Digitising the Linguistic Atlas of Scotland: Scope and Potential of a new Corpus

John Kirk

The origins of the present work lie with the various attempts by the author to digitise the material of the Linguistic Atlas of Scotland (LAS), vols. 1 and 2 (Mather & Speitel 1975, 1977) which culminated in several publications (Kirk 1994, 1996, 2001; Kirk & Kretzschmar 1992; Kirk & Munroe 1989; Kirk, Munroe & O’Kane 1994) and Master Theses (Cooper 2001, Doherty, 2001, O’Kane 1990, O’Neill 2002, Robinson 2002). With preliminary funding by the University of Vienna, the project has been resurrected in 2019 and begun anew. Since the first endeavours, there has been a revolution in geographical information systems, the interactivity of databases, and automatic cartography, so that the present endeavour has in effect to start afresh.

A second interest in the original first project was to initiate the quantitative analysis of dialect data, inspired by Goebl (1982) (cf. also Viereck 1985, as applied to the data of the Survey of English Dialects) and by work in the USA which culminated in Kretzschmar & Schneider (1996). Burkette & Kretzschmar (2018) have recently shown, on the basis of the item ‘parlour’, how complexity theory provides a good explanation for dialect survey data, in particular showing “that most linguistic data are, in fact, not normally distributed. The type of distribution we most often encounter for language data is that of the hyperbolic asymptotic curve (A-curve)” (Kretzschmar & Burkette, in prep. p. 5). This hypothesis lends itself to testing by the LAS data as it is well-known that many Scots lexical items are pan-Scottish (i.e. widely distributed throughout the Lowland Scots-speaking areas) whereas many others are quite localised. However, to test this thoroughly, a digitised atlas would be required.

The present paper begins with a critical look at the onomasiological approach displayed in LAS1 and LAS2 with the preference for an ortho-lexical, quasi-phono-lexical categorisations, which excludes oncers, and which avoids physical maps. The present approach seeks to recategorise the data in terms of purely lexical (i.e. etymologically-lexical) types and thus is able in to include the many omitted oncers listed in the atlas appendices. The new atlas uses physical base maps and makes uses of interactive mapping techniques, including displays for age and gender. Preliminary examples appear in Hessle (2019).

A further aim is a re-evaluation of the distribution of lexical types per map item; and to account afresh for the origins of these items. Hessle (2019) shows that for many speakers along the Fife littoral, for ‘splinter’ many informants are recorded as offering both skelb and skelf. Far from being simply phonological variants, skelb is shown to have Gaelic origins whereas skelf appears to have come from Middle Dutch. Although traditionally regarded as a Scots-speaking area, Fife was early a stronghold of Gaelic settlement, as shown extensively in the five volumes of The Place-Names of Fife (Taylor & Markus 2006-2013). In the medieval period, for reasons of trade, many Flemish immigrants settled along the Fife littoral, as reflected in the Dutch architecture of Culross (cf. Price 2013), and contributing loanwords to the vocabulary of Scots, most strikingly the word for what became Scotland’s national sport: golf (cf. Murison 1971).

The paper will report ongoing developments by the time of the conference. It will meet all the workshop’s requirements by being corpus-based; by triangulating lexical analysis, cartography and socio-cultural history; by taking a quantitative approach (following the lead being shown by Kretzschmar & Burkette); by making a new use of an old resource with new and innovative software; and by drawing comparison with similar lexical dialectological projects (e.g. in the US: cf. Burkette & Kretzschmar 2018).

References


