses to notion words are discussed at length, with, for example, two young male Middlesbrough informants arguing that they would never use the variant *boney* for the notion word 'infringue', it being an 'old person's word', and they would never use *cammy-laying* being used by gals, using themselves to use recor, *reark, fire and lark. Thus, the informal speech which can be analysed phonologically and grammatically contains a mass of data on knowledge and use of lexical items; attitudinal information on dialectal variants; views on etymologies, changing societal attitudes to lexical items and perceptions of and actual gender and age variation in usage. In this way a multi-levelled bank of data is produced through use of the SRNs.

3.3 Identification Questionnaire

Combined with the three SRNs, an Identification Questionnaire (IQ) is included in the interview. The IQ is given to the informant, with the three SRNs, prior to the interview, thus forming the interview pack. The questions posed in the IQ of the core interview are listed below in Figure 4. The IQ can be expanded for use in a given area with questions which are locally relevant being used additionally, as in the Middlesbrough example, see Appendix 4.

The core IQ comprises 15 questions whose aim is to elicit attitudinal data and metalinguistic comment from the informant. Information on identities, identifications, orientations, affiliations, and perceptions of language variation can be elicited through the identification questionnaire. The benefit of these types of data is that the context of the speech can be analysed in order to gain insight into the linguistic behaviour of the informants and the individuality inherent in the linguistic forms of interest.

As communities and boundaries are often symbolic, it is problematic to impose a definition of speech community onto a geographical area and a group of people, even when an investigator is a native of the geographical area to be studied. The similarities and differences
information will give comparable data across regions and may reveal differing attitudes towards areas and dialects, as well as revealing possible gender, age and class variation in a
given area.

This then is the core method designed to elicit a sample of informal speech for use in
studies of sociolinguistic variation and change. From the speech sample, data are obtained for
a multi-levelled analysis of phonological, grammatical and lexical variation, as well as
allowing for the elicitation of information on identities, allegiances, orientations, perceptions
of variation and the indiscernibility inherent in language. More formal word lists can be included
to observe intra-speaker variation and to control for linguistic environment, and, because of
the larger nature of data necessary for grammatical analyses, and the structural limitations of
the interview which places 'pragmatic and discourse constraints on syntactic structure'
(Milroy 1987b: 56), a more formal grammatical element can be included in the methodology.

Although some, such as Hazen (2001: 777), would argue that including specific
questions about language should appear as part of the sociolinguistic interview (or module);
'it' sociolinguistic module should focus on matters of importance for the subjects and have a
series of language questions, including specific questions about language variation patterns of
the area'. Others, such as Tagliamonte, see the elicitation of personal narratives and topics as
central to the sociolinguistic interview. Indeed, Tagliamonte (2005) suggests that discussion
of language should be used almost as a last resort, 'if you are going to include a module on
Language, always put it at the very end of the interview when your informant has exhausted
all the more personal topics' (2006: 39). She further stresses that the content of the interview is
of little consequence 'keep in mind that you are not asking questions to get information:
you are asking questions that reach the 'real' sentiments of your speakers and which elicit
natural, spontaneous speech' (2006: 43). The method of the data elicitation presented here is built
around the belief that accessing the attitudes to and perceptions and knowledge of linguistic
variation allows greater insights into motivations for linguistic variation, the causes of
sociolinguistic variation, and the indiscernibility inherent in language forms. It should therefore
explicitly investigate the social identities of the speakers and how they are negotiated and
maintained through linguistic behaviour and knowledge of linguistic variation.

4. Use of the method in a given locality: the Middlesbrough study

As noted, the core methodology can be used in a study of sociolinguistic variation of a given
area. This can be either in its core form or in an expanded form. For use in the Middlesbrough
study several additions were made.

4.1 Language Questionnaire

In order to obtain awareness of informants' perceptions of the nonstandard grammatical
features found in the area, a Language Questionnaire was included in the Middlesbrough
interview pack (see Appendix 3). The responses to the questionnaire can then be compared to
and correlated with the informant's actual usage of nonstandard grammatical features in
informal speech. The Language Questionnaire is based on the type of questionnaire used by
Chastain, Edwards and Whittle (1989) for the Survey of British Dialect Grammar. The
majority of the sentences used in the Language Questionnaire are authentic, having been taken
from the recordings of 'free' conversations for a pilot study of Middlesbrough (Llamas 1994).
Many grammatical features included in the questionnaire are features associated with
urban varieties of British English, e.g. multiple negation, them as demonstrative adjective,
what as subject relative pronoun, present participle was, never as past tense negator. Given the
geographical position of Teesside, certain sentences have been included to ascertain whether features associated with a variety from further north, namely Northumbrian English, are used. Questions 55, 36, and 27 have been taken from Beal (1993). Other sentences contain features associated with Yorkshire English, e.g., questions 10, 15, 17.

4.2 Middlebrough Identification Questionnaire

The IQ used in the Middlebrough study has four additional questions (these can be seen in Appendix 4, questions 7, 9, 10 and 13). Due to the study's focus on the transitional nature of the area, its changing local identity, and convergent and divergent linguistic trends, questions are asked to elicit responses about attitudes towards the changes in county boundaries and the transitional location of the area. Also, the informants are asked their attitudes towards perceived misidentifications as Geordie or Yorkshire.

The responses to the IQ give an indication of the level of allegiance and orientation to their area and their accentual/identity feelings. Responses to the questions on the IQ give an insight into whether the informant feels positively, negatively or neutrally towards their language variety and their area. Responses to certain questions can be judged to be: 'positive', in that they express positive feelings towards the area and the variety; 'negative', in that the informant expresses a desire to live elsewhere and disbelieve in the variety; or 'neutral', in that the informant feels neither positive nor negative, or perhaps a combination of the two, such as a section of Labov's (1972: 39) work in his Martha's Vineyard study. Not all of the responses to the questions in the IQ lead themselves to a classification of his kind (e.g., density of social networks, perceptions of boundaries), but many do (e.g., whether or not the informants like their accent; whether they would live in another area; how they perceive themselves in other areas). As overall assessment, therefore, can be made by the researcher as to whether the informant responded positively, negatively or neutrally to their language and their area in the IQ as a whole.

4.3 Identification Score Index

To construct the subjectivity of this decision somewhat, a more objective and quantifiable Identification Score Index has been devised (see Appendix 6). This is an adapted and extended version of the Identity Score Index used by Underwood (1988) in his study of Texan English. This was an attempt to derive a new Le Page and Tabacco-Keller's (1985) theory of a factor to account for linguistic variation in Texas. Le Page had intended his theory of acts of identity to be 'universally applicable' (1985: 182); and it had previously been used by Trudgill (1993) to account for the variation in British pop song pronunciations. Underwood (1988: 410) constructed an 'index of Texan Identification' with which he scored responses to three questions designed to test the level of local affiliations. When analysed, the use of the localised variant under consideration was found to have no linear relationship with social variables, the localised variant appearing dominant in all groupings. There was, however, a clear linear relationship between scores on the index and the use of the localised variant, i.e., the closer the informant identified with the group in question, the Texan, the higher the use of the localised variant.

This idea has been used in the present study of Middlebrough English. The Identification Score Index comprises seven questions designed to test how clearly or how loosely tied to the area the informants feel. The Identification Score Index includes direct questions about how the informants feel towards other people from their home town, and questions which test in-group preference. The Identification Score Index is not designed to elicit any linguistic data, but simply compares seven multiple choice questions. The questions are short and an immediate response is sought. Therefore the questions are included in the brief section of the interview in which informants provide details for the biographical information sheet (see Appendix 1).

Each of the multiple choice responses given in answer to the question carries a score of 1, 2, or 3, with a score of 3 indicating the strongest level of local affiliation. The scores are added and categorised into three groups, broadly termed positive, neutral, and negative. The linguistic variables under consideration are then analysed in terms of the three groupings, and non-linguistic variables can also be correlated with the scores.

5. Further Refinements

As noted earlier, the method is currently being used in a number of doctoral studies of variation and change in varieties of British English. Further refinements to the method of data elicitation mainly involve the addition of locally relevant questions to the IQ and locally relevant example sentences to the Language Questionnaire (for further details see Asquoy, Buurman-Hiltendo and Wallace 2006). An additional element has been added to the Language Questionnaire which seeks to elicit informants' perceptions of the regional and social distribution of grammatical structures; in addition to informants indicating their use of particular grammatical forms and their awareness of the existence of the forms in the speech community under investigation, informants are asked to indicate their perception of the regional distribution of forms and any association of the forms with, for example, male/female speakers, young/old speakers etc. (for further details see Pircher forthcoming). As well as extended versions of the method, simplifications can be used as evidenced through its use in the Voices project. In the Voices Recordings, a simplified version of the SRQs, known as 'spider diagrams', were used to elicit data in the collection of 300 recordings of varieties of British English made between 2004-2005. Such adaptations demonstrate the flexibility and usability of the data elicitation method which can be modified and/or extended for use in a particular study, a large or small, yet can retain its central core which allows for comparability and transparency.

6. Conclusion

This paper has presented a methodology designed to elicit data which are comparable regionally and socially for a large-scale study of language variation. The primary aim of the methodology is to obtain informal data from socially-stratified samples of speakers. The core of this methodology consists of the SRQs whose design and techniques of administration permit the elicitation of data which are analyzable on three levels of variation: phonological, grammatical and lexical. Additionally, the core IQ is available to provide additional data and non-linguistic common making information on perceptions of variation and the identity of linguistic forms available to assist in interpreting linguistic trends uncovered.

Beyond the core elements of the method, supplementary features can be added as demonstrated through the additions to the method used in the Middlebrough study. Such additions consist of a Language Questionnaire, acting as a more formal grammatical element, and an extended IQ, to address questions of local relevance in terms of orientations towards perceived centres of gravity and delimitation of the area and the variety, and an Identification Score Index, to ascertain a numerical indication of the informant's level of local affiliation.
These additional elements demonstrate how the core methodology can be adapted and extended for use in an individual study to address particular questions of interest.

Raw data obtained through use of the methodology have been presented in the appendix of the paper. A Middletown informant’s biographical data, completed SIEs, responses to the Language Questionnaire and Identification Score Index responses are given.

The methodology presented is currently being used by a number of other researchers either in its core form, or in an adapted or extended form. It is hoped that more researchers will utilise the method. Thus, a bank of directly comparable data will be collected, with the methodology from which the data are derived remaining consistent and explicit.

Notes

1. The Survey of Regional English (SURE) is a combined project proposed by the universities of Leeds and Sheffield. For further information see Kerawill, Llanasa and Upton (1999).

References


Data collection for regional and social language variation studies


Carmen Llamas
Department of Language and Linguistic Science
University of York
Heslington
York
YO10 5DD.

email: cll58@york.ac.uk
Appendix One

**BIOGRAPHICAL INFORMATION**

name .......................................................... [redacted]
sex .......................................................... female
age (DOB) .................................................. 31.6.43
place of birth .................................................. Middlesbrough
birthplace of mother ........................................... Hartlepool
birthplace of father ............................................ Hartlepool
birthplace of grandparents .................................. Hartlepool, Hartlepool
birthplace of grandparents .................................. Stockill, Seaham
ethnic group .................................................. White
occupation (current / usual) ................................... Teacher, Lecturer
assessment of social class .................................... Working Class
housing ......................................................... Flat
education ...................................................... Higher
Identification score index ................................... 3.9

Appendix Two

**First Name** ................................................. [redacted]

**Place of birth** ............................................. Middlesbrough

**Other places you have lived and for how long**

- Teesdale 1 year
- Durham 3 years (College)

- Please complete the sheets with words you think are dialect words or are local to the area you are from.

- Try to put down the first thing that comes to your mind, words you use every day when talking with friends, for example.

- After that, think about it for a while and note down any other examples of words local to the place you live which come to mind.

- Feel free to discuss the words with other people from the same area as you. But try to keep a note of who you discuss the words with (especially if you note down their suggestions).

- Put down more than one word, if you like. Also, feel free to use expressions as well as single words.

- Use the sections called "any others" to note down any extra words or expressions you think of (yourself, or in discussion with others). If these are words for things not listed on the sheet, please put down what you think they mean, or what someone not necessarily from your area would understand by them.

- Have a look through the questions about your language and your area, which we’ll also be talking about (there is no need to answer these questions on the sheet).

- Complete the Language Questionnaire by putting ticks in the appropriate boxes.
Appendix Four

Your Language

- What accent would you say you had, and do you like it?
- Can you recognize the accent of Middlesbrough (e.g., heard on the radio or TV)? If so, how?
- Do you think older and younger people talk the same here (pronounce things the same and use the same words)?
- Have you ever been in a situation where you've deliberately changed the way you talk? If so, why?
- Do you think there's a difference between how males and females speak here?
- Where, geographically, would you say people stop talking the same as you and start sounding different?
- What would you think if your accent was referred to as Geordie or Yorkshire?

Your Area

- If you were watching a regional news programme, what places would you expect to hear news from?
- Do you remember when the county of Teesside was formed and Middlesbrough was no longer in Yorkshire? Do you think this change made a difference?
- Would you consider Teesside to be in a larger 'north-eastern' part of the country or a larger 'Yorkshire' part of the country? Why?
- What image or description of Middlesbrough would you give to someone who didn't know it?
- If you wanted a day out shopping, where would you go?
- Do you think Middlesbrough is a fashionable place to be?
- What do you consider the local football derby to be?
- If you could, would you change where you came from? Why/why not?
- What do you consider the best and worst things about growing up and living in Middlesbrough?
- Have you ever seen Middlesbrough on a national TV programme (e.g., a documentary)? If so, how was it portrayed?
- If an outsider was complaining about Middlesbrough, would you defend it even if you agreed with what she was saying? Why/why not?
- How many friends, relations, and work/school/college mates do you have in the neighborhood (not more than about 3 mins. away) who you see regularly?
Appendix Five
Language Questionnaire

Tick (✓) this box if you would use this type of sentence yourself in speech.
Tick (✓) this box if you would use this type of sentence when writing to a friend.

1. [ ] [ ] [ ] [ ]
   - He was just set there by himself.
   - They can't do nothing without you saying.
   - There's a job going at our place if you're two want to go for it.
   - We all talk different.

2. [ ] [ ] [ ] [ ]
   - You weren't stood there, were you?
   - Just say what you want, itself.
   - They said they were coming back on Monday and they never.
   - That's the best one what she's got on.

3. [ ] [ ] [ ] [ ]
   - You're leased on them items for 80 days.
   - He's working 9 while 6 this week.

4. [ ] [ ] [ ] [ ]
   - I'm going down London next week.
   - I don't fancy going up Stockton.
   - The shirts were only two foot long.
   - I seen Sarah at work yesterday.

5. [ ] [ ] [ ] [ ]
   - I know a bloke who were doing speech therapy.
   - We was walking alon the road when it happened.
   - It were too cold to go out.
   - We usually gun down the pub on Thursday's.

6. [ ] [ ] [ ] [ ]
   - I've never heard of him like.
   - He said it wasn't scary but, mind you, he is about 45.

7. [ ] [ ] [ ] [ ]
   - They proper hurt you when you cash.
   - The cope ain't gonna do anything.
   - They isn't gonna pull you up.
   - It's the only life decent night out we have, isn't it?

8. [ ] [ ] [ ] [ ]
   - He wouldn't could've worked, even if you had asked him.
   - Will I put the bottle on?
   - My last needs washed.
   - I'm opening another account me.

9. [ ] [ ] [ ] [ ]
   - If you're left-handed, you're more clever.
   - I've forgot my money can you buy me a pint.