Studies in Language Variation 27

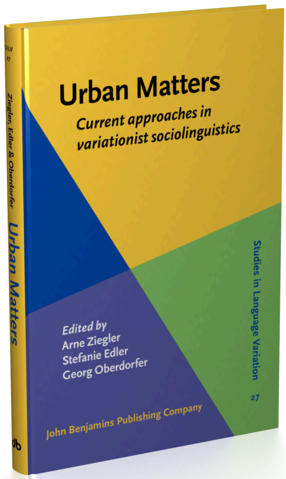
# Areal microvariation in German-speaking urban areas (Ruhr Area, Berlin, and Vienna)


Simon Pröll | Ludwig-Maximilians-Universität München
Stephan Elspaß | Universität Salzburg
Simon Pickl | Universität Salzburg



https://doi.org/10.1075/silv.27.10pro





Pages 227–252 of
**Urban Matters: Current approaches in variationist sociolinguistics**
Edited by Arne Ziegler, Stefanie Edler and Georg Oberdorfer
[**Studies in Language Variation**, 27] 2021. x, 280 pp.

# Areal microvariation in German-speaking urban areas (Ruhr Area, Berlin, and Vienna)

Simon Pröll[1], Stephan Elspaß[2] and Simon Pickl[2]
[1]Ludwig-Maximilians-Universität München / [2]Universität Salzburg

This chapter focuses on spatial and social variation in the colloquial everyday speech of three German-speaking urban areas. The basic question is whether geographical variation (the kind that is commonly observed in studies on dialect variation in large areas) is detectable in the reported everyday speech of the spatially much smaller, yet denser urban areas as well. To this end, we use methods of quantitative dialectology to examine the patterns of variation in the Ruhr Area, Berlin, and Vienna. The respective analyses of the spatial patterns of 245 linguistic variables from a total number of 4,251 informants yield clear – but remarkably distinct – patterns of areal variation for each of these three urban areas. To account for those differences, social variables are factored in.

**Keywords**: language variation, urban language, quantitative dialectology, sociolinguistics, colloquial German

## 1. Introduction

While many studies on areal variation focus on large regions or entire countries, research on areal variation on a geographically smaller scale is rare. This may be due to the assumption in the linguistic tradition that urban space may be defined more adequately as a social rather than a geographic entity. In dialectological studies, for instance, cities and towns have mainly been viewed as centres of diffusion of linguistic innovations (e.g. in the *gravity model*, cf. Chambers and Trudgill 1998: 178–185). In principle, however, it seems plausible that urban space itself also has an areal dimension.

This assumption is worth looking at from a variationist perspective. In this explorative study, we investigate small-scale spatial variation in German-speaking urban areas. By applying quantitative methods to fine-grained data from surveys on reported colloquial German, we examine such small-scale spatial variation in

https://doi.org/10.1075/silv.27.10pro

three densely populated urban areas in the German-speaking countries – the Ruhr Area and Berlin in Germany as well as Vienna in Austria. Based on the empirical results of our study, we will discuss a central issue that arises from adopting an areal-linguistic perspective of variation in the city: Do these findings point to the existence of geographically conditioned variation in urban spaces, or are these spatial structures in the linguistic variation of cities symptoms of another underlying type of (social) variation?

## 2. Aims of the study

This chapter investigates variation of everyday language data in urban settings from an areal-linguistic point of view. It has three main objectives, all of which are interconnected: The first one is of a primarily documentary nature, as we concentrate on portraying and analysing variation patterns of three distinct German-speaking urban areas (the Ruhr Area, Berlin and Vienna). This leads to the second objective: Can variation in the reported everyday language use in urban spaces be traced back to areal patterns, i.e., does urban variation reflect patterns known from traditional dialectology? The third objective of this chapter is to find out whether there is a lower limit to the detection of areal variation in terms of geographical scale: Small-scale dialectometrical studies (< 50 km) are rare, and the few studies in this field (see especially Stanford 2012) have mostly been limited to cases of base-dialectal variation and/or geographic obstacles (e.g. Jeszenszky et al. 2017), where substantial systemic differences were to be expected.

Thus, the study is both interested in how narrow the geographical scope can get and whether urban everyday language carries enough variation to exhibit a signal strong enough for quantitative variationist research.

## 3. Data, areas and methods

### 3.1 Atlas zur deutschen Alltagssprache (AdA)

The data for our study come from the *Atlas zur deutschen Alltagssprache* (AdA) ('Atlas of colloquial German') (Elspaß and Möller 2003ff.; cf. Elspaß 2007; Möller and Elspaß 2014, 2015). The AdA is a linguistic atlas based on internet surveys of contemporary colloquial German, particularly as spoken by the younger urban generation. It was originally targeted at almost 500 cities and towns in the German-speaking countries and regions (Germany, Austria, Switzerland, Liechtenstein, Luxembourg and the German-speaking parts of northern Italy, eastern Belgium and the Alsace and Lorraine regions in eastern France).[1] The data were obtained through internet questionnaires. The participants were requested to name variants in the colloquial speech in their towns and cities, i.e. the kind of speech 'one would normally hear' in these places, 'be it more dialect or more standard German'.[2] It is important to note that the participants act as *informants*. They are not meant to state what their individual language use is, but what people in their city or town usually say. Due to the diversity of the varietal spectra in the German-speaking countries, including diglossic and diaglossic configurations (cf. Auer 2005), everyday colloquial speech ('Alltagssprache') can mean the local dialect (e.g. in German-speaking Switzerland), some kind of intermediate variety (e.g. in many parts of Austria and southern Germany) or a regional form of language close to a standard German variety (particularly in the north of Germany, but also in many of the bigger cities).

So far, twelve surveys have been completed, the results of which amount to some 600 linguistic maps. All maps are available online; the webpage[3] also contains the questionnaire for the current survey. Whereas in the first survey round 1,763 answers were obtained, later survey rounds snowballed to more than 20,000 responses.

The participants were asked (1) to provide local variant(s) in lexis, pronunciation, grammar, and idioms and routine formulae, or (2), in some cases, to state how common/uncommon a certain variant or construction is in their hometown.[4] In the first case (1), the individual tasks typically follow an onomasiological approach – starting from a given concept (presented in a short description and/or in a picture), participants are asked for the expression(s) 'normally' used for this concept in their town/city. They can choose from a list of given variants or provide a variant which is not listed. Figure 1 shows two examples from questionnaire no. 12.

---

1. Data from Alsace and Lorraine have been included beginning in the ninth survey round, which started in 2012.

2. "Bitte geben Sie bei den folgenden Fragen jeweils an, welchen Ausdruck man in Ihrer Stadt normalerweise hören würde – egal, ob es mehr Mundart oder Hochdeutsch ist." (Elspaß and Möller 2003ff. – http://www.atlas-alltagssprache.de/runde-12-fragebogen/; http://www.atlas-alltagssprache.de, 28 February, 2021)

3. http://www.atlas-alltagssprache.de/ (28 February, 2021).

4. One AdA survey (survey no. 6 from 2006) was aimed at eliciting perceptual data (cf. Möller 2012). The AdA data collection is still ongoing and is presently going into its thirteenth survey round.

1. Wie nennt man bei Ihnen …

(a)... selbstgemachtes Weihnachtsgebäck?

☐ *Plätzchen*
☐ *Plätzle*
☐ *Keks*
☐ *Kekserln*
☐ *Kekslan*
☐ *Brötle (Bretle, ...)*
☐ *Guetsle*
☐ *Biskuit*
☐ *Guetzli*
☐ *Loible*
☐ *Krapferl*
☐ *Weihnachtsbäckerei*
anders, und zwar:

(b)... ein Stück eines Apfels, wie im Bild links?

☐ *Apfelschnitz (/Appel-/öpfelschnitz u.ä.)*
☐ *Apfelspalte (/Appel-/Öpfelspalte u.ä.)*
anders, und zwar:

**Figure 1.** Examples for tasks from AdA questionnaire no. 12 (variants for (a) ‘cookie’, (b) ‘piece of an apple’)

The resulting maps show either one or two colour-coded dots per location: In the latter case, the bigger dot represents the most frequently reported, i.e. the dominant, variant at the location. A smaller dot next to the big dot indicates, firstly, that there is variation in this location, and, secondly, it symbolizes the second most common variant in this area, cf. Figure 2 (variants for ‘potato’) from survey round 9 (see online map for full colour version).

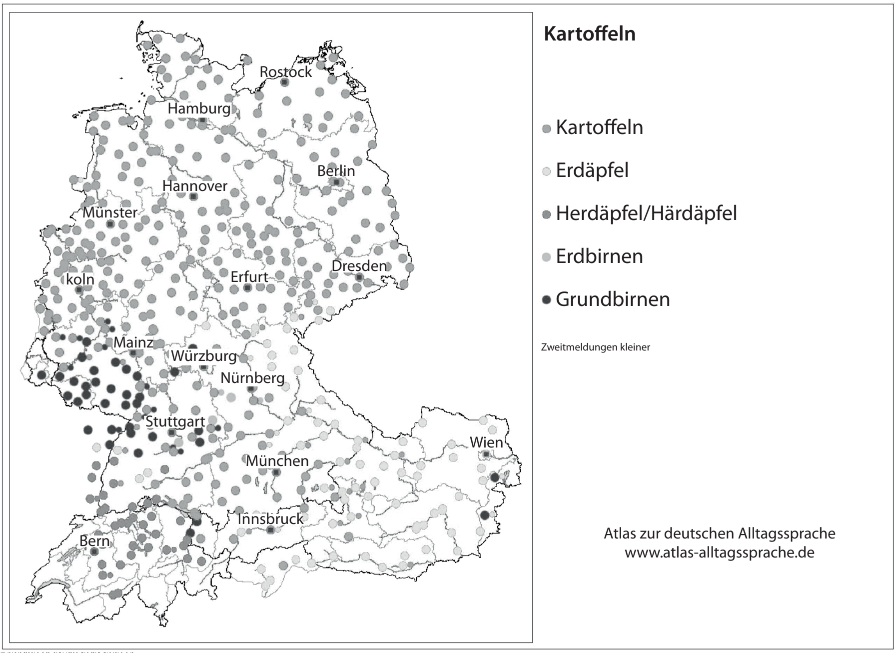



**Figure 2.** Map no. 1 e, variants of 'potato' in colloquial German, form AdA survey round 9[5]

## 3.2 German-speaking urban areas (three different ones)

We have chosen three urban areas (see Figure 3) with different sizes and population densities as testing grounds. The following list depicts these areas in decreasing size and increasing density:

- The *Ruhrgebiet* (*Regionalverband Ruhr, Ruhr Area*) is a metropolitan area in the west of Germany that is economically centred on a formerly thriving coal and steel industry. It is inhabited by approximately 5.1 million people (4,435 km$^2$), population density: 1,646/km$^2$.[6]
- *Berlin* is the capital of present-day Germany, with approximately 3.8 million inhabitants (891 km$^2$), population density: 4,206/km$^2$.[7]
- *Vienna*, the capital of Austria, has approximately 1.9 million inhabitants (415 km$^2$), population density: 4,326/km$^2$.[8]

5. See http://www.atlas-alltagssprache.de/kartoffeln/ (28. February, 2021).

6. Source: https://en.wikipedia.org/wiki/Ruhr (28 February, 2021).

7. Source: https://en.wikipedia.org/wiki/Berlin (28 February, 2021).

8. Source: https://en.wikipedia.org/wiki/Vienna (28 February, 2021).


**Figure 3.** Overview map

Note that, unlike the two capitals Berlin and Vienna, the Ruhr Area is a polycentric urban area. Although the Ruhr Area consists of administratively independent cities and towns, it is interlinked by continuous urban settlement, which allows for a comparison with Berlin and Vienna. Historically, Berlin and Vienna also developed from formerly independent municipalities which were merged as the two capitals grew in size.

Whereas Vienna has been an urban centre since the Middle Ages, Berlin and the Ruhr Area mainly developed into urban areas during the Industrial Revolution in Germany (18th and 19th centuries).

### 3.3 Methods

The material we use in this study originates from four AdA questionnaires, survey rounds seven to ten (2009–2013/14). Since the launch of the AdA in 2003, when potential informants in 500 cities and towns of the German-speaking countries were addressed directly by e-mail and pointed to the website of the first survey, the number of participants had risen to over 5,000 in survey round five. The AdA has a high participant retention rate, as the sending out of a new questionnaire is timed to coincide with the publication of the mapped results of the previous round. Due to the ensuing immense popularity of the AdA and the distribution of its findings in traditional and digital media, the number of participants snowballed to over 20,000 participants in survey round ten. The number of variables in survey round seven to ten amounts to 245 (out of a total of over 600 in the entire AdA).[9] Most of these variables are of lexical (cf. Figure 2, other examples include *Kissen* / *Polster* 'pillow', *Rechen* / *Harke* 'rake', or *g/kuck!* / *schau!* / *lueg!* 'look!'),[10] morphological (e.g. past participle of *einschalten* 'to switch on' – *eingeschaltet* or *eingeschalten*),[11] morphosyntactic (e.g. auxiliary verb *sein* or *haben* in perfect tense forms of posture verbs)[12] or phonological nature (e.g. short of long vowel in *Spaß* 'fun').[13]

The original data from the surveys are organised in tables with nominal-scaled answers from individual participants, i.e. a database containing a row for each participant and a column for every variable asked, with the individual cells containing the respective variants in plain text (cf. Table 1).

**Table 1.** Original data structure of the AdA

| **informant id** | **postal code** | **variable 1** | **variable 2** | … |
|---|---|---|---|---|
| 1 | 86150 | *variant a* | *variant b* | … |
| 2 | 86167 | *variant a* | *variant a* | … |
| 3 | 86167 | *variant b* | *variant c* | … |
| … | … | … | … | … |

For quantitative analyses, we transformed them to relative frequencies of variants (columns) per location (rows) (cf. Table 2). Thus, the reported local variation is interpreted as usage frequency on site. This differs from most 'classical' dialectometrical analyses of dialect data, in which the ideology of homogeneity of varieties is still very much reflected in both data collection and data representation methods:

9. We use the term *variable* in the sense usually employed in variationist sociolinguistics and dialectology, i.e. meaning "a linguistic unit with two or more variants involved in covariation with other social and/or lingustic variables" (Chambers & Trudgill 1998: 50).

10. See http://www.atlas-alltagssprache.de/r10-f2c, http://www.atlas-alltagssprache.de/r10-f3f and http://www.atlas-alltagssprache.de/r10-f13h (28 February, 2021).

11. See http://www.atlas-alltagssprache.de/runde-7/f10a-e/, map 10 c (28 February, 2021).

12. See http://www.atlas-alltagssprache.de/hilfsverb/ (28 February, 2021).

13. See http://www.atlas-alltagssprache.de/spass/ (28 February, 2021).

One location (i.e. its base dialect) is represented by one 'monolingual' representative informant through one single data point/variant (cf. e.g. Goebl 1984).

Of course, as Smakman and Heinrich (2018), among others, point out, this is not an overly realistic scenario for variation in the city (and, incidentally, not for rural areas either). As a consequence of these different concepts of what kind of variation to expect at a single site, our results (based on usage frequencies) may not be entirely comparable with dialectometrical studies carried out in the 'classical' fashion. We will revisit this thought in Section 5.

**Table 2.** Transformed data structure, relative variant frequencies per location

| **location** | **relative frequency of variant a of variable 1** | **relative frequency of variant b of variable 1** | **relative frequency of variant a of variable 2** | … |
|---|---|---|---|---|
| *location name 1* | 0.10 | 0.90 | 0.05 | … |
| *location name 2* | 0.25 | 0.75 | 0.33 | … |
| *location name 3* | 0.00 | 1.00 | 0.40 | … |
| … | … | … | … | … |

The relative variant frequencies per location are then used to calculate a (Euclidean) distance matrix of the linguistic dissimilarity of all pairs of locations. This matrix serves as the basis for all following quantitative analyses.

## 4. Areal variation

First, we approach the issue from a descriptive statistical angle: Do the three urban areas chosen for this study show spatial variation, and if yes, how is it structured? To answer this question, we use two quantitative methods that are common in both sociolinguistics and dialectometry, namely multidimensional scaling (MDS) and cluster analysis.[14]

14. In short, both are methods designed to reduce multidimensional variation (which – for cognitive reasons – cannot be categorized into accessible subgroups by the human brain) to a smaller number of dimensions that can be understood rather intuitively. A not entirely correct, but illuminating, analogy to multidimensional scaling would be taking a (two-dimensional)

## 4.1 Ruhr Area

Rounds seven to ten of the AdA contain 1,707 responses from 50 different locations in the Ruhr Area. However, not all of these 50 locations feature responses from all four rounds. We thus pooled some smaller, neighbouring locations with missing data into 'virtual' locations comprising more than one data location, resulting in a total of 34 locations for analysis (see Figure 4). The maximum distance between two locations is about 100 kilometres.

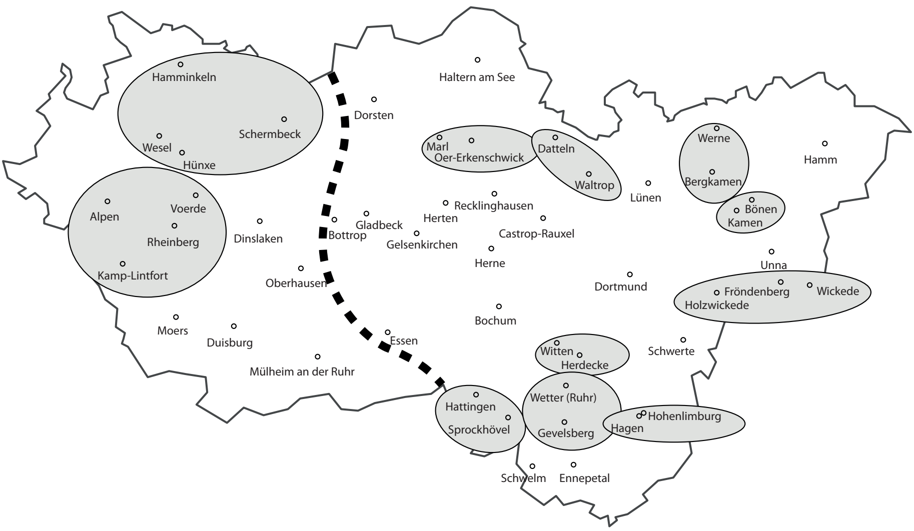



**Figure 4.** All Ruhr Area locations with responses in AdA rounds seven to ten; collapsed, 'virtual' locations marked by circling; the dotted line approximates the base dialect divide between Low Franconian (west) and Westphalian (east)

picture of a (three-dimensional) object from an angle that still allows the viewer of the picture to infer the original space. Naturally, this always comes at the cost of losing some of the original detail.

In the past, two of the authors of this study published several papers that (implicitly and explicitly) argued for using factor analysis for spatial exploration of dialect data instead of multidimensional scaling or hierarchical cluster analysis (cf. Pickl 2013; Pröll 2015; Pickl and Pröll 2019). We still believe that this is the preferable approach to big data sets with high intrinsic, primarily geographic variation, a scenario that is common for traditional rural dialects. The situation of small-scale variation of everyday language in cities, however, is different: (a) there is a lot of random variation and (b) the actual systemic differences are relatively small. Thus, it is preferable to employ methods that concentrate on extracting the hierarchically highest structures in these noisy datasets.

In combining individual locations, we made sure that we did not conflate data from the two main dialectal regions of the Ruhr Area, namely the Low Franconian area in the west and the Westphalian area in the east. A cluster analysis (using Ward's method) of the (Euclidean) distances in the data set suggests a division into three clusters (Figure 5).

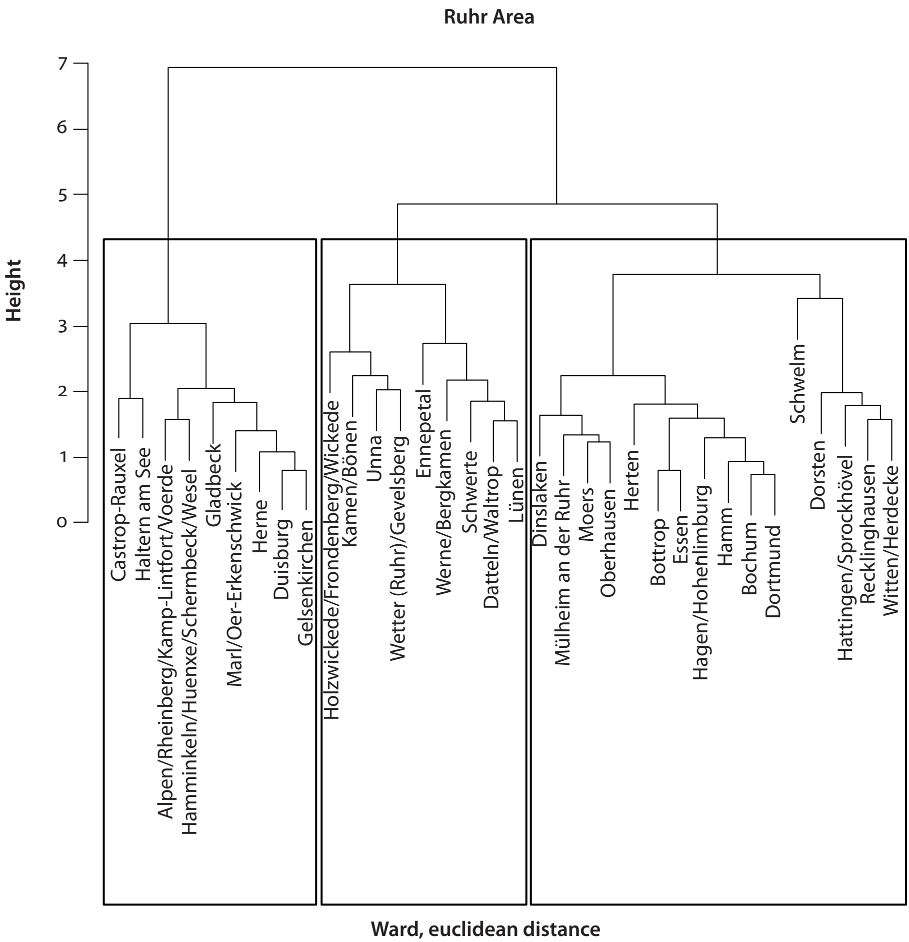



**Figure 5.** Cluster dendrogram of the Ruhr Area, Ward's method

If we shade a map of the Ruhr Area according to this cluster analysis, we arrive at the map displayed in Figure 6.

Clearly, there is regional variation in the data that is not random: Locations that are geographically closer to each other tend to be linguistically similar. The three clusters correspond to three geographical areas in the west (dark grey), the


**Figure 6.** Cluster map of the Ruhr Area

centre (white) and the east (light grey), while the individual spatial 'outliers' may be indicative of gradual or even patchy variation rather than clear areal patterns. From a dialectological perspective, this is hardly surprising. However, we are not dealing with base-dialectal data of NORM speakers here, but with the reported everyday language use of socially highly heterogeneous groups of speakers.[15]

Nonetheless, the shaded map, based on a cluster analysis, reveals some similarities with the dialectal structure of the Ruhr Area in that it reflects a general west-east pattern, although it does not quite match the division of Low Franconian dialects in the west and Westphalian dialects in the east, showing a three-fold rather than a two-fold division.

Judging from this cursory visual analysis that uses methods from quantitative geolinguistics, the variation of everyday language as reported in the rather large Ruhr Area looks analogous to dialectal variation when whole cities are treated as the smallest item. But will that result remain consistent if we change the scale, use smaller overall distances and treat districts of a city (rather than whole cities) as the smallest item?

15. Note that the focus here is not on the heterogeneity of the informants, but on that of the speakers in the respective towns and cities about whose everyday language the informants provide information.

## 4.2 Berlin

To answer this, the second case study focuses on Berlin, which has a higher average population density. The maximum travel distance within the boundaries of Berlin is about 45 kilometres. 1,354 responses from Berlin were collected in AdA rounds seven to ten, for which the informants provided the postal codes of their home addresses. We used these postal codes to assign each answer to one of the twelve present-day districts of Berlin (see Table 3).

**Table 3.** Districts of Berlin (as established in 2001)

| District number | Name |
|---|---|
| 01 | Mitte |
| 02 | Friedrichshain-Kreuzberg |
| 03 | Pankow |
| 04 | Charlottenburg-Wilmersdorf |
| 05 | Spandau |
| 06 | Steglitz-Zehlendorf |
| 07 | Tempelhof-Schöneberg |
| 08 | Neukölln |
| 09 | Treptow-Köpenick |
| 10 | Marzahn-Hellersdorf |
| 11 | Lichtenberg |
| 12 | Reinickendorf |

If we run a multidimensional scaling analysis (two dimensions) of the variation in the Berlin data and plot the results, we arrive at Figure 7. This analysis accounts for 35.69% of the variance of the original data. There is noteworthy variation to be seen, but the underlying pattern seems unclear at first.

Applying the same type of cluster algorithm as above (Ward's method, Euclidean distance) to the Berlin data yields the dendrogram in Figure 8. Again, a division into three (or alternatively two) clusters seems to be the most natural.

Figure 9 visualises these cluster results in the form of a map (a two-cluster solution would collapse the white area and the dark grey area into one). Similar to the Ruhr Area data, it is obvious that there is some order in the spatial distribution of variation in reported everyday speech. It is not random, but clearly follows a pattern that is geographically conditioned, i.e. a west-east as well as a north-south division.

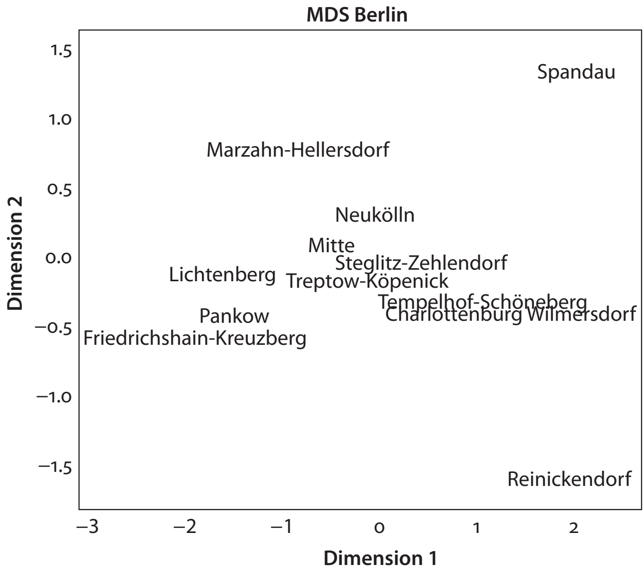



**Figure 7.** MDS of Berlin

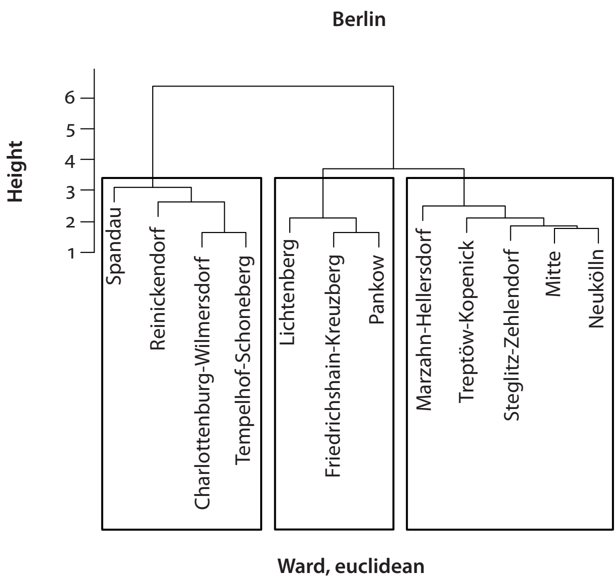



**Figure 8.** Cluster dendrogram of Berlin's districts, Ward's method


**Figure 9.** Cluster map of Berlin's districts

## 4.3 Vienna

In our third and last case study, we focus on Vienna. The population/area ratio is the highest here, i.e. Vienna is more densely populated on average than the Ruhr Area and Berlin, with a maximum distance of about 26 kilometres within the inhabited parts of the city limits. We received 1,190 responses from Vienna. Again, we used the postal codes of the informants to assign them to Vienna's 23 districts (see Table 4).

**Table 4.** Districts of Vienna

| District number | Name |
|---|---|
| 1 | Innere Stadt |
| 2 | Leopoldstadt |
| 3 | Landstraße |
| 4 | Wieden |
| 5 | Margareten |
| 6 | Mariahilf |
| 7 | Neubau |
| 8 | Josefstadt |
| 9 | Alsergrund |
| 10 | Favoriten |

**Table 4.** (*continued*)

| District number | Name |
|---|---|
| 11 | Simmering |
| 12 | Meidling |
| 13 | Hietzing |
| 14 | Penzing |
| 15 | Rudolfsheim-Fünfhaus |
| 16 | Ottakring |
| 17 | Hernals |
| 18 | Währing |
| 19 | Döbling |
| 20 | Brigittenau |
| 21 | Floridsdorf |
| 22 | Donaustadt |
| 23 | Liesing |

Multidimensional scaling of the data, as seen in Figure 10, accounts for 19.89% of the underlying variation. While its interpretation is not obvious at first glance, the plot shows that there is a reasonably high amount of variation in the data – less so than in Berlin, but that is to be expected due to (geographic) size and number of speakers/inhabitants. Also, the number of speakers per location is lower, as we are dealing with almost twice the number of districts and, at the same time, only about half the number of speakers.

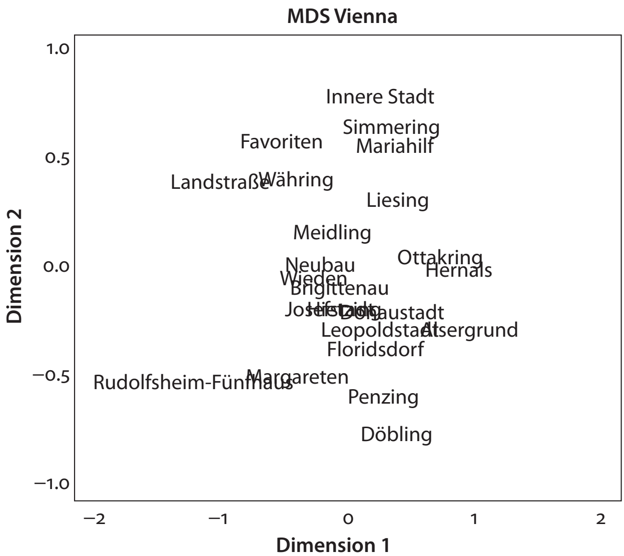



**Figure 10.** MDS of Vienna's districts

A cluster analysis again yields three major clusters, visualized in Figure 11 as a dendrogram with an accompanying map in Figure 12.

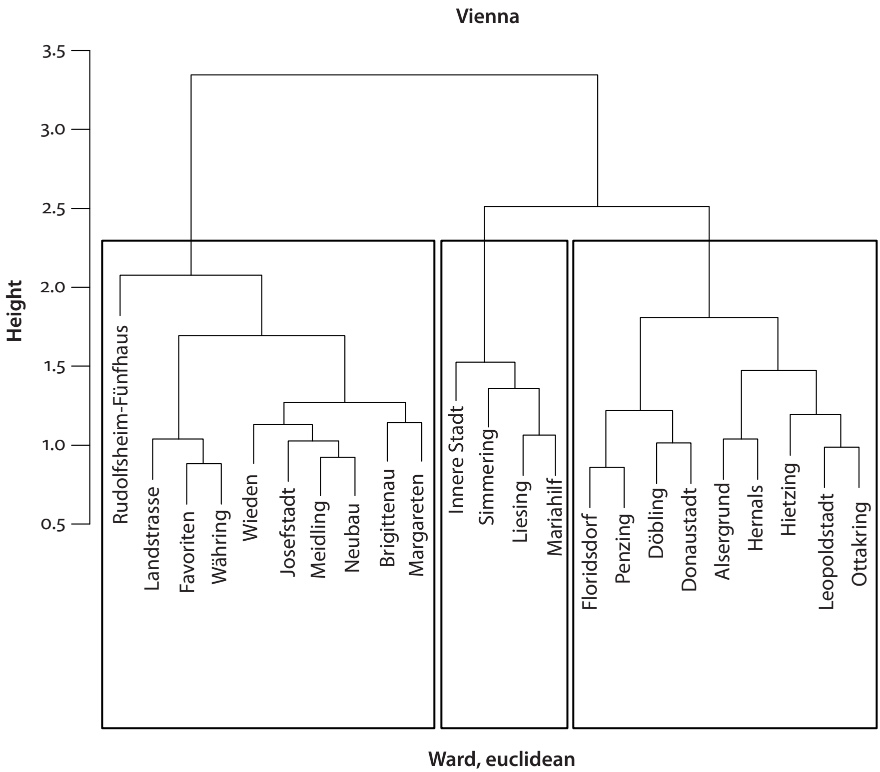



**Figure 11.** Cluster dendrogram of Vienna's districts, Ward's method

Here, we observe a centre/south vs. inner ring vs. suburbs/north configuration. A two-cluster solution would collapse the light grey area and the dark grey area.

If we now revisit the MDS plot (Figure 8) and add the cluster information to it, the picture becomes clearer – the areal pattern is now visible in the MDS plot as well (Figure 13).


**Figure 12.** Cluster map of Vienna's districts

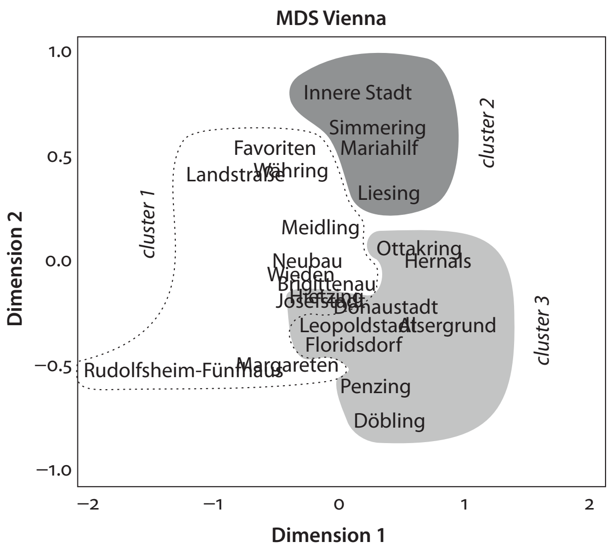



**Figure 13.** MDS of Vienna's districts, with colour added for cluster assignment

## 5. Discussion: The interplay of space and social factors

We have demonstrated here that methods devised for areal variation in rural base dialects enable us to detect variation in German-speaking urban areas as well, even if the underlying variability of the data is much smaller. While the Ruhr Area might be comparable in size to smaller dialect areas, even a very narrow scope – e.g. of only 26 km, as in the case of Vienna – still reveals geographical patterns.

Now that it is clear that urban areas do show spatial variation, the question remains: Is this geographical variation of the same type as regular (base-)dialectal variation, only on a smaller scale? The structure of geographically conditioned variation is often assumed to be what is usually referred to as a *dialect continuum* (cf. Chambers and Trudgill 1998), where diffusion of innovations is considered to travel according to a wave model. In this view, areal language variation is a function of geographical distance to a large extent (cf. Séguy 1971; Nerbonne 2010; Stanford 2012). Accordingly, typical dialect atlases aim at covering large areas of space while deliberately limiting the social scope. This focus on NORM speakers is motivated by an interest in geographical variation alone, while excluding social variation – as far as possible – from the picture.[16]

The interest in geographical variation has led to the notion that social variation would distract from the spatial picture and obscure the spatial signal. Some dialect atlases try to include social variation in a systematic way, e.g. with respect to the parameter of age (e.g. the *Mittelrheinischer Sprachatlas*, cf. Bellmann, Herrgen and Schmidt 1994–2002). More recent methods of data collection, such as crowdsourcing (e.g. Kolly and Leemann 2015) or crawling of social media like Twitter (e.g. Grieve, Nini and Guo 2018 or the ongoing *Tweetolectology* project, cf. Willis, Leemann, Blaxter and Gopal 2017–2020), have led to socially diverse datasets, while potential social factors such as age, income, etc. are often not available and can therefore not be linked to the linguistic data.

The AdA belongs to the latter category: There is no restriction in terms of age, gender, social background etc. of the informants, and data collection is not balanced for these parameters, i.e. it is not checked for these factors. In addition to the elicitation of colloquial language use, personal data are collected to a limited extent. The AdA data are therefore socially heterogeneous, while also partially documenting the social diversity.

In this section, we will explore the relevance of social factors for the use of colloquial German in the Ruhr Area, Berlin, and Vienna, as documented in the AdA, but we will restrict our analyses to averaged census data for reasons of simplicity.

---

**16.** While modern geolinguistics recognises that even 'purely' geographical variation entails local variation (cf. e.g. Pickl 2013: 52), the idea that local base dialects are somehow uniform and do not exhibit internal variation is not uncommon, and is reflected in the methodological decision to allow only one variant per variable in dialectometry (cf. Schneider 1984).

For all three test cases, we factor the parameters income, unemployment rates, age, and migration background into the forthcoming inferential statistics, using the most recent numbers available.

## 5.1 Ruhr Area

As seen above (4.1), the Ruhr Area shows signs of a west-east divide, which can be traced back to the two main old dialect areas, Low Franconian and Westphalian, to some extent.

In order to evaluate the relevance of social factors for language variation as well, we look for correlations between the MDS dimensions and census data for the individual locations, using Pearson's correlation coefficient (Table 5). The strongest association is found for available income and MDS dimension 1, amounting to 7% explained variance. None of the correlations are statistically significant for either of the dimensions. A comparison with the corresponding results for the locations' geographical coordinates, however, yields a value close to that of income for latitude and dimension 1, and the overall highest correlation found was for longitude and dimension 1, the latter being statistically significant.[17] This is not surprising given the linguistic west-centre-east partition of the area shown by the cluster analysis (Figure 6).

**Table 5.** Statistical correlations between MDS dimensions and census data/geographical coordinates for the Ruhr Area. Correlations in bold print are statistically significant ($\alpha = 0.05$, two-tailed $t$-test)

| **Pearson correlation** | **MDS dimension 1** | | **MDS dimension 2** | |
|---|---|---|---|---|
| | $r$ | $R^2$ | $r$ | $R^2$ |
| available income 2015[a] | 0.27 | 0.07 | −0.17 | 0.03 |
| unemployment rate 2016[b] | −0.09 | 0.01 | −0.13 | 0.02 |
| average age 2016[c] | 0.13 | 0.02 | −0.13 | 0.02 |
| non-German population 2017[d] | 0.03 | 0.00 | −0.15 | 0.02 |
| locations latitude | −0.23 | 0.05 | 0.13 | 0.02 |
| locations longitude | **0.47** | **0.22** | 0.07 | 0.00 |

a. Source: Regionalverband Ruhr. 2017.
b. Source: https://www.wegweiser-kommune.de/statistik/ (28 February, 2021).
c. Source: https://www.wegweiser-kommune.de/statistik/ (28 February, 2021).
d. Source: Regionalverband Ruhr. 2019.

17. The orientation of the coordinate system along a north-south and an east-west axis is, of course, arbitrary in geolinguistic terms. Since the coordinate system is orthogonal (as are the MDS dimensions), however, the sum of the two dimensions' explained variances is not dependent on rotation. Therefore, the total explained variance of latitude and longitude is independent of the rotation of the coordinate system and gives a good (if somewhat simplified) impression of the overall relevance of geography for reported language use, as captured by the MDS analysis.

These results show that, for the Ruhr Area, geography appears to be a more relevant predictor of colloquial language variation than social differences, accounting for a total of 29% explained variance in the MDS dimensions. This seems plausible, given the large geographical area covered, but it is interesting to see that social differences do not have a stronger impact on colloquial language variation in such a heavily urbanised area.

## 5.2 Berlin

For Berlin, both a historical and an economic explanation for variation in reported colloquial language seem plausible. Firstly, the rather complex, hyphenated names for some of the twelve modern districts hint at their history: They have been merged from smaller historical towns and villages. In addition, during the years 1948–1990, when Germany was divided into a western and an eastern state (according to the occupation zones set at the Potsdam Conference regulating the administration of Germany after World War II), Berlin (as the capital of the former German Reich) was divided into a western and an eastern part as well.

If we shade the districts according to their respective affiliation to either West and East Berlin during the period of division (1948–1990) in an MDS plot (cf. Figure 7), we arrive at Figure 14. (The two districts Mitte and Friedrichshain-Kreuzberg were constructed in the 2001 reform from both former east and west districts to deliberately bridge the past division of the city. We account for this by shading them using both black and white.)

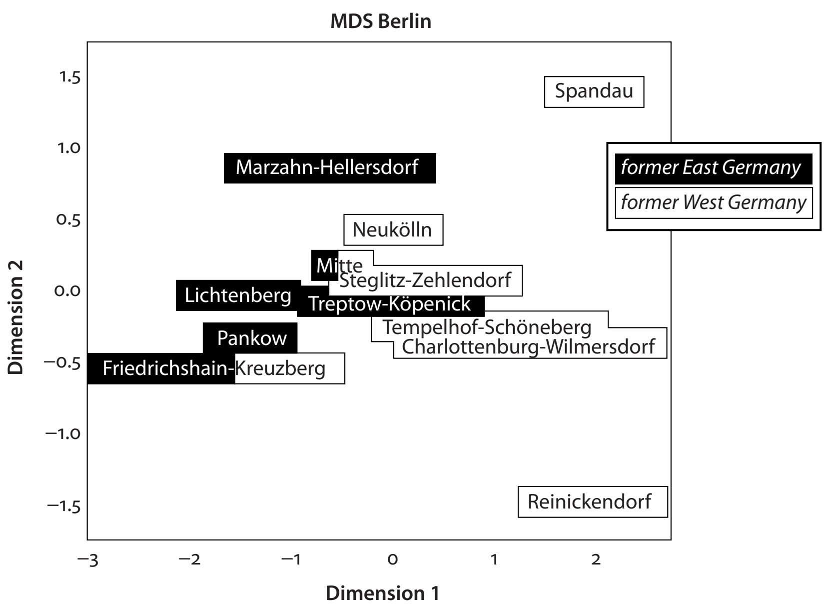



**Figure 14.** MDS of Berlin (divided)

Here, the interpretation is obvious: The first dimension of the MDS captures the old east-west divide, which demonstrates that it is still relevant for reported colloquial language nowadays. If we explore the associations between social factors and MDS dimensions (Table 6), we find generally higher values of correlation between both dimensions and social factors compared to the results for the Ruhr Area. The highest values are found for the correlation between average age and dimension 1 at 0.66 with 44% explained variance, which is statistically significant, and 0.46 (21%) for income and dimension 2.

**Table 6.** Statistical correlations between MDS dimensions and census data/geographical coordinates for Berlin. Correlations in bold print are statistically significant ($\alpha = 0.05$, two-tailed *t*-test)

| Pearson correlation | MDS dimension 1 | | MDS dimension 2 | |
|---|---|---|---|---|
| | $r$ | $R^2$ | $r$ | $R^2$ |
| average income 2015[a] | −0.05 | 0.00 | 0.46 | 0.21 |
| unemployment rate March 2019[b] | 0.23 | 0.05 | −0.04 | 0.00 |
| average age 2018[c] | **0.66** | **0.44** | −0.00 | 0.00 |
| migration background 2018[d] | 0.18 | 0.03 | 0.03 | 0.00 |
| locations latitude | −0.02 | 0.00 | 0.27 | 0.07 |
| locations longitude | **−0.63** | **0.40** | −0.03 | 0.00 |

a. Source: Michel (2019).
b. Source: Bundesagentur für Arbeit (2019).
c. Source: Statistik Berlin Brandenburg (2019: 5).
d. Source: Michel (2019).

For geography and the MDS results, we find a strong link as well, namely for MDS dimension 1 and longitude, representing 40% explained variance, which reflects the relevance of the east-west divide as discussed above.

In contrast to the results for the Ruhr Area, however, social factors appear to play a more important role than geography in Berlin. This points to greater social and (perceived) sociolinguistic stratification in Berlin, which manifests itself in reported colloquial language use in the different districts.

### 5.3 Vienna

Vienna does not show the Berlin pattern of a clear west-east contrast. The division of Vienna (and likewise of Austria as a whole) into four occupation zones only lasted for ten years (1945–1955). Its centre/south vs. inner ring vs. suburbs/north configuration can be at least partially explained by socioeconomic factors: The central and southern districts, with the exception of the inner city district, (marked in white in Figure 12) are economically weaker, with relatively low income rates,[18]

**18.** Source: Redaktion DER STANDARD (2014).

and they include some of the districts with the highest numbers of first and second generation immigrants, measured by the percentage of schoolchildren with a language other than German as L1.[19]

Interestingly, social factors do not appear to play much of a role in colloquial language variation in Vienna, as captured by the two MDS dimensions (Table 7). The highest correlations are found between dimension 1 and migration background ($r = -0.19$; $R^2 = 4\%$) and for dimension 2 and income ($r = 0.22$; $R^2 = 5\%$).

**Table 7.** Statistical correlations between MDS dimensions and census data/geographical coordinates for Vienna. Correlations in bold print are statistically significant ($\alpha = 0.05$, two-tailed *t*-test)

| **Pearson correlation** | **MDS dimension 1** | | **MDS dimension 2** | |
|---|---|---|---|---|
| | $r$ | $R^2$ | $r$ | $R^2$ |
| average net income 2015[a] | 0.17 | 0.03 | 0.22 | 0.05 |
| unemployment rate 2015[b] | −0.13 | 0.02 | 0.04 | 0.00 |
| average age 2015[b] | 0.15 | 0.02 | 0.14 | 0.02 |
| migration background 2015[b] | −0.19 | 0.04 | −0.00 | 0.00 |
| locations latitude | 0.26 | 0.07 | **−0.48** | **0.23** |
| locations longitude | −0.07 | 0.00 | 0.21 | 0.04 |

a. Source: MA 23 (2016).
b. Source: https://www.wegweiser-kommune.de/statistik/ (28 February, 2021).

Like for the Ruhr Area, geography has a stronger impact on reported language use than any of the social factors, but here the results point to a north-south rather than an east-west divide or gradient: The MDS dimension 2 is negatively correlated with the locations' latitude at $r = -0.48$ ($R^2 = 23\%$, statistically significant).

This might be linked to the social housing policy in Vienna, which has aimed at offering subsidised, publicly owned living spaces for low-income households across districts since the 1920s, and is often cited as exemplary. This may have prevented starker contrasts between more affluent and poorer districts from developing, and might in turn have led to socially diverse, economically mixed neighbourhoods in each of the districts.

On the other hand, the cluster map of the Vienna data (Figure 12) is clearly reminiscent of the distribution of the social factors in question across Viennese districts.[20] It is quite possible that the two MDS dimensions do not capture enough of the underlying variation in the Vienna data (a little less than 20%) to reflect patterns

19. Source: Redaktion Die Presse (2009).

20. E.g. Redaktion DER STANDARD (2014) or Redaktion Die Presse (2009).

of social variation. Additionally, in contrast to the east/west dichotomies of the Ruhr Area and Berlin, the geographically complex inner ring vs. suburbs configuration of Vienna cannot aptly be captured with just two geographical dimensions alone (although latitude captures at least that part of the structure that can be expressed by a north/south opposition).

## 6. Summary

The present study started with the assumption that urban language variation does not only reflect social structures, but that it may also display areal patterns.

In sum, the three urban areas under investigation show three different patterns: The Ruhr Area shows signs of a spatially conditioned, west-centre-east partition, which may be linked to the two main old dialect areas, Low Franconian and Westphalian, to some extent, as well as the relatively large geographical area covered. Social factors do not seem to play a crucial role in this conurbation, especially when compared to geography. Berlin shows a division along the west-east axis as well, but here it is clearly due to the decades-long division of the city into a western and an eastern half, which made everyday communication between people in the two parts virtually impossible. In addition to the west-east divide, age and perhaps income distribution seem to have an effect on language variation in Berlin. Vienna shows an areal-linguistic pattern which reflects the social structure of the city: To a certain extent, the areal configuration of the cluster analysis matches the social structure of Vienna, where the central and southern districts have comparatively low average income and particularly high percentages of non-native German speakers. This pattern, however, does not seem to be captured by MDS analysis, and geography (here along a north-south axis) appears to have much stronger relevance for colloquial language variation than social factors.

On the whole, it becomes clear that geographical variation does play a role in language variation, even in geographically small and socially diverse entities such as urban areas, and it does so consistently in all three test cases, though with different spatial orientations that reflect different historical influences. Interestingly, social structure has a (slightly) stronger impact on reported language use than geography only in Berlin; in the two other cases, geography was the single strongest factor. The lower geographical threshold for spatial variation in urban areas does not seem to be reached even within Vienna, the smallest of the three test cases. This means that in urban spaces with diameters of 30 km and less, there is still geographical variation to be found in addition to social variation.

## References

Auer, Peter. 2005. Europe's sociolinguistic unity, or: A typology of European dialect/standard constellations. In Nicole Delbecque, Johan van der Auwera and Dirk Geeraerts (eds.), *Perspectives on variation. Sociolinguistic, historic, comparative* (TiLSM 163), 7–42. Berlin/New York, NY: de Gruyter Mouton. https://doi.org/10.1515/9783110909579.7

Bellmann, Günter, Joachim Herrgen and Jürgen E. Schmidt (eds.). 1994–2002. *Mittelrheinischer Sprachatlas (MRhSA)*. 5 vols. Tübingen: Niemeyer.

Bundesagentur für Arbeit. 2019. Arbeitsmarkt im Überblick (Berichtsmonat März 2019, Mitte, JC). https://statistik.arbeitsagentur.de/Navigation/Statistik/Statistik-nach-Regionen/SGBII-Traeger/Berlin-Nav.html. (17 July, 2019.)

Chambers, Jack K. and Peter Trudgill. 1998. *Dialectology*, 2nd edn. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511805103

Elspaß, Stephan. 2007. Variation and change in colloquial (standard) German. *The Atlas zur deutschen Alltagssprache (AdA) project*. In Christian Fandrych and Reinier Salverda (eds.), *Standard, Variation und Sprachwandel in germanischen Sprachen/Standard, variation and language change in Germanic languages* (Studien zur deutschen Sprache 41), 201–216. Tübingen: Narr.

Elspaß, Stephan and Robert Möller. 2003ff. *Atlas zur deutschen Alltagssprache*. http://www.atlas-alltagssprache.de/. (28 February, 2021.)

Goebl, Hans. 1984. *Dialektometrische Studien: Anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF* (Beihefte zur Zeitschrift für romanische Philologie 191–193). Unter Mitarbeit von Siegfried Selberherr, Wolf-Dieter Rase und Hilmar Pudlatz, 3 vols. Tübingen: Niemeyer.

Grieve, Jack, Andrea Nini and Diansheng Guo. 2018. Mapping lexical innovation on American social media. *Journal of English Linguistics* 46(4). 293–319. https://doi.org/10.1177/0075424218793191

Jeszenszky, Péter et al. 2017. Exploring global and local patterns in the correlation of geographic distances and morphosyntactic variation in Swiss German. *Journal of Linguistic Geography* 5(2). 86–108. https://doi.org/10.1017/jlg.2017.5

Kolly, Marie-José and Adrian Leemann. 2015. *Dialäkt Äpp*: Communicating dialectology to the public – crowdsourcing dialects from the public. In Adrian Leemann et al. (eds.), *Trends in phonetics and phonology. Studies from German-speaking Europe*, 271–285. Bern: Lang.

MA 23 (Wirtschaft, Arbeit und Statistik). 2016. Wien. Bezirke im Fokus. Statistiken und Kennzahlen. https://www.wien.gv.at/statistik/pdf/bezirke-im-fokus-1-23.pdf. (28 February, 2021.)

Michel, Anke. 2019. So viel verdienen Berliner und Brandenburger im Vergleich. *rbb 24*. https://www.rbb24.de/wirtschaft/beitrag/2019/01/einkommen-beitragsbemessungsgrenze-berlin-brandenburg.html. (17 July, 2019.)

Möller, Robert. 2012. Der Sprachgebrauch „bei uns". Arealbildung in Karten des Atlas zur deutschen Alltagssprache, objektive Grenzen und subjektive Räume. In Sandra Hansen et al. (eds.), *Dialectological and folk dialectological concepts of space. Current methods and perspectives in sociolinguistic research on dialect change* (linguae & litterae 17), 96–118. Berlin/Boston, MA: de Gruyter. https://doi.org/10.1515/9783110229127.96

Möller, Robert and Stephan Elspaß. 2014. Zur Erhebung und kartographischen Darstellung von Daten zur deutschen Alltagssprache online. Möglichkeiten und Grenzen. In Fabio Tosques (ed.), *20 Jahre digitale Sprachgeographie*, 121–131. Berlin: Humboldt-Universität. http://www2.hu-berlin.de/vivaldi/tagung/. (28 February, 2021.)

Möller, Robert and Stephan Elspaß. 2015. Atlas zur deutschen Alltagssprache. In Roland Kehrein, Alfred Lameli and Stefan Rabanus (eds.), *Regionale Variation des Deutschen. Projekte und Perspektiven*, 519–540. Berlin/Boston, MA: de Gruyter. https://doi.org/10.1515/9783110363449-022

Nerbonne, John. 2010. Measuring the diffusion of linguistic change. *Philosophical Transactions of the Royal Society B* 365(1559). 3821–3828. https://doi.org/10.1098/rstb.2010.0048

Pickl, Simon. 2013. *Probabilistische Geolinguistik. Geostatistische Analysen lexikalischer Variation in Bayerisch-Schwaben* (ZDL-B 154). Stuttgart: Steiner.

Pickl, Simon and Simon Pröll. 2019. Geolinguistische Querschnitte und Tiefenbohrungen in Bayern und darüber hinaus. In Sebastian Kürschner, Mechthild Habermann and Peter O. Müller (eds.), *Methodik moderner Dialektforschung. Erhebung, Aufbereitung und Auswertung von Daten am Beispiel des Oberdeutschen* (Germanistische Linguistik 141–143), 11–43. Hildesheim/Zürich/New York, NY: Olms.

Pröll, Simon. 2015. *Raumvariation zwischen Muster und Zufall. Geostatistische Analysen am Beispiel des Sprachatlas von Bayerisch-Schwaben* (ZDL-B 160). Stuttgart: Steiner.

Redaktion DER STANDARD. 2014. Einkommen nach Bezirken. *DER STANDARD*. https://derstandard.at/2000006966085/Einkommen-nach-Bezirken. (28 February, 2021.)

Redaktion Die Presse. 2009. Wien: In manchen Klassen sitzen ausschließlich Migranten. *Die Presse*. https://diepresse.com/home/panorama/oesterreich/465824/Wien_In-manchen-Klassen-sitzen-ausschliesslich-Migranten. (17 July, 2019.)

Regionalverband Ruhr. 2017. Verfügbares Einkommen Ruhrgebiet Zeitreihe seit 2004. https://www.rvr.ruhr/daten-digitales/regionalstatistik/arbeit-und-soziales/. (17 July, 2019.)

Regionalverband Ruhr. 2019. Ausländischen [sic] Bevölkerung 2018 und 2017. https://www.rvr.ruhr/daten-digitales/regionalstatistik/bevoelkerung/. (17 July, 2019.)

Schneider, Edgar. 1984. Methodologische Probleme in der Dialektometrie. In Hans Goebl (ed.), *Dialectology* (Quantitative Linguistics 21), 314–332. Bochum: Brockmeyer.

Séguy, Jean. 1971. La relation entre la distance spatiale et la distance lexicale. *Revue de Linguistique Romane* 35 (139–140). 335–357.

Smakman, Dick and Patrick Heinrich. 2018. Introduction: Why cities matter for a globalising sociolinguistics. In Dick Smakman and Patrick Heinrich (eds.), *Urban sociolinguistics. The city as a linguistic process and experience*, 1–11. Abingdon/New York, NY: Routledge.

Stanford, James N. 2012. One size fits all? Dialectometry in a small clan-based indigenous society. *Language Variation and Change* 24(2). 247–278. https://doi.org/10.1017/S0954394512000087

Statistik Berlin Brandenburg. 2019. Statistischer Bericht. Einwohnerinnen und Einwohner im Land Berlin am 31. Dezember 2018. https://www.statistik-berlin-brandenburg.de/publikationen/stat_berichte/2019/sb_a01-05-00_2018h02_be.pdf. (28 February, 2021.)

Willis, David, Adrian Leemann, Tam Blaxter and Deepthie Gopal. 2017–2020. *Tweetolectology. Investigating the diffusion of morphosyntactic innovations using social media*. http://tweetolectology.com/. (28 February, 2021.)