

# Employing Geographical Principles for Sampling in State of the Art Dialectological Projects

Isabelle Buchstaller<sup>1\*</sup> and Seraphim Alvanides<sup>2</sup>

<sup>1</sup> School of English, Leipzig University, Germany

<sup>2</sup> Geography and Planning, Northumbria University, Newcastle, UK

The aims of this paper are twofold. First, we locate the most effective human geographical methods for sampling across space in large-scale dialectological projects. We propose two geographical concepts as a basis for sampling decisions: Geo-demographic classification, which is a multidimensional method used for the socio-economic grouping of areas; we also develop an updated version of functional regions that can be used in sociolinguistic research. We then report on the results of a pilot project that applies these models to collect data regarding the acceptability of vernacular morphosyntactic forms in the North East of England. Following the method of natural breaks advocated for dialectology by Horvath & Horvath (2002), we interpret breaks in the probabilistic patterns as areas of dialect transitions. This study contributes to the debate about the role and limitations of spatiality in linguistic analysis. It intends to broaden our knowledge about the interfaces between human geography and dialectology.

## 1. Introduction

Most sociolinguistic research, in the UK as well as elsewhere, has not been cognisant of the recent advances in human geography (Britain, 2009, 2010). More specifically, current dialectological research tends not to be informed by rigorous geographical sampling methods or relies on geographical methods from the 1980s and early 1990s, such as the CURDS functional regions algorithm used to great effect by Cheshire, Edwards & Whittle (1989, 1993). Britain, who has been at the forefront of sociolinguistic theorising of the concept of space in dialectology, proffers three main points of criticism of the variationist enterprise, which we will represent here in full:

Firstly, variationism has at worst largely ignored spatiality and at best treated it quite distinctly and separately from other social factors until relatively recently. Secondly, when it has engaged with space, it has tended to be a social devoid, Euclidean, distance-is-all type of space, rather than a socially rich spatiality, which recognises that “the fact that social processes take place over space and in a geographically-differentiated world affects their operation” (Massey, 1985: 16), again until relatively recently. And thirdly, space has not, yet again until recently, seen the sort of critique in sociolinguistics that has been witnessed by concepts such as style (...). (Britain, 2009:142)

Indeed, the majority of multi-locality sociolinguistic work can be described as spatially naïve, using geographical space merely as a canvas—unanalysed and undertheorized—onto which the results of linguistic analysis can be mapped. However, since the 1970s and 1980s, human geographers have started to conceptualise regions—and places within them—as dynamic entities which warrant more flexible and emically driven multifactorial approaches. Contemporary human geography, having moved beyond static, a priori approaches to space, aims at investigating “the construction of human geographies, the social production of space and the restless formation and reformation of geographical landscapes” (Soja, 1989:10–11). Since little of this work has permeated into sociolinguistics, this paper sets out to investigate the ways in which the discipline can fruitfully draw on models created within the framework of human geography. More specifically, we investigate the benefits of using geographically informed parameters for sampling in multi-locality studies.

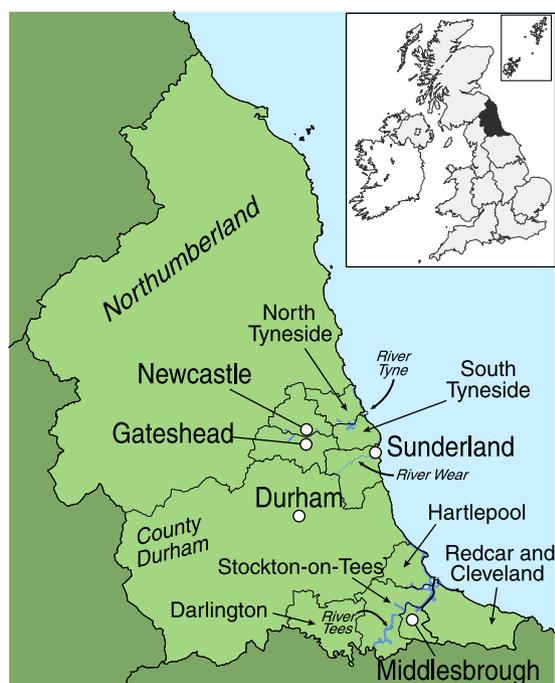
In this paper we put forward a model that embraces a socially sensitive approach to space as a sampling criterion. We also report on a pilot analysis that used a range of human geographical methods for sampling across the extreme North East of England (consider Map 1).

## 2. Geographical approaches for sampling in dialectological projects

The primary concern of most multi-locality dialectological projects, especially of older studies but also many recent

---

\*Address for correspondence: Isabelle Buchstaller, Beethovenstrasse 15, 04107 Leipzig, Haus 4, Zi. 311; +49 (0)341 97 37 314 (Email: i.buchstaller@uni-leipzig.de)



**Map 1.** The North East of England (from Buchstaller et al. 2011:3, based on two outline images: UK and Ireland [NordNordWest, 2011 CC BY SA 3.0, [http://commons.wikimedia.org/wiki/File:United\\_Kingdom\\_NUTS\\_location\\_map.svg](http://commons.wikimedia.org/wiki/File:United_Kingdom_NUTS_location_map.svg)] and North East England [Nilfanion, 2011 CC BY-SA 3.0, [http://commons.wikimedia.org/wiki/File:North\\_East\\_England\\_districts\\_2011\\_map.svg](http://commons.wikimedia.org/wiki/File:North_East_England_districts_2011_map.svg)]).

ones, has been social (rather than geo-demographical) representativeness. Great care is generally taken to investigate and/or control for variability along the classic factors of gender, socioeconomic class, and age (plus sometimes attitudinal and/or networks factors). As such, “social variables of the local dialect speakers [are] ... homogenized as much as possible in order to examine geographical variation” (Barbiers, Cornips & Kunst, 2007:60). Space, however, the object of investigation, tended to be treated as carrier material, a blank slate over which linguistic variability was superimposed. Britain (2009:144) comments that “there was actually very little that can be considered truly geographical, let alone spatially sensitive in the work of the traditional dialectologists,” and to a great extent there still is not.<sup>1</sup> And so, Labov’s summary paper (1982:42) rightly states that “the study of the heterogeneity in space has not advanced at the same tempo as research in single communities.”

At the start of the 21<sup>st</sup> century, dialectology—and with it the theorisation and manipulation of space as it pertains to linguistic analysis—seems to undergo an upswing. Two large overview volumes have recently appeared (Auer & Schmidt, 2010; Lameli, Kehrein,

& Rabanus, 2010). Critical reflections on space are underway and published more widely in the literature, (Buchstaller, 2008; Britain, 2004, 2009, 2010; Horvath & Horvath, 2001, 2002; Stuart-Smith, 2002–5). Also our descriptive base has been broadened with the recent collection of a range of large-scale multi-locality data-sets, many of which aim at spatial and human geographical representativeness, leaving outdated grid-based models behind or at least supplementing them with more socially sensitive sampling methods. Let us investigate the sampling strategies of a number of recent large-scale projects in order to trace their conceptualisation and manipulation of space as well as the notion of representativeness that underlies these methods.<sup>2</sup>

An ever-increasing number of atlas projects are coming out of ‘socio-syntax’, a new linguistic sub-discipline that investigates syntactic micro-variability by sampling across larger geographical areas. We briefly discuss the sampling methods underlying the *Dynamic Syntactic Atlas of the Netherlands Dialects* (SAND, <http://www.meertens.knaw.nl/sand/zoeken/>), which—under the auspices of the European Science Foundation funded Edisyn project—functions as a hub for similar dialect syntax projects ([http://www.dialectsyntax.org/wiki/About\\_Edisyn](http://www.dialectsyntax.org/wiki/About_Edisyn)).<sup>3</sup>

The SAND approach to sampling is combinatory: It relies on tessellation via a grid model—250 cells of variable size for the whole of the Netherlands including both urban and rural localities—to ascertain overall spatial representativeness. But it is also sensitive to human geographical factors such as political borders, demographic changes, (counter)urbanisation and isolation: Certain types of locations received a higher sampling density, namely (i) those that are relatively isolated (e.g. (former) islands) and (ii) locations in transitional areas (e.g. between Frisian and Low-Saxon, German and Dutch and along the Germanic-Romance language border). The same also holds for locations in areas of which pilot projects or the linguistic literature revealed more dialectological variation (cf. Lekakou and Barbiers, p.c. 29 March, 2009). SAND does not sample according to population size, “but two important criteria for the selection of the locations were: 1. the villages should have some history. [...]. Very recent locations which are fast growing due to industrial or administrative developments were excluded like Zoetermeer or Almere (which is a recently founded village near Amsterdam and thus fast growing), 2. the location didn’t undergo very fast demographic changes recently” (Cornips, p.c. 6 April, 2009). While the inclusion of such socio-demographic sampling parameters is a huge step forwards, they appear to be administered on a case-by-case basis rather than based on principled parameters rooted in geographical practise. What is more, similar to the early days of dialectology, instead of investigating

the effects of certain geosensitive types of human activity—such as in-migration—SAND excludes areas that are the locus of such changes.<sup>4</sup>

The *Atlas of North American English* (TELSUR) (Labov, Ash & Boberg, 2006), which is based on 417 speakers across the territory of English-speaking North America, “was designed with the goal of representing the largest possible population, with special attention to those speakers who are expected to be the most advanced in processes of linguistic change” ([http://www.ling.upenn.edu/phono\\_atlas/sampling.html](http://www.ling.upenn.edu/phono_atlas/sampling.html)). The project has incorporated contemporary human geographical concepts such as urbanisation and newspaper readership catchment areas into its sampling design. Three types of areas are sampled: Central Cities (CC), Zones of Influence (ZI), and Urbanized Areas (UA). CCs are defined on the basis of population distribution, with at least 200,000 inhabitants in the 1990 census. ZIs are derived from data on newspaper circulation from the 1992 County Penetration Reports of the Audit Bureau of Circulations (ABC); they consist of counties with the highest circulation of a city’s newspaper(s), compared to the circulation of all other cities’ newspaper(s) in that city (cf. [http://www.ling.upenn.edu/phono\\_atlas/sampling.html](http://www.ling.upenn.edu/phono_atlas/sampling.html)). Finally, UAs are used as a way to sample at a geographically refined and more meaningful scale, unconstrained by political and administrative boundaries. UAs consist of a core CC (or a group of nearby cities) and the surrounding densely settled territory, with a combined population of at least 50,000. Various population density measures are used to incorporate contiguous census block groups (rather than whole counties) around each core in order to form distinctive UAs. In the design of the TELSUR/Atlas sample, if a speaker is a native of any place within a UA, s/he is taken to be linguistically representative of the respective city’s speech community. In order to differentiate the amount of sampling to be carried out in smaller cities within each ZI, the CCs are further divided into types by population of the corresponding UA (above one million, between 200,000 and one million, or below 200,000 inhabitants) and by physical area of the ZI (with 5,000 square miles as a cut-off).

Hence, TELSUR achieves broad geographical coverage and is based on a well-defined, geographically sensitive sampling strategy. However, it is restricted to urban speech (see Milroy & Gordon, 2003:21). Obviously, focusing on either the urban OR the rural dramatically reduces the demographic representativeness of the study to just this settlement type—a rather narrow sampling universe in Sankoff’s (1980) term.<sup>5</sup>

Furthermore, while the sampling strategy of newspaper readership catchment areas might provide an adequate profile of speakers’ ideological belonging in

the United States, this may not be an appropriate approach in other national contexts. In the UK for example, regional newspapers have a limited following and newspaper readership is class-based rather than geographically distributed (although socio-economic class is obviously not distributed evenly across space). Hence, other measures are needed in order to “cut (...) through the connective tissue of the world in such a way that its fundamental [social] integrities are retained” (Gregory, 1985:328).

Geographers have drawn our attention to the fact that “it is flows between places and not places themselves that matter” (Dorling, 2004:104). The only dialectological project we are aware of that applies flow-based geographies is the Survey of British Dialect Grammar (Cheshire et al., 1989, 1993). This project, which aims at collecting a large-scale data-base in the British Isles, is quite radical in its adaptation of human geographical models to sociolinguistics and in many respects it functions as a methodological precursor to this study. Cheshire et al. (1989, 1993) conducted a large-scale investigation into vernacular morpho-syntax based on questionnaires sent to schools across the UK. They relied on the functional regions system produced on the basis of the 1981 census data by the Centre for Urban and Regional Development Studies (CURDS) at Newcastle University for the classification of their data-points. Functional regions are defined as areas with some geographical coherence, usually measured via parameters such as an area’s socio-economic profile, commuting flows by working age population and in/out-migration patterns (Coombes, Dixon, Goddard, Openshaw & Taylor, 1982; Masser & Scheurwater, 1980). They divide the country into a set of urban centres, based on statistical information regarding employment and retailing opportunities. The surrounding areas attached to these urban centres are defined on the basis of commuting patterns, resulting in 228 Functional Regions for the UK, consisting of cores, rings, and outer and rural areas. These cores are described as the “pivotal nodes of economic activity and social life” (Champion & Coombes, 1983), while their surrounding areas are defined in relation to commuting patterns, reflecting the degree to which their residents depend on the cores for their jobs.

The CURDS functional regions have been widely used by economic geographers and regional scientists for the analysis of economic and social change in a variety of urban and regional scales in Great Britain. Cheshire et al. (1989, 1993) did not sample according to these parameters, but they categorize the 87 schools whose questionnaire responses they analyse in terms of their geographical location into cores, rings, and outer and rural areas. Since 75% of their responses were from the core areas, the Survey of British Dialect

Grammar is biased towards the urban centres. Cheshire et al. (1993:63) conclude that “the CURDS system is potentially of great value for research into patterns of linguistic variation and change in the British Isles since it identified important patterns of social communication between people from different geographical areas, on the basis of their economic activity”. In this paper we use an approach that reflects the concept of functional zones pioneered by CURDS as a sampling strategy.

### 3. Towards a sampling model for the British Isles

The first and to date only large scale atlas project in England, the *Survey of English Dialects* (SED, Orton et al. 1962–1971) conducted between 1950 and 1961, covers an impressive number and geographical spread of sampling points: 313 localities in England, the Isle of Man and some areas of Wales close to the English border. A contemporary investigation of dialectal differences in the UK could follow two, often conflicting, principles, namely diachronic comparability with the SED or synchronic geo-demographical representativeness of the area, both of which we discuss in turn.

We could aim for the former and take the sampling points of the *Survey of English Dialects* as starting points. However, given that the selection process that led to the choice of the SED localities was rather ad hoc<sup>6</sup> (see Chambers & Trudgill’s 1998 criticism), the representativeness of the data is questionable and—we would argue—not defensible. Indeed, even diachronic comparability is debatable, since several sampling points that were once rural isolated localities (such as the former mining villages Earsdon and Washington) have become commuter villages/towns as a result of counterurbanisation.

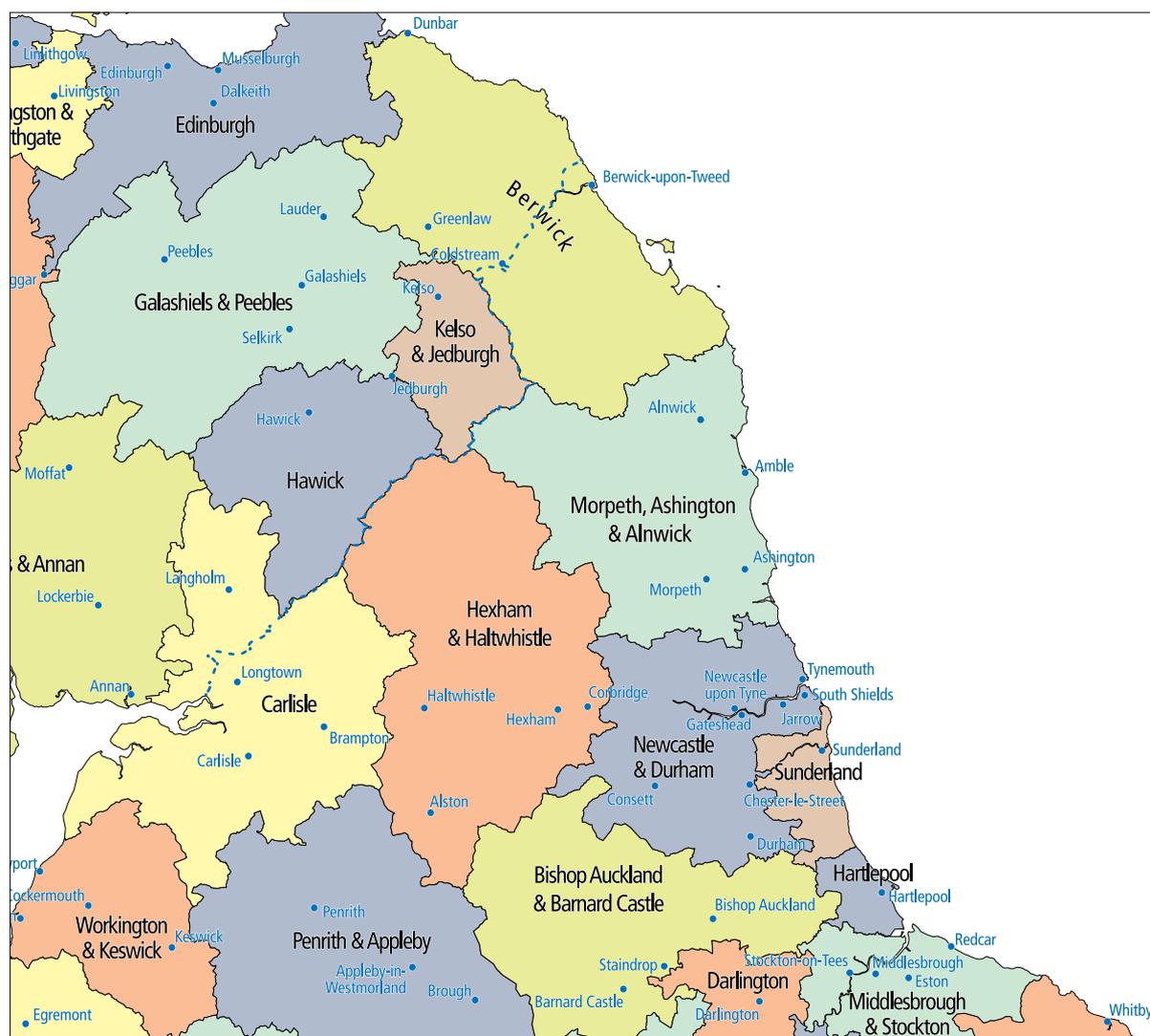
Even if we get around this issue—by sampling nearby localities for example—the problem persists that such a sampling strategy is arbitrary and not based on *bona fide* socio-spatial parameters. What is needed is a dialectology that is rooted in the everyday reality of the people who live in the area investigated and thus cognisant of the fact that “space and spatiality in general is socially constructed (...). [and] constantly evolving” (Allen, Massey & Cochrane, 1998:138). A geographically informed sampling method for a dialectological project would thus aim to represent human activity across space, leading to the appropriation and manipulation of geographies. Indeed, sociolinguists such as Britain (2002) and Kerswill (2009b) remind us that dialectological researchers need to orient our understanding of space to the socio-geographical day-to-day practises of people. More specifically, our research needs to be sensitive to the fact that the

geographies and histories of our social networks and those of the social, economic, and political

institutions which guide our daily lives in the West are played out, routinised, and reproduced within functional zones (...) [Consequently] the socio-geographical trajectories of speakers and their institutions are often strongly guided by past practices, by attitudinal considerations and by physical factors, and hence regions are formed. (Britain, 2009:151)

Dialectology thus needs to develop sampling criteria that are sensitive to the everyday flows of human interaction and routinised activities.<sup>7</sup>

As we discussed above, the concept of functional regions, the “pivotal nodes of economic activity and social life” (Champion & Coombes, 1983), has been used to great effect by Cheshire et al. (1993) to classify the schools participating in their Survey of British Dialect Grammar. In this paper, we will use functional regions as a parameter for sampling across space rather than as a descriptive element post hoc. Our unit of analysis, the Office for National Statistics travel-to-work areas (TTWAs), are based on up to date information from the 2001 census, yet they also reflect the concept of CURDS functional regions in that they group smaller areas into larger ones according to the strength of flows between them. TTWAs are defined by the following criteria, which were laid out in 2007 using 2001 census data on commuting (home and work addresses/postcodes, see <http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/other/travel-to-work-areas/index.html>): (i) At least 75% of the resident economically active population work in the area; (ii) at least 75% of everyone working in the area live in the area; and (iii) the minimum size is a working population of 3,500 (Coombes & Bond, 2008). This means in effect that geographical units (in our case, 2001 census areas) “‘organise themselves’ on the basis of their mutual commuting links” (Mooney & Carling, 2006:71) within the group with which they had the strongest mutual coherence. As such, TTWAs satisfy the criteria of minimum “population size and self-containment” (Shortt, Moore, Coombes & Wymer, 2005:2715), but they rely solely on optimising commuting flows, making them straightforward to conceptualise in a sociolinguistic framework. In other words, TTWAs do not pose an additional level of complexity to the functional regions.<sup>8</sup> Based on these criteria, the whole of the British Isles subdivides into 243 TTWAs. These areas of routine movement host the most fundamental grooves of daily interaction, based on commuting to and from work, often subsuming school runs, shopping trips and evening entertainment on the way and thus leading to the creation of space time zones. They are thus inherently meaningful in terms of people’s daily routines and interactions.

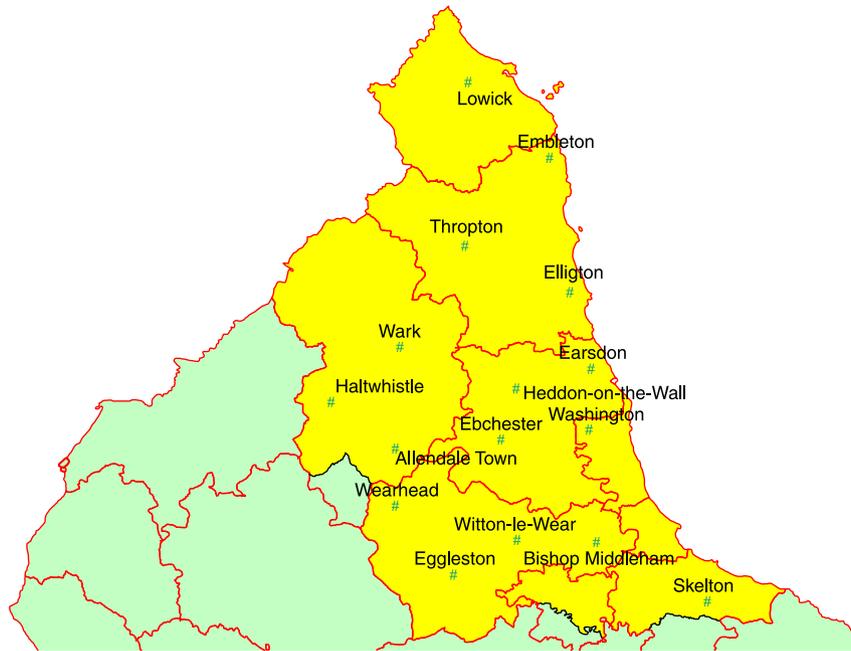


**Map 2.** TTWAs, settlement, and the Scottish border.

**Source:** <http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/other/travel-to-work-areas/index.html>

What makes sampling via TTWAs inherently superior to approaches that are based on grids or stationary political boundaries is the fact that they are the fundamentally local outcomes of people constructing their 'own' place (Kerswill, 2009b). They also conform entirely to Giddens' (1984:376) concept of routinisation as "the habitual, taken for granted character of the vast bulk of activities of day-to-day social life". For example, TTWAs have been used in geographical research to compare patterns of commuting in relation to employment opportunities and to develop employment policies. We propose that they are an ideal starting point for dialectological work since they (i) provide a stringently controlled sampling framework that is based on contemporary geographical methods, (ii) are widely available and easily accessible at <http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/other/travel-to-work-areas/index.html>, and (iii) readily lend themselves to a range of applications within the field of dialectology.

Taking zones of routinised every-day movement as a starting point thus results in a geography that is based on the human appropriation of space. Map 2 shows that the TTWAs in the North of England/South of Scotland are fundamentally independent of, and indeed criss-cross, political boundaries. For example, the TTWA centring around Berwick-upon-Tweed stretches on both sides of the political border. The special status of Berwick in the history of the English-Scottish border is reflected in the gestalt of the TTWA, with commuters from both directions flocking into Berwick-upon-Tweed (Glauser, 1974).<sup>9</sup>



**Map 3.** TTWAs in the North East of England with SED sampling points superimposed.

Britain (2010) points out that while routine interaction creates spaces of various kinds and shapes, a geographically sensitive approach to space also needs to account for a wealth of intimately inter-correlated socio-demographical factors. Indeed, once we have chosen the fundamental basis of spatial analysis, the next, rather thorny, question is the issue of where to sample inside of a TTWA while ensuring socio-demographic representativeness. We will briefly discuss the repercussion of using SED localities before proposing several socio-demographic parameters that could be used as sampling criteria. Using GIS (Geographical Information Systems) for manipulating the socio-economic information, we then embark on overlay analysis of TTWAs with socio-economic areal characteristics in order to define the sampling areas.

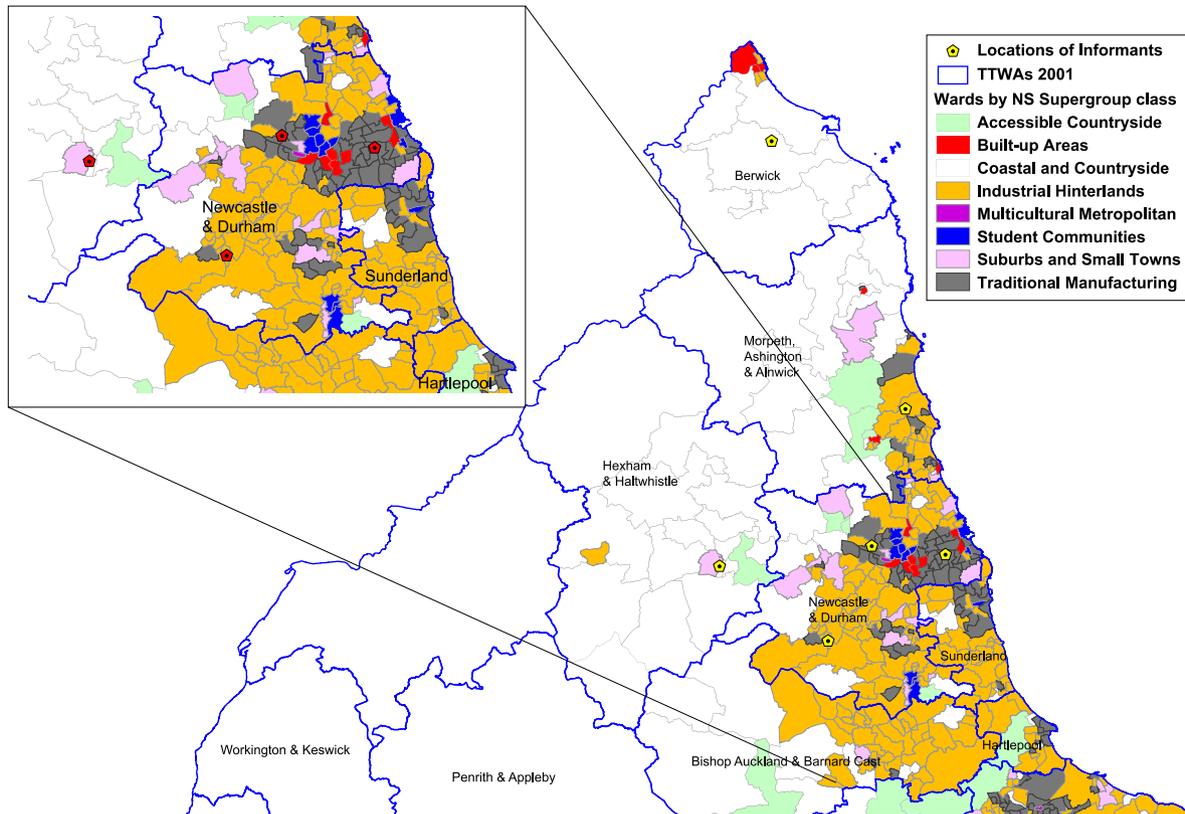
Above, we have argued against using SED sampling points due to their ad hoc character and lack of diachronic comparability. Map 3 provides another argument against the use of SED sampling points: It would lead to oversampling in certain areas (i.e. Wark, Haltwhistle, and Allendale in the Hexham & Haltwhistle TTWA) as well as undersampling in others (no sampling points in the Hartlepool or Darlington TTWAs).

What is needed is a principled method for selecting localities within these TTWAs. We would like to argue that the sampling points of any dialectological project that aims to be representative of the area it covers should correspond to the socio-demographic and economic make-up of the area. Since the TTWAs are obviously heterogeneous in this respect (given the

emphasis on commuting criteria for their construction), we need to investigate their socio-demographic characteristics. Such an analysis fundamentally relies on the concept of socio-economic area classification, or SEAC, a key concept in social geography, in relation to area profiling and geo-demographics (Harris, Slight & Webber, 2005). Geo-demographic classifications use socio-economic data from national censuses and other governmental and commercial databases to “group together geographic areas according to key characteristics common to the population in that grouping” (<http://www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/ns-area-classifications/index/available-geographies/index.html?format=print>). In the context of the British Isles, it “distils key results from the 2001 Census for the whole of the UK at a fine grain to indicate the character of local areas” (<http://areaclassification.org.uk/getting-started/getting-started-what-is-the-output-area-classification/>). This results in a categorization of areas of variable sizes (from local authorities to wards down to very small census output areas) according to a range of socio-demographic and economic components that were included in the census. The main dimensions of these components are demographic, household composition, housing, socio-economic, employment and industry sector.

Cluster analysis of the 2001 census data has revealed that the British Isles can be grouped into 9 socio-economic “supergroups,” i.e. areas with characteristic socio-economic and demographic profiles. These





**Map 4.** SEAC of the extreme north East of England superimposed on TTWAs (thick boundaries), with our sampling points.

why wards were selected as the unit of analysis is that they facilitate communication between researchers, fieldworkers and subjects during the sampling and recruiting process. In short, it is easier to seek subjects—and communicate the exact space requirements to them—from a list of qualifying wards (that people can relate to), rather than a much longer and complex list of postcode areas (or even specific streets). However, it is noted here that this method can be fine-tuned by using a finer level of areal sampling units (such as census output areas or even full postcodes) if a sufficiently large number of informants is to be recruited.

We then superimposed the SEAC-based supergroup ward profiles on the TTWAs of the North East of England; Map 4 is the result of this procedure. It reveals the diversity of the North East region in terms of socio-demographic make-up: From the predominantly “coastal and countryside” areas south of Berwick (which itself is classified as a “built up area”) and the Northumberland countryside we move south to the urban conurbation of Newcastle and Gateshead, which is predominated by wards classified as “industrial hinterland” and “traditional manufacturing”.

The fundamental advantage of the SEAC classification—apart from the fact that it is readily available online—is that it is population-sensitive and

encapsulates a wealth of socially relevant variables that have been selected on the basis of multivariate analysis (see <http://www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/national-statistics-area-classifications/national-statistics-2001-area-classifications/methodology-and-variables/wards/index.html> for methodology and the full set of variables). Thus, using a SEAC-based sampling strategy not only gives us an overview about the socio-economic make-up that our TTWAs are composed of, it also allows us to make representative sampling decisions on the basis of a wealth of socio-demographic information.

The question of how many data points are needed is obviously fundamentally dependent on a range of factors, including the research questions, focus and scale of the project in terms of time and financial resources, and thus cannot possibly be decided a priori. Here, we report on a small-scale pilot project that tests the usefulness of the methods described above for dialectological research. We decided to sample in the four northernmost TTWAs of the North East of England, namely “Berwick”, “Morpeth, Ashington and Alnwick”, “Hexham and Haltwhistle” as well as “Newcastle and Durham”. Our sampling points were chosen on the basis of geo-demographic and

population-based representativeness and—as a further consideration (if possible)—the existence of an old SED point in the wider area. For the northernmost TTWA, we sampled in Lowick, a former SED sampling point, which—being a hamlet of only 560 inhabitants—is wholly representative of a TTWA that is classified as predominantly coastal and countryside. The inland of the Morpeth, Ashington & Alnwick TTWA is coastal and countryside. Most of the population live along the coastline, in a stretch of area classified as industrial hinterland and traditional manufacturing. We sampled in Linton Colliery, a small ex-mining village of only a few hundred inhabitants about 1.5 miles from the old SED sampling point Ellington. In the Hexham and Haltwhistle TTWA, which is also mostly classified as countryside, we sampled just outside of Hexham.

The geo-demographic profile of the heavily populated Newcastle and Durham TTWA was slightly more complex, with the majority of wards classified as traditional manufacturing (52 wards comprising 394,834 inhabitants) and industrial hinterland (67 wards comprising 392,995 inhabitants). We aimed at a sampling strategy that captures this diversity. We thus chose two traditional manufacturing sampling points, namely Westerhope and Jarrow, which are situated north and south of the Tyne within the perimeters of the urban conurbation. We also chose one industrial hinterland sampling point further south, the ward of Delves Lane, a village of about 1,300 inhabitants. This also gives us the opportunity to investigate whether the traditional isoglosses that earlier research has revealed to run south of the urban conurbation (see Glauser, 1974, 2000; Kolb, 1966; Kolb, Glauser, Elmer & Stamm, 1979)<sup>11</sup> still hold in 2009.

#### 4. Putting theory into practice: Applying the new method to a dialectological project

We now discuss an application of the model we have developed for sampling across space in the context of the British Isles. Given that the aim of this pilot study is to test the socio-geographically sensitive method outlined above, we restricted our sample to only one slice of the population: Older (40+) speakers with comparatively little formal education (none of our informants went to university or received any form of higher postsecondary education). We sampled one man and one woman per location, all of whom share either kinship or friendship networks with their paired partners and maintain dense social networks in their local community. The informants were born in the locality and have lived in the same ward or in an adjacent one provided that it has the same socio-demographic profile at least until the age of 18 and most of their adult lives.

One fundamental restriction of our sample is thus that it only includes the informants commonly used

in dialectological research. Note in this respect that previous research has established that different socio-demographic groups have different geographies; restricting one's sampling universe to one or more groups can only give us access to one amongst a multitude of intersecting spatialities (see i.e. the geographies of age Hopkins & Pain, 2007, gender Bondi, 1996; McDowell, 1992, ethnicity Bonnett, 1996, 1997 or disability Imrie, 1996). We have thus decided to control for a maximum of social factors. A larger follow-up project will need to include speakers with a range of different speaker profiles in order to get a picture of the full socio-demographic reality of the area covered.

We report on the results from an indirect grammaticality judgement task.<sup>12</sup> Informants were asked to rate sentences by assigning them a number that corresponds to a verbal descriptor (see Labov, 1996). We used the following four-point scale:

- 1 This type of sentence would never be used here—it seems very odd.
- 2 This type of sentence is not very common here but it doesn't seem too odd.
- 3 I have heard this type of sentence locally but it's not that common.
- 4 People around here use this type of sentence a lot.

Example (1) illustrates a sentence as it was administered in our questionnaire. All sentences to be judged were marked in bold and embedded in a short contextualising text of two to three sentences to make them pragmatically more acceptable (see also Buchstaller & Corrigan, 2011).

- (1) Example of the Indirect Grammaticality judgment task  
Please rate these sentences as described above.

The local supermarket got robbed and the police were looking for a witness. They were asking a group of children whether they had seen anything. Suzie pointed at a little girl. She said "**That's the girl seen it**".

1-----②-----3-----4

Altogether there were 149 sentences (74 experimental sentences, 75 fillers) which alternated in randomised order. We divided these sentences into two questionnaires, of which we constructed 2 randomisations each. Every informant thus completed 2 questionnaires with a lengthy break in-between—half of the informants filled out the first randomisation and the others filled out the second.

The linguistic features included into this pilot project are so-called typical "Northern" features, i.e. variants that are traditionally associated with either

**Table 1.** Average ratings for 2<sup>nd</sup> person *yous* in 6 localities in the North East of England.

	Lowick	Linton	Hexham	Westerhope	Jarrow	Delves Lane
Overall	2.06	2.31	3.75	3.5	3.75	2.25
Subject	2.25	2.38	3.88	3.5	4	2.63
Object	1.88	2.25	3.63	3.5	3.5	1.88
	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

Tyneside English and/or Scottish English as described in Beal (1993, 2004) and Miller (2004) *inter alia*. We illustrate them briefly in turn.

The non-standard second person plural pronoun, often spelled *yous*, is a feature of both Tyneside and Scottish English (Beal, 1993:205; Beal & Corrigan, 2004; Miller, 2004:49; Buchstaller & Corrigan, to appear). We tested for the effect of syntactic position on respondents' rating, namely subject versus object position (2a. and 2b. respectively).

- (2) a. **You**s could share some pasta.  
 b. I want to play my song to **yous**.

Multiple negation is widely regarded as being "one of the most stigmatized features of non-standard English" (Beal & Corrigan, 2005:145). We investigated respondents' acceptance of two types of non-standard negation, multiple negation with Standard English lexis—verbal negation and negation with negative polarity items (in 3a.-b.)—as well as the presence of a vernacular negator, Scots *dinnae* where Standard English calls for *don't* or *do not* (in 3c.), and the Tyneside English equivalent *divven't* (in 3d.).

- (3) a. I **didn't** see **nobody**.  
 b. **Nobody** bought **nothing**.  
 c. I **dinnae** eat steak.  
 d. She **divven't** read novels.

We also considered the acceptance rates of relative clause markers used in subject, animate, restrictive sentences. The vernacular variants examined were *as* (4a.), *what* (4b.) and *zero* (4c.).<sup>13</sup> Ball (1996:243) states that "there is no vernacular norm for either BrE or AmE with respect to the distribution of relative markers." Indeed, speakers of non-standard varieties of English tend to show locally specific patterns in their usage of vernacular strategies, which tend to be at the expense of marking with *WH*-elements (Poussa, 1985; Tagliamonte, Smith & Lawrence, 2005).

- (4) a. It's my mother **as** needs them.  
 b. He's the man **what** bought it.  
 c. That's the man **Ø** helped me.

Finally, we investigated the Northern Subject Rule (henceforth NSR, as in 5a.-b.), a phenomenon whereby

verbs attract an *-s* suffix even when the subject NP is not third person singular in function (Beal, 2004:122).<sup>14</sup> Little is known about the geographical scope of its use, and the extent to which its constraints are stable across space. We tested for the NP/PRO constraint, which "marks a verb with *-s* if its subject is anything but an adjacent pronoun" (Montgomery, 1994:86, see 4a). We also analyse the acceptability of NSR with conjoined nouns forming the subject (as in 4b., see Beal & Corrigan, 2000; Godfrey & Tagliamonte, 1999; Buchstaller, Corrigan, Holmberg, Honeybone & Maguire, 2013).

- (5) a. **The children** says they will return your kindness when **they** go **Ø** out there ... (Fitzpatrick, 1994:350)  
 b. **My mother and father** hides in the garden.

We now move on to describe our results in a series of tables which depict the acceptability ratings of these four constructions by linguistic environment and locality. The two-dimensionality of these tables conceals a north-south axis from Lowick in the north, over Linton Colliery to Hexham, Westerhope and Jarrow and finally to Delves Lane in the south, and an east-west axis, which will become particularly important with respect to the location of Hexham, west of the urban conglomeration.<sup>15</sup> Following Horvath & Horvath (2002), we consider the linguistic conditioning of these variables across geography, interpreting breaks in the probabilistic patterns of these variables as areas where "one pattern of sociolinguistic variability gives way to another pattern of sociolinguistic variability at some point in space" (Horvath & Horvath, 2001:47). Hence, according to the cartographical method of natural breaks advocated by these authors, the loci of quantitative or qualitative differences in the constraints that govern these linguistic variables can be interpreted as areas of dialect transitions or—if they are found to cluster in space—even dialect boundaries.

Table 1 depicts the ratings for 2<sup>nd</sup> person plural *yous*, a feature that has been described as a Northern variant more widely (Beal & Corrigan, 2004). Indeed, the ratings from the 6 localities—while variable in terms of overall acceptability—confirm that *yous*, overall as well as in both syntactic positions, is generally recognized as being used by people across the localities sampled in the North East.

**Table 2.** Average ratings for negation strategies in 6 localities in the North East of England.

	Lowick	Linton	Hexham	Westerhope	Jarrow	Delves Lane
Overall	3.84	3.03	3.75	2.84	3.28	2.84
Neg Pol. Item	4	2.75	3.75	2.63	3.88	2.88
Verbal Neg.	4	2.88	3.88	3.25	3.75	2.63
<i>Divven't</i>	3.38	3.25	3.63	<b>3.38</b>	<b>3.75</b>	<b>3.13</b>
<i>Dinnae</i>	<b>4</b>	<b>3.25</b>	<b>3.75</b>	2.13	1.75	2.75
	n.s.	p < .05	n.s.	n.s.	p = .001	n.s.

Note, however, that informants in Delves Lane, Linton Colliery, and Lowick, the three non-urban localities at the north and the south of the periphery of our sample are less accepting of the feature. It is to be investigated whether this is an indication of *yous* being associated with urban speech communities (as suggested in Beal & Corrigan, 2004), especially since the feature receives high acceptability rates in the rural locality close to Hexham.<sup>16</sup>

The preferred linguistic context of *yous* is subject position, and while this constraint is not significant—probably due to low token numbers—the overall direction is the same everywhere, except for Westerhope, where *yous*-ratings are independent of syntactic position. But even in Westerhope, including more speakers into the analysis (as we have done on the basis of a follow-up study) and thus increasing token numbers results in the same pattern of subject over object. Overall, the results for vernacular 2<sup>nd</sup> person *yous* form a relatively homogenous picture where all localities share the same constraint hierarchy. Differentiation across space starts to show when we look at negation in Table 2.

Hughes, Trudgill and Watt (2005) claim that multiple negation is used frequently in Northern and Scottish dialects (but see Anderwald 2004). Indeed, all of our informants identified multiple negation as a feature that is used in the North East, albeit with different degrees of acceptability. Importantly, Table 2 displays the transition from the Scottish *dinnae* to the typical Tyneside *divven't* as incremental changes in mean ratings from one locality to the next. Unsurprisingly, *dinnae* received the highest possible acceptability rating, 4, in the northernmost locality, Lowick, where informants are generally very accepting of vernacular negation. Some 43 miles further south, in Linton Colliery, the acceptability of *dinnae* has shrunk to 3.25, but it is still rated as the highest negative variant. Conversely, informants in the urban Newcastle-Gateshead area prefer the Tyneside form, *divven't* (see Beal, 1993; Glauser, 1974). Note the very low ratings for *dinnae*, particularly in Jarrow.

Note also that Delves Lane, the southernmost sampling point, manifests reduced ratings of *divven't* and increased acceptability for *dinnae*. We suspect that this is due to the phonetic similarity of *dinnae* to another localised northern form, *dinnet* (attested south of the Tyne by Ellis, 1889, for South Shields and by Orton, 1933, for Byers Green, a mere 16 miles from Delves Lane, see also Beal, Burbano-Elizondo & Llamas, 2012)<sup>17</sup>, which we did not test for in this pilot study. There is thus a clear north-south gradation in terms of preference of forms, from *dinnae* in the North over *divven't* in the urban Newcastle-Gateshead conurbation to (we assume) *dinnet* further south. We interpret these results as the perceptual probabilistic outcome of a fan (Glauser, 2000).<sup>18</sup> Note however, that the east-west dimension also matters in this respect: Informants just outside of Hexham, which is about at the same latitude as Newcastle, have roughly equal ratings for *dinnae* and *divven't*. Further research is needed to ascertain whether this finding is an expression of the fact that the *dinnae*-territory spreads further south in the rural TTWA west of Newcastle or whether Hexhamites also regard themselves as users of *dinnet* (or of other local forms). Overall, it seems that the urban Newcastle-Gateshead conurbation is the clear geographical stronghold of *divven't*, whereas nasal variants reach much higher acceptability rates elsewhere. Let us now tackle the ratings for relativisation in Table 3.

In line with Hughes et al. (2005) and Cheshire et al. (1989), *what* is rated highest in all our localities (except Delves Lane, where vernacular relatives receive relatively even ratings). Note, however, contra to claims in the literature, high acceptability of *what* is not restricted to urban localities: the form achieves high scores in Linton Colliery, Lowick and amongst the informants close to Hexham. Note also that, in spite of the fact that *what* is readily accepted in Linton Colliery/Lowick, and has been recorded in Glasgow (Miller, 1993:62), the variant is not traditionally found in Scottish dialects, and Poussa (1985) has suggested that it has spread upwards from the south.

**Table 3.** Average ratings for relativisation strategies in 6 localities in the North East of England.

	Lowick	Linton	Hexham	Westerhope	Jarrow	Delves Lane
Overall	2.75	1.5	2.46	1.71	1.38	2.46
Zero	3	1.13	2.13	1.75	1.13	2.63
As	1.88	1.25	2.38	1.00	1.13	2.25
What	<b>3.38</b>	<b>2.13</b>	<b>2.88</b>	<b>2.38</b>	<b>1.88</b>	2.5
	n.s.	n.s.	n.s.	p < .05	n.s.	n.s.

Note in this respect that Edwards & Weltens' review (1985) suggests that—especially in the North—speakers prefer other vernacular relativisation strategies. In our study, however, *as* is generally rated relatively low and does not follow any consistent pattern (see also Tagliamonte et al., 2005; Kortmann, 2004). Zero relatives only achieve acceptability ratings that surpass *what*-ratings amongst informants in Delves Lane, the southernmost locality. The only other locality with reasonable acceptability ratings for zero relatives is Lowick in the extreme North East (and to a certain extent Hexham). Note that the zero form, which has been in use ever since Old English (Traugott, 1972) has been found in the Southern Scottish Borders by Murray (1873:194), who commented that “an ellipsis of the relative is extremely common.” It has also been attested in both Tyneside and Sheffield in *The Survey of English Dialects* (Orton et al., 1962–1971), *The Newcastle Electronic Corpus of Tyneside English* (<http://www.ncl.ac.uk/necte>) and *The Survey of Sheffield Usage* (see Beal & Corrigan, 2007; Buchstaller & Corrigan, to appear). Given the lack of comparative diachronic quantitative data across the North Eastern area, it is not entirely clear whether our finding might be taken as an indication that the geo-spatial locus—at least synchronically—of zero relatives in the North East is more in the peripheral areas. More data, also from younger age groups, is needed in order to establish the complex competition amongst relativisation strategies in the North East of England.

Finally, let us consider the linguistic conditioning of the NSR across the six localities in Table 4. Historically, as we pointed out above, verbal *-s* has been reported to be conditioned by the NP/PRO constraint. Also conjoined nouns tend to favour the occurrence of verbal *-s*. Synchronically, however, these constraints seem to be undergoing locally specific reinterpretation (see Buchstaller et al., 2013).

The acceptability ratings in Table 4 reveal localised patterns. Three localities, Westerhope, Delves Lane and Linton Colliery, display a binary constraint hierarchy whereby conjoined NPs receive considerably higher ratings than subjects that consist of full non-3<sup>rd</sup> person singular NPs or pronouns, which are rated least

acceptable. The rating of NPs over pronouns is a synchronic reflex of the NP/PRO rule. The preference of conjoined NPs over full NPs is fully in line with Visser (1963), Beal & Corrigan (2000) and Godfrey & Tagliamonte (1999). Indeed, Buchstaller et al. (2013) have suggested that this pattern might be due to reanalysis of the 2<sup>nd</sup> conjoint of the complex subject NP as a 3<sup>rd</sup> person sg. subject.

Note, however, that Jarrow and Hexham display a slightly different pattern whereby full NPs favour the acceptance of the NSR over conjoined NP with pronouns coming last as in the other localities. We might want to argue that in these localities, whereas the original NP/PRO constraint is still firmly in place, the reinterpretation of the 2<sup>nd</sup> conjoint has not taken place. Indeed, research in Hawick, a small town in the Scottish borders has revealed similar results (see Buchstaller et al., 2013; Childs, 2013).

Finally, the informants in Lowick, while displaying the conjoined NP effect, rate pronouns—the lowest ranked environment anywhere else—higher than single NPs. Hence, it seems that informants in Lowick do not orient to the NP/PRO constraint at all. Obviously, given the small number of informants sampled in these localities, the variability in Table 4 might be due to orthogonal social/attitudinal or even idiosyncratic factors and these results need to be confirmed on the basis of larger data base. However, the findings reported here support research by Buchstaller et al. (2013) conducted in Westerhope and Hawick that is based on a larger number of participants.

We suggest that there are two possible explanations for the geographically differentiated outcome in Table 4—assuming they are not sampling artefacts: The variability could be the result of the locally specific adaptation of a bundle of linguistic constraints that are currently changing across a wider spatial area. As Buchstaller et al. (2013) point out, the NSR seems to be in the process of undergoing major reanalysis—and our data from 6 different localities across the North East suggests that this process results in geospatial diversity synchronically. Alternatively, it might well be that the NSR, even historically, has never

**Table 4.** Average ratings for 2<sup>nd</sup> person NSR in 6 localities in the North East of England.

	Lowick	Linton	Hexham	Westerhope	Jarrow	Delves Lane
Overall	2.17	1.34	2.39	1.86	1.71	2.56
Conjoined	<b>2.57</b>	<b>1.71</b>	2.29	<b>2.57</b>	1.93	<b>3.79</b>
Full NP	1.94	1.33	<b>2.78</b>	1.56	<b>2.67</b>	2.61
Pronoun	2.15	1.27	2.17	1.33	1.46	2.35
	n.s.	n.s.	n.s.	p < .001	p < .05	p < .001

had the geographical uniformity it has been portrayed as having. Rather, it might have always been subject to localised constraints. Historical treatments tend to be based on impressionistic and/or small-scale studies and past empirical research lacks systematic geographical coverage. More data—synchronic as well as diachronic—is needed from a range of localities in order to ascertain the mechanisms lying behind the results shown in Table 4 (see Pietsch, 2005; Ramisch, 2008).

Overall, Tables 1–4 display probabilistically gradient acceptability ratings of linguistic variability. Conceptualising these ratings within two dimensions, namely space—north to south and east to west—and place—the urban conurbation Newcastle–Gateshead versus various rural locations (Horvath & Horvath, 2001)—reveals that some variables are more systematically patterned than others. Indeed, using the natural break pattern allows us to examine the “dialect landscapes” (Britain, 2010:72) across the North East that fall out of a socio-demographically informed sampling strategy.

The northern sampling points, Lowick and Linton Colliery, give high ratings to *dinnae*, whereas the urban Tyneside complex has particularly high ratings of *divven’t*. Delves Lane, the most southern locality, while not categorically different from any of the other sampling points, manifests the influence of another nasal variant that has been associated with more southern localities. These ratings give support on the perceptual level to the description of the English–Scottish border as a fan. However, orthogonal to space, place effects are also operational in the ratings for vernacular negatives: Hexham, which is at the same latitude as Newcastle, garners relatively high *dinnae* ratings. It thus seems as if preponderance for *divven’t* is associated mainly with the conurbation Newcastle–Gateshead. We also detected a possible urban predominance for *you*s, which received much higher ratings in localities within the boundaries of the urban conurbation—in Jarrow and Westerhope—compared to the rural countryside, except for Hexham. Similarly, the higher acceptability ratings for zero forms on the northern and southern periphery might be due to the preponderance of its competitor form, *what*, in the urban centre. The ratings for the NSR, on the other

hand, seem to be the locally specific manifestations of a phenomenon that has been described as generally northern (Murray, 1873) but the constraining factors of which seem to vary from place to place (see Buchstaller & Corrigan, to appear).

## 5. Conclusion

Chambers and Trudgill (1998:30) point out that “the future of dialect studies will have to be directed towards more representative populations.” Indeed, we have argued that considerations of geo-demographical representativeness have not received the kind of attention they warrant in large-scale dialectological work. Britain, similarly, finds that much of dialectological research has either tended to “carefully control (...) [space] out of the study” (2009:143) or turned it into a “homogenised, historically-, socio-economically-, and institutionally blind blank canvas” (2010:87). In this paper, we investigate the interface between human geography and sociolinguistics with an eye on methods that have the potential to enrich dialectological research. We focus on concepts and models that inform sampling decisions in multi-locality dialectological research.

The first sampling parameter we propose relies on tessellation via travel to work areas (TTWAs), although other types of zones, borders or “functional” regions that encapsulate regular human flows/activities can also be used, depending on the study context and data availability. Such functional regions, we argue, are the fundamentally local outcomes of routinised day-to-day behaviour and therefore inherently meaningful for the understanding of spatialised practices, linguistic as well as others. As such, socially sensitive tessellation has a major advantage over the standard dialectological sampling criteria, which focused, if at all, on “fairly long-term mobility rather than that of the taken-for-granted everyday kind” (Britain, 2010:87). However, we have to point out that the UK TTWAs we used have been defined to treat all commuting flows equally, ignoring personal circumstances such as teleworking and gendered employment opportunities and practices. A more sensitive approach to tessellation might take into account factors such as age, ethnicity, industry, occupation and

gender, to name a few, which is also possible from an analytical perspective, depending on relevant data availability.

Within these TTWAs we rely on a formal socio-economic area classification for choosing sampling points that correspond to the socio-demographic and economic make-up of the region they represent. A model that relies on a combination of travel to work areas and socio-economic area classification has a number of assets: Both parameters are readily available online and easily applied to a sampling universe of any size and location within Britain (and—contingent on the availability of census data—also elsewhere). They have been tested in a wide range of geo-demographical research, which has the added benefit of interdisciplinary convergence. Finally, they hand dialectological researchers a ready-to-use sampling tool that is fully cognisant of the recent advances in human geography. This is also true for many other countries with regular population censuses (e.g. US, Australia) or detailed citizens' registers (e.g. Germany, Sweden). Such population-wide data may have different names and consist of different variables, but they are widely acceptable for geo-demographic research. For example, functional regions derived from the census (i.e. similar to the TTWAs) are known as Traffic Analysis Zones (TAZ) in the US and Travel Zones in Australia, available with Journey to Work (JTW) data. Although different parameters may have been used in their construction, the resulting regions and purpose are similar to the TTWAs. At the smaller geographical level, there is a range of governmental or proprietary data "products" providing socio-economic classifications (also known as population segmentations), similar to the one used here. Official census classifications are available freely, while commercial population segmentations attract a premium, although sometimes data companies are prepared to negotiate lower prices or make available free data for research purposes. In contexts where such official or commercial products are not available, researchers should be in a position to obtain a number of census-derived variables (e.g. age, employment, housing tenure of population) for larger areas and produce such classifications themselves, using standard statistical techniques (e.g. cluster analysis in SPSS). Such a task is more demanding and requires some understanding of spatial statistics, but it is still feasible to produce classifications from raw population data for sociolinguistic sampling. In any case, we encourage sociolinguists to consider such geographical approaches and discuss their data needs with colleagues from geography and planning.

Applying these two methods as sampling parameters to the extreme North East of England leaves us with a measure of socio-demographic representativeness and

geographically informed coverage. We chose six localities with varying geospatial and socio-economic profiles to investigate a range of linguistic features. Acceptability judgement tasks conducted across these sampling points reveal a dialectological landscape constrained by linguistic, geographic and human geographic factors. By themselves, the results of our pilot study are only a small jigsaw piece amongst innumerable linguistic geographies which continue to develop and evolve and, as such, fundamentally limited in scope. But we hope to have demonstrated that the combinatory approach we propose is an adequate tool for dialectological research.

Obviously, given the fluidity of development both on the linguistic and the geographical plane, geo-spatially sensitive research needs to keep abreast of newer census data as it becomes available (the raw data for the 2011 UK census has just been released for larger areas and local authorities). Indeed, a comparison between census outputs from different years, while often marred by varying tessellation units, can provide important information about the changing nature of routines, linguistic as well as geographical ones. Furthermore, while space precludes us from showcasing such an analysis here, a research design that samples different age bands would also be in the position to investigate important questions such as the extent and direction of levelling or dialect supralocalisation.<sup>19</sup> In future research, such questions need to be investigated on the basis of both perception and production data. Another unanswered question is whether the findings reported here converge with patterns formed on the basis of phonology. Moreover, while we have refrained from doing so due to the relatively low number of speakers per tessellation unit, the statistical concept of standard deviation could be used as an analytical parameter, providing researchers with a fruitful diagnostic of focusing (Le Page, 1978), especially when comparing younger and older speakers.

To conclude, applying a combinatory human geographical sampling method to investigate linguistic variability in the extreme North East has resulted in a socio-demographically informed snapshot of socio-geographical patterns of language variation. We hope the method we propose has brought us one step further towards a "spatially sensitive dialectology, one which recognises and synthesises the ever evolving physical, social and perceptual spaces we live in and by, it places the spaces created, maintained and changed by *interaction* at centre stage" (Britain, 2010:69, emphasis in original).

#### Acknowledgements

We gratefully acknowledge the support of Newcastle University's Faculty Research Fund, as well as our

fieldworker Laura Steventon. The digital maps used hold Crown Copyright from EDINA Digimap, a JISC supplied service. © Crown Copyright/database right 2013. An Ordnance Survey/EDINA supplied service.

## Notes

- <sup>1</sup> For a discussion of early attempts of using space as a sampling criterion in dialectological projects see Britain (2009, 2010) as well as Chambers & Trudgill (1998).
- <sup>2</sup> This overview does not cover data-bases that were collected on the basis of prefabricated materials, such as the Freiburg Corpus of English Dialects (FRED), the Scots corpus or the Origins of New Zealand (ONZE) corpus, since their sampling criteria are largely based on—or at least majorly informed by—the availability of pre-recorded data rather than concerned with the question of how to conceptualise and sample across space.
- <sup>3</sup> Note that other dialect atlas projects, such as SCANDIA-SYN and ASIS (the *Syntactic Atlas of the Italian Dialects*) employ rather different sampling strategies. None of them rely on grids, and while ASIS pursued a more traditional dialectological focus on rural communities, SCANDIA-SYN sampled in urban and rural areas (<http://asis-cnr.unipd.it/> and [http://www.tekstlab.uio.no/nota/scandiasyn/dialect\\_data\\_collection.html](http://www.tekstlab.uio.no/nota/scandiasyn/dialect_data_collection.html)).
- <sup>4</sup> We would like to thank an anonymous reviewer for pointing out that traces of this aversion to sample sites affected by heavy immigration are also evident in many later variationist sociolinguistics studies, where, for example, the avoidance of non-natives is still well-ingrained.
- <sup>5</sup> Sociolinguists (Britain, 2010; Kerswill, 2009a) have argued that the binary division between urban vs. rural speech community type is much too simplistic and needs to be elaborated by considering socio-demographic, human geographical, and ideological factors. This argument parallels similar discussions in human geography debating the binary nature of the urban-rural divide, which is sensitive to matters of scale and hence fundamentally arbitrary (Cloe & Little, 1997; Shucksmith, 2000).
- <sup>6</sup> Field sites were at the minimum 10–12 miles apart but “no prior consideration was given to the social history of individual localities. Ultimately, the final selection was left to the individual fieldworkers themselves.” (<http://www.yorkshiredialect.com/>, accessed 22 October, 2012). Linguistic historicism led to the favouring of rural communities, especially those involved in agriculture, at the expense of urban localities. However, some urban areas were interspersed into the sample, albeit in a rather unprincipled way. “The literature usually refers to the “four urban sites” of Hackney, Leeds, Sheffield, and York. The survey does seem to have been generally more urban-focused in West Yorkshire (...). Outside of London and West Yorkshire, nowhere near to a large city was examined.” (<http://www.leeds.ac.uk/english/activities/lavc/PDFs/SEDIM.pdf>, accessed 8 May, 2009). Human geographical parameters (in/out-migration and new town formation) did play a role in the research design, albeit a negative one. They were used as the basis for *excluding* certain areas from the investigation: “Newly developed localities were to be avoided (presumably to evade the possible linguistic influence of in-migratory groups from other dialect areas). Preference was to be given to communities containing a stable population of approximately 500 inhabitants for a hundred years or more” (<http://www.yorkshiredialect.com>, accessed 22 October, 2012). Overall, SED sampling seems not to have been based on principled geographical or human geographical parameters.
- <sup>7</sup> Obviously inter-regional mobility and global/supra-local flows (Kerswill, 2003; Milroy, Milroy & Hartley, 1994; Stuart-Smith, 2006; Watt, 2002) interact with local/regional developments (Buchstaller & D’Arcy, 2009; Meyerhoff, 2009). In this paper we focus on capturing the mundane, day-to-day activities that shape our most fundamental understanding and appropriation of space, bearing in mind that a fully comprehensive model of social space will also have to include these supra-local flows.
- <sup>8</sup> Travel to work areas were used by Corrigan (1997) and Sayers (2009) to investigate dialect diversity.
- <sup>9</sup> The independence of TTWA and institutionalised boundary is also evident by the fact that the “Hexham and Haltwhistle” TTWA stretches outside the North East region (compare Maps 1 and 2).
- <sup>10</sup> Consider in this regard neighbourhood-based approaches within sociolinguistics, which focus on very local level socio-scapes (such as social networks and communities of practise). These micro-approaches place the appropriation of place via the development of linguistic and other practices at the centre of investigation (Eckert, 2000; Labov, 2001; Milroy & Milroy, 1985). While such approaches conceptualise spatial practices as emergent—networks are “formed, maintained and renewed across social space” (Britain 2010:16)—and are thus able to account for the continuous becoming of local geographies, it is less clear how they can be employed in multi locality, atlas-type projects.
- <sup>11</sup> Thanks to Warren Maguire for pointing out that the appendices of the Computer developed *Linguistic Atlas of England* (2<sup>nd</sup> vol., Viereck & Ramisch, 1997) show isoglosses “separating the northern two or three Durham locations and Northumberland from everything further south” (p.c., 10<sup>th</sup> December 2009). These isoglosses are based on the use of morphological (i.e. *I am* vs. *I is*, see also Ellis, 1889), on lexical (i.e. *wrap* vs. *lap*, *shank* vs. *shaft*) and phonological differences (rhoticity in #C contexts and [ej] in *five*, *Friday*).
- <sup>12</sup> We also collected spontaneous conversational data. However, due to the relative infrequency of the investigated morphosyntactic features in spontaneous interaction (see also Buchstaller & Corrigan, 2011), here, we have chosen to report on the perception data. The test we chose, the indirect grammaticality judgement task, has the benefit of placing relatively little prescriptive pressure on informants.

Indeed, by asking them to state how frequently vernacular features are being used by other people in their locale—rather than whether they themselves use them—this task gives informants the opportunity to distance themselves from these features. Furthermore, as we have argued elsewhere, indirect grammaticality judgement tasks are relatively simple to convey to informants—once they have mastered the notion of gradable acceptability (see Buchstaller & Corrigan, 2011). Bearing in mind the reservations voiced by Fasold (1984), they also produce results that are readily quantifiable (Coward, 1997:72).

- <sup>13</sup> While zero forms with object function are acceptable in Standard English, their use with subject function is restricted to colloquial or vernacular varieties (see Quirk, Greenbaum, Leech & Svartvik, 1985).
- <sup>14</sup> Despite its name, the phenomenon is (i) not categorical but variable and (ii) has also been found to occur in dialects outside the North of England (Godfrey & Tagliamonte, 1999; Montgomery, Fuller & DeMarse 1993; Ramisch, 2008).
- <sup>15</sup> The question of how to visualise findings adds another complex angle to the interplay between language and geography: How to display such fine grained, multi-factorial differences on three or many more dimensions? We suggest that, rather than isoglosses or percentage based maps, recent human geographical methods of visualisation, such as the generation of surfaces, might be a future avenue for displaying the full multidimensional complexity of linguistic variability across space.
- <sup>16</sup> We need to point out that the low ratings in Lowick are due to the female informant categorically rejecting all instances of *yous*, rating them as 1. We are uncertain how to interpret this result since this informant rated all other forms variably acceptable, in line with the other Lowick informants.
- <sup>17</sup> Heslop (1903) mentions three forms for Northumberland, *dinna*, *dinnet* and *divnt*. Glauser (1974) reports that *dinnet* is not attested in his data; it seems to occur south of his sampling universe.
- <sup>18</sup> The term ‘fan’ is used metaphorically in order to describe the phenomenon whereby isoglosses are split up and run almost parallel, resulting in a linguistic continuum. In extreme cases such as the Rhenish Fan described in Bloomfield (1933), “splay out like the spokes of a fan” (Chambers & Trudgill, 1998:92). The net effects resulting from such a continuum are cumulative systemic differences between the linguistic systems on both sides of the fan.
- <sup>19</sup> While this paper does not investigate important socio-demographic factors such as immigration, aging, or counter-urbanisation, we propose that the method described here can provide a useful template via which such factors can be investigated. SEAC profiles can be built to include a wealth of socio-demographic factors—basically anything that was asked in the census. However, migration patterns test the boundary of its usefulness since the UK census only asks for where people lived last year and where they were born with no information about their whereabouts in-between.

## References

- Anderwald, Liselotte. 2004. The morphology and syntax of the varieties of English spoken in the Southeast of England. In B. Kortmann & E. Schneider (with K. Burridge, R. Mesthrie & C. Upton) (eds), *A handbook of varieties of English, vol. 2: Morphology and syntax*. Berlin & New York: Mouton de Gruyter, 175–195.
- Allen, John, Doreen Massey & Allan Cochrane. 1998. *Rethinking the region*. London: Routledge.
- Auer, Peter & J. Schmidt (eds.) 2010. *Language and space: An international handbook of linguistic variation*. Berlin: Mouton de Gruyter.
- Ball, Catherine. 1996. A diachronic study of *relative* markers in spoken and written English. *Language Variation and Change* 8:227–258.
- Barbiers, Sjef, Leonie Cornips & Jan Pieter Kunst. 2007. The syntactic atlas of the Dutch dialects: A corpus of elicited speech and text as an on-line dynamic atlas. In J. C. Beal, K. P. Corrigan & H. Moisl (eds), *Creating and digitizing language corpora, vol. 1: Synchronic databases*. Hampshire: Palgrave Macmillan, 54–90.
- Beal, Joan C. 1993. The grammar of Tyneside and Northumbrian English. In J. Milroy & L. Milroy (eds), *Real English: The grammar of English dialects in the British Isles*. Harlow: Longman, 187–213.
- Beal, Joan C. 2004. The morphology and syntax of English dialects in the north of England. In B. Kortmann & E. Schneider (eds), *A handbook of varieties of English*. Berlin: Mouton de Gruyter, 114–141.
- Beal, Joan C., Lourdes Burbano-Elizondo & Carmen Llamas. 2012. *Urban North-Eastern English: Tyneside to Teeside*. Edinburgh University Press.
- Beal, Joan C. & Karen P. Corrigan. 2000. Comparing the present with the past to predict the future for Tyneside British English. *Newcastle and Durham Working Papers in Linguistics* 6:13–30.
- Beal, Joan C. & Karen P. Corrigan. 2004. The impact of nineteenth century Celtic English migrations on contemporary Northern Englishes: Tyneside and Sheffield compared. Paper presented at Celtic Englishes IV, University of Potsdam.
- Beal, Joan C. & Karen P. Corrigan. 2005. No, nay, never: Negation in Tyneside English. In Yoko Iyeyri (ed.), *Aspects of negation in English*. Amsterdam: John Benjamins, 139–158.
- Beal, Joan C. & Karen P. Corrigan. 2007. ‘Time and Tyne’: A corpus based study of variation and change in relativization strategies in Tyneside English. In S. Elspaß, N. Langer, J. Scharloth & W. Vendenbussche (eds), *Language history from below- Linguistic variation in the Germanic languages from 1700–2000*. Berlin/New York: Walter de Gruyter, 1–16.
- Bloomfield, Leonard. 1933. *Language*. Holt, Rhinehart and Winston. New York.
- Bondi, Liz. 1996. Geography degrees and gender – interpreting the patterns. *Area* 28:221–224.
- Bonnett, Alastair. 1996. Constructions of ‘race’ place and discipline: Geographies of ‘racial’ identity and racism. *Ethnic and Racial Studies* 19:864–883.

- Bonnett, Alastair. 1997. Geography, race and whiteness: invisible traditions and current challenges. *Area* 29:193–199.
- Britain, David. 2010. Conceptualisations of geographic space in linguistics. In A. Lameli, R. Kehrein & S. Rabanus (eds), *The handbook of language mapping*. Berlin: Mouton de Gruyter, 69–97.
- Britain, David. 2009. Language and space: The variationist approach. In P. Auer & J. Schmidt (eds), *Language and space: An international handbook of linguistic variation*. Berlin: Mouton de Gruyter, 142–162.
- Britain, David. 2004. Geolinguistics – diffusion of language. In U. Ammon, N. Dittmar, K. Mattheier & P. Trudgill (eds), *Sociolinguistics: International handbook of the science of language and society*. Berlin: Mouton De Gruyter, 34–48.
- Britain, David. 2002. Space and spatial diffusion. In J. K. Chambers, P. Trudgill & N. Schilling-Estes (eds), *The handbook of language variation and change*. Oxford: Blackwell, 603–637.
- Buchstaller, Isabelle. 2008. The localization of global linguistic variants. *English World-Wide* 29:15–44.
- Buchstaller, Isabelle & Corrigan, Karen P. 2011. How to make intuitions succeed: Testing methods for analysing syntactic microvariation. In W. Maguire & A. McMahon (eds), *Analysing variation in English: What we know, what we don't, and why it matters*. Cambridge: Cambridge University Press, 30–48.
- Buchstaller, Isabelle & Karen P. Corrigan. to appear. Northern English (morpho)syntax. In R. Hickey (ed.), *Northern Englishes*.
- Buchstaller, Isabelle, Karen P. Corrigan, Anders Holmberg, Patrick Honeybone & Warren Maguire. 2013. 'T-to-R' and the 'Northern Subject Rule'. *Journal of English Linguistics* 17:85–128.
- Buchstaller, Isabelle, Karen P. Corrigan, Adam Mearns & Hermann Moisl. 2011. The talk of the toon: Stories from the people of the North East, past and present, in their own words. Newcastle upon Tyne: Newcastle University.
- Buchstaller, Isabelle & Alexandra D'Arcy. 2009. Localized globalization: A multi-local, multivariate investigation of *b like*. *Journal of Sociolinguistics* 13:291–331.
- Chambers, J. K. & Peter Trudgill. 1998. *Dialectology*, 2nd edn. Cambridge: Cambridge University Press.
- Champion, T. & M. G. Coombes. 1983. *Functional Regions: Definitions, Applications, Advantages*. Centre for Urban and Regional Development Studies, University of Newcastle: Factsheet 1.
- Cheshire, Jenny, Viv Edwards & Pamela Whittle. 1989. Urban British dialect grammar: The question of dialect levelling. *English World-Wide* 10:185–225.
- Cheshire, Jenny, Viv Edwards & Pamela Whittle. 1993. Non-standard English and dialect levelling. In J. Milroy & L. Milroy (eds), *Real English. The grammar of English dialects in the British Isles*. London: Longman, 53–96.
- Childs, C. 2013. Verbal -s and the Northern Subject Rule: Spatial variation in linguistic and sociolinguistic constraints. In X.-A. Álvarez Pérez, E. Carrilho & C. Magro (eds), *Current approaches to limits and areas in dialectology*. Newcastle upon Tyne: Cambridge Scholars, 311–344.
- Cloke, Paul & Jo Little (eds). 1997. *Contested countryside cultures*. London: Routledge.
- Coombes, M. G. & S. Bond. 2008. *Travel-to-work areas: The 2007 review*. London: Office for National Statistics.
- Coombes, M. G., J. S. Dixon, J. B. Goddard, S. Openshaw & P. J. Taylor. 1982. Functional regions for the population census of Great Britain. In D. T. Herbert & R. J. Johnston (eds), *Geography and the urban environment*. London: Wiley, 63–112.
- Corrigan, Karen P. 1997. The syntax of South Armagh English in its socio-historical perspective. Unpublished PhD dissertation, University College Dublin.
- Cowart, Wayne. 1997. *Experimental syntax: Applying objective methods to sentence judgments*. Thousand Oaks, CA: Sage Publications.
- Dorling, Danny. 2004. Healthy places, healthy spaces. *British Medical Bulletin* 69:101–114.
- Eckert, Penelope. 2000. *Linguistic variation as social practice*. Oxford: Blackwell.
- Edwards, Viv & Bert Weltens. 1985. Research on non-standard dialects of British English: Progress and prospects. In W. Viereck (ed.), *Focus on: England and Wales*. Amsterdam: Benjamins, 97–139.
- Ellis, A. J. 1889. *The existing phonology of English dialects compared with that of West Saxon Speech, pt. V: On early English pronunciation*. London: English Text Society.
- Fasold, Ralph. 1984. *The sociolinguistics of society*. Oxford: Blackwell.
- Fitzpatrick, David. 1994. *Oceans of consolation. Personal accounts of Irish migration to Australia*. Ithaca: Cornell University Press.
- Giddens, Anthony. 1984. *The constitution of society: Outline of the Theory of Structuration*. Cambridge: Polity Press.
- Glauser, Beat. 1974. *The Scottish-English linguistic border*. Bern: A Francke AG Verlag.
- Glauser, Beat. 2000. The Scottish/English border in hindsight. *International Journal of the Sociology of Language* 145:65–78.
- Godfrey, Elizabeth & Sali Tagliamonte. 1999. Another piece for the verbal -s story: Evidence from Devon in southwest England. *Language Variation and Change* 11:87–121.
- Gregory, Derek. 1985. Suspended animation: The stasis of diffusion theory. In D. Gregory & J. Urry (eds), *Social relations and spatial structures*. London: Macmillan, 296–336.
- Harris, Richard, Peter Slight & Richard Webber. 2005. *Geodemographics, GIS and neighbourhood targeting*. Chichester: Wiley.
- Heslop, R. O. 1903. Dialect notes from northernmost England. *Transactions of the Yorkshire Dialect Society* 1:7–31.
- Hopkins, Peter & Rachel Pain. 2007. Geographies of age: Thinking relationally. *Area* 39:287–294.
- Horvath, Barbara M. & Ronald J. Horvath. 2001. A multilocality study of sound change in progress: the case of /l/ vocalization in New Zealand and Australian English. *Language Variation and Change* 13:37–57.
- Horvath, Barbara M. & Ronald J. Horvath. 2002. The geolinguistics of /l/ vocalization in Australia and New Zealand. *Journal of Sociolinguistics* 6:319–346.
- Hughes, Arthur, Peter Trudgill & Dominic Watt. 2005. *English accents and dialects: An introduction to social and regional*

- varieties of English in the British Isles*, 4th edn. London: Hodder Arnold.
- Imrie, Robert. 1996. Ableist spaces, disablist geographies: Towards a reconstruction of Golledge's geography of the disabled. *Transactions of the Institute of British Geographers* New Series 21:397–403.
- Kerswill, Paul. 2003. Dialect levelling and geographical diffusion in British English. In D. Britain & J. Cheshire (eds), *Social dialectology. In honour of Peter Trudgill*. Amsterdam: Benjamins, 223–243.
- Kerswill, Paul. 2009a. New motors or change in an old world speech community: Minority-led innovation in Britain. Plenary lecture given at ICLAVE, University of Copenhagen.
- Kerswill, Paul. 2009b. Supralocalisation equals deracination; innovation equals striking new roots. Paper presented at FRIAS, Freiburg University.
- Kolb, Eduard. 1966. *Phonological atlas of the northern region*. Bern: A. Francke Verlag.
- Kolb, Eduard, Beat Glauser, Willy Elmer & Renate Stamm (eds.). 1979. *Atlas of English sounds*. Bern: A. Francke Verlag.
- Kortmann, Bernd. 2004. Synopsis: Morphological and syntactic variation in the British Isles. In B. Kortmann & E. Schneider (with K. Burridge, R. Mesthrie & C. Upton) (eds), *A handbook of varieties of English, vol. 2: Morphology and syntax*. Berlin & New York: Mouton de Gruyter, 1089–1103.
- Labov, William. 1982. Building on empirical foundations. In W. Lehmann & Y. Malkiel (eds), *Perspectives on historical linguistics*. Amsterdam and Philadelphia: John Benjamins, 17–92.
- Labov, William. 1996. When intuitions fail. *Papers from the 32nd Regional Meeting of the Chicago Linguistics Society* 32, 76–106.
- Labov, William. 2001. *Principles of linguistic change: Social factors*. London, New York: Blackwell.
- Labov, William, Sharon Ash & Charles Boberg. 2006. *Atlas of North American English: Phonology and phonetics*. Berlin: Mouton de Gruyter.
- Lameli, Alfred, Roland Kehrein & Stefan Rabanus (eds.). 2010. *The handbook of language mapping*. Mouton de Gruyter.
- Le Page, Robert. 1978. Projection, focussing, diffusion, or, Steps towards a sociolinguistic theory of language, illustrated from the sociolinguistic survey of multilingual communities. Stages I: Cayo District, Belize (formerly British Honduras), and II: St. Lucia. *Society for Caribbean Linguistics Occasional Paper* No.9.
- Masser, I. & J. Scheurwater. 1980. Functional regionalization of spatial interaction data: an evaluation of some suggested strategies. *Environment and Planning A* 12:1357–1382.
- Massey, Doreen. 1985. New directions in space. In D. Gregory & J. Urry (eds), *Spatial relations and spatial structures*. London: Macmillan, 9–19.
- McDowell, Linda. 1992. Doing gender: Feminisms, feminists and research methods in human geography. *Transactions of the Institute of British Geographers* 17:399–416.
- Meyerhoff, Miriam. 2009. Replication, transfer and calquing: Using variation as a tool in the study of language contact. *Language Variation and Change* 21:1–21.
- Miller, Jim. 1993. The grammar of Scottish English. In J. Milroy & L. Milroy (eds), *Real English: The grammar of English dialects in the British Isles*. Harlow: Longman, 99–138.
- Miller, Jim. 2004. Scottish English: Morphology and syntax. In B. Kortmann & E. Schneider (eds), *A handbook of varieties of English*. Berlin: Mouton de Gruyter, 47–72.
- Milroy, James & Milroy, Lesley. 1985. Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21:339–384.
- Milroy, Lesley & Matthew Gordon. 2003. *Sociolinguistics: Method and interpretation*. Oxford: Blackwell.
- Milroy, James, Lesley Milroy & Sue Hartley. 1994. Local and supra-local change in British English: The case of glottalisation. *English World-Wide* 15:1–33.
- Montgomery, Michael. 1994. The evolution of verbal concord in Scots. In A. Fenton & D. MacDonald (eds), *Studies in Scots and Gaelic*. Edinburgh: Edinburgh University Press, 81–95.
- Montgomery, Michael, Janet Fuller & Sharon DeMarse. 1993. 'The black men has wives and sweet hearts (and third person plural-s) jest like the white men': Evidence for verbal-s from the written documents on 19<sup>th</sup> century African American speech. *Language Variation and Change* 5:335–357.
- Mooney, John & Jon Carling. 2006. *Spatial analysis of economic flows in North East England*. <http://www.nerip.com>
- Murray, James. 1873. *The dialect of the southern counties of Scotland: Its pronunciation, grammar, and historical relations with an appendix on the present limits of the Gaelic and Lowland Scotch, and the dialectical divisions of the Lowland tongue, and a linguistic map of Scotland*. London: Asher and Co.
- Orton, Harold. 1933. *The phonology of a South Durham Dialect*. London: K. Paul.
- Orton, Harold, et al. (1962–1971). *Survey of English dialects (B): The basic material, vol. 4*. Leeds: E. J. Arnold.
- Pietsch, Lukas. 2005. *Variable grammars: Verbal agreement in northern dialects of English*. Tübingen: Niemeyer.
- Poussa, Patricia. 1985. Historical implications of the distribution of the zero-pronoun relative clause in Modern English dialects: Looking backwards towards OE from Map55 of the linguistic atlas of England. In S. Jacobson (ed.), *Papers from the third Scandinavian symposium on syntactic variation*. Stockholm: Almqvist & Wiksell, 99–117.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Ramisch, Heinrich. 2008. The Northern Subject Rule and its 'northernness' – A geolinguistic perspective. Paper presented at the Sociolinguistics Symposium 17, Free University of Amsterdam.
- Sankoff, Gillian. 1980. A quantitative paradigm for the study of communicative competence. In G. Sankoff (ed.), *The social life of language*. Philadelphia: University of Pennsylvania Press, 47–49.
- Sayers, David. 2009. Reversing Babel: Declining linguistic diversity and the flawed attempts to protect it. Unpublished PhD thesis, University of Essex.

- Shortt, Niamh K., Adrian Moore, Mike Coombes & Colin Wymer. 2005. Defining regions for locality health care planning: A multidimensional approach. *Social Science & Medicine* 60:2715–2727.
- Shucksmith, Mark. 2000. *Exclusive countryside? Social inclusion and regeneration in rural Britain*. York: Joseph Rountree Foundation.
- Soja, Edward. 1989. *Postmodern geographies: The reassertion of space in critical theory*. London: Verso.
- Stuart-Smith, Jane. 2006. The influence of media on language. In C. Llamas, P. Stockwell & L. Mullany (eds), *The Routledge companion to sociolinguistics*. London: Routledge, 140–148.
- Stuart-Smith, Jane. 2002–5. Economic and Social Research Council: Contributory factors in accent change in adolescents, Grant no. R000239757.
- Tagliamonte, Sali, Jennifer Smith & Helen Lawrence. 2005. No taming the vernacular! Insights from the relatives in northern Britain. *Language Variation and Change* 17:75–112.
- Traugott, Elizabeth. 1972. *A history of English syntax*. New York: Holt, Rinehart, and Winston.
- Vickers, Dan & Phil Rees. 2007. Creating the national statistics 2001 output area classification. *Journal of the Royal Statistical Society Series A* 170:379–403.
- Viereck, Wolfgang & Heinrich Ramisch. 1997. *Computer developed linguistic atlas of England (CLAE), vol. 2*. Tübingen: Max Niemeyer Verlag.
- Visser, Fredericus. 1963. *An historical syntax of the English language*. Leiden: Brill.
- Watt, Dominic. 2002. 'I don't speak with a Geordie accent, I speak, like, the Northern accent': Contact-induced levelling in the Tyneside vowel system. *Journal of Sociolinguistics* 6:44–63.