Onomastic lexicography

Patrick Hanks & Richard Coates

Abstract

Modern scholarship and techniques of data analysis have shown that even the best current dictionaries of surnames in Britain are full of errors, oversights, guesswork, fudges, and omissions. In this paper we present a new project designed to rectify this situation. A database has been compiled containing entries for all the family names in Britain. Entries in this database for family names with more than 100 bearers — and for many less frequent names, too — are being systematically compared with data on medieval surnames and with a geodemographic analysis of the 1881 census. The associations between surnames and localities are explored systematically, with results that quite often have a profound effect on our understanding of the origins and etymology. The English, Celtic, French, and Scandinavian etymologies of native names are investigated using the best techniques of historical linguistic scholarship. The national identity of recent immigrant names is explained.

1. Names vs. words

The great European historical dictionaries of the 19th and 20th centuries (Grimm, OED, Tresor, etc.) applied comparative and historical linguistic methods to the elucidation of the history of the forms and meanings of each word in a national language. As a matter of policy, personal names and place-names were not included. This omission was justified on the grounds that names are different from words: names are devices for referring to individuals in context, rather than classes of entities or events.

As far as names dictionaries are concerned, a small number of scholarly accounts of place-names have been compiled—although there are still many serious gaps; for example, there is still no satisfactory dictionary of Scottish place-names. Dictionaries of first names tend, especially in America, to be "Name your baby" books—often promulgating nonsensical associations of qualities with names (e.g. "**Richard** denotes a strong and courageous person") although a few more scholarly works were published. In the field of surname studies, lexicography in European and other languages has lagged far behind other kinds of lexicography and indeed other kinds of names studies.

In the 21st century, this state of affairs is changing. Very large electronic databases of names records are becoming available, statistical methods of analysis are being applied, and scholars are beginning to explain systematically the origins and history of surnames as well as words and place-names.

In this paper we focus on a new development in surname lexicography in Britain. We present the goals and methodology of a major new research project called *Family Names of the United Kingdom* (abbreviated here as FaNUK).

2. Goals and Methods of the FaNUK project

Several thousand long-established UK family names remain unexplained; many others are wrongly explained in the standard literature. The main goal of FaNUK is to rectify these deficiencies by creating a database of family names containing an evidence-based account of the geographical and linguistic origins, history, and demography of each name.

Should the researcher work forwards or backwards in time? I.e. should he/she collect as many examples as possible of early bearers of names in medieval data and work forwards, or start with an inventory of modern surnames and seek analogues in the historical data? The systematic investigation of Middle English surnames would be colossally expensive and would probably not yield the results that scholars might have hoped for. For one thing, it would focus attention on many thousands of names that have not survived to the present day. FaNUK intends to produce a database focused on exploring the origins of current family names. It will emphasize family names as linguistic and historical entities, rather than focusing on genealogy and family history. Nevertheless, it is crucially necessary to take account of the work of genealogists and family historians – especially the Guild of One-Name Studies.[1] FaNUK will bring together medieval evidence with the distributional evidence of modern online geodemographic tools.

The first duty of a lexicographer is to compile a comprehensive or at least a systematically selected inventory. But existing dictionaries of surnames in Britain contain no such inventory: there are no entries for common names such as *Alderson* (northern English), *Blair* (Scottish), *Critchley* (Lancashire), *Perks* (West Midlands), *Pringle* (Northumberland), *Sneddon* (Dumfriesshire), and over 20,000 others. At the opposite end of the spectrum, the standard work, the *Dictionary of English Surnames*, by P.H. Reaney and R. M. Wilson; 3rd edition 1991; cited here as R&W) contains over 3000 entries for defunct surnames such as Dogshanks. This is because it is essentially a dictionary of medieval surnames without declaring this in the title.

FaNUK's entry list is based on a comprehensive set of modern family names in Britain supplied by Professor Richard Webber of University College London, a geodemographic analyst with access to very large source files.

Having established the inventory, the next task is to explain each entry in it. Most of our modern English family names have their origin in a name that became hereditary at some time between the 13th and the 16th century. Hereditary post-medieval surnames continued to change, mostly in pronunciation and spelling, but also more radically through folk etymology. Many surnames became garbled in speech, mis-heard and mis-written in the records, and re-interpreted to fit known names or name-patterns. The precise forms of many of our surnames have only become permanently fixed in the twentieth century with the spread of literacy through all social classes. Consequently, modern surname forms can be quite misleading as regards their likely relation to each other and to their medieval originals. R&W's identification of modern surname variants and their allocation to Middle English etyma are largely based on superficial resemblances and are not safe unless confirmed by historical evidence—and in the long run, genealogical evidence. Reaney systematically gives Old English etymons for names that could only have been derived from Middle English forms, and he fudges the distinction between the meaning of an Old English or Anglo-Scandinavian topographical descriptor and the derivation of a surname from a place that had been named with such a descriptor several centuries before surnames were thought of.

Other **English**-oriented works include Basil Cottle's *Penguin Dictionary of Surnames* (1967, 1978), third, revised, edition by John Titford (2009). This is heavily dependent on R&W for its explanations. *A Dictionary of Surnames* by P. Hanks and F. Hodges (cited below as H&H, 1988), like its successor *Dictionary of American Family Names* (cited below as DAFN, 2003), is a general resource aimed mainly at the American market, containing much material relevant to the UK and foreshadowing FaNUK in that its dataset has a broad ethnic and etymological scope. However, DAFN's etymologies are to a large extent derived from R&W and are not supported by citations of medieval bearers.

The main **Irish** resources, likewise showing their age, are Patrick Woulfe, *Irish names and surnames* (1923) and Edward MacLysaght, *The surnames of Ireland* (1985, 6th edn). Both

H&H and DAFN include information that is generally both extensive and reliable on the etymologies of Irish surnames, thanks chiefly to the contributions of the late Tomas de Bhaldraithe and of Kay Muhr. None of the works mentioned here provide supporting evidence for early bearers of Irish names. Such evidence exists, most notably in the form of Tudor Fiants (Nicholls 1994), which show surnames in transition from their Irish language forms to their conventional anglicized form.[2] In FaNUK, for each Irish surname, evidence from Fiants and other sources will be selected and presented. While the Republic of Ireland is not part of the UK, it would be unthinkable to omit Irish names, not least because of the lengthy interaction and population exchanges between our two countries. Prompted by FaNUK, Dr Paul Ell of the Centre for Data Digitisation and Analysis, Queen's University, Belfast, has succeeded in getting a small grant from the Marc Fitch Foundation for an exploratory project to digitize the Tudor Fiants, and we are already able to make use of some of his work.

Despite any reservations that we may have, we can and indeed must make use of all the works just mentioned. Lexicography is typically accretive, each new work building on foundations laid by its predecessors, which offer hypotheses for confirmation or correction, as the case may be, in the light of newly available evidence. FaNUK evaluates and builds on the hypotheses of its predecessors, supplemented by geographical and statistical evidence.

The UK's multiethnic character will be addressed by including stub entries for most **immigrant** names (principally **Huguenot, Jewish,** and more recent arrivals with more than 100 current bearers). A "stub entry" gives information about the ethnic and linguistic origin of a name, together with an etymology if one can be obtained from a reliable source. However, it does not attempt to undertake primary research or to identify early bearers of the name in the land or culture of origin.

3. Databases of source materials

Another aim of FaNUK is to encourage and support the digitization and processing of relevant historical resources such as medieval tax returns, patent rolls, and court records, parish registers (in which christenings, marriages, and deaths were recorded in the parishes of England from 1538 onwards), 17th-century protestation returns and hearth tax returns, and so on. In each case, the aim is to compile an index of the place-names and personal names in the document(s), which in turn will support and be supported by a "medieval ←→ modern index". In this way, the remarkable variety of medieval name forms can be related to the preferred modern spelling of each surname and, if no modern spelling exists, surnames that have died out before the present day can be accurately identified.

The Poll Taxes of 1377, 1389, and 1381 (ed. C. Fenwick) already exists in the form of an electronic database, and it is hoped that this will be made publicly available, in addition to the published book. Dr Fenwick has compiled a medieval ←→ modern index of place-names, and the FaNUK team is currently compiling similar indexes of family names and given names in the Poll Tax data, linking the medieval names to their modern forms. This is a non-trivial task, given the vast number of variant spellings in both medieval and modern data.

4. Database software and links

The database software for the project was built by information scientists at the Faculty of Informatics, Masaryk University (FIMU), Brno, Czech Republic, customizing their DEB ("Dictionary Editor and Browser") platform. FIMU was selected because among its staff are

some of the world's leading experts in creating computational tools for speedy and reliable analysis, editing, and browsing of very large linguistic databases, text corpora, and dictionaries. They provide substantial and ongoing software support in response to FaNUK researchers' requests. Funding by Oxford University Press allowed the project to be piloted in 2007-9, and the database began by digitizing the entries in R&W and integrating them with Webber's headword list. Links were then set up to the electronic text of DAFN; other links are currently being established, in conjunction with The National Archine (TNA in Kew (formerly the Public Record Office).

FaNUK has also established a link to the vast International Genealogical Index (IGI), which has over 100 million transcribed records from parish registers, noncormist 'circuits', and other sources, prepared under the auspices of the Church of Jesus Christ of Latter Day Saints (LDS; "Mormons") and the Federation of Family History Societies. The IGI is available through Family Search International, an agency of the LDS Church with whose principals we have established a fruitful collaboration.[3] We have created a version of IGI in which the place-names have been corrected and made consistent, while records of doubtful reliability have been 'quarantined'. A selection of reliable, validated IGI evidence for each surname is appended as a spreadsheet to each surname entry in the FaNUK database. In some cases, there are many thousands of records for a single spelling of a surname. We are currently just beginning to work on developing statistical techniques for identifying significant associations between surnames and localities. It seems likely that this resource will have a profound effect on our understanding of population movement over time, among other things. It is well known that many surnames have an 'epicentre', being statistically strongly associated with a particular locality. It is less well known that in some cases, the epicenter itself may move from century to century. This phenomenon deserves further attention, ideally with links to DNA studies of family relationships.

On geographical distribution of names, FaNUK collaborates with **Steven Archer**, who has mapped the geographical distribution and frequencies of every surname recorded in the 1881 national census. A surname whose association with a particular locality is statistically significant may well have originated there, and this possibility needs to be investigated before other possibilities are considered. We say this with confidence, because although people move around, there is ample evidence that many surnames still cluster around a point of origin. This enables us to show that, for example, the surname *Harmison* originates in Hermiston in Roxburghshire rather than (as previously thought) in Harmston in Lincolnshire. Similarly, *Rochester* is commonly said to be from the city of Rochester in Kent, but the evidence of geographical distribution in 1881 and again in 1997 indicates that most if not all modern bearers of the surname derive it from one of three much smaller places of the same name in Northumberland.

6. Task and Targets

FaNUK's primary target is, then, to create reliable explanations, supported by evidence of historical bearers, for approximately 40,000 surnames in the UK that have more than 100 current bearers. These are in most case long-established English, Scottish, Irish, Welsh, and Cornish surnames. A secondary target is to add explanations for unproblematic names of lower frequency. A tertiary target, as explained above, is to add entries for about 4,000 names of recent immigrant origin, saying at least where they came from, what (if anything) is known about their meaning, and wherever possible giving information relevant to their UK status. Data from recent electoral rolls and censuses shows that there are over 370,000 different surnames in Britain today, but the vast majority of them are extremely rare, being borne by

only a handful of bearers. To the surprise of many, no doubt, over 300,000 of these are the names of recent immigrants from a vast number of countries including but by no means restricted to the countries of the former British empire. That leaves the 43,000 surnames referred to above.

These figures give an impression of the magnitude of the task. It is enough to indicate both what we can do successfully and what some of the major difficulties are. We have a lot to learn, but we expect to achieve a considerable amount. The project should radically improve the basis on which dictionary entries are researched and presented for all UK surnames, and the online database should provide a productive framework for new surname research long after the present project has come to an end.

Notes

[1] Their web presence is at http://www.one-name.org/, tested in October 2011.

[2] *Fiants* were authorizations to the Court of Chancery in Ireland for the issue of letters patent under the Great Seal of English monarchs in the 16th and 17th centuries.

[3] It is well known that a number of individual entries in IGI have shortcomings, e.g. obvious and less obvious transcription errors. However, the value of IGI lies partly in the statistically addressable mass of data which it offers and partly in the fact that at least 100 million of the transcribed records are demonstrably accurate, correct, and usable.