Two Large-scale and Long-term Language Variation Surveys: 
a Retrospective and a Plan

CLIVE UPTON and CARMEN LLAMAS
School of English
The University of Leeds
Leeds LS2 9JT
England
c.s.upton@leeds.ac.uk
enym11@leeds.ac.uk

ABSTRACT

Of the several regional dialect surveys which have been carried out in the British Isles, the Survey of English Dialects (SED) is without doubt the best known and the most widely used primary data for language variation studies of many kinds. As we take stock of our subject at the turn of the Millennium, this paper takes the opportunity to put on record the SED method, briefly evaluates its past and continuing contribution, and offers an indication of how some in the English dialectological and sociolinguistic communities may move on from data half a century old with a hope of doing for regional variation today what SED did in mid-century. (KEYWORDS: dialectology. methods in dialectology. English dialects. linguistic atlases).

RESUMEN

De los muchos estudios sobre dialectos regionales que se han llevado a cabo en las Islas Británicas, el Survey of English Dialects (SED) es, sin duda, el más conocido y el que más ampliamente se ha utilizado en diversos estudios sobre variación lingüística. Este trabajo quiere hacer balance de la utilidad del SED al final del milenio, dejar constancia del método empleado por sus compiladores y evaluar brevemente su pasado y su continua contribución. Por otro lado, también apunto al modo cómo en la actualidad se puede conseguir hacer por el estudio de la variación regional lo que el SED hizo o mitad de este siglo para las comunidades dialectológicas y lingüísticas inglesas. (PALABRAS CLAVE: dialectología, métodos dialectológicos. dialectos del Inglés, atlas lingüísticos)

I. SED: Employing the 'fundamental instrument of the Survey'

The SED method begins with the Questionnaire, the 'fundamental instrument of the Survey' (Orton 1962: 15). This is of the 'direct interview with direct questioning' type (Johnston 1985: 82), containing 1092 numbered questions which expand with transformations to 1322 questions in total, and is structured to obtain specific and comparable data from the 313 localities.
surveyed. It is important to recognize that the Questionnaire did not spring into use fully formed, but rather that it evolved over a period of some seven years to achieve its finally published form, the sixth version. Regular refinement as a result of practical testing, surely a tenet of any practical field investigation, characterized the SED questioning technique. The fact of gaps appearing in the evidence presented in the Yorkshire data shown, for example, in maps M3-5, 7 and 8 of The Linguistic Atlas of England (Orton et al. 1978), testifies to this refinement, and is a reassurance of quality control rather than creating a great gap in the record.

Although it is published in the Introduction to the Survey (Orton 1962), the SED Questionnaire is perhaps now best known in the piecemeal form in which it appears in the Survey's other publications. For this reason, and to enable readers to trace an entire 'run' of
data through the processes of collection and publication. One questionnaire page is reproduced as Figure 1. Immediately following this, as Figure 2, is a page of the fieldwork recording book for locality 5La.7 (Thistleton, Lancashire), showing responses to some of the questions asked on the relevant questionnaire page and, in fainter script, the SED editors' marks: the SED recording books are located in the Special Collections department of Leeds University Library. (Figure 3, presented later, continues the theme, being the Basic Material entry for to the quick for the Northern Counties, of which Lancashire is one).

![Figure 2: SED recording book for locality 5La.7, Thistleton, Lancashire (responses to questions VI.7.2 to VI.7.13b). Reproduced with permission of the Librarian, Leeds University Library.](image-url)
The detail in which the questions are set out, the uniformity with which they were meant to be administered by fieldworkers, and the narrowness of the phonetic record, create some measure of certainty with which the responses elicited by the Survey's eleven fieldworkers can be considered to be comparable. (The fact of 'fieldworker boundaries', where differences in the transcribed data may be suspected to be artifacts of differences between fieldworkers' practices rather than in informants' responses, is usefully addressed in Trudgill 1983:38-41. That such boundaries are not present, when the record may lead one strongly to suspect that they are, can also be remarked, as it is in Upton 1985: 392).

It was Orton's hope (1962: 15) that the data would, upon publication of the Survey's findings, prove to be 'genuine vernacular' and the nonstandard nature of the responses is to be seen in all subsequent Survey publications. Johnston, however, argues convincingly that, at least at the phonological level, much of what was prompted was 'canonical style' speech, equivalent in the spoken unscripted mode to the well-known 'word list style' of the social dialectologists (Johnston 1985: 84). Readers may make their own judgement on this from the sample of responses reproduced in Figure 3. We may fairly conclude that we have, in SED, a survey whose data-gathering device searches out linguistic data in considerable detail, and permits comparison locality by locality throughout the network, albeit that the data may be considered to lie within a formal style-range for the informants chosen.

### Figure 3: SED Basic Material (Orton et al. 1962-71): Volume 1 Part 2. entry for question V1.7.9

<table>
<thead>
<tr>
<th>Q. Somebovshaxanhabitof bringtheirnailsdown [g]...</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rr.</strong> (INTO TILL THE QUICK INTO THE RED QUICK)</td>
<td></td>
</tr>
<tr>
<td>Note—See also IV.2.1 and IV.3.6 for additional forms of QUICK.</td>
<td></td>
</tr>
<tr>
<td>1 N 1nta da hwik'</td>
<td>2 ta da wik'</td>
</tr>
<tr>
<td>2 Cu 1 ta twik'</td>
<td>2 ta-t wik'</td>
</tr>
<tr>
<td>3 Da 1 ta t wik'</td>
<td>2 inta da wik'</td>
</tr>
<tr>
<td>4 We 1 inta t wik'</td>
<td>2 ta t wik'</td>
</tr>
<tr>
<td>5 La 1 int t wik'</td>
<td>2 ta twik'</td>
</tr>
<tr>
<td>6 Y 1 ta twik'</td>
<td>2 inta twik'</td>
</tr>
<tr>
<td>Man 1 ta da hwik'</td>
<td>2 g.w. ta da hwik'</td>
</tr>
</tbody>
</table>

*SED Basic Material (Orton et al. 1962-71): Volume 1 Part 2. entry for question V1.7.9.*
II. SED: ‘NORMs’

Probably the most often-rehearsed fact concerning SED is that its informants were largely older members of rural communities. This, together with the fact that the majority (some 88 per cent) were men and that a principal qualification was that a speaker should have had little or no residence away from their home locality, has led to their characterization as ‘non-mobile older rural males’, or NORMs (Chambers and Trudgill, 1980: 33). Such a description necessarily masks social diversity - the age-spread alone is across six decades, and age has recently been used as a free variable by Easson (1998) to good effect in an SED-based study (Chianibers, personal communication) - but in broad terms it fairly reflects the informant type, and the acronym is a fixture in the discipline anyway. That NORM should have become, for some commentators, a term of adverse criticism is, however, unfortunate. The orientation of Orton, and of his SED co-founder Eugeri Dieth, was firmly diachronic: what better way could there be of studying reflexes of Middle English and older English linguistic forms than by the well-established route of searching the speech of those in a community who most closely preserve those forms? Although an element of social variation study was in place in America in the 1930s (Kurath 1939). such study was not that in which Dieth and Orton chose to participate. To criticize them for this is to miss the point of their especial orientation, and to divert attention from their achievement of their actual goal. (For a copent discussion of major themes and methods of regional dialectologists, including a defence of their orientation. see Davis 1983: 16-68).

III. SED: Publications

Employment of the questionnaire with the informants in an overwhelmingly rural locality network resulted in the accumulation of an unrivalled British English dialectal database. Much of this, the responses to the questionnaire questions together with such additional ‘Incidental Material’ interview information as was found to be directly relevant to those questions, was published as the Basic Material (Orton et al. 1962-71), and this has recently been reprinted (1998). Initial publication in this phonetic-list form rather than as an atlas was forced upon the Survey by financial constraints. However, in the rapidly-growing and diversifying discipline that is Dialectology, this form of publication has proved to be of the greatest value. permitting as it does the utilization of the information in ways which the Survey’s founders could not have expected to have foreseen.

The original goal of a dialect atlas was achieved with The Linguistic Atlas of England (Orton et al. 1978): this too has now been reprinted (1996) To this can be added a variety of other atlases (Kolb 1966, Orton and Wright 1974, Kolb et al. 1979, Anderson 1987, Upton et al. 1987, Viereck and Ramisch 1991 and 1998, Upton and Widdowson 1996). Also beyond the original design, but constituting a further presentation of its findings from within the Survey organization, is the Dictionary and Grammar (Upton et al. 1994). Atlas and dictionary material relevant to quick are summarized as Figures 4, 5 and 6: 4 is the Linguistic Atlas map for question V1.7.9 (to the) quick; 5 is the dictionary entry for quick (various senses), and 6 is the dictionary ‘core entry’ for loose skin, at which synonyms for question VI.7.11, including quick, are summarized. Current work to make existing and as yet unseen and unheard SED data available in electronic form includes that on the SED sound-recordings by Juhani Klemola, and digitization of the Basic Material (Elmer and Rudin 1997, Elmer fc).
Figure 4. SED The Linguistic Atlas of England (Orton et al. 1978): map Ph212, (to the) quick, question VI.7.9.

To these dedicated publications must now be added an array of others which have made use of Survey of English Dialects material to good effect in many different ways. Selection for mention of particular books and papers which have made use of the SED data would be both invidious and otiose: much good work has been done in this regard, and familiarity with any important works on English accent varieties, grammatical variation, dialect contact or other aspects of dialectology, diachronic or synchronic, regional or social, will have made the reader aware of the purposes to which the Survey has been and continues to be put.

IV. THE CHALLENGE SET BY SED

At this point it would be possible to draw up lists of successes and failures relating to the Survey, to point to the imaginativeness of its conception and to the flaws in its design, to the vigour of its prosecution and to errors in its execution. To do this, however, would be merely to revisit apologies and critiques which have been formulated at length elsewhere. And to do so would surely be to miss the fundamental point, that in SED we have a record of mid-twentieth century speech variation which continues to be visited by linguistic scholars of a great diversity of interests. In the mid 1940s, Dieth and Orton set out to plan a survey which would answer questions of language maintenance and change in which they were fundamentally interested, but which, through the flexibility of its design and in its wide-ranging scope, was destined to serve scholars of very diverse interests.

Whilst SED is a splendid lasting tool to have available, however, the Survey’s very existence must also be seen as a challenge. We can use SED, and its findings can be set against those of geographically-restricted monographs whose areas to a greater or lesser degree coincide with its locality distribution (see for example Trudgill 1986: 110-119; Williams and Kerswill 1997; Stoddart, Upton and Widdowson forthcoming). But what of the future? It is as well to remember that every England-wide map of regional variation which is produced today, and which more besides, is dependent upon the SED data. If we continue to draw on this, and if we pass on into the twenty-first century by way of innovation is monographs which are of restricted geographical scope and the product of fragmented methodologies, what chance do future linguists have of obtaining the overview of end-of-century English which SED permits us to have of that of the mid-century?

V. SuRE: A Survey of Regional English

When one considers the complex of social variables which today’s dialectologist and sociolinguist is obliged to consider, two things become immediately apparent. Firstly, it is inevitable that the geographical range of the individual scholar is likely to be severely restricted. If many informants are to be studied in each locality. Secondly, lexical data, which are notoriously time-consuming to collect, are likely to take second place to more readily-gathered phonological and grammatical material: this discrepancy is exacerbated by the fact that phonology and grammar, unlike lexis, permit of that statistical analysis which is central to the social dialectologist’s method.

Conscious of the fact that we have no up-to-date and immediately comparable information on regional variation and that, beyond the methodological similarities of the Survey of English Dialects and the Survey of Anglo-Welsh Dialects, comparability is lacking even between existing surveys within Britain and Ireland, dialectologists at the Universities of Leeds, Reading and Sheffield have been moving towards the creation of a new Survey of Regional English (SuRE). In this they have been encouraged by very many other dialect scholars throughout Europe, who have recognized the desirability of such an undertaking. That the desirability has been recognized is not to say that the logistical and methodological difficulties have been underestimated, but the team has been heartened by others’ support, and has progressed to the point where a core method is being trialled.

A SuRE method will, of course, bear little resemblance to that of SED: the development of social dialectology since SED was devised has ensured this. Since any modern survey must take account of a range of speaker profiles, and social sampling of informants in locality after locality over a wide geographical area produces a mix far more complex than that of Dieth and Orton’s one-speaker-per-question study. This, the need to locate for quantification systematically-occurring variables (Francis 1983: 19ff.), and the time-consuming nature of fieldwork under any circumstances, has led to a marked concentration on phonological variation of late. Whilst the uneven treatment of different types of variation is to be deplored and should be countered as vigorously as possible, it is inevitable that labour-intensive methods of data collection have now to be rethought if the speech of a variety of inhabitants of large territories is to be surveyed. And the overwhelmingly rural and diachronic orientation of SED must now be set aside. since societal change, and the variationist linguistic orientation towards the accumulation of data from a large and varied population, now demand an essentially

**Cuadernos de Filología Inglesa, vol. 8, 1999, pp. 291-304**
Despite inevitable changes in focus and technique, however, such an undertaking as SuRE must collect data which can be analyzed on all three levels of possible variation, phonological, grammatical and lexical. To discount any of these levels would be to obtain an incomplete picture of regional variation found throughout the British Isles at the turn of the Millennium. So we are faced with an interesting task: how do we set in motion a survey which will have a wide geographical sweep, can sample speech from a wide range of informants, and can do justice to all three levels of variation, and yet which can be embarked upon with a very reasonable chance of producing results within a quite short space of time and which will not readily lose momentum for the longer haul? We consider that these requirements, challenging though they are, are by no means insurmountable, and have set out to develop a method for the purpose. That method is predicated on our belief that there exists a body of scholars who will be willing to join us in the venture, perhaps not committing themselves wholly to it but prepared to 'super-add' a simple technique to their own field-studies, and that we can evolve just such a simple technique which is easy to administer, enjoyable to participate in, and productive in its linguistic yield.

The essentials of the SuRE approach (which, it should be emphasized, is still taking shape) are as follows. The rather formal context of the fieldworker asking set questions to elicit grammar or lexis in an extremely lengthy interview, as in the SED, is now considered to be inappropriate, as it would be impossible to undertake phonological analysis of casual speech from the data obtained. So too is the fieldworker asking questions to elicit involving personal narratives (Lahov, 1972), or allowing the informants to 'chat' in pairs, with or without the fieldworker present (Docherty et al. 1997): the possibility of obtaining any comparable data on lexis in particular through such methods would be nil. Instead, the SuRE fieldworker 'leads' a conversation around linguistic domains, with socially paired informants, permitting interaction to be more like a conversation than an interview. The fieldworker prompts informants to discuss their 'dialect' words, during which discussion data of phonological and, to some extent, grammatical significance is recorded too. In the course of the conversation, how much the speakers are actually aware of variation, as well as interesting social and attitudinal information on dialect, are also revealed.

The principal tool to permit rapid, focussed, and enjoyable interviews to be carried out is the Sense Relation Network sheet (SRN). This device is inspired by the idea that there exists a 'web of words', or a series of interconnected networks, which define, delimit and store linguistic expressions in the mind (Aitchison 1994). SRNs are built around domains of language, and in this regard are akin to the grouping of questions by subject matter in the SED questionnaire. Three are now in use for early SuRE interviews, one of which is represented as Figure 7.

The SRN domain is broken down into subdivisions, with Standard English notion words given as an initial prompt. Space is left for the informant to provide nonstandard synonyms for the standard notion word. Notion words are directly offered because interviews which use indirect elicitation techniques are more time-consuming than those which use direct ones. Also, indirect questioning may feel more like interviewing than conversation, so skewing speech style towards the formal. The sheets are made as visually pleasing as possible, each printed in a different colour. Informants are given the sheets a few days in advance of the interview, allowing them time to consider the words they use and eliminating the possibility of the mind going blank during discussion. This technique has the added benefit that, since the informants
are forewarned as to what is to be discussed, the 'testing' element of the exercise is lessened.
The desire for the informant to enjoy the interview and to feel comfortable at all times is vital,
both in terms of finding people willing to be informants and in accessing the informant's least
overly careful or monitored speech style.

Figure 7: A Sure Sense Relation Network sheet (Leeds informant).

When the informants have had some days in which to study the sheets, and to discuss
them with others should they so wish (differentiating between their own and others’ responses),
their response to the exercise is tape-recorded in the ensuing paired discussion. The
fieldworker thus secures both the written record of an informant’s responses on the sheets, and
the digitized spoken record of the responses on mini disc, available for phonological,
grammatical, and further lexical analysis. (In this last regard, of course, the method promotes
study of the added dimension of non-standard orthography.) In discussion, other lexical items
not given on the sheets may also be revealed, with informants becoming aware only when they
hear someone else use it that they themselves use a particular word, or using dialectal variants
without necessarily being aware they are doing so. As regards phonological and grammatical
data, informal conversation is produced, with informants seeming willing to talk at length
about lexis, and about attitudes towards lexical items and awareness of variation.

This, then, is the essence of the SuRE interview—swift to administer, unimimidating
and arresting for the informant, and, from early indications, productive in collecting data of
all the required kinds. Other devices can, of course, be employed alongside the SRNs, as the
individual researcher wishes. At Leeds we are employing an Identification Questionnaire,
fifteen questions designed to yield more extended talk, valuable information on people’s
attitudes towards language and identity (Le Page & Tabouret-Keller. 1974), and information
on people’s perception of language areas and boundaries (Preston. 1988). Also included in the
first SuRE fieldwork exercise are a word list to permit study of stylistic variation, and a more
formal grammatical element, similar to those used by Cheshire. Edwards and Whittle for the
survey of British dialect grammar (1993). An identity score index is also being developed for
use in the study of Teesside English: this is to be an adapted and extended version of the idea
used by Underwood (1988) in his study of Texan accent and identity, and will give an
indication of how closely or how loosely tied to the specific area the informant feels, this to
be correlated with linguistic and other non-linguistic variables.

CONCLUSION: Something to build on, and to use

It is greatly to be hoped that this methodology, with its rapid but productive central element
to which can be added any other elements—or none—will prove to be attractive to very many
students of English Language variation. Indeed, only if we can get agreement on this or some
other basic method on which to collaborate can we hope to make progress in surveying speech
variation over a wide geographical area, whilst keeping in touch with issues of social speech-
difference too.

But how does the amassing of comparable data amount to a ‘Survey’? The answer is
quite simple. The digitized recordings of interviewees discussing the Sense Relation Network
sheets, the comparable ‘core’ of the data, are centrally held on computer, tagged simply but
methodically for date, place, and biodeata. All those linguists who contribute to the building of
the bank of data will, by that act, have earned the right of access to it as it grows. A request
from one of them for, say, all data concerning ‘feniales in Northern England between the ages
of 15 and 50’, or ‘men and women in Liverpool’ (or ‘in Britain’ or, ultimately ‘worldwide’) will
result in the delivery, of a set of recordings for analysis. (Ultimate delivery is anticipated to be
via the Internet.) There will be no need for a central administration to process the
information beyond the initial tagging, no need for centralized decisions to be made as to what

Cuadernos de Filología Inglesa, vol 8, 1999, pp. 291-304
is or is not significant for publication. And there will be no need for the survey to stop. Methods may change somewhat over time, and the technology can be expected rapidly to improve, leading to quicker delivery of material of enhanced quality. But the principled collection of recordings which, in essence, remain comparable, and which therefore allow of phonological, grammatical, and lexical study over real time, will have been set in motion, for our benefit and for that of scholars who will follow us, into the unforeseeable future.

REFERENCES


