Dialectologia. **Special issue, I** *(2010), 161-173.*
ISSN: 2013-22477

CURRENT TRENDS IN BRITISH GEOLINGUISTICS
LINKING THE PAST WITH THE PRESENT

Heinrich RAMISCH
University of Bamberg
heinrich.ramisch@uni-bamberg.de

Abstract

The general aim of this presentation will be to discuss some major developments in British geolinguistics, especially as far as methodological aspects are concerned. Before discussing important cartographical and computational procedures for the production of linguistic maps within our own project *The Computer Developed Linguistic Atlas of England* (CLAE, cf. Viereck/Ramisch 1991 and 1997) I will give a short description of the database for the atlas, namely the *Survey of English Dialects* (SED). The SED is still the best known and most widely used British dialect survey that has served as a database for a variety of linguistic atlases on the regional dialects of England. The main objective of the SED was to collect linguistic data on the traditional regional dialects of England. Its questionnaire comprises over 1,300 items and is concerned with different spheres of rural life such as farming, animals, nature, housekeeping, weather and various social domains. Fieldwork was carried out in 313 localities all over England between 1950 and 1961, mainly in small villages with a population of less than 500 people.

As for the production of linguistic maps it is clear that modern computer technology offers considerable advantages. Firstly, large amounts of data can be processed and analysed. Secondly, computer cartography is very flexible in itself, offering a great variety of mapping techniques. It is possible, for example, to produce complex symbol maps to display detailed dialectological information. In contrast to the *Linguistic Atlas of England* (Orton et al. 1978) symbol maps are used in the CLAE to depict regional variation. Symbol maps have the particular advantage that they are able to show transitional zones and they are therefore a more accurate representation of linguistic reality. Another characteristic feature of the CLAE is that the comments by the informants and fieldworkers are integrated into the symbols. The symbolization itself follows several basic principles. If a certain form occurs frequently, it is assigned a relatively simple, that is, strictly geometrical symbol. If a form is rather rare the symbol is more complex. Additionally, differences in frequency are indicated by the size of the symbols. With frequently occurring forms the symbols are printed in a smaller format, whereas larger symbols are used for rarer forms. As a result, they should attract the user's eye more directly.

With respect to computational procedures it seems advisable to use standardised programmes both for the production of the maps and the legends. As these programmes can be adapted to one's individual

©Universitat de Barcelona

needs, any time-consuming programming is generally avoided. Modern word-processing programmes easily allow the integration of graphical elements, while mapping programmes such as PCMAP include options to insert textual elements. For the basic data collection and the legends we use a standard text processor, MS Word. This means, above all, that we can take full advantage of the formatting facilities of MS Word for the production of the legends. Moreover, graphical elements can easily be included into a text file produced by MS Word. With the help of our mapping programme PCMAP it is possible to produce so-called 'thematic maps', which are automatically produced by changing a base map into a map whose features depend on the information found in a basic data file.

Finally, I will report on some other recent projects in British geolinguistics. In cooperation with the University of Leeds the 'BBC Voices Project' was set up to study variation in British English by including aspects such as urban centres, age, gender and ethnic group as well as social and geographical factors. It provides a forum on the Internet, where people from the general public can find information on regional forms of speech. They are also encouraged to send in lexical items from their home region to build up a lexicon of regional speech. Other projects include the *Atlas of English Surnames*, edited by colleagues from Bamberg, and the digitisation of Joseph Wright's *English Dialect Dictionary* which is currently undertaken at the University of Innsbruck (Austria).

Key words

geolinguistics and techniques of electronic data processing, *Computer Developed Linguistic Atlas of England* (CLAE), *Survey of English Dialects* (SED), symbol maps vs. isogloss maps, mapping techniques, PCMAP as a mapping programme, MS Word for basic data collection and production of legends, use of macros in MS Word, base map and thematic map in PCMAP, BBC Voices Project, Atlas of English Surnames, Wright's *English Dialect Dictionary* in electronic form

Introduction

There can be no doubt that as a result of the rapid advancement in computer technology over the last few decades, the field of geolinguistics has experienced substantial changes. A new generation of linguistic atlases has come into existence which is profoundly influenced by modern computing. Two aspects seem to be particularly noteworthy in this context. First, it is possible to record and to store large amounts of data in the form of data bases. Secondly, the data can be searched automatically, be processed and be visualised effectively by computer cartography. The history of our own project in Bamberg, the *Computer Developed Linguistic Atlas of*

©Universitat de Barcelona

England (CLAE, cf. Viereck/Ramisch 1991 and 1997), also reflects this general progress in geolinguistics.

1. Using the *Survey of English Dialects* as a database

Before discussing some major dialectological and computational aspects of the atlas, I would like to give a short description of the database that we employed, that is the 'Basic Material' of the *Survey of English Dialects* (or SED). The SED is still the best known and most widely used English dialect survey. The project was originally designed by Eugen Dieth (University of Zurich) and Harold Orton (University of Leeds). The main objective of the SED was to collect linguistic data on the traditional regional dialects of England. Fieldwork was carried out by trained linguists in 313 localities all over England between 1950 and 1961, mainly in small villages with populations of less than 500 people. The SED can indeed be regarded as a continuation of earlier research in English dialectology, but it is also markedly different, especially with respect to methodology. Like many previous dialectological studies, the general orientation of the SED is first and foremost diachronic. The founders of the project were linguistic historians who were concerned with nonstandard dialectal forms as manifestations of earlier forms of the language. Accordingly, it was the objective of the SED to record the oldest kind of vernacular speech. Local pronunciations of the item 'eyes' such as *een* [i:n] for example, are of particular interest as they can inform us about certain historical vowel developments or the formation of the plural in this case. The occurrence of pronouns such as *thou/thee* in various parts of England provides evidence for the survival of an older pronominal system. Or the northern use of the lexical item *lake* [le:k] for to 'play' in Standard English represents an example of the Scandinavian influence on English.

In order to collect the oldest and most localized data, it was clear that the SED was primarily interested in informants that were influenced as little as possible by other social, educational or geographical factors. The informants used for the project have commonly been described as NORM-informants, the individual letters of this acronym standing for non-mobile, older, rural, males. It goes without saying that the informants that were interviewed for the SED definitely do not constitute a representative sample of the whole population of England. Most of the 955 informants were farm workers of the

©Universitat de Barcelona

older generation. Consequently, the SED data cannot be analysed in a meaningful way with respect to factors such as age, gender or social class.

As far as methodology is concerned the SED project was rather rigorous in its fieldwork techniques to make sure that the collected data was as reliable and as comparable as possible. For one thing, the data was not collected in an indirect way, for example, by a postal questionnaire or with the assistance of voluntary helpers. The SED fieldworkers were trained linguists, who had learnt the different questioning techniques and phonetic transcription and who knew about the reasons for each question. The fieldworkers visited informants in their home environment and carried out on-the-spot interviews. The whole SED questionnaire includes over 1,300 items and is concerned with different spheres of rural life such as farming, animals, nature, housekeeping, weather and a variety of social domains.

2. Mapping procedures of the *Computer Developed Linguistic Atlas of England*

In the series of atlases that have used the SED material, the CLAE was the first one to make use of computer technology. The first two volumes of our atlas include 150 lexical, 121 morphological and 50 syntactic maps. Additionally, Volume 2 includes an appendix with dialectometrical contributions, using various quantitative methods to make generalisations from the SED data and to try to define linguistic areas. The studies include, among others, multidimensional scaling, dendrographic classifications, a multivariate analysis as well as a study entitled “selective dialectometry” where the data is subjected to a specially developed gravity centre method. In fact, two participants at this symposium wrote articles for the second volume. Hans Goebl presented a dialectometrical analysis of the CLAE data, dealing with isoglosses and dialect integration. Chitsuko Fukushima examined processes of standardization of a great variety of morphological forms to be found in the data.

Probably the most essential feature of the CLAE is that only symbols are used to depict regional variation. In contrast to the *Linguistic Atlas of England* [LAE, Orton et al. 1978] no isoglosses appear on the maps. The major advantage of symbol maps is that they are able to show transitional zones and are therefore a more accurate representation of linguistic reality. Isoglosses can certainly be useful as a way of abstracting from this

©Universitat de Barcelona

reality, but they run counter to the general principle, namely to influence the user of the atlas as little as possible. Every map of the CLAE is accompanied by a complete documentary list giving the individual responses for each of the 313 localities of the SED. These lists also inform the reader which forms are subsumed under one symbol or are not mapped at all. Moreover, the respective question of the SED questionnaire is cited and in a number of cases pictures help to identify a given item.

Another characteristic feature of the CLAE is that the comments by the informants and fieldworkers are integrated into the symbols. The following seven status categories are distinguished: 1) 'usually, familiarly'; 2) 'rare, occasionally, less common'; 3) 'older, obsolete'; 4) 'modern, newer'; 5) '(strong) pressure, suggested form/word'; 6) 'preferred'; 7) 'excerpted from incidental material'. Each of these categories is represented by slightly altering the basic symbol (cf. for example the map *Rind* L 16, CLAE vol. 2). The cartographical system made it possible to integrate two of the categories into one and the same symbol, if necessary. The symbolization itself follows several basic principles. If a certain form occurs frequently, it is assigned a relatively simple, that is, strictly geometrical symbol. If a form is rather rare the symbol is more complex. Additionally, differences in frequency are indicated by the size of the symbols. With frequently occurring forms the symbols are printed in a smaller format, whereas larger symbols are used for rarer forms. As a result, they should attract the user's eye more directly. Evidently, highly detailed and complex symbol maps of this type can only be produced with the assistance of computer technology.

A good example to illustrate the relevance of comments by informants or fieldworkers is the map '*Rind'* (L 16, CLAE vol. 2) One can notice that the form *sward* or *swath* is still present in the north and the west of England. But it frequently co-occurs with *rind*, indicated by the double symbols. *Rind* was indeed given as a first response by many informants. The horizontal lines inside the symbols show that the forms *sward/swath* were sometimes suggested by the fieldworker. Additionally, *sward/swath* are described as 'older' or 'preferred' by some informants, showing that they are regarded as the more traditional dialect words. Some readers may argue that the combination of 'suggested word' and 'preferred' with respect to one response seems somewhat unusual, if not contradictory. But such a combination is not uncommon in the SED material. Of course, the informants knew that they were being interviewed about their traditional dialect. Even if a certain form was suggested to them they could still

©Universitat de Barcelona

confirm that this form was the genuine (that is, preferred) form in their own local dialect. As *sward/swath* were often suggested or described as 'older', one can assume that they have become more and more obsolete and will probably be replaced by the standard word *rind*. All the maps in the CLAE are accompanied by a full documentary list, which is placed on the left hand side in the atlas. The list presents the individual responses for all the localities. In these lists one can also check which responses are summed up under one and the same symbol. A notorious problem in linguistic cartography is the treatment of possibly related forms. In fact, the linguist/cartographer has to decide whether these items are assigned just one symbol or two similar symbols or two completely different symbols. In the map *Rind*, the editors decided to regard the item *sward* and *swath* as two separate but lexically related forms. Accordingly, they are assigned similar symbols (a full circle for *sward* and an indented circle for *swath*).

Generally, it seems advisable to restrict the number of different symbols on a dialect map. A map that includes too many different symbols may in the end become confusing and difficult to read. In the CLAE project the number of symbols is normally limited to a maximum of seven or eight. In a few cases it was decided to present the information not just on one, but on two base maps. Responses that occur less than five times are usually not mapped at all.

3. Computational procedures of the *Computer Developed Linguistic Atlas of England*

Originally the CLAE was a joint project of the Chair of English Linguistics and Medieval Studies at the University of Bamberg and the University of Marburg, where the German linguistic atlas ('Deutscher Sprachatlas') is based. Our colleagues in Marburg carried out the software development for the first two volumes. For CLAE 1 they wrote their own programmes for a mainframe computer in 'Fortran'; for CLAE 2 they worked on a PC and used C + + as a programming language (cf. Händler/Marx 1997: XII). After the completion of CLAE 2 we had to find new computational procedures for the production of linguistic maps. At the same time this gave us the opportunity to incorporate recent software developments into our project.

©Universitat de Barcelona

The basic concept of our colleagues in Marburg was a largely unitary software system (cf. Händler 1991: 12) that produced both the textual (legends) and the graphical elements (maps and symbols). Such a procedure required a substantial amount of programming. In considering present-day software packages it seems far more practical to abandon the idea of a unitary system. Modern word-processing programmes easily allow the integration of graphical elements, whereas mapping programmes such as PCMAP also include options to insert textual elements. Our approach has therefore been to use different programmes for the productions of legends and maps. In proceeding like this, we can indeed make good use of existing, standard programmes and adapt them to our needs, if necessary. For the basic data collection and the legends we use a standard text processor, MS Word. This means, above all, that we can take full advantage of the formatting facilities of MS Word for the production of the legends. Moreover, graphical elements can easily be included into a text file produced by MS Word. For example, the legend for the item *'Rind'*, which includes a variety of symbols, can entirely produced in MS Word.

Figure 1 shows an example of a basic data file in MS Word. The basic data files form the basis for both the legends and the maps. Essentially, a basic data file is a table in MS Word, organised in columns and rows. The initial table already includes Columns A, B and D as these columns need not to be altered. In Column A the 313 SED localities are identified by numbers. These numbers are subsequently employed by the mapping programme PCMAP, in which each locality is assigned a particular number. Column B includes the locality code for the first response. Nb1, for example, stands for locality 1 in Northumberland, Cu2 for locality 2 in Cumberland. The locality codes are later to be used in the legend (cf. Figure 1). In the cells in Column C the first response at a particular locality can be entered. Normally, this can be done in a semi-automatic way, as the different forms can be inserted by MS Word macros. Column D again includes the locality codes for a possible second response. Second responses occur quite frequently in the SED Basic Material. They can be entered in Column E.

Column F is produced by an automated searching process in Columns C and E. Column F is crucial for the automatic insertion of symbols by PCMAP, which just requires figures for their identification. 20, for example, stands for the ordinary circle, 60 for the indented circle, and 50 for the cloud symbol. The figure 2050 at locality Cu2 represents a combined symbol (ordinary circle plus cloud). Additionally, within

©Universitat de Barcelona

PCMAP it is necessary to determine the size of symbols. This is done by the figures in Column G. The figures in Column G are automatically generated from the figures in Column F.

A locality	B locality code	C 1st response	D locality code	E 2nd response	F type of symbol	G size of symbol
1	Nb1,	hɪɔˣs	Nb1,		60	45
2	Nb2,	hɪɔˣs	Nb2,		60	45
3	Nb3,	hɪɔʁəs	Nb3,		60	45
4	Nb4,	hɪɔˣs	Nb4,		60	45
5	Nb5,	hɪəˣs	Nb5,		60	45
6	Nb6,	hɪəs	Nb6,		60	45
7	Nb7,	hɪɔˣs	Nb7,		60	45
8	Nb8,	hɪɔˣs	Nb8,		60	45
9	Nb9,	hɪəs	Nb9,		60	45
10	Cu1,	hɪərs	Cu1,		60	45
11	Cu2,	həːs	Cu2,	jəs	2050	331
etc.	etc.	etc.	etc.		etc.	etc.

Figure 1. Extract of a Basic Data File (in MS Word): item '*hearse*'

Another important procedure is the automatic production of the legends from the basic data files with the help of a rather complex MS Word macro (cf. Figure 2). The key element within this macro is a so-called 'while-wend cycle', which makes sure that a series of commands is repeatedly carried out, until all the instances of a particular response have been transferred to the legend. The feature analysed in the example is h-dropping (the non-pronunciation of [h-] in initial position, cf. Wells 1982: 252ff.). The macro first searches for the next occurrence of a response that does not have an [h-] in initial position. It then deletes the response and goes back to the preceding cell in which the locality code is found. The locality code is copied and transferred to a new file. After going back to the basic data file the same process can start anew.

©Universitat de Barcelona

```
Public Sub MAIN()
Dim x
WordBasic.ScreenUpdating 0
WordBasic.NextWindow
WordBasic.StartOfDocument
x = -1
While x = -1
WordBasic.EditFind Find:="xyz", Direction:=0, MatchCase:=1,
WholeWord:=1, PatternMatch:=0, SoundsLike:=0, Format:=0, Wrap:=1
x = WordBasic.EditFindFound()
If x = -1 Then WordBasic.EditClear: WordBasic.PrevCell: WordBasic.EditCopy:
WordBasic.NextWindow: WordBasic.EditPaste: WordBasic.NextWindow
Wend
WordBasic.NextWindow
WordBasic.EditClear -1
WordBasic.InsertPara
WordBasic.InsertPara
WordBasic.Insert Chr(9)
WordBasic.ScreenUpdating 1
End Sub
```

Figure 2. Macro (MS Word) for the production of a legend

For the production of computer-drawn maps with PCMAP first a base map has to be digitised. The actual procedure is to load a scanned image of the SED network (e.g. in the form of a Windows bitmap) into the programme. Then one can zoom into the right place and use the mouse on-screen to digitise lines, areas and localities. It is, therefore, no longer necessary to use a digitising board or any other special equipment for producing computer-drawn maps. Textual elements can be integrated into the map, using any font installed under Windows, including phonetic symbols. It is a special feature of PCMAP that different graphical 'types' can be defined, which can be

©Universitat de Barcelona

switched ‘on’ and ‘off’ as is required. On some maps it may be useful, for example, to delete county boundaries or the names of counties.

The next step is to change the base map into a so-called ‘thematic map’, which is automatically produced by the programme, depending on external data. In our project this is a symbol map that is drawn according to the information in an ASCII file, based on Columns A, F and G of the basic data file. Other options for the production of a thematic map in PCMAP include: area colouring, area hatching, different colours of symbols or different symbol hatchings. The last-mentioned mapping technique is particularly useful for quantificational maps. Moreover, the base map could be directly linked to other databases. For instance, a DDE (direct data exchange) connection can be established with MS Excel or MS Access. With an ODBC (open database connection) it is equally possible to connect localities with other Windows applications to display text files, for example, or to play audio files. In this way a ‘speaking’ atlas can be accomplished within PCMAP. One feature of PCMAP is particularly relevant for existing projects, including our own, namely that symbols can be generated freely. Any graphical element can indeed become a symbol.

4. Other recent projects

A new type of research project in British geolinguistics has been initiated by the BBC in cooperation with Sally Johnson and Clive Upton from the University of Leeds. So far, the ‘BBC Voices Project’ has collected over 700 hours of regional English speech (cf. <*http://www.bbc.co.uk/voices/*>). The aim of the project is to study variation in British English by including aspects such as urban centres, age, gender and ethnic group as well as social and geographical factors. 51 BBC journalists recorded the speech of 1,200 people from all over the UK in 2004 and 2005. Additionally, the informants were given questionnaires to state which words they use for everyday concepts such as illness, wealth, relationships and clothing. At the same time, the project takes advantage of the interactive facilities of the Internet. The regional distribution of lexical items can be displayed in the form of “word maps” on your personal computer or one can listen to audio clips from different areas. People have been asked to send in words from their own region with the objective to build up a

©Universitat de Barcelona

detailed lexicon of regional vocabulary. The project will also study how language is reported in the media. For this purpose a large number of news articles has been collected, ranging from reports about slang or swear-words to articles that describe how language items have appeared in the news. The website of the BBC Voices Project has been well received as a forum of discussion. Clive Upton explains people's general interest in the project in the following way, "It's about their identity. It's about who they are and how they express that – and these are things which people fell strongly about."

A geolinguistic project of a rather special nature is the *Atlas of English Surnames* that was compiled by colleagues from Bamberg. The atlas includes over 240 maps of surname variants, and more than 150 tables and figures that can be used for further research on local or occupational surnames and nicknames. Each surname is discussed with respect to its etymology, historical background and geographical distribution. The data that provides the basis for the atlas was mainly extracted from telephone directories that were published on CD-ROM. Other sources of information were census records and parish registers.

Finally, I would like to report on a project that is currently carried out by Manfred Markus and his colleagues at the University of Innsbruck (Austria). The aim of this project is to digitise and to assess the six large volumes of Joseph Wright's *English Dialect Dictionary* [EDD]. The dictionary, which was published between 1896-1905, is still the most comprehensive and reliable source on regional dialects used in the 18th and 19th centuries, even providing more information on dialectal forms than the well-known *Oxford English Dictionary* (OED). Once the electronic form of the EDD is available, it will be an important research tool not just for dialectological studies but also for English historical linguistics, the use of spoken and written English and related fields (for more information on this project, cf. <*http://www.uibk.ac.at/anglistik/projects/speed/index.html*)>.

5. Conclusion

The various projects described in this article amply illustrate the prominence of electronic data processing in recent geolinguistic studies in Britain. New ways of analysing large dialect corpora have become possible and one can only be pleased to see

©Universitat de Barcelona

that there is such a strong and renewed interest in the regional variation of English. On the other hand, it is also true to say that studies on regional variation have been somewhat neglected in the past. In Britain, the sociolinguistic research paradigm has been particularly strong over the last few decades. From this point of view, one may indeed regret, for example, that there have been no substantial research initiatives to work on regional dialect atlases, as has been the case in other European countries such as France or Germany. But linguists certainly agree that geolinguistic data frequently provide fascinating insights into our own cultural and linguistic history, thus truly linking the past with the present.

6. References

BARKER, Stephanie et al. (2007) *An Atlas of English Surnames*, Frankfurt: Peter Lang.

HÄNDLER, Harald (1991) "Computational Aspects", in: W. VIERECK, in collaboration with H. RAMISCH, 1991, 9-15.

HÄNDLER, Harald and Christian MARX (1997) "Computational Aspects", in: W. VIERECK and H. RAMISCH, 1997, XII.

ORTON, Harold et al. (1962-71) *Survey of English Dialects. The Basic Material*, Leeds: E. J. Arnold. [SED]

ORTON, Harold et al. (1978) *The Linguistic Atlas of England.* London: Croom Helm. [LAE]

RAMISCH, Heinrich (1997) "Dialectological and cartographical features of the *Computer Developed Linguistic Atlas of England (CLAE)*", in: Alan R. THOMAS (ed.) *Issues and Methods in Dialectology*, Bangor: Department of Linguistics, University of Wales Bangor, 224-233.

RAMISCH, Heinrich and Wolfgang VIERECK (2006) "Recent Developments in Computer Cartography", in: H. GRABES and W. VIERECK (eds.), *The Wider Scope of English*, Frankfurt: Peter Lang, 67-78.

UPTON, Clive and J. D. A. WIDDOWSON (2006) *An Atlas of English Dialects*, Oxford: OUP (2nd edition).

VIERECK, Wolfgang, in collaboration with Heinrich RAMISCH (1991) *The Computer Developed Linguistic Atlas of England 1*. Computational production: Harald Händler, Petra Hoffmann, Wolfgang Putschke, Tübingen: Niemeyer. [CLAE 1]

VIERECK, Wolfgang and Heinrich RAMISCH (1997) *The Computer Developed Linguistic Atlas of England 2*. Computational production: Harald Händler and Christian Marx. With

©Universitat de Barcelona

dialectometrical contributions by: Sheila Embleton, Chitsuko Fukushima, Hans Goebl, Harald Händler, Fumio Inoue, Guillaume Schiltz, Alan R. Thomas, Wolfgang Viereck and Eric Wheeler. Tübingen: Niemeyer. [CLAE 2]

WELLS, John (1982) *Accents of English*, Cambridge: CUP (3 vols.).

WRIGHT, Joseph (1898-1905) *The English Dialect Dictionary*, Oxford: Clarendon Press. [EDD]

©Universitat de Barcelona