



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Mapping The Existing Phonology of English Dialects

Citation for published version:

Maguire, W 2012, 'Mapping The Existing Phonology of English Dialects' *Dialectologia et Geolinguistica*, vol 20, no. 1, pp. 84-107. DOI: 10.1515/dialect-2012-0006

Digital Object Identifier (DOI):

[10.1515/dialect-2012-0006](https://doi.org/10.1515/dialect-2012-0006)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Dialectologia et Geolinguistica

Publisher Rights Statement:

© Maguire, W. (2012). Mapping The Existing Phonology of English Dialects. *Dialectologia et Geolinguistica*, 20(1), 84-107doi: 10.1515/dialect-2012-0006

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Mapping *The Existing Phonology of English Dialects*

Warren Maguire

Abstract

Given its early date, breadth of coverage (geographical and linguistic) and the huge amount of data it contains, Alexander Ellis's *The Existing Phonology of English Dialects* marks an extremely significant episode in British dialectology. Despite this, there has been very little in the way of detailed linguistic analysis of Ellis's survey, and no attempt has been made to construct a linguistic atlas from the data it contains, although several studies have included a few preliminary maps based on it. Why is this so, and what might we discover if we did investigate this early survey of the dialects of English and Scots in more detail? The aim of this paper is to begin such an investigation and, in particular, to demonstrate that there is considerable mileage (and benefit) in mapping the data in Ellis (1889).

1. Introduction

The *Survey of English Dialects* (SED; Orton and Dieth 1962–71) and the *Linguistic Atlas of Scotland* (LAS; Mather and Speitel 1975, 1977, 1986) give us a detailed insight into aspects of the traditional dialects of England and Lowland Scotland in the 1950s, and in many ways represent a culmination of dialectological research in Britain which had been going on for over a century. Petyt (1980: 78) points out, though, that these nationwide surveys of English and Scots dialects (as opposed to individual studies at single locations) were conducted at a much later date than similar surveys in continental Europe and beyond (see Petyt 1980: 40–3 and Chambers and Trudgill 1980: 18–23 for discussion). Thus, for example, Georg Wenker began his survey of German dialects in 1876 (see the *DiWA* website, <http://www.diwa.info>), Jules Gilliéron started fieldwork for the *Atlas Linguistique de la France* (Gilliéron 1902–10) in 1896, Karl Jaberg and Jakob Jud published the results of their survey of Italian dialects in Italy and Switzerland between 1928 and 1940 (Jaberg and Jud 1928–60), and work for the *Linguistic Atlas of New England* was begun in the early 1930s (Kurath et al. 1939–43). Prior to the SED and LAS, nothing on the same scale, using techniques of the same sophistication and producing comparable and sizable data sets for such

a wide range of locations, had been carried out in Britain.¹ This is important given the dramatic decline in traditional dialect forms which had been ongoing since at least the middle of the 19th century. But as Petyt himself discusses (Petyt 1980: 70–88), there were in fact *two* nationwide surveys of English and Scots dialects towards the end of the 19th century which, despite methodological inadequacies and rather restricted data sets, give us a detailed insight into the dialectological landscape of Britain decades before the SED and the LAS. These were Alexander Ellis’s *The Existing Phonology of English Dialects* (Ellis 1889; henceforth EPED), and Joseph Wright’s *English Dialect Grammar* (Wright 1905; henceforth EDG). Ellis’s survey, which covered all of England and lowland Scotland, and a few points in Wales and Ireland, was conducted in the 1870s, making it contemporaneous with Wenker’s survey. Wright’s survey had a similar geographical coverage, and was conducted not long after Ellis’s. As is discussed further below, Wright’s survey, whilst containing much original material, is heavily dependent on Ellis’s, and is, consequently, less important in the history of British dialectology despite being better known and more accessible. Ellis’s survey, on the other hand, was entirely original and, given its early date, breadth of coverage (geographical and linguistic) and the huge amount of data it contains, it marks a significant and important episode in British dialectology which rivalled the efforts of continental dialectologists and dialectologists in Britain in the mid 20th century.

Despite this, there has been very little in the way of detailed linguistic analysis, as opposed to brief statements of comparison in later studies, of Ellis’s data (but see Kökeritz 1932, Jones 2002 and Maguire 2012 for notable exceptions), and no attempt has been made to construct a linguistic atlas from Ellis’s survey, although several studies have included a few preliminary maps based on it (Dieth 1946, Anderson 1987 and Trudgill 2004). Why is this so, and what might we discover if we did investigate this early survey of the dialects of English and Scots in more detail? The aim of this paper is to begin such an investigation and, in particular, to demonstrate that there is considerable mileage (and benefit) in mapping the data in Ellis (1889). In Section 2, I discuss the nature of Ellis’s survey and data, consider previous reactions to his work, and examine reasons why we might want to study it in more detail. In Section 3, I detail the issues and methods involved in mapping Ellis’s data, provide some illustrative maps, and assess the results of these first steps in the construction of a linguistic atlas of *The Existing Phonology of English Dialects*. It will be seen that such an atlas will allow us to more fully explore and appreciate this monumental work on

¹ But see Kurath and Lowman (1970) and Rydland (1998) for notable moves in that direction.

the dialects of Britain in the second half of the 19th century and will be an essential tool for dialectologists and historical linguists concerned with variation and change in English and Scots.

2. The Existing Phonology of English Dialects

2.1. *What is 'The Existing Phonology of English Dialects'?*

As its name suggests, the EPED was a survey of the phonology (by which was meant pronunciation as related to the historical phonology of English) of non-standard dialects of English (including Scots). The data for the survey were mostly collected between 1873 and 1879 (Ellis 1889: 1–3). Although the survey was primarily concerned with phonetics and phonology, the way in which the data were elicited meant that the EPED also contains some information on dialect morphology, syntax and, to a minor extent, lexis. Ellis's interests were, like most traditional dialectologists, historical in that he was interested in determining the endogenous local developments of the earlier phonology of English in an attempt to learn more about the historical phonology of the language. As Ellis puts it (Ellis 1889: 1):

The object of this treatise is to determine with considerable accuracy the different forms *now*, or *within the last hundred years*, assumed by descendants of the same original word in passing through the mouths of uneducated people, speaking an inherited language, in all parts of Great Britain where English is the ordinary medium of communication between peasant and peasant.

As was the case in later studies of traditional dialect, Ellis directed his attentions towards those people in society whose speech was least likely to have been standard or influenced by standard patterns – older, uneducated, often illiterate members of the (primarily rural) 'peasant' class – so that the native sound patterns of each dialect could be identified. In order to determine these, Ellis required a phonetic transcription of a selection of everyday English words in the dialects under investigation. These words were organised in three different ways:

- 1) A *Dialect Test* (dt.), a short paragraph which contained 94 words representative of the major historical phonemic classes and a small range of morphological and syntactic variables;
- 2) A *Comparative Specimen* (cs.), a longer paragraph (405 words) which contained a wider selection of words and morpho-syntactic variables;

3) A *Classified Wordlist* (cwl.) of 971 words covering in detail many historical phonemic classes (including words derived ultimately from Old English, Norse, French and Latin); an earlier, shorter version of this *Wordlist* (wl.) was used at some locations, and in the majority of cases transcriptions were only given for parts of the (c)wl.

In addition, transcriptions of various, usually short, texts (e.g. sentences, poetry) at a number of localities were made.

Ellis often used the ‘indirect method’ in the collection of his data. As well as transcribing data directly from the natural speech or conscious dictation of native dialect speakers (several fieldworkers aided Ellis in this task), Ellis relied on intermediaries to whom he sent his dt., cs. and (c)wl. These intermediaries, often educated people such as school teachers and vicars, were required to phonetically ‘translate’ Ellis’s examples into the dialect of their area, using their own knowledge, experience, and the help of native speakers. Ellis (1889: 3–4) explains why this procedure was adopted:

But why not go to the peasantry at once? Why not learn from word of mouth, so that errors would be limited to the writer’s own appreciation? Where possible, this mode of obtaining information has been followed. But I have myself been able to do so in very few cases. There are many difficulties in the way. First the peasantry throughout the country have usually two different pron[unciations]., one which they use to one another, and this is that which is required; the other which they use to the educated, and this which is their own conception of r[ecieved]p[ronunciation]., though often remarkably different from it, is absolutely worthless for the present purpose. If I, having no kind of dialectal speech, were to go among the peasantry, they would of course use their “refined” speech to me. I have therefore not attempted it.

In fact, Ellis did gather some of his data directly, either from uneducated native speakers (as in the case of quite a few of his locations in Northumberland, Dialect Division 32), or from students from rural backgrounds in the Whitelands Training College in Chelsea. Furthermore, much of the data in EPED was gathered directly from dialect speakers by three fieldworkers (Ellis 1889: 4–5): Mr. C. Clough Robinson, Mr. J. G. Goodchild, and, especially, Mr. Thomas Hallam.

As noted above, Ellis intended his survey to cover “all parts of Great Britain where English is the ordinary medium of communication between peasant and peasant” (Ellis 1889: 1). Ellis included Scots in his definition of English so that, in addition to all of England, his survey covered Lowland Scotland, Shetland and Orkney. He excluded from his survey most areas where Celtic languages were still or were recently spoken – i.e. the Highlands and western isles of Scotland (which were also excluded from the

LAS), most of Wales, and almost all of Ireland (but see Section 3.3. below for further discussion). He included the Isle of Man and Cornwall, however. Within the area included in his survey a large number of locations were sampled,² although the amount of data gathered at each location varies greatly and coverage in some parts of the country was better than others.

In the case of the dt. and cs., Ellis presented his data in the form of inter-linear glossed phonetic transcriptions. For the (c)wl., he provided numbered lists of phonetic transcriptions. In all cases, the data are given in a phonetic alphabet called the *Palaeotype*, which was designed to be able to represent the sounds of English dialects and other languages (see Ellis 1889: 76*–88* for discussion and explanation). This at first sight confusing and unfamiliar phonetic notation is as complex as the International Phonetic Alphabet (IPA), though grounded to an extent on principles which are no longer current in our understanding of articulatory phonetics (see Eustace 1969: 35–36). As Eustace (1969) and Local (1983) discuss, the Palaeotype is a sophisticated phonetic alphabet and it is in most cases straight-forward to translate it into the more familiar IPA (although certain vowel symbols and diacritics are somewhat more difficult to interpret and are sometimes used idiosyncratically).

Finally, Ellis arranged his data into hierarchical *Dialect Districts* (the main divisions being between Southern, Western, Eastern, Midland, Northern, and Lowland Scots dialects), defined by shared linguistic features, and he defined ten *Transverse Lines* (isoglosses) which were particularly important for defining relations between these dialects (see Ellis 1889: 3*). At 923 pages of dense discussion and data, the EPED is the largest unified collection of dialect material ever assembled for English *and* Scots dialects, one which demands to be taken into account in any investigation of the history of language and its study in Britain and Ireland.

2.2. *Evaluations of the EPED*

Evaluations of the data in EPED have been both positive and negative, and are often mixed – see Anderson (1977) and Shorrocks (1991) in particular for discussion. Dialectologists in the past have pointed out various deficiencies, especially its use of the indirect method (e.g. Petyt 1980: 73, Wakelin 1972: 50), difficulty in interpreting the Palaeotype (e.g. Sweet 1877, Wakelin 1972: 50, Petyt 1980: 73), and the quality of the resulting data (e.g. Wettstein 1942: xviii). Most famously, Wright (1892) and Dieth (1946)

² 494 in total (not 1145, as is reported in Petyt (1980: 72) and Shorrocks (1991: 322). Of these, 1 was in Ireland, 3 in the Isle of Man, 13 in Wales, 30 in Scotland, and 447 in England.

were scathing in their opinions on the worth of EPED. Wright (1892: 174) wrote the following, despite later praising the worth of the survey and drawing extensively on it in the EDG:

If his rendering of the dialect test of other dialect speakers is as inaccurate as that of the Windhill dialect, the value of these tests for phonetic and philological purposes is not very great.

Dieth (1946: 76) offered the following negative comments, before going on to demonstrate that mapping Ellis’s data produced very coherent results:

This book may well be termed a tragedy: a stupendous piece of work lasting fourteen years, born of a great vision, but carried out with inadequate means; a huge store of information which every dialectologist consults, but, more often than not, rejects as inaccurate and wrong

Conversely, other dialectologists have praised the quality of the EPED, whilst acknowledging its methodological shortcomings. For example, Wakelin (1972: 50) noted that:

It should be said in conclusion, in spite of all objections, that Ellis’s results are sometimes quite impressive, and even though he was working with inadequate tools, modern dialect research often confirms his findings ... For all Ellis’s inadequacies, his book is an indispensable source of reference to the dialectologist who wants earlier confirmation of his own findings, and it was at least a beginning on which later research could build. Neither, when one gets to know it, is the book as formidable as it first appears.

Likewise, Anderson (1977), Rydland (1982) and Shorrocks (1991) conclude that on balance the data in the EPED compares very favourably with data from later surveys such as the SED, and Shorrocks (1991: 323) notes that “the palaeotype is certainly not the completely impossible system that some have made it out to be”. Crucially, a number of dialectologists have produced analyses of Ellis’s data which illustrate that the EPED is not without value. Anderson (1987: 58, 62, 66, 68, 70, 100, 108, 111) compares his maps derived from the SED data with eight maps constructed from Ellis’s data, and the two datasets show very close similarities indeed. Jones (2002) analyses the geographical distribution of phonetic variants of reduced forms of the definite article in northern England, and demonstrates the importance of Ellis’s data for understanding the history of this phenomenon. Trudgill (2004) shows convincingly that Ellis’s data is crucial for explaining the dialectal origins of New Zealand English in Britain. He uses comparative maps of the distribution of features in Ellis (1889), Wright

(1905), Kurath and Lowman (1970, data gathered in the 1930s) and the SED to illustrate the likely source regions of features of the colonial dialect.

2.3. *Why study the EPED?*

Given the reservations expressed by some dialectologists on the value of Ellis's data, it is reasonable to ask why anyone should bother to consider the EPED further, especially when we already have large quantities of data for English and Scots dialects from subsequent dialect studies such as the SED and LAS. What can be gained by studying the EPED? In addition to the fact that some dialectologists have found Ellis's data to be in close agreement with data from later studies, there are several other key reasons why we might want to study Ellis's data in detail and which make the task of constructing a linguistic atlas from it a worthwhile endeavour.

Ellis collected the data for EPED in the 1870s, which means that it predates the data in the SED and LAS by 80 years (or, approximately, three generations). Since Ellis's data were collected from the most conservative, old-fashioned speakers in Britain, it is likely that they represent an even earlier form of speech than was generally current when the data were collected. We can be sure that language in Britain did not stand still in this period, especially when we consider the far-reaching social and economic changes which occurred in the second half of the 19th and first half of the 20th centuries (including massive urban expansion, immigration from Ireland and elsewhere, two world wars, compulsory education, universal suffrage, and a breakdown of the rigid social hierarchies of Victorian Britain). In other words, Ellis's data should, assuming that they are not altogether invalid, provide us with an insight into the linguistic landscape of mid-19th century Britain in a way that later surveys cannot do. They may also allow us to project even further backwards in time, especially when we compare them with later data from, for example, the SED. A comparison of Ellis's data with later studies may reveal trajectories of change, or may reveal that there has been little or no change in the distribution of a particular feature. Thus, my comparison of the distribution of Pre-R Dentalisation in the EPED and the SED (Maguire 2012) revealed that in the eight decades separating the two studies the feature has all but disappeared from large areas of northern and north Midland England, suggesting that it was once more widespread in England and, in turn, offering a possible explanation for its appearance in varieties of Irish English. As noted above, Trudgill (2004) uses data from Ellis (1889) to reveal earlier distributions of features in English dialects which formed part of the input in the development of New Zealand English (which was formed in the first half of the 19th century). Conversely, Ellis's data reveal that the Ribble-Humber dialect boundary (see

Wakelin 1984: 71, 73 and Figures 2, 5 and 6 in this article) was essentially the same in the mid 19th century as it was in the mid 20th century, suggesting that this dialect boundary was both ancient and stable.

The EPED also contains information on many linguistic features which were not targeted in later surveys such as the SED and LAS, making it, along with the EDG, a unique source of information. Although the coverage of historical phonological word groups in the later surveys is good, they are not exhaustive and they do not contain any information on many individual words which featured in the earlier studies. For example, the vast amount of SED data contains almost no information about words of the NURSE lexical set (Wells 1982: 137–40) which contained late Middle English /er/ (e.g. *certain*, *clergy*, *service*), and the words which it does give us details of are not typical members of this class (*fern*, SED question IV.10.13, seems to be something of an oddity and may not have a long history in many English dialects, and the vowel in *learn/learned/learnt*, often given in answer to SED question III.13.17, was probably subject to early lengthening before a homorganic cluster). Similarly, the SED is not of much use if we want to determine the development of clusters such as initial /gn/ (as in *gnaw*) and medial /mb/ (as in *timber*), or the development of words such as *book*, *heaven*, *long* and *you* (pl.), all of which are of interest to historical phonologists and dialectologists alike. Since detailed studies of the traditional dialects of England and lowland Scotland are few and far between (especially outside of northern England), the EPED and EDG are often our only direct sources of evidence for the pronunciation of particular words in most varieties.

It is worth pointing out at this juncture that the EPED is in many ways a superior source of information on 19th centuries dialects than the EDG. Although there is a great deal of original material in Wright (1905), much of the data therein (perhaps as much as a third) is taken directly from Ellis (1889). Thus, for example, Wright’s data for middle-east and southeast Northumberland and for south Durham are re-transcribed examples from the EPED. If we exclude such locations, the geographical coverage of the EDG is decidedly patchy (and is not helped by general labels such as ‘Cheshire’ and ‘south Staffordshire’) compared with the large number of narrowly defined locations in the EPED (see Figures 1–7 in this paper). Furthermore, Wright’s data is in a much broader, near phonemic transcription which ignores many interesting details (such as Pre-R Dentalisation; see Maguire 2012).

Ultimately, a thorough analysis of Ellis’s data is necessary if we want a deeper understanding of its value and of its flaws. Detailed investigations of the data for particular locations and comparison of them with data from the same areas in later studies is a worthwhile, if time-consuming, way of doing so (see Kökeritz 1932 and Maguire 2003 for examples). Dialectometrical

analysis of comparable datasets within the EPED (e.g. the dt. transcriptions) might also be enlightening. However, the approach pursued in this paper is an examination of cross-section samples of Ellis's data, in the form of maps of particular linguistic features. This method has the advantage that it not only allows us to assess the validity of Ellis's data (in terms of its internal consistency and its comparability with patterns from other surveys), but also provides an analysis of the data which is immediately more useful than the rather impenetrable body of information that constitutes the EPED.

3. Mapping the EPED

Unlike Wright (1905), Ellis (1889) gives the exact location (town, village, even farm) for each data point in the survey, allowing for precise mapping using the British and Irish national grid reference systems (see Figures 1–7).³ In addition, the strictly comparable nature of much of the data in the EPED lends itself to cartographic representation. There are 126 locations for which the EPED provides a transcription of the dt., 105 with transcriptions of the cs., and 358 with transcriptions of at least part of the (c)wl. The data for each of these is easily searched, since the dt. is divided into seven short sentences, the cs. into 16 short sections, and the transcriptions in the (c)wl. are numbered.^{4,5}

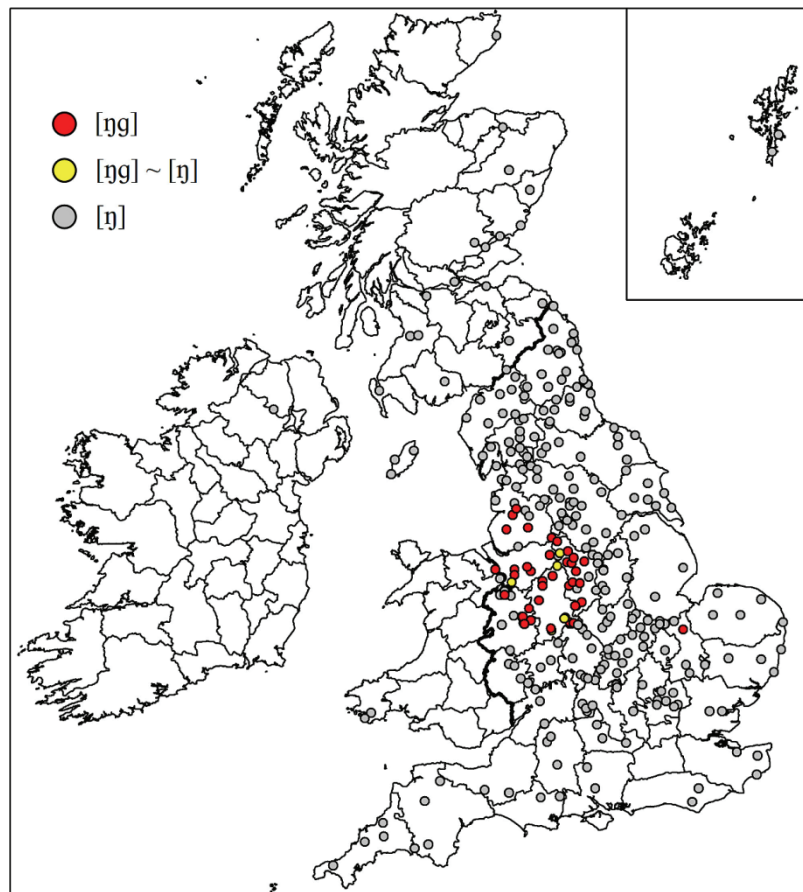
³ All maps in this paper have been generated using Alan Morton's DMAP mapping software (www.dmap.co.uk/).

⁴ In order to make the analysis of the (c)wl. data easier, I have compiled an Excel database which indicates whether a transcription of each word in the (c)wl. is present or not for each location. Since many words are missing at particular locations, this provides a quick way of reducing (often considerably) the number of searches of the original data in the EPED.

⁵ The EPED provides data for only one point in Ireland – a reconstruction of the extinct (even in the second half of the 19th century) divergent dialect of Forth and Bargo in County Wexford. In an earlier volume of Ellis's *Early English Pronunciation*, of which the EPED is a part, Ellis gives quite a bit of detail concerning the varieties of English spoken in Belfast and Cork (Ellis 1869: 1230–1243), although data for these locations is not available for the features analysed in this paper. I have, however, included data in this analysis from a point in Ireland which is not found in Ellis's works at all. These data, given in the Palaeotype and consisting of the cs. and a part of the cwl., are found in Staples (1896) and were, according to the author, intending for inclusion in the EPED but were ultimately left out for reasons of time and space. Since the data are fairly copious, are reasonably consistent with those found in the EPED, and come from an area otherwise left out of Ellis's survey (the border of Counties Tyrone and Londonderry), they provide crucial information about the English dialects of Ireland in the late 19th century. It is interesting to note that Staples' data are the source of Wright's Ulster (Uls.) data in the EDG.

It is relatively straight-forward, for example, to map the occurrence of final [ŋg], in Palaeotype (qg), as opposed to [ŋ], palaeotype (q), in the word *wrong* in the dt. (where it occurs in the phrase *the door of the wrong house*) and in the cs. (where it occurs in the phrase *ought not to be wrong on such a point*). The word *wrong* also occurs in the (c)wl. (word number 64), but because only some of the (c)wl. was recorded at most locations the pronunciation of this word is missing for 210 of them. Accepting that there is necessarily some missing data, we can nevertheless map these too. The results of this analysis are given in Figure 1.

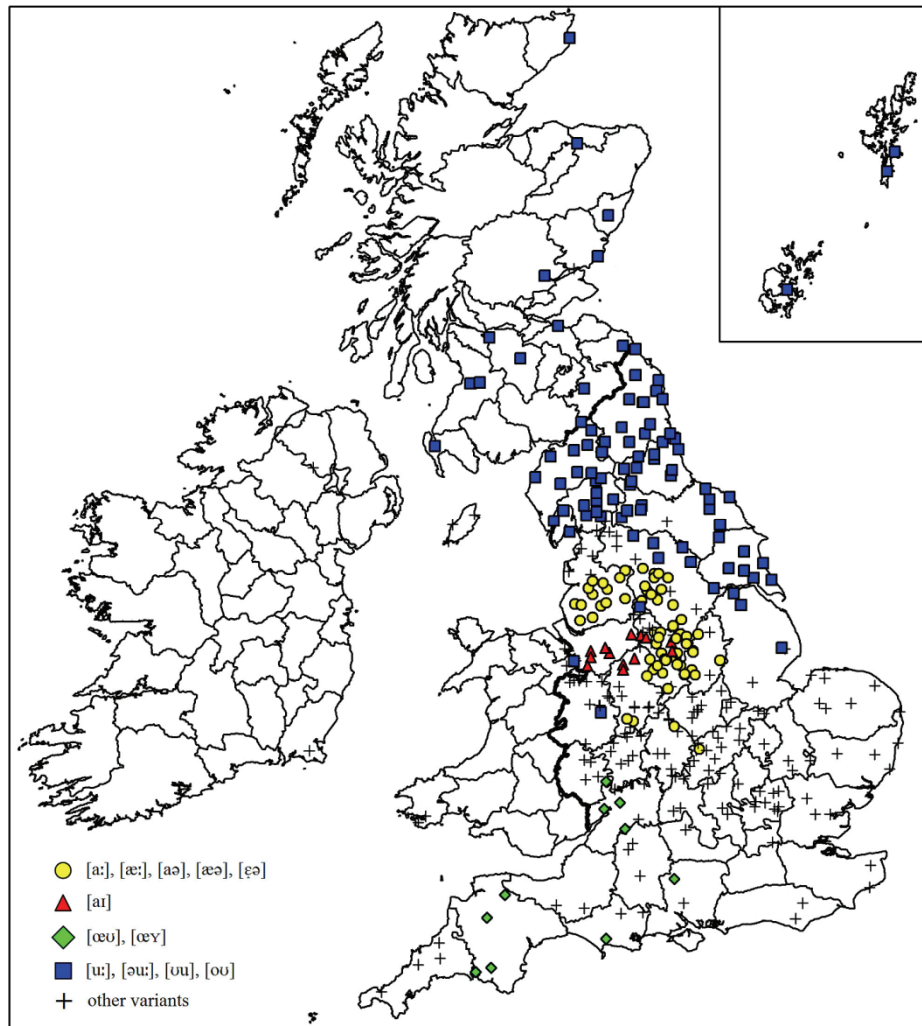
Fig. 1: Final -ng in *wrong* in the EPED



Other than a single outlier in north Cambridgeshire (which location is aberrant in several other respects), [ŋg] is confined to northwest England, specifically southern Lancashire, Cheshire, Staffordshire, most of Derbyshire, and parts of Shropshire and Warwickshire. This is almost identical to the distribution of [ŋg] in *tongue* in the SED (see Orton et al. 1978: Map Ph242), indicating that for this feature at least the EPED is as accurate as

the later dialect survey and that this pronunciation has been relatively stable between the 1870s and the 1950s.

Fig. 2: The pronunciation of the vowel in *down* in the EPED



This is a straight-forward case of course. It involves a simple opposition between presence or absence of [g], involves consonants, which, unlike vowels, are less likely to have gradient realisations, and involves only a single word (which is present at the majority of locations). But how accurately did the EPED record subtle vowel distinctions, and how suited are the data for analysis of structural similarities and differences between varieties?

Figure 2 illustrates pronunciations of the word *down*, which occurs in the dt., cs. and (c)wl. (word number 658) in the EPED. This word can be taken as representative of the development of ME /u:/. As expected, there is a profusion of variants, and these have been grouped for display purposes into

five broad categories (this is an issue with mapping, not with Ellis's data). These categories are: (1) monophthongal [u:] (uu) and high diphthongs of the type [əu:] (əuu), [ʊu] (uu) and [oʊ] (ou); (2) monophthongs and centring diphthongs of the type [a:] (aa), [æ:] (ææ), [aə] (aə), [æə] (æə) and [ɛə] (eə); (3) an [aɪ] (ai) diphthong; (4) diphthongs with front rounded starting point, e.g. [œʊ] (əu) and [œʏ] (əy); (5) other variants, essentially [aʊ] (au), [æʊ] (æu), [ɛʊ] (Eu), [əʊ] (əu) and [ʌʊ] (ɛu).⁶

If we compare Figure 2 to maps of words which contained ME /u:/ in the SED data (see in particular Orton et al. 1978: Maps Ph149–Ph152, and Anderson 1987: 50–55), the parallels are striking. The high monophthong and diphthongs are found in Scotland, Northumberland, Durham, Cumberland, Westmorland, North Yorkshire, East Yorkshire, northern West Yorkshire, and Lincolnshire (the diphthongal variants are typical of parts of Durham, Cumberland, Westmorland and mid-Lincolnshire). Apart from the mid-Lincolnshire point, this is exactly the same as the location of these variants in the SED, with the distribution following the well known Ribble-Humber line. Note in particular the extension of [u:] into north Lincolnshire and the lack of this pattern in the northwest tip of West Yorkshire.⁷

The low monophthongs and centring diphthongs, especially characteristic of southern Lancashire, parts of West Yorkshire, Derbyshire and parts of Nottinghamshire are found in almost exactly the same areas as they were recorded in the SED, with one notable exception: London. But this is only an apparent exception, since these variants are recorded in other words with ME /u:/ in the EPED London data alongside [æʊ] (which is the only variant recorded for *down*). This is exactly the same kind of variation in this lexical set that we find in the SED data for Hackney.

The [aɪ] diphthong variant is restricted to Cheshire and north Staffordshire, exactly where it is recorded in the SED, and north Derbyshire, where it is not. That this represents a change rather than an error is suggested by the fact that the EPED data for north Derbyshire were gathered directly from native informants by Thomas Hallam, himself a native of that county.

The diphthong variants with a central or front rounded starting point (e.g. [œʊ]) parallel the data in the SED less closely. In the SED, these variants are essentially restricted to Devon and adjacent areas of Somerset and Cornwall, and even here they are in variation with diphthongs with un-

⁶ Slightly simplified versions of Ellis's Palaeotype symbols given in parentheses.

⁷ The few exceptions to this pattern involve [oʊ]-type diphthongs (mid-Lincolnshire and the Welsh Marches) which probably represent either transitional variants or occasional higher variants of the unmarked case (+ on the map), or vowels and diphthongs (e.g. [i:], [iə]) which look like unexplained reflexes of ME /o:/ rather than /u:/ (North Yorkshire and Dundee).

rounded starting points. The EPED data for Devon is consistent with the SED, but other locations in the southwest, especially in Gloucestershire, also have [œʊ] in *down*. The small number of locations involved, and the profusion of variants recorded in this part of England in the SED make it difficult to draw any further conclusions, and further study of this feature in this area is required.

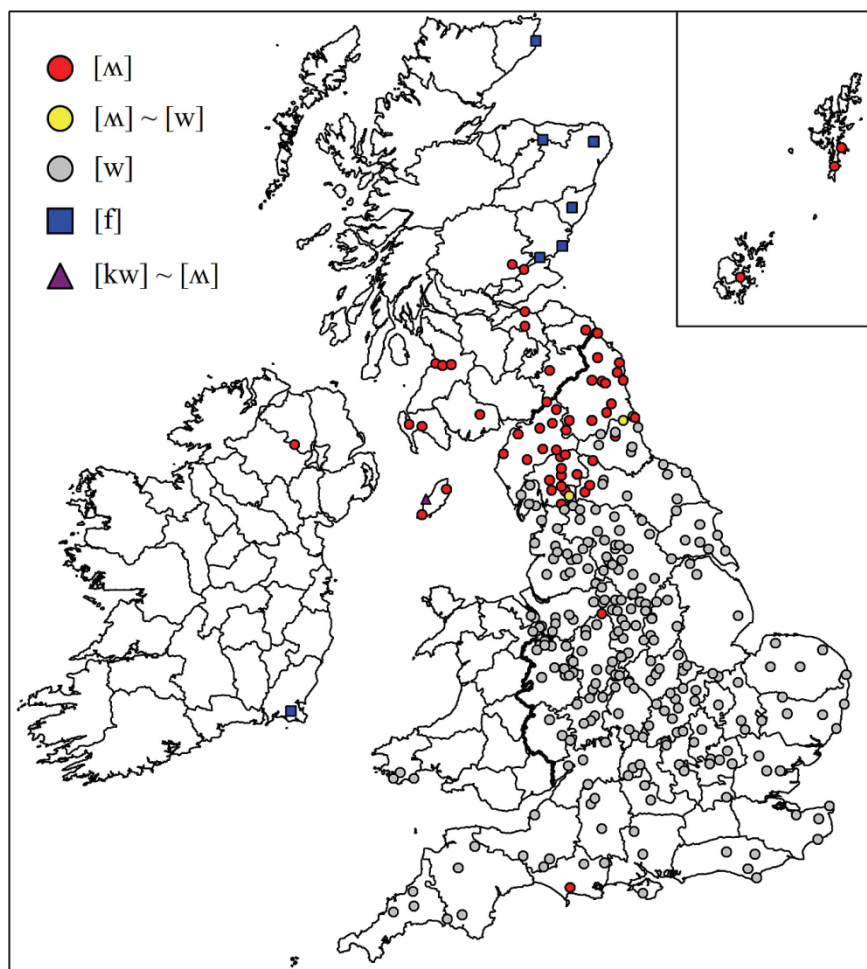
The other variants of this vowel (all grouped under the ‘other’ category) are not distinguished on the map because of their profusion and complex distributions, but they also compare well with the distribution of similar variants in the SED. Although there are exceptions and some variation at individual locations between variants, diphthongs of the type [ɛʊ] in the EPED predominate in East Anglia, the East Midlands, Kent and east Sussex, which is consistent with, though more restricted than, the pattern found in the SED. [aʊ]-type diphthongs are particularly typical of northern Lancashire, the north-western part of West Yorkshire, and parts of the north Midlands, just as they are in the SED. [əʊ] and [ʌʊ] are found elsewhere. This investigation of the pronunciation of the vowel in *down* in the EPED reveals that in this respect, at least, Ellis’s work is comparable to the SED, and suggests that Ellis’s work also captures the complexity of vowel variation in English and Scots dialects.

Although Ellis’s example words, especially in the short dt. and cs. texts, were selected to be representative of historical phonemic classes, analysis of only one word at a time runs the risk of the pattern being disrupted by blips (as a result of missing data, error, or atypical historical developments), and fails to take into account variation at single locations and the fact that variants are characteristic of groups of words, not just single lexical items. As such, it is useful to construct frequency maps (see Chambers and Trudgill 1980: 130, and Anderson 1987 for examples), which give us a deeper understanding of the distributions of variants and are particularly suited to identifying phonemic mergers and oppositions in geographical space. In order to construct a frequency map, we need more than one token, and preferably quite a few more than that, containing the linguistic variable under investigation. This linguistic variable must have at least two variants so that the relative frequency of occurrence of each at every location can be determined. Anderson (1987) is an excellent illustration of how this approach can be applied to traditional dialect data (see also Maguire 2012 and Maguire and McMahon 2011). Although the same approach can be adopted for the cartographic analysis of the EPED (see below), there are issues with Ellis’s data which means that it doesn’t always work very well in practice. If we wish to compare dt., cs. and (c)wl. data, it may well be the case that we have large numbers of relevant tokens for some locations (e.g. those with a relatively complete (c)wl.), and very little data, perhaps only one or

two tokens, for others (e.g. those with a dt. or only a handful of items in the (c)wl.). Creating a frequency profile for a location based on only one or two tokens is pointless and, instead, simpler forms of representing the variation which is present are required.

These issues can be exemplified by an examination of the distribution of variants corresponding to historical /hw/ (spelt <wh>) in word-initial position in the EPED. In the dt., there is only one relevant word, *where*. In the cs., there are six words (*what*, *when*, *where*, *while*, *whine* and *why*) of which two are represented by more than one token (two of *what*, three of *when*). In the (c)wl., there are nine words (*what*, *wheat*, *when*, *where*, *whew*, *which*, *while*, *whine* and *why*), although they are not all present at all locations. In a one-word-at-a-time approach, the only candidate word is *where*, and the distribution of major variants of the initial of this word is given in Figure 3.⁸

Fig. 3: The pronunciation of *wh-* in *where* in the EPED



⁸ For the purposes of this map, /hw/ covers Ellis’s (wh) [ʍ] and (kwh) [kw].

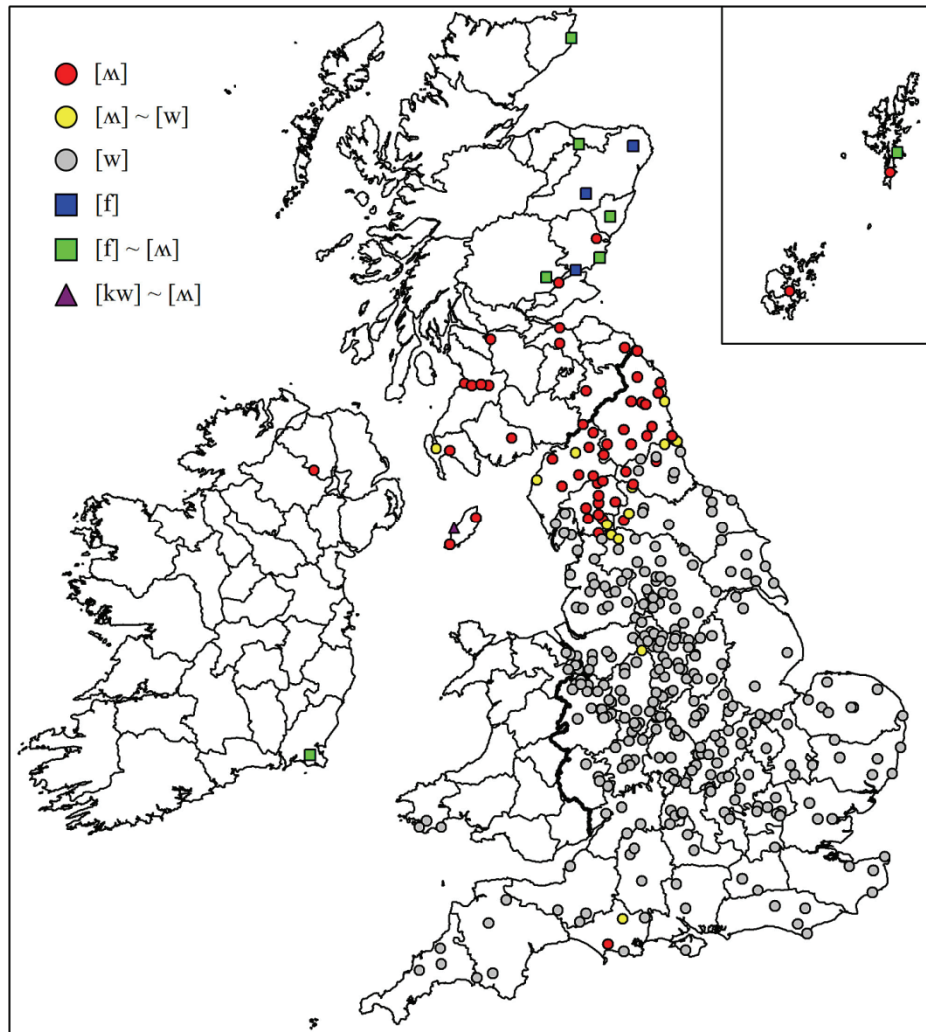
The change from historical /hw/ to [f] is recorded in the EPED in the extinct Forth and Bargy dialect of Wexford (where it is indicated in the spelling), and in mainland Scotland north of the Tay, which is consistent with what we know about these dialects from later studies (see McColl Millar 2007: 61-62). [kw] is recorded in variation with [ɹ] in Peel in the Isle of Man, as it is in *wh*- words in the southern SED location on that island. [ɹ] is recorded by Staples (1896) in Ulster, and in the EPED [ɹ] and [xw] are recorded throughout the rest of lowland Scotland, and [ɹ] is recorded in the northern English counties of Northumberland, Cumberland, Westmorland, and adjacent areas of Durham and Yorkshire. It is also recorded in two outlying locations, one in north Staffordshire and one in Dorset. In the SED, [ɹ] in *wh*- words is found in the Isle of Man, Northumberland (especially away from Tyneside), in northern Cumberland, north-western Durham, and sporadically in Westmorland (see Orton et al. 1978: Maps Ph222 and Ph223, and Anderson 1987: 145). Thus, the distribution of [ɹ] in the EPED and SED is rather different – ignoring the outliers for now, the EPED records [ɹ] in *where* throughout Cumberland, Westmorland and in the north-western fringe of Yorkshire, but not in west Durham. But apart from the two southern outliers, these areas aren't random – they form a continuous block with Northumberland and Scotland and, given that the SED records sporadic examples of [ɹ] in Westmorland, it looks as if there has been a change in the distribution of this feature in the 80 years separating the two surveys, with it disappearing from most of the far northwest of England. Note that Hirst (1906) consistently recorded /hw/ in *wh*- words in Kendal, Westmorland, confirming that [ɹ] has subsequently been lost from this area.

But this analysis ignores lots of relevant data in the EPED (nine *wh*- tokens in the cs. and up to nine tokens in the cwl.). Does an analysis of all the data give us a better understanding of the development of *wh*- in English and Scots dialects, and does it confirm the presence of [ɹ] in Westmorland and Cumberland, and in the two southern outliers? Figure 4 is a map based on the dt., cs. and (c)wl. data regardless of the number of tokens involved. Because of the relatively small numbers of tokens, it is not a frequency map, but it does capture some of the variation which is present in the data.

Figure 4 looks promising. It reveals essentially the same picture as Figure 3, but in more detail. It shows, for example, that [ɹ] is indeed consistently found in Cumberland and Westmorland, and is also found in west Durham (as the SED would lead us to expect), but also that there is some degree of variation between initial [ɹ] and [w] in varieties on the edge of the [w]-only zone. It also reveals that retention of [ɹ] is only variable in the north Staffordshire location ([ɹ] in three out of four tokens, including the dt.), and that there are in fact two locations in Dorset with some degree of

initial [ʌ], suggesting that this south-western county was a last southern stronghold for the historical distinction between *w*- and *wh*-.

Fig. 4: [ʌ] and [f] in all *wh*-words in the EPED



So far so good, but it must be remembered that Figure 4 includes the dt. data, which involve only a single token at each location (by necessity giving either 0 or 100% [ʌ]), and that, for the (c)wl. data in particular, several locations have very little data indeed. As a result, Figure 4 involves a rather broad categorisation of locations into those with 100% [ʌ], those with 0% [ʌ], and those with some variation between [ʌ] and [w] (and likewise for [f]). Whilst locations with cs. data consistently have between 7 and 9 *wh*-tokens, Table 1 reveals that there is much more variation for (c)wl. locations with at least some degree of initial [ʌ] or [f].

Although many of these locations have as many as six, seven, eight or nine tokens, making the percentage calculation at least indicative, others have very few tokens, and several have only one. This means that, like the dt. data, the (c)wl. data may over- or understate the degree to which [ʌ] (or [f]) is found in *wh*- words. Nevertheless, taking all of the relevant data in the EPED into account and establishing the broad categories indicated in Figure 4 helps to fill out the picture of the geographical distribution of variants, as long as we are careful in our interpretation of the results.

Table 1: Numbers of tokens of *wh*- in the (c)wl. data

Location	[w]	[ʌ]	[f]	%
D1 Forth and Bargo	0	1	2	66.7 [f]
D4 Hanford	2	2	0	50
D4 Winterborne Came	0	7	0	100
D23 Lazyre	0	1	0	100
D23 Peel	0	2	0	40 [ʌ], 60 [kw]
D23 Rushen	0	2	0	100
D26 Flash	1	2	0	66.7
D31 Chapel-le-Dale	6	1	0	14.3
D31 Horton-in-Upper-Ribblesdale	2	5	0	71.4
D31 Muker	1	5	0	83.3
D31 Laithkirk	1	6	0	85.7
D31 Dent	6	1	0	14.3
D31 Howgill	0	7	0	100
D31 St. John's Weardale	0	7	0	100
D31 Middleton-in-Teesdale	0	1	0	100
D32 Brampton	0	7	0	100
D32 South Shields	0	3	0	100
D32 Wark-on-Tyne	0	6	0	100
D32 Pitmatic	0	4	0	100
D32 Warkworth	2	4	0	66.7
Lowland Hawick	0	2	0	100
Lowland Liddesdale Head	0	7	0	100
Lowland Lothian	0	1	0	100
Lowland Chirnside	0	3	0	100
Lowland Ayr	0	3	0	100
Lowland Coylton	0	6	0	100
Lowland Glasgow	0	3	0	100
Lowland Cumnock	0	1	0	100
Lowland Ochiltree	0	7	0	100
Lowland Glenluce	0	4	0	100
Lowland Kirkpatrick Durham	0	3	0	100
Lowland Perth	0	3	1	25 [f]

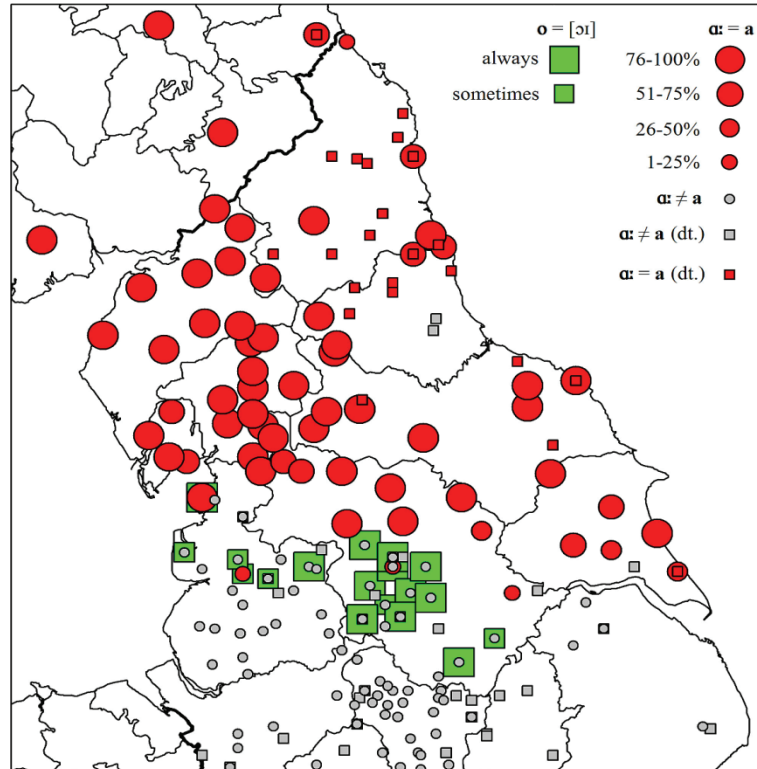
Lowland Glenfarquhar	0	3	4	57.1 [f]
Lowland Brechin	0	1	0	100
Lowland Keith	0	1	8	88.9 [f]
Lowland Buchan	0	0	4	100 [f]
Lowland Cromar	0	0	2	100 [f]
Lowland Wick	0	2	4	66.7 [f]
Lowland Orkney	0	3	0	100
Lowland Shetland	0	6	1	14.3 [f]
Staples (Ulster)	0	2	0	100

It is possible to go further than the broad categories in Figure 4 though. Figure 5 illustrates the geographical distribution of the different developments of Old English (OE) /ɑː/, found in words such as *home* and *stone*. English and Scots dialects north of the Ribble-Humber line merged OE /ɑː/ with the Middle English Open Syllable Lengthening (MEOSL) of /a/ (as in *name*), whilst south of this line it merged with the result of MEOSL of /o/ (as in *coat*), except for a small area of northwest England where the three source vowels remained distinct (see Wakelin 1984: 74 and Anderson 1987: 97–102).

The dt. contains three potential tokens of interest (*so*, *road*, *gone*), but only one of these (*gone*) is of use, since *so* is often unstressed and only occasionally has reflexes consistent with northern ME /ɑː/, and *road* never has a vowel which reflects northern ME /ɑː/, suggesting that it is not a typical member of the STONE lexical set. As discussed above, a single token is problematic from a frequency map perspective and, as a result, the dt. locations are indicated separately, as small squares, in Figure 5. The cs. includes ten tokens and the (c)wl. up to 27 tokens which definitely contained OE /ɑː/, so that in most cases there are sufficient tokens to establish a reliable frequency of variants. Variants of STONE have been grouped as follows for the purposes of the map: those which are consistent with merger (in northern ME /ɑː/) with the MEOSL of /a/ (i.e. the same as the vowel in words such as *name*) and those which are not. Locations with some degree of merger with MEOSL /a/ are indicated in red, those without in grey. Because of the number of relevant tokens in the cs. and (c)wl., the categories are narrower than those in Figure 4 (merger with MEOSL /a/ absent; 1–25% merger; 26–50% merger; 51–75% merger; and 76–100% merger).

Figure 5 also indicates (in green) locations where OE /ɑː/ has remained distinct not only from MEOSL /a/ but also from MEOSL /o/. These are locations which have developed an [ɔɪ]-type diphthong in words with MEOSL /o/ (no tokens in the dt., two in the cs., up to five in the (c)wl.), indicating that, although OE /ɑː/ and MEOSL /a/ have not (usually) merged, OE /ɑː/ has not merged with MEOSL /o/ either.

Fig. 5: The frequency of merger of MEOSL /a/ and OE /ɑ:/, and the development of [ɔɪ] in MEOSL /o/, as evidenced in the EPED.

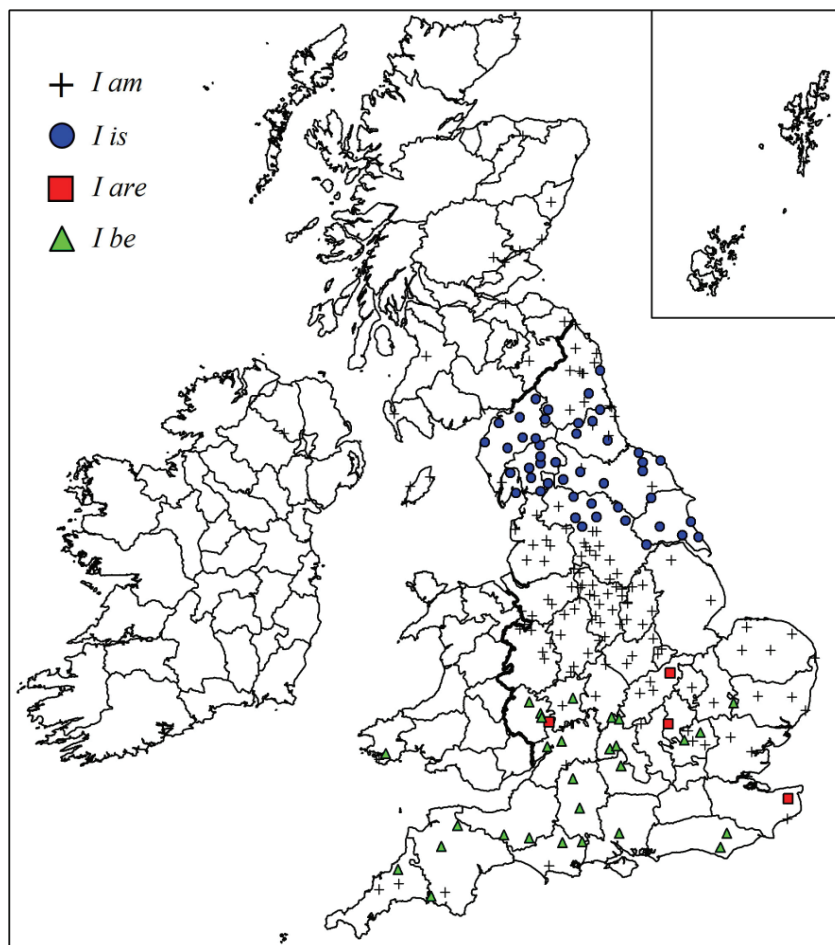


Since there are only small numbers of tokens in the cs. and (c)wl. (none in the dt.), a three-way categorisation is given in Figure 5: [ɔɪ]-type diphthong absent (no symbol); [ɔɪ]-type diphthong variably present (smaller green square); [ɔɪ]-type diphthong always present (larger green square).

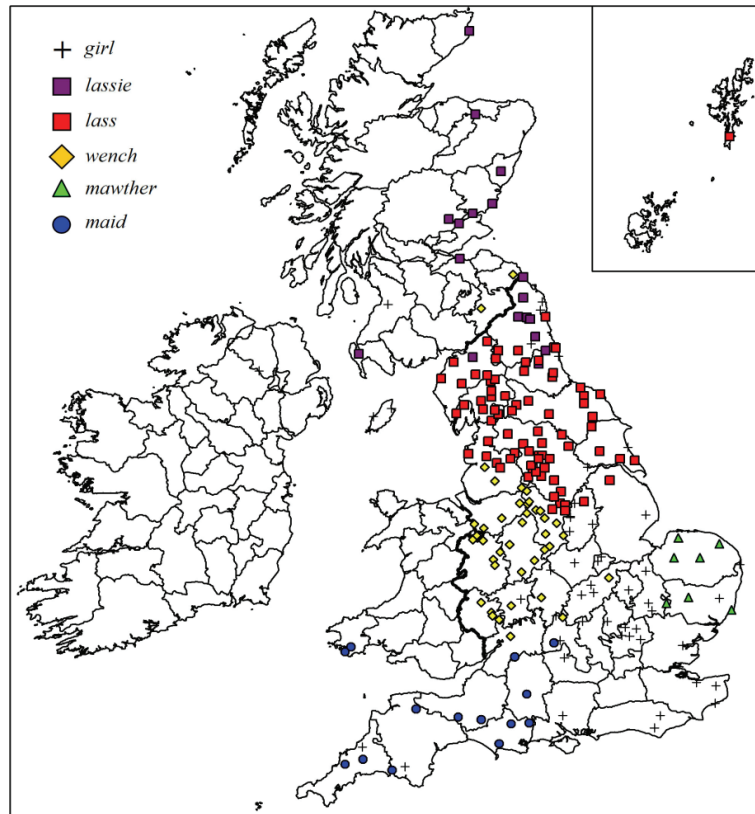
Figure 5 indicates that merger of OE /ɑ:/ with MEOSL /a/ is characteristic of dialects north of the Ribble-Humber line, as expected, with only two stray tokens being found further south, in Bolton and Windhill, perhaps representing borrowings from dialects further north, old survivals, or errors. Otherwise this pattern is almost identical to that found in the SED (see Anderson 1987: 112), and the dt. data for *none* fit the over-all pattern too. Although the pattern is less pronounced, the EPED parallels the SED in showing that the merger is less common in East Yorkshire, around Tyne-side and on the Northumberland coast than elsewhere in northern England. Note also that the merger is uncommon in Berwick-upon-Tweed just south of the Scottish/English border. This matches the findings of the LAS (Mather and Speitel 1986: 190–1), where only a single token (*clothes*) out of over 20 had the merger. Equally striking is the distribution of the [ɔɪ]-type diphthong in MEOSL /o/ words, which stretches from the mid-Lancashire coast into southern West Yorkshire as far as the Derbyshire and

Nottinghamshire borders. Again, this almost exactly matches the distribution of the same variant in the SED (see Anderson 1987: 114). Figure 5 indicates that the EPED is a faithful record of traditional English and Scots dialects, at least with respect to these developments, and that some of the EPED data can, with care, be represented in the form of frequency maps.

Fig. 6: Variants of *I am* in the EPED Dialect Tests and Comparative Specimens



Although the EPED is essentially a survey of traditional dialect phonetics and phonology, it does contain information on other linguistic levels. The *Dialect Test* and, particularly, the *Comparative Specimen* contain a range of morphological and syntactic variables, as well as a few lexical items, which vary from region to region. The *(Classified) Wordlist* contains various past tense and participle verb forms, pronouns, deictics and lexical items which similarly vary. Just to give an indication of the kinds of patterns that can be found in these data, Figure 6 illustrates a morpho-syntactic variable (*I am*) and Figure 7 a lexical one (*girl*).

Fig. 7: Words for *girl* in the EPED Dialects Tests and Comparative Specimens

In both cases, only data from the dt. and cs. are considered since the (c)wl. doesn't contain phrases such as *I am*, and the way the (c)wl. was elicited meant that informants could give a pronunciation for the word *girl* even if it wasn't their usual term for that meaning.

Four variants are shown in Figure 6: *I am*, *I is*, *I are*, and *I be*. *I is* is restricted to northern England north of the Ribble-Humber line and disappears again in the far northeast. *I be* is a southern variant, found especially in the southwest of England. *I are* is a rare southern and south Midland English variant. If we compare these patterns with data from the SED (Orton et al. 1978: Map M1), we find almost exactly the same pattern. Especially noteworthy are the absence of *I is* from north Northumberland, and the presence of *I are* in Kent and Bedfordshire, where it is also found in the EPED.

Six terms for 'girl' are found in the EPED, as indicated in Figure 7: *girl*, *maid*, *wench*, *mawther*, *lass*, and *lassie*. Although no map for this meaning is given in Orton et al. (1978), it is obvious from a perusal of the SED Basic Materials (question VIII.1.3) that the distributions of variants in it and in the EPED are very similar. In the SED, *mawther* is found only in Norfolk and Suffolk, *wench* predominates in the west Midlands and is found sporadically elsewhere, *maid* (or *maiden*) is restricted to the southwest of England, *lass* is

typical of the northern English counties, and *lassie*, typical of Scotland, is also found in Northumberland (see Upton et al. 1994, entry *lassies*).

4. Conclusions

The *Existing Phonology of English Dialects* is a vast compendium of information on English and Scots dialects. The mass of data it contains is both intriguing and perplexing – on the one hand it promises to enlighten us as to the nature of vernacular speech in the middle of the 19th century and on the other it seems impossible to discern any useful patterns in the morass of details. Undoubtedly there are problems with the EPED, given its narrow focus on the most traditional forms of speech, the indirect nature of much of its data collection, the small fragments of data which were collected at many locations, and the complexity of the presentation and representation. But that doesn't mean that the EPED is useless or that it is not worthy of investigation in its own right. The only way we can ever hope to determine whether the EPED has anything to tell us, and what that message might be, is to analyse it. An outline of one kind of analysis has been proposed in this paper, and this investigation reveals that such an approach is both possible and worthwhile if done with care. The map-based analysis in this paper shows that the EPED data contain consistent, historically coherent patterns which compare very favourably with later surveys which, whilst not without their own flaws, may be trusted to give a faithful representation of certain kinds of speech in the middle of the 20th century. For every feature examined in this article in the EPED, consonantal, vocalic, morpho-syntactic or lexical, the pattern is either identical to that revealed by later data or is consistent with an earlier stage in its history. Assuming that these cases are typical of the whole dataset, and I have no reason to think that they are not since they were not specially selected from a mass of ‘less convenient’ patterns, the conclusion must be that the EPED is a valuable source of information. However, this is only the first step in the construction of a linguistic atlas of the EPED, and much more work remains to be done.⁹ It is hoped that this ongoing work will contribute to a recognition of the importance of this unique and monumental work for the history of dialectology in Britain and Ireland and for the history of the languages spoken there.

⁹ An online atlas of the EPED is under development, and may be accessed at www.lel.ed.ac.uk/EllisAtlas. This atlas contains many more maps based on Ellis's data than it has been possible to provide here, and it can be seen that the maps contained in this paper are not atypical.

References

- Anderson, Peter. 1977. A new light on Early English Pronunciation. *Transactions of the Yorkshire Dialect Society* 77, 32–41.
- Anderson, Peter. 1987. *A structural atlas of the English dialects*. North Ryde, New South Wales: Croom Helm.
- Chambers, J. K. and Peter Trudgill. 1980. *Dialectology*. Cambridge: University Press.
- Dieth, Eugen. 1946. A new survey of English dialects. *Essays and studies* 32, 74–104.
- Ellis, Alexander. 1869–1889. *On early English pronunciation*. New York: Greenwood Press.
- Ellis, Alexander. 1889. *The existing phonology of English dialects, compared with that of West Saxon speech*. New York: Greenwood Press.
- Eustace, Sinclair. 1969. The meaning of the Palaeotype in A. J. Ellis's On Early English Pronunciation 1869–89. *Transactions of the Philological Society* 68, 31–79.
- Gilliéron, Jules. 1902–1910. *Atlas linguistique de la France*. Paris: Champion.
- Hirst, Tom Oakes. 1906. *A grammar of the dialect of Kendal (Westmorland)*. Heidelberg: Winter.
- Jaberg, Karl & Jacob Jud. 1928–1940. *Sprach- und Sachatlas Italiens und der Südschweiz*. Zofingen: Ringier.
- Jones, Mark. 2002. The origin of Definite Article Reduction in northern English dialects: evidence from dialect allomorphy. *English language and linguistics* 6, 325–345.
- Kökeritz, Helge. 1932. *The phonology of the Suffolk dialect: Descriptive and historical*. Uppsala: A.-B. Lundequistska Bokhandeln.
- Kurath, Hans, Miles Hanley, Bernard Bloch & Guy Lowman. 1939–1943. *Linguistic atlas of New England*. Providence, RI: Brown University Press.
- Kurath, Hans & Guy Lowman. 1970. *The dialectal structure of southern England*. Tuscaloosa: University of Alabama Press.
- Local, John. 1983. Making a transcription: The evolution of A. J. Ellis's Palaeotype. *Journal of the International Phonetic Association* 13, 2–12.
- Maguire, Warren. 2003. "Mr. A. J. Ellis – the pioneer of scientific phonetics in England" (Sweet 1877, vii): an examination of Ellis's data from the northeast of England. Paper presented at the *Henry Sweet Society for the History of Linguistic Ideas annual colloquium*, August 2003. Trinity College, Dublin.
- Maguire, Warren. 2012. Pre-R dentalisation in northern England. *English language and linguistics* 16, 361–383.
- Maguire, Warren & April McMahon. 2011. Quantifying relations between dialects. In Warren Maguire & April McMahon (eds.) *Analysing variation in English*, 93–120. Cambridge University Press.
- Mather, James & Hans-Henning Speitel. 1975, 1977, 1986. *The linguistic atlas of Scotland, Scots section*. Beckenham: Croom Helm.
- McColl Millar, Robert. 2007. *Northern and insular Scots*. Edinburgh: University Press.
- Orton, Harold and Eugen Dieth (eds.). 1962–1971. *Survey of English dialects*. Leeds: Arnold & Son.

- Orton, Harold, Stewart Sanderson & John Widdowson (eds.). 1978. *The linguistic atlas of England*. London: Routledge.
- Petyt, Keith. 1980. *The study of dialect: An introduction to dialectology*. London: Deutsch.
- Rydland, Kurt. 1982. *Vowel systems and lexical-phonemic patterns in south-east Cumbria: A study in structural dialectology*. Bergen: University of Bergen.
- Rydland, Kurt. 1998. *The Orton corpus: A dictionary of Northumbrian pronunciation, 1928–1939*. Oslo: Novus Press.
- Shorrocks, Graham. 1991. A. J. Ellis as dialectologist: A reassessment. *Historiographia linguistica* XVIII, 321–334.
- Staples, J. H. 1896. Notes on Ulster English dialect for comparison with English dialects by the late A. J. Ellis, F.R.S., with samples in Palaeotype, comparison specimen and wordlist. *Transactions of the Philological Society* 23, 357–398.
- Sweet, Henry. 1877. *A handbook of phonetics*. Oxford: Clarendon Press.
- Trudgill, Peter. 2004. *New-dialect formation: The inevitability of colonial Englishes*. Edinburgh: University Press.
- Upton, Clive, David Parry & John Widdowson. 1994. *Survey of English dialects: The dictionary and grammar*. London/New York: Routledge.
- Wakelin, Martyn. 1972. *English dialects: An introduction*. London: Athlone Press.
- Wakelin, Martyn. 1984. Rural dialects in England. In Peter Trudgill (ed.), *Language in the British Isles*, 70–93. Cambridge: University Press.
- Wells, John. 1982. *Accents of English*. Cambridge: University Press.
- Wettstein, Paul. 1942. *The phonology of a Berwickshire dialect*. Bienne: Schüler.
- Wright, Joseph. 1892. *A grammar of the dialect of Windhill in the West Riding of Yorkshire*. London: Kegan Paul, Trench, Trübner & Co.
- Wright, Joseph. 1905. *The English dialect grammar*. Oxford et al. Henry Frowde.
- DiWA – Digital Wenker-Atlas. <http://www.diwa.info/> (accessed 17/09/2012).
- DMap – Distribution mapping software. <http://www.dmap.co.uk/> (accessed 17/09/2012).
- Ellis Atlas website – <http://www.lel.ed.ac.uk/EllisAtlas> (accessed 17/09/2012).

Warren Maguire • University of Edinburgh • w.maguire@ed.ac.uk