Clive Upton
University of Leeds (GBR)
C.S.Upton@leeds.ac.uk

Blurred Boundaries: the Dialect Word from the BBC

For the dialectologist, words are unruly and hard to codify. That is not to say that, for the English-language specialist, excellent English dialect dictionaries are not available (PENHALLURICK 2009; 2010). Online versions in particular now have much to offer, taking the user a long way into understanding the non-standard lexicon and, crucially, the inter-connections between related lexical items. Even the seminal century-old *English Dialect Dictionary* (WRIGHT 1898-1905), in paper form a mine of information which could be stubbornly reluctant to yield its full riches to a user, is showing its full potential through the Innsbruck-based SPEED initiative (MARKUS/HEUBERGER 2007; ONYSKO/MARKUS/HEUBERGER 2009; MARKUS 2010; PRAXMARER 2010). But such dictionaries, typically the product of questionnaire-based fieldwork and textual research, tend not to capture informal spontaneous usage. Nor do they contain quantifiable data on independent social variables of the kinds required by the social dialectologist. If the aim is to study language collected in an informal manner from a range of speakers identified by independent variables such as location, age, or gender, usable data are not to be found in works deriving from formal interviewing and the gleaning of written sources. They will best emerge from the informal speech of speakers who, as much as possible, have not been subjected to the steering of professional linguists. Also, lexical variants need to be gathered in quantities sufficient to offer the likelihood of reliable quantification of their use.

STRANG (1968, 215) was quick to recognise the marginalisation of lexis, this caused in large part by the difficulties attendant on its capture for modern research purposes. Rule-governed phonemic inventories and grammatical paradigms provide environments in which comparisons between variants can be readily made. And data-capture strategies are relatively unproblematic here: find a topic about which a biographied informant can talk fluently and unreservedly, record them discussing it, and a complete phonemic inventory and a wide range of grammatical structures will readily emerge, to be avaliable for comparative social-dialectological investigation. The same cannot be said of lexis. A quite elaborate mechanism must be devised if an informant's unforced lexical usage is to be documented, a mechanism which is likely to trigger the observer's paradox and which might be expected to preclude its being used repeatedly, for comparative purposes, across a large population in a feasible space of time.

The British Broadcasting Corporation's *Voices* initiative of 2005-2007 offered a way to bring lexis into the frame. Using a technique devised by Llamas (KERSWILL/LLAMAS/UPTON 1999), thirty-eight lexical variables were placed centre-stage as 'concept words', to prompt members of the public to enter their personal variants for these on a dedicated BBC website. Such was the take-up for this operation that responses were volunteered at an average of more than 16,000 per concept, together with information on contributors' age, gender, and (post-coded) location. The resultant database makes it apparent that vernacular lexis allows of and justifies exploration, not least in terms of its untidiness.

Orthographic representation of the non-standard is an area offering itself for immediate exploration in this regard. Anyone trying to make use of a 'dialect' dictionary quickly becomes aware of this: just what written shape will the dictionary compiler have given to a word which, in normal use, only has a spoken form? And where will it be alphabetised in consequence? We are in a situation analogous to that of Middle English in the fluidity of the orthography. More remarkably, even Standard or other well-recognised spellings are registered only lightly by very many contributors to a database such as *Voices*. It is not that such contributors are language-unaware: after all, they have voluntarily responded to a broadcast invitation to offer up their information. Nevertheless, the prompt PLAY TRUANT, for example, results in spellings of 'skive' such as <scive>, <sceive>, and <skaive>, alongside many still more exotic forms, pointing up the primacy of the spoken medium in modern society.

Lexicographers customarily assign style labels to non-standard and some standard lexical items, these acting as both adjuncts to definitions and usage guides. Prominent amongst these are 'dialect', 'slang', 'colloquial' and the like. Clearly, it is necessary to offer guidance on lexical usage. However, a comparison of the application of such discrete labels to the same items across a number of dictionaries shows how randomly they tend to be applied, sometimes within the same work and frequently between different ones. The *Voices* exercise cuts through the nice distinctions and confusions of style labels, with respondents offering up their information on local and personal speech regardless of where it lies on a cline of formality from the most elevated Standard English lexis, through the colloquial, the obviously regional, local and idiolectal, to the most vulgar and obscene. The *users* of the language often emerge as being as blissfully unaware of artificially-imposed stylistic distinctions in their language as they are of the niceties of English spelling.

It must be accepted that, as far as informing denotative meaning is concerned, the *Voices* database is a blunt instrument. Contributors to the website were simply asked to give their words for thirty-eight different prompts, and subtle semantic nuances are lost: are the 7,792 instances of 'living room' analogous in every case to the 4,958 instances of 'lounge', the 2,387 of 'sitting room', and the 2,260 of 'front room'? This is highly unlikely. And where do the 53 instances of 'den' lie here? We can only speculate, engaging with the fuzziness inherent in such data. But when we come to

the social connotations of *Voices* lexis we have material to work with, since responses are accompanied by age and gender tags that provide an opportunity to investigate in depth, and words carry socio-cultural messages.

Alongside data-tagging for age and gender is that for geographical location, this identified through postcode data, both the broader postcode area information relating to 'post towns' and more finely tuned postcode district information. Dialectometric analyses now underway predictably indicate regional patterning of lexis, while warning against too-ready acceptance of such certainties as that of the neat 'dialect area'. Even a relatively simple sequence of cluster-analysis maps shows a shifting pattern as delineated areas alter depending on the data selected. The selection of competing models prompts still more questioning of complex lexical distributions, and might even suggest some answers.

As with mapping, so with orthography, style-labelling and denotative and connotative meaning: data available in the mass raises serious questions to challenge any too-ready assumptions we might be inclined to make.

KERSWILL, PAUL / CARMEN LLAMAS / CLIVE UPTON (1999): The first SuRE moves: early steps towards a large dialect database'. In: UPTON, CLIVE / KATIE WALES (eds.): DIALECTAL VARIATION IN ENGLISH. PROCEEDINGS OF THE HAROLD ORTON CENTENARY CONFERENCE 1998. Leeds Studies in English Vol. 30, 257-269.

MARKUS, MANFRED / REINHARD HEUBERGER (2007): The Architecture of Joseph Wright's English Dialect Dictionary. Preparing the Computerised Version. International Journal of Lexicography 20 (4), 355-368.

MARKUS, MANFRED (2010): Introduction to Part II: The structure of Joseph Wright's English Dialect Dictionary. In: MARKUS, MANFRED / CLIVE UPTON / REINHARD HEUBERGER (eds.): Joseph Wright's English Dialect Dictionary and Beyond. Frankfurt/Main, 77-89.

ONYSKO, ALEXANDER / MANFRED MARKUS / REINHARD HEUBERGER (2009): Joseph Wright's English Dialect Dictionary in Electronic Form. A Critical Discussion of Its Parameters and Query Routines. In: RENOUF, ANTOINETTE / ANDREW KEHOE (eds.): Corpus Linguistics. Refinements and Reassessments. Amsterdam/New York, 201-220.

PENHALLURICK, ROBERT (2009): Dialect Dictionaries. In: COWIE, A. P. (ed.): The Oxford History of Lexicography Volume II: Specialized Dictionaries. Oxford, 290-313.

PENHALLURICK, ROB (2010): The dialect dictionary: what is it good for? In: HESELWOOD, BARRY / CLIVE UPTON (eds.): Proceedings of Methods XIII. Papers from the Thirteenth International Conference on Methods in Dialectology, 2008. Frankfurt/Main, 133-142.

PRAXMARER, CHRISTOPH (2010): Dialect relations in the English Dialect Dictionary. In: HESELWOOD, BARRY / CLIVE UPTON (eds.): Proceedings of Methods XIII. Papers from the Thirteenth International Conference on Methods in Dialectology, 2008. Frankfurt/Main, 153-159.

STRANG, BARBARA (1968): Modern English Structure. 2nd edn. London.

WRIGHT, JOSEPH (ed.) (1898-1905): The English Dialect Dictionary. Oxford.