

The State of the Art in Linguistic Cartography

Stefan Rabanus, Università degli Studi di Verona

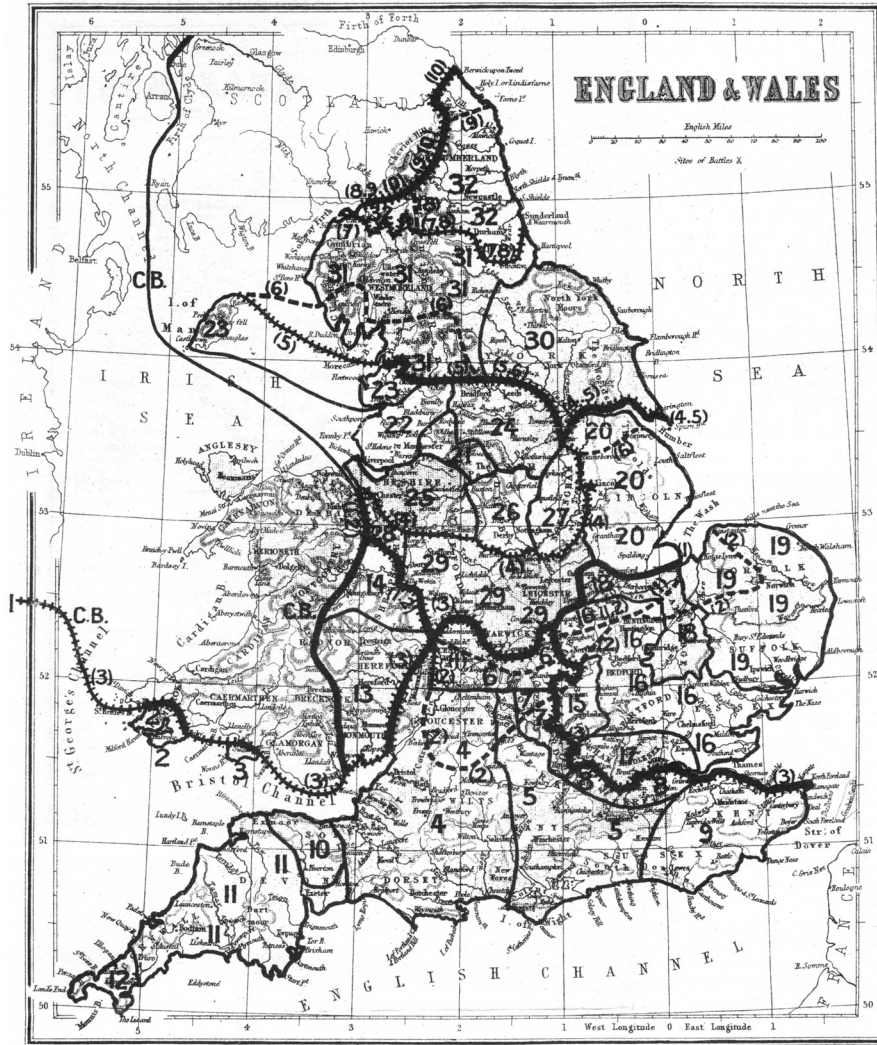
1. Introduction

Linguistic cartography has a solid tradition which dates back to the first half of the nineteenth century, when the first maps of languages in the modern sense were produced, i.e., geographic maps showing the delimitations of languages (*Sprachenkarten* in German, e.g. the languages of Asia by Klaproth 1823, the Slavic languages by Šafařík 1842, the Romance languages of Europe by Fuchs 1849). In the second half of the nineteenth century, dialectology became the main field of application for linguistic cartography, inspired by national dialect atlas project like the *Atlas linguistique de la France* (*ALF*) and the *Sprachatlas des Deutschen Reichs* (Wenker 1889–1923). Figure 1 reproduces Ellis' (1889) subdivision of English dialects into districts as an example of a *Sprachenkarte*, a map that synthesizes language areas from linguistic data. Figure 2 is a *Sprachkarte*, i.e., a linguistic map in the narrow sense (*Atlas linguistique de la Wallonie* [*ALW*], vol. 1, map 16): the pronunciation variants of the word *cher* 'expensive, dear' in the Belgo-Romance territory are shown using point symbols. Maps like these represent a well-established standard in dialectology and language typology. Classical dialect surveys focus on phonetic, morphological and lexical data and depict the results in point-symbol maps (e.g., *ALW*, Figure 2), point-text maps (e.g., *ALF*) or area-text maps (e.g., Wenker 1889–1923, cf. Section 7). Typological representations of language situations use isoglosses, as in Figure 1, and/or colors in order to mark language areas (see the language maps of the Ethnologue for instance, <<http://www.ethnologue.com>> [21.06.2011]). For an exhaustive description of this traditional linguistic cartography the reader is referred to the first two parts of the *Handbook of Language Mapping* (Lameli, Kehrein & Rabanus 2010).

This article will concentrate on new and innovative forms of language mapping that have mainly been developed recently, after the so-called “spatial turn” in the humanities notably enlarged the circle of scholars interested in and practicing linguistic cartography. Hence, the following aspects are presented and exemplified in order to picture the “State of the Art in Linguistic Cartography”:

- mapping of features from linguistic subfields that have received less attention in the history of linguistic cartography (section 2);

ENGLISH DIALECT DISTRICTS, by Alexander J. Ellis, 1887.



*To illustrate Alex. J. Ellis's "EXISTING PHONOLOGY OF ENGLISH DIALECTS,"
and "ENGLISH DIALECTS—THEIR SOUNDS AND HOMES."*

Figure 1. Ellis' subdivision of English dialects into districts.

- mapping aggregate data, i.e., techniques in linguistic cartography that go beyond the representation of single features (section 3);
- probabilistic mappings, i.e., geostatistical techniques for dealing with data gaps (section 4);
- mapping socially stratified data (section 5);
- geographically unfaithful mappings (section 6);
- user-defined mappings (section 7).

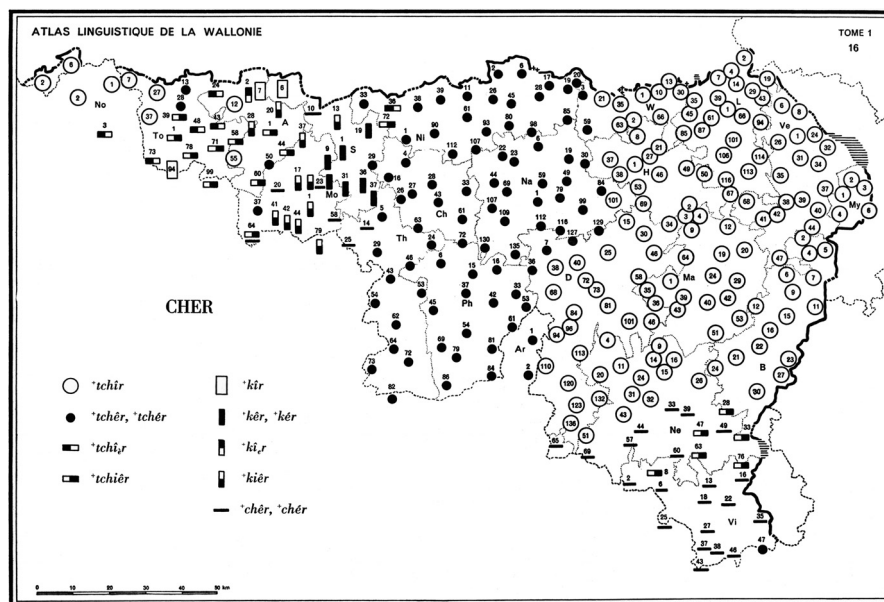


Figure 2. Pronunciation variants of *cher* ‘expensive, dear’ in Walloon and Picard dialects (*ALW*, vol. 1, map 16).

2. Mapping linguistic features other than phonetics, morphology or lexicon

As indicated in section 1, traditional dialect cartography focused very much on sounds (phonetics), forms (morphology) and words (lexicon). In this section, maps devoted to additional topics will be presented, i.e., syntax, pragmatics and phonology. Syntax was not systematically included in the major dialect surveys of the past. This situation has changed significantly in the last ten to fifteen years. There has been a “syntactic turn” in dialectology on the one hand, and a “dialect turn” in syntactic theory on the other (cf. Kortmann 2010, p. 838), resulting in a considerable number of linguistic map projects, especially in the Germanic area (e.g., *SAND*). Figure 3 is a map from the *Syntaktischer Atlas der Deutschen Schweiz* projekt (Bucheli Berger & Glaser 2004, map 1). The map topic is the agreement of the predicative adjective in the sentence *weil er nass ist* ‘because he is wet’, where the adjective *nass* ‘wet’ gets the agreement marker *-ä* in the southern Swiss dialects marked by triangles (*wül er nass-ä isch*). Figure 3 is a point-symbol map like Figure 2 (the isogloss marks the northern boundary of the feature area in a former linguistic atlas of Switzerland). But the important difference to Figure 2 is that size and color (white, gray, black) of the triangles give us information about the number of informants who used the construction with the agreement marker (*nass-ä*, ‘wet-M.SG’). Thus, the point symbols are proportional. This different symbolization is attributable to the fact that, while the phonetics and morphology of

dialects are usually very fixed (at least in areas like Switzerland where the dialect is still the everyday language), their syntax is much more flexible and usually allows the coexistence of different variants (here, the adjective *nass* with or without an agreement marker). Hence, syntax maps that simply depict *yes/no*-information (form is attested vs. form is not attested) are highly problematic.

Figure 4 is a point-symbol map taken from the morphology volume of the German *Mittelrheinischer Sprachatlas* (*MRhSA*, vol. 5, map 682). However, it can also be considered a pragmatic map since its topic is the differentiation of the direction adverbs *hinein* and *herein* (both meaning ‘into’, but towards or away from the speaker, respectively). The circles in the south-east of the area represent a distinction between the two adverbs (as in standard German) whereas the bars in the north-west indicate a suspension of the contrast (cf. Girth 2010, p. 104), i.e., dialects in which the speaker’s position in space is not considered (as in standard English). Pragmatic maps are very rare but it is worth noting that a map with a very similar topic (usage of adverbs in the deictic field) was already included in Bruno Schweizer’s hand-drawn linguistic atlas of Cimbrian (Schweizer 1954, p. 112).

Pronunciation or sound maps (*Lautkarten* in German) are the most frequent type of linguistic map. In traditional dialectology they usually visualize the different resultant forms of a single sound from a historical reference system, e.g., Latin *cl-* in today’s Romance languages, West Germanic *p-* in German dialects, Proto-Slavic **y* after labials in Sorbian, etc. Maps of this type can hardly be considered “phonological” from a synchronic point of view. Real phonological maps are extremely rare. An example is the map featuring complete consonant phoneme inventories in ten southern Low German dialects drawn by Wagener (1985, p. 155 [figure 3]). The data is, necessarily, highly condensed. Each location features a box containing 28 fields, one for each of the consonants that have phoneme status in at least one of the dialects of the surveyed area. The actual phoneme system of a local dialect is then visualized by blacking out the relevant fields. The map is a good data source for scholars interested in the individual systems. But it is very difficult—if not impossible—to recognize areal patterns.

Research into areal variation of prosody and, especially, intonation has increased significantly in recent years (cf. Gilles & Siebenhaar 2010). As with segmental features, it is necessary to distinguish realizational differences from systemic, genuinely phonological ones. An important contribution to the study of systemic areal differences is Peters’ (2006) book on the intonation of German regional languages. Since almost no data for the identification of systematic intonational differences (“intonation grammars”) are available from past dialect surveys, and hence very intense fieldwork is required, the location network is as yet too wide-meshed for detailed linguistic cartography (Peters 2006 considers only six German cities). There is, however, a small series of (isogloss and point-symbol) maps on the nuclear contours used in questions in Palatinate dialects in southwestern Germany (Peters 2006, pp. 385, 388, 390).

3. Mapping aggregate data

Most dialectological research focuses on the distribution of isolated features; this holds true even for the innovative features presented in the previous section. Scholars attempting to define language and dialect areas thus face the difficult task of choosing which features single out significant differences between languages and dialects. Whatever the approach taken, it becomes necessary to assign different weights to the various features when trying to identify significant areas. In German dialectology, special weight has always been assigned to the isoglosses of the second Germanic sound shift, giving rise to the “Rhenish Fan” that is considered the border between Low and High German dialects. In Romance dialectology, the “La Spezia-Rimini Line” is regarded as the decisive boundary between *Western* and *Eastern Romania*, bundling morphological isoglosses (e.g., formation of noun plurals) and phonetic lines (e.g., treatment of voiceless intervocalic stops). The two approaches presented here try to avoid having to make a choice and consider as much data as possible in order to display the general similarities and dissimilarities of languages or dialects. The pictures that arise from their analyses can be both interpreted in terms of dialect areas and correlated to extralinguistic features.

Probably the most representative maps recording the similarities/dissimilarities between linguistic varieties are the choropleth maps of the “Salzburg school of dialectometry” (also called “Visual Dialectometry”; for details cf. Goebel 2010). The basic idea behind this approach is the transformation of the qualitative features associated with geographical points or areas that are recorded in traditional linguistic atlases into what are known as “taxates” (Goebel 2010, p. 437), i.e., units whose quantitative similarities can be calculated and mapped using different colors for the reference points or areas. The typical cartographic outcome of the operation is known as a “similarity map”. The series of similarity maps that Goebel and his team computed using data from the *Atlas linguistique de la France (ALF)* provide a good example of this mapping technique, see Figure 5 (Goebel 2010, map 2204; cf. Goebel 2010 and <<http://www.dialectometry.com>> [21.06.2011] for further full-color examples). The map was generated as follows: first the 641 *ALF* survey locations were polygonized, thus creating an areal representation of France subdivided into survey locations. Then the phonetic, morphological and lexical data from all (1687 working) *ALF* maps were transformed into “taxates”. Finally the degree of similarity between a reference location and all other survey locations with respect to all of the features considered were mapped by coloring the polygons according to a well-defined color scheme. In Figure 5, the reference location is *ALF* point 885, Gréoux (Département Basses-Alpes), the white polygon in the southeast. It is surrounded by red polygons representing the dialects with the highest degree of similarity. Goebel uses a scale of up to eight colors from deep red for the highest degree of similarity (“warmth”) to deep blue for the lowest (“coldness”). A picture in which the reference location is surrounded by red po-

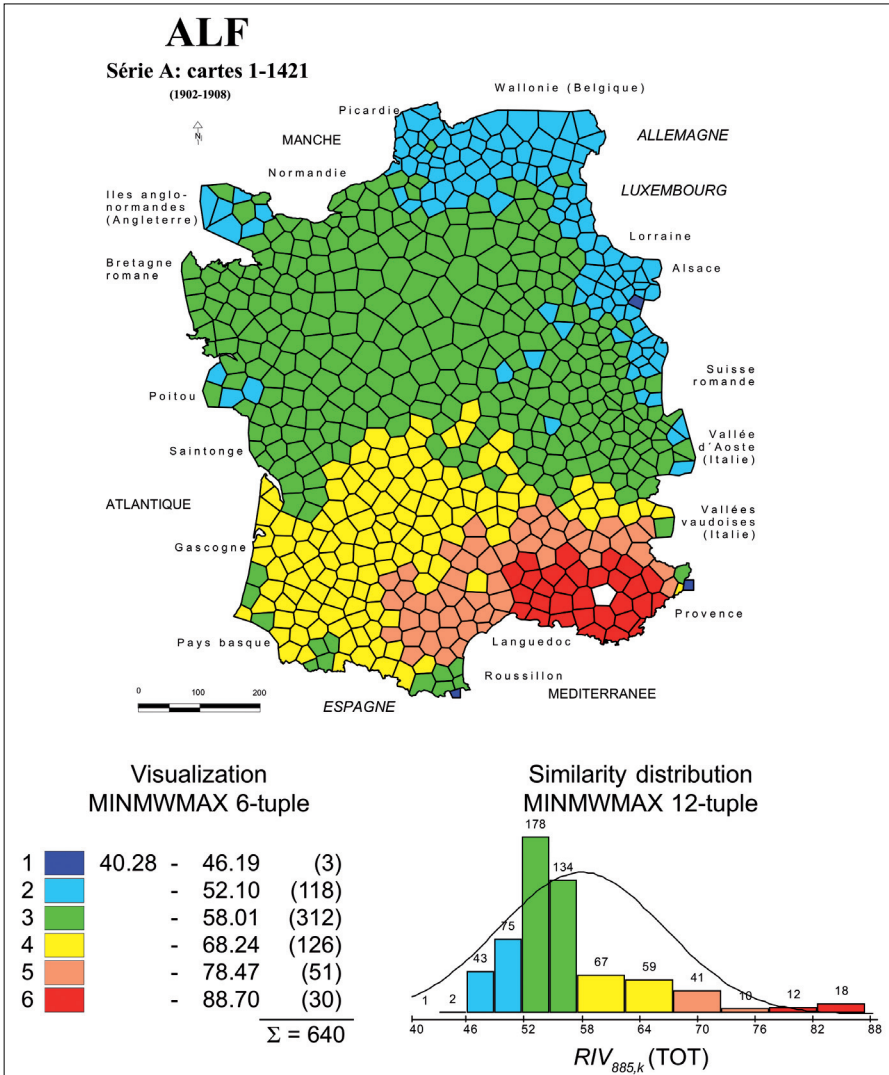


Figure 5. Similarity map showing the similarity values for *ALF* point 885, Gréoux, Département Basses-Alpes (Goebl 2010, map 2204).

lygons while increasing geographic distance leads first to orange, then yellow, green, light blue and finally deep blue polygons (i.e., the larger the geographic distance the “colder” the color) could be considered the approach’s null hypothesis (and is visualized by “proximity” maps such as Goebl 2010, map 2206). Deviations from this picture are due to factors such as different substrata, different sound laws, discontinuous communication barriers, etc. This is the case in Figure 5: the distribution of colors is not simply an effect of geographic distance. The geographically distant Gascogne dialects are more similar to the Gréoux dialect than the dialects of east-central France, which are geographical-

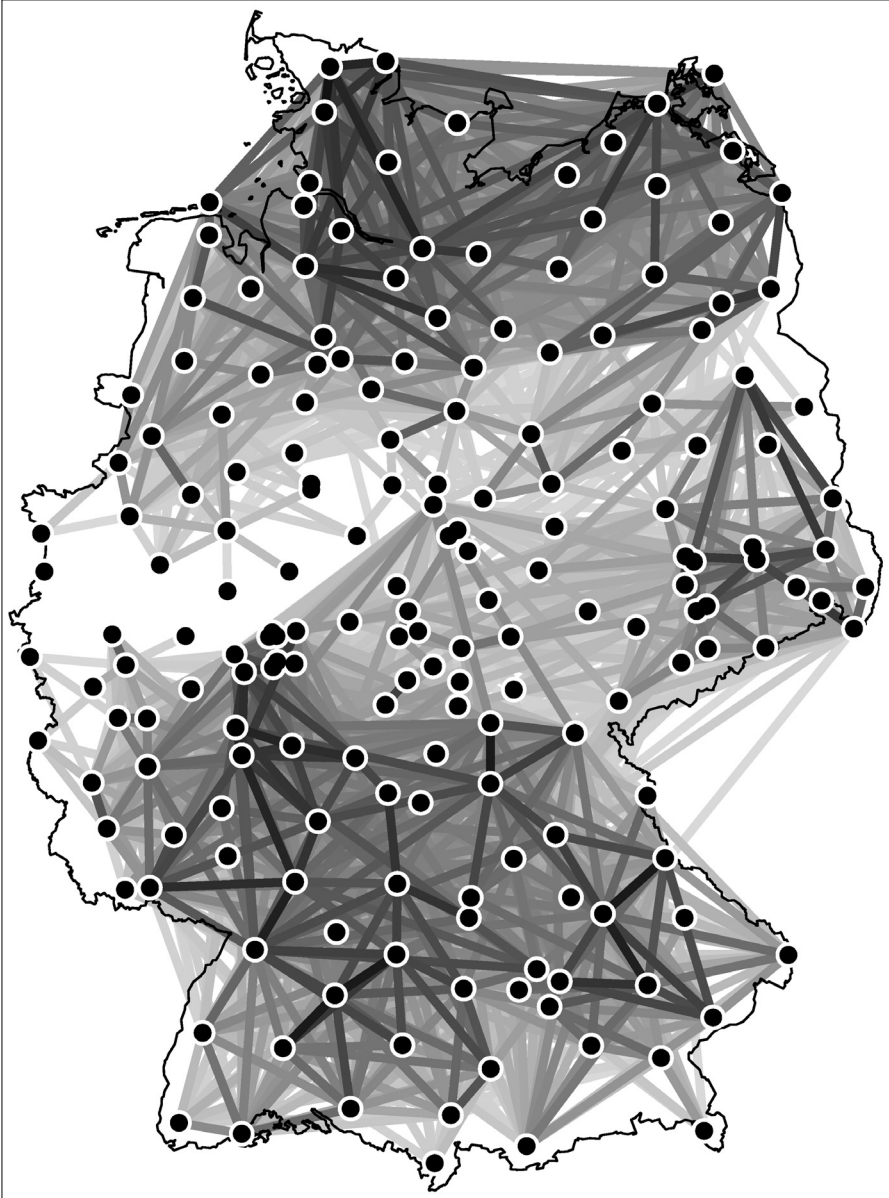


Figure 6. Network map of pronunciation similarities between German dialects (Nerbonne 2010, map 2402A).

ly closer. This result is in accord with the traditional linguistic bipartition of France into a northern and a southern area: the warm-colored polygons around Gréoux cover the *Domaine d'Oc* (Southern French, Occitanian) whereas the cold-colored areas in the north represent the *Domaine d'Oil* (Northern French; cf. Goebel 2010, p. 443).

The other influential method for calculating dialect similarities and dissimilarities to be introduced here is the “Groningen approach”. Figure 6 (Nerbonne 2010, map 2402A) is a network map which shows the degrees of similarity between dialects in Germany with respect to pronunciation. It is based on 201 phonetic features extracted from materials collected for the *Phonetischer Atlas Deutschlands* project (cf. Göschel 1992). In contrast to the Salzburg similarity maps, there is no reference location: each of the 186 locations is compared to every other location, making 20,100 pairs of locations which are checked for differences in the 201 phonetic features using Levenshtein distance (cf. Nerbonne 2010, pp. 480–482). The cartographic result of the procedure is the network map in Figure 6, in which each site is connected to every other site and the degree of darkness of the lines corresponds to the similarity of the local dialects: the darker the line the more similar they are. The “aggregate perspective” Nerbonne (2010) argues for, shows, on the one hand, the well-known north–south contrast in Germany (Low vs. High German). On the other, however, Figure 6 also shows that the pronunciation contrast between north and south is much sharper in the west (no visible connecting lines) than it is in the east. Aggregate mapping thus visualizes a little known aspect that is worthy of more detailed study.

4. Probabilistic mappings

Dialectologists are often interested in the identification of areas, be they dialect areas as discussed in the previous section or just feature areas. However, the task of identification and delimitation confronts them with two absence-of-information problems. First, every dialect survey includes locations in which it was not possible to collect data, so the data coverage across the location network is almost never uniform and complete (see, for instance, the hook symbols in Figure 4, which represent locations without data). Second, even where the data are almost evenly distributed across the territory, the question of where exactly to draw the boundary lines between the identified areas arises. In traditional dialectology, the first problem was usually ignored and the second one solved intuitively by the map drawer, e.g., by putting the isogloss half-way between two locations with attested data. In this section, a technically more advanced solution will be illustrated. Figure 7 (a reproduction of Wattel & van Reenen 2010, map 2502) is a contour-line map which visualizes the spelling variants *o* and *a* in *af* ‘off’ in fourteenth-century Middle Dutch charters. The map shows the variant *o* in brown, mainly in a stripe along the coast between Bruges and Amsterdam, while variant *a* is attested mainly in a triangle between Brussels, Zwolle and Maastricht, marked in green. Figure 7, however, visualizes a geostatistical elaboration of the original data. Firstly, with respect to the completeness problem, although the charters are distributed very unequally across the territory (e.g., high frequency of charters around Maastricht and low

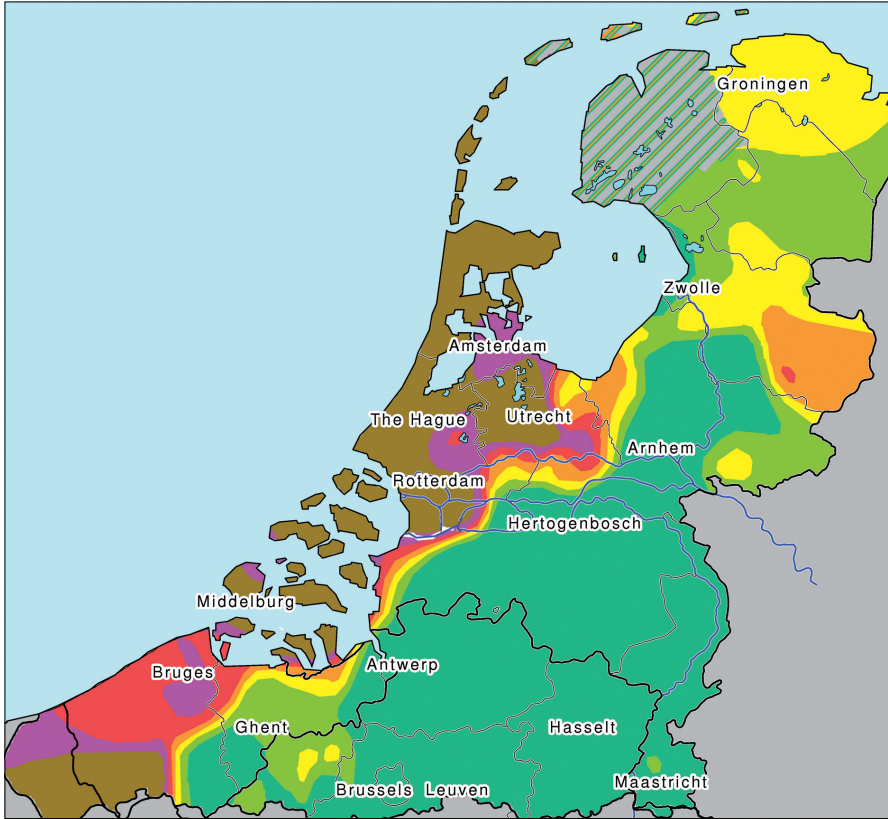


Figure 7. Spelling variants *o* and *a* in *af* ‘off’ in Middle Dutch charters (Wattel and van Reenen 2010, map 2502).

frequency around Antwerp; cf. Wattel & van Reenen 2010, map 2501), the map levels these differences so that the areas can be clearly visualized (Maastricht and Antwerp lie in the same [green] area). Secondly, and with respect to the delimitation problem, transition zones are inferred between the main areas of the *o* and *a* variants (pink, red, orange and yellow areas in Figure 7). Transition zones reflect linguistic reality much better than sharp boundaries. In this instance, the transition zones represent variation between *o* and *a* spellings in not only the same locations but even the same charters. Both solutions are achieved using an interpolation procedure to calculate unattested values. “By making a distinction between the *value* of a dialect observation [...] and the *weight* of the dialect observation [...], the extrapolation procedure calculates the influence of the value on its surroundings by diminishing the weight over distance” (Wattel & van Reenen 2010, p. 499; for details on the algorithm cf. Wattel & van Reenen 2010, pp. 499–502). That is why Figure 7 is called a “probabilistic map”: every pixel of the image is assigned a color representing a linguistic value, although the value is really attested for only a very small percentage of the pixels,

and even in those cases only in the form of *yes/no* data (*o* or *a*, never transition values). The values for all other pixels have been computed and are hence only probable values. The upper-left corner of the map, the Frisian area (west of Groningen), shows the limits of the method. It is striped because the corpus does not contain any charters from the Frisian area. The color patterns (green and yellow) are thus extrapolations from outside the Frisian area and therefore do not have a high degree of probability.

5. Mapping socially stratified data

Traditional dialect surveys take the data of one or two informants as representative of a local dialect which is considered homogeneous. Consequently, on most linguistic maps, one form per location is depicted. We can take for granted that even very traditional dialectologists know that the “one form per location” principle is not correct. However, whereas today’s monographs devoted to single dialects almost always consider social or contextual variation, for practical reasons (limited resources) this less frequently occurs in large-area surveys. As early as 1902, the authors of the *ALF* were well aware of the existence of social variation in local dialects: in the *Notice servant à l’intelligence des cartes* (Gilliéron & Edmont 1902) Edmont pointed out individual or age-related variation in comments like “Les personnes très âgées remplacent *j* par *h*” (Gilliéron & Edmont 1902, p. 22). But the different variants were not recorded on the maps. The first atlas which systematically took social variation into account was the *Linguistic Atlas of New England* (*LANE*, 1939–1943) by Hans Kurath and colleagues. There is a direct connection to the French tradition here, in that Jakob Jud, a student of Gilliéron’s and one of the authors of the Italian *Sprach- und Sachatlas Italiens und der Südschweiz* (*AIS*), participated in the design of the *LANE* (cf. Lameli 2010, p. 583). Like the Romance atlases, the *LANE* features point-text maps. But unlike them, *LANE* maps provide at least two variants per variable, one from an older informant with little formal education and the other from a younger informant with more formal education. In bigger cities additional data is represented, with up to four variants per variable. The *LANE* maps thus provide good data on social variation but are difficult to read. In the point-text maps of the French atlas, in which only one form per location is shown, it is not easy to recognize clusters of locations that share the same form; in the maps of the *LANE*, it is almost impossible. That is why, more recently, point-symbol maps have been used to depict socially stratified linguistic data. (It should be noted for the sake of completeness that the *LANE* Handbook [Kurath 1939] does feature twenty point-symbol and two isogloss maps [labeled “charts”] of selected features.)

An advanced example of such “pluridimensional cartography” is Harald Thun’s *Atlas lingüístico Diatópico y Diastrático del Uruguay* (*ADDU*; cf. also Thun 2010). As illustrated in Figure 8 (Thun 2010, map 2602, occurrences of

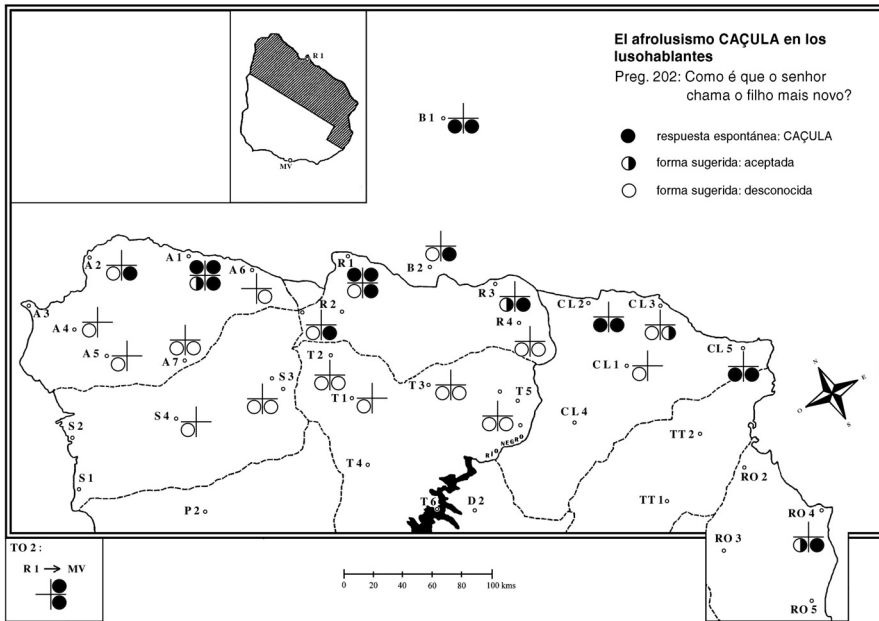


Figure 8. Attestations of the word *caçula* ‘youngest son’ in northern Uruguay and two locations of southern Brazil (pluridimensional representation; Thun 2010, map 2602).

the Afro-lusism *caçula* ‘youngest son’ in northern Uruguay and two southern Brazilian locations), complex symbols are used: a cross creates four fields that depict what Thun calls the “four standard groups”. Each group represents a specific combination of the social variables already used in the *LANE*, i.e., ‘age’ and ‘formal school education’. “Thus, the two fields above the horizontal line of the cross designate the groups of informants who have at a minimum completed Uruguayan, Argentine, Brazilian or Paraguayan secondary education, and those below the bar record the responses of groups whose level of formal education ranges from zero (illiteracy) to a completed primary education. The two left-hand fields are reserved for the groups of older informants (GII, sixty years old and above); the two right-hand fields are those of the younger groups (GI, from 18 to 36 years old)” (Thun 2010, p. 518). The symbols themselves are subdivided into sectors: in Figure 8 a completely black circle indicates a spontaneous positive response (attestations of the usage of *caçula*) whereas a half-black half-white circle stands for the acceptance of the word only after its suggestion by the interviewer. It is true that the symbology is complex. However, the map is far easier to interpret than one of *LANE*’s point-text maps. The reason for preferring the pluridimensional map becomes obvious when Figures 8 and 9 are compared. Figure 9 (Thun 2010, map 2601) is a traditional point-symbol map of the same lexical variable (*caçula* ‘youngest son’) in the speech of the older generation in the same area. The map shows just three attestations (black circles). In contrast, Figure 8 shows the picture that emerges

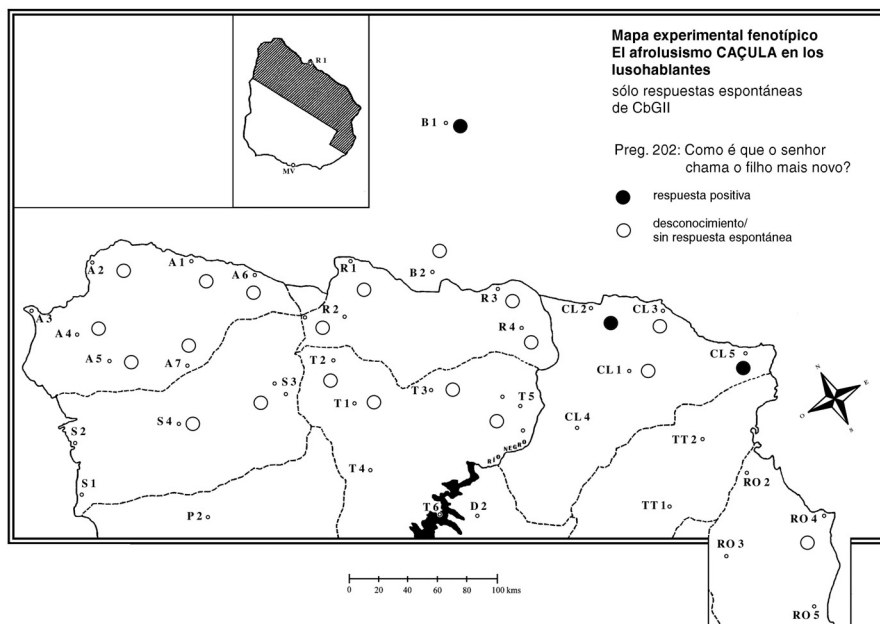


Figure 9. Attestations of the word *caçula* ‘youngest son’ in the speech of the elder generation in northern Uruguay and two locations of southern Brazil (monodimensional representation; Thun 2010, map 2601).

when the survey is enlarged to include all of the four standard groups. We see the three attestations from Figure 9 in the lower left-hand field of the cross symbols in Figure 8. The overview of the socially stratified data in Figure 8 allows us to characterize the situation as follows:

1. The word *caçula* is attested at both Brazilian locations and all near-border locations in northern Uruguay. Some attestations are even found further south.
2. The word is more frequently attested among speakers of the younger generation (right-hand column).
3. The word is always attested among speakers with a higher formal education (upper row; N.B.: speakers with higher formal education were interviewed at only two locations).

Ergo, the form is a genuine linguistic innovation, which has spread from Brazil to the Uruguayan border locations and which will probably continue to spread, both geographically (from north to south) and socially (from the younger, better-educated to older, less-educated speakers). This conclusion would not have been possible on the basis of Figure 9, which could on the contrary actually suggest that *caçula* is an archaism, attested only among the older generation.

6. Geographically unfaithful mappings

At first glance it may seem strange to include a section on maps that are not geographically correct. However, it is worth noting, first, that the history of linguistic cartography from the nineteenth century to the present has not been one of continuous improvement in geographic correctness. On the contrary, Wenker's *Sprachatlas des Deutschen Reichs* (Wenker 1889–1923) was geographically more accurate than the German regional dialect atlases published at the end of the twentieth century. Second, the deliberate incorrect positioning of locations or areas may make sense, either so as to abstract from irrelevant geographical detail or to link the representation to the result of statistical procedures. An example of the former is Figure 10 (Wikle & Bailey 2010, map 1201). It shows the percentages of intervocalic voiceless *s* in the words *to grease* and *greasy* in the United States, which Hempl (1896) subdivided into four sectors at the state level: “the North (extending from New England to the Dakotas), Midland (St. Louis to New York), South (southward from the Mason Dixon Line) and West (St. Louis to California)” (Wikle & Bailey 2010, p. 255). Representations like this have proved to be a good starting point for further research.

An example of the second type is the multidimensional scaling (MDS) plot in Figure 11 (Nerbonne 2010, p. 488 [figure 24.3]). As in Figure 6 (section 3), the input data are materials collected for the *Phonetischer Atlas Deutschlands* project. The symbols represent 186 local dialects. The shapes of the symbols correspond to the results of a cluster analysis that recognized nine large dialect groups (Nerbonne 2010, p. 484). Their positions in the MDS plot are not geographically correct since the boxes stand for the northern and eastern Low German dialects and the plus signs for the western and southern Low German dia-

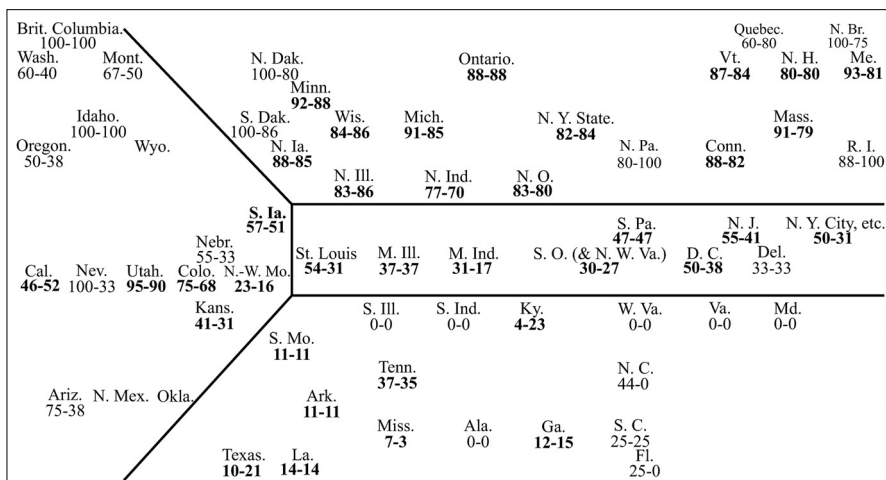


Figure 10. Percentages of voiceless *s* (as in *sin*) in the United States: the first number refers to *grease*, the second number to *greasy* (Wikle & Bailey 2010, map 1201).

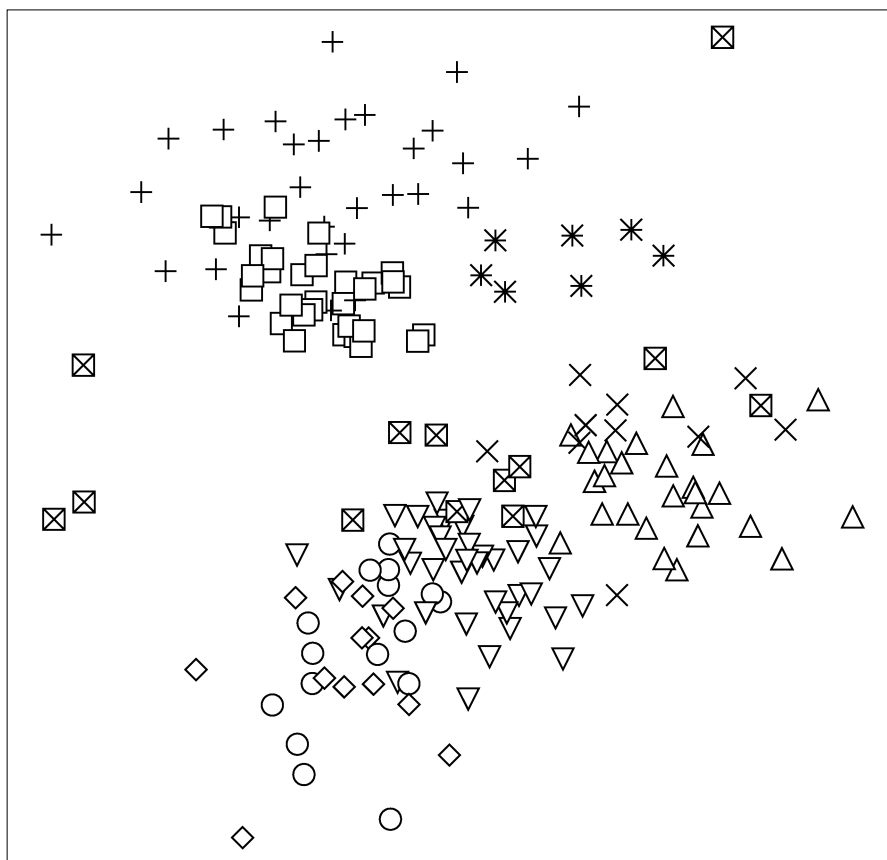


Figure 11. Pronunciation distance data for German dialects in two MDS dimensions (Nerbonne 2010, p. 488 [figure 24.3]).

lects (cf. Nerbonne 2010, map 2403)—while in Figure 11 most plus signs are “north-east” of the boxes. However, the MDS plot indicates the distinction between Low and High German in that the Low German varieties (boxes, plus signs and asterisks) are separated fairly well from the High German varieties (circles, diamonds, crosses, triangles, inverted triangles). The only problematic group are the Palatinate and Ripuarian varieties, symbolized by crossed boxes that do not fit well in the overall picture since the crossed boxes can be found everywhere (cf. Nerbonne 2010, pp. 487–489).

7. User-defined mappings

The topic of this final section is somewhat different from those of the previous sections because the innovation here does not regard the maps themselves, which are often very traditional linguistic maps. It is the process of their cre-

ation and/or representation which is innovative because it is user-defined. Additionally, the maps are always electronically represented, being stored on CD/DVD or, most often, on Internet servers. The examples discussed here shall be grouped in two classes: (i) user-defined visualizations of static maps, i.e., of maps that also exist in printed form, and (ii) user-defined mappings of data stored in databases (no static maps).

The examples of the first type come from the *Digitaler Wenker-Atlas* (*DiWA*) which is and will probably remain the most elaborated example of this type of user-defined mapping. Its core is the Internet publication of Georg Wenker's *Sprachatlas des Deutschen Reichs* (Wenker 1889–1923), whose hand-drafted maps remained unpublished for a whole century because of the huge size of the map sheets and the multitude of colors Wenker and his collaborators used to represent the linguistic data. In a second stage, other linguistic maps (mostly from German regional linguistic atlases from the second half of the twentieth century) were added to *DiWA* to form a corpus of more than 3,300 maps. The sheer size of the map files was a challenge for the Internet publication, since files of up to 1.6 GB (for Wenker's historical maps) are difficult to handle via the Internet even if we expect a continuous increase in bandwidth. Another important issue was the faithful representation of the historical maps. On the one hand, the digital atlas had to be a functional tool for dialectological research on German dialects. On the other hand, the original maps also had to be faithfully reproduced as a preservation of cultural heritage. These challenges and their technical solutions are described in detail in Rabanus, Kehrein & Lameli (2010). The central aspect of the technical solution is the geocoding of the map files. 'Geocoding' means that geographic coordinates are assigned to each pixel of the graphics files. This enables the maps to be added to a Geographic Information System (GIS), which is the necessary precondition for all functions described in the following. In addition to various capabilities for browsing, zooming and searching the historical maps, the following functions render the atlas 'user-defined' (<<http://www.diwa.info>> [freely accessible, no registration required] for the original full-color maps which are only poorly reproduced in the screenshots in Figures 12–14):

1. Maps of the same feature from atlases of different areas can be loaded in the same window and viewed alongside to each other in space. This function makes it possible to circumvent atlas boundaries, which are often administrative boundaries that transect homogeneous dialect areas or, worse, transition zones that need to be viewed in their entirety to understand the dynamics of regional speech. Figure 12 shows a view of the Tyrolean Alps in which maps of the second person plural subject pronoun *ihr* 'you' from three German dialect atlases are superimposed (the light area to the left from *VALTS* [vol. 5, map 209], the light area in the upper-right corner from *SOB* [vol. 4, map 26] and the dark-gray area from *TSA* [vol. 3, map 107]). The superimposition is possible thanks to the geocoding mentioned above: geocoded maps are automatically

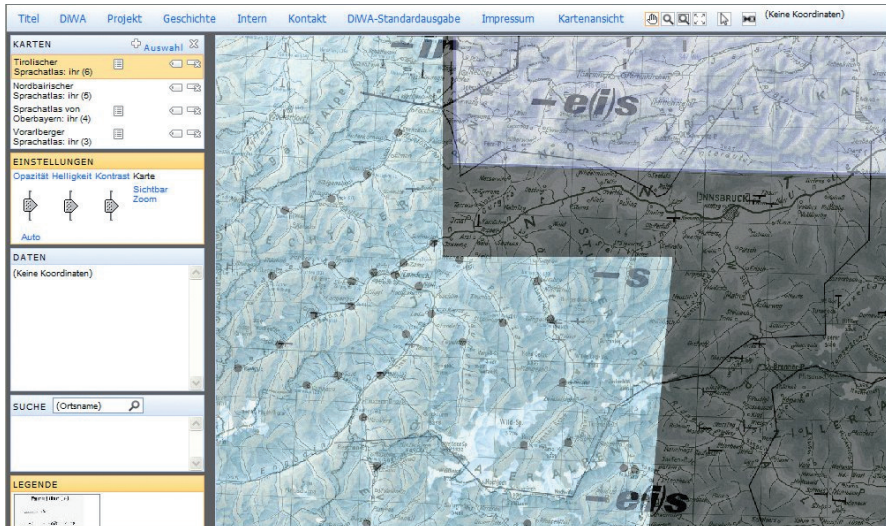


Figure 12. Overlay of maps of the 2nd person plural subject pronoun *ihr* ‘you’ from *TSA* (vol. 3, map 107), *VALTS* (vol. 5, map 209) and *SOB* (vol. 4, map 26) in a *DiWA* window.

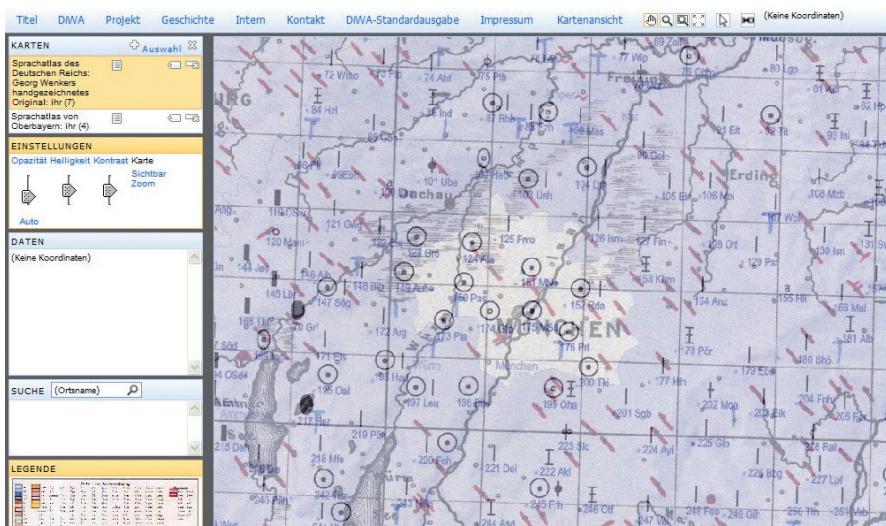


Figure 13. Overlay of maps of the 2nd person plural subject pronoun *ihr* ‘you’ from *SOB* (vol. 4, map 26) and Wenker (1889–1923, map 397) in a *DiWA* window.

assigned to the right location when viewed in a GIS window, as here. Hence, the whole area can be viewed and studied, no matter which atlas projects the subregions belong to.

2. Maps of the same area but of different periods can be superimposed in order to study language change, as illustrated by Figure 13. In Figure 13 we see a section (the Munich area) of the map of the second person plural subject

pronoun *ihr* ‘you’ from Wenker’s historical atlas (Wenker 1889–1923, map 397; data collection 1887, map drawn 1901–1902) and of the corresponding map from the *Sprachatlas von Oberbayern (SOB)*, vol. 4, map 26; data collection 1991–1998, map published 2008). The picture thus depicts a cross-section in time of around 100 years. In the mapped area, two morphological types of the pronoun occur: the Bavarian type *es* (one of the so-called Bavarian “Kennwörter”) and the standard-like type *ihr*. In Figure 13, *es* is the default form on Wenker’s nineteenth-century map: dots (small open circles) without additional symbols and dots with red backslashes represent location with *es*. In the *SOB*, the *es* forms are symbolized by vertical bars. The standard-like form *ihr* is indicated by blue “T”-symbols on Wenker’s map (very few, e.g., behind *SOB* locations 176 Prl and 107 Wpk) and by big open circles in the *SOB*. Unfortunately, the two layers are not easily distinguishable in the reduced resolution shown here. In the original Internet view, however, it is clearly visible that the *es* area was very homogeneous and even included most of the locations in the Munich city area (light area in Figure 13) in the nineteenth century. In the *SOB* the situation has changed completely. Whereas *es* forms (vertical bars) persist in the south-east and north-west of the map section, in the city area of Munich—and spreading out from Munich towards north and south—the *es* forms have been almost completely replaced by *ihr* forms (circles). This is an example of a typical language change: locations in urban areas adopt standard-like forms.

3. Linguistic maps can be superimposed onto non-linguistic maps in order to find correlations between linguistic and non-linguistic facts and, hence, explanations for the distribution of linguistic features. Figure 14 shows two

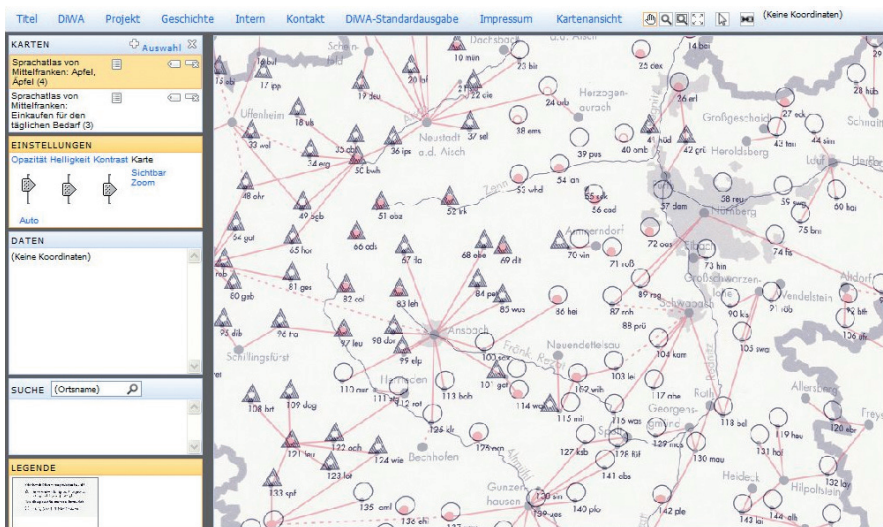


Figure 14. Overlay of two *SMF* maps (vol. 1, map 6 and vol. 7, map 4) in a *DiWA* window.

maps from the *Sprachatlas von Mittelfranken (SMF, northern Bavaria)*. The morphological map (*SMF, vol. 7, map 4*) records the formal marking of the plural of the word *Apfel* ‘apple’ using point symbols. The circles in the east stand for dialects that mark number distinction by Umlaut as in standard German (e.g., *Apfel* ‘apple[SG]’ vs. *Äpfel* ‘apple[PL]’; glossing according to the Leipzig Glossing Rules, cf. <<http://www.eva.mpg.de/lingua/resources/glossing-rules.php>> [21.06.2011]). The triangles in the west symbolize locations in which there is no formal expression of number difference because *Apfel* always has Umlaut, even in the singular. The network map in the background (*SMF, vol. 1, map 6*; red line symbols) records people’s shopping behavior: it shows the shopping centers and the locations associated with them. There is a rather good correspondence between the linguistic division and the spatial division in shopping behavior, which mirrors social contact on a general level.

The second type of user-defined mapping can be illustrated with the map function of the Nordic Dialect Corpus (cf. Johannessen et al. 2009). It is a database freely accessible on the Internet (<<http://www.tekstlab.uio.no/nota/scandia-syn>> [21.06.2011], registration required), which consists of spontaneous speech data from the dialects of the North Germanic languages that have been transcribed, linked to audio and video files and grammatically tagged. The tagging makes it possible to search the corpus not only for words but also for grammatical features, e.g., cases. The result of the search is, in first instance, a list of all sentences that contain the words tagged with the features in question.

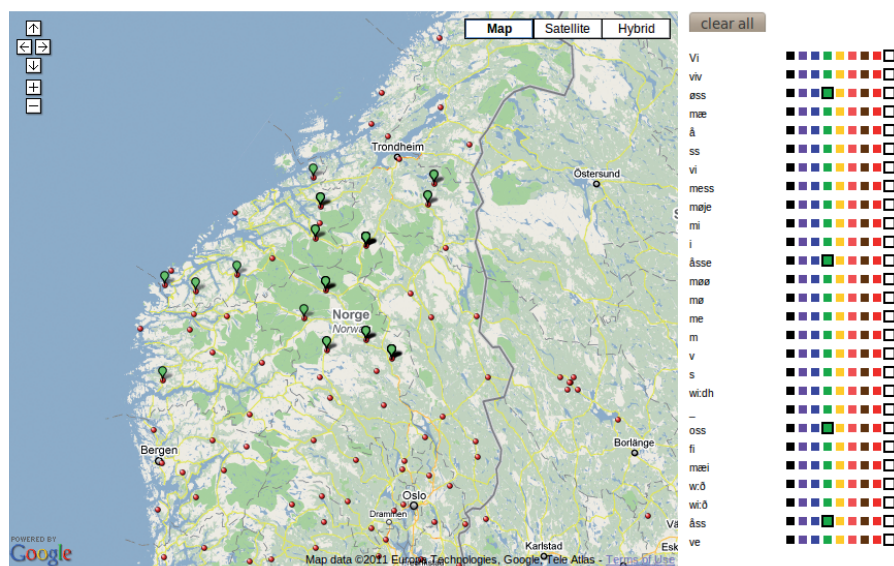


Figure 15. Attestations of the 1st person plural object pronoun *oss* ‘us’ in subject function (‘we’) in the Nordic Dialect Corpus.

With the map function it is then possible to visualize the geographic locations of the occurrences in a personalized Google Maps window. Figure 15 visualizes the occurrences of the pronoun *oss* ‘us’ (first person plural object pronoun in standard Norwegian [Bokmål and Nynorsk]) used in subject function (i.e., *oss* ‘we’). In order to generate the picture (a fairly homogeneous area south of Trondheim) it is necessary to search the database for all forms of the subject pronoun *vi* ‘we’ and then pick from the right-hand column the pronunciations which can be considered variants of the type *oss* (I chose green color to mark the *oss* locations on the map). Since the Nordic Dialect Corpus contains spontaneous speech the feature is not necessarily restricted to the marked points. It might also occur at additional locations, perhaps less frequently, so that it is not attested in the thirty minutes of spontaneous speech recorded at each location (obviously, the map cannot visualize more than the information in the database). However, the map gives a first indication and proves that the feature (*oss* as a subject pronoun) does indeed occur in spontaneous speech. Another example of a user-defined map generated from a database is the *Dynamic Syntactic Atlas of Dutch Dialects (DynaSAND)*. Since the database contains materials from questionnaire sessions, the occurrence or non-occurrence of a feature on a *DynaSAND* map is more likely to reflect the actual presence or absence of that feature than is the case with a Nordic Dialect Corpus map (i.e., in *DynaSAND* ‘absence of symbol’ really does mean ‘absence of feature’). On the other hand, *DynaSAND* has a more limited feature set, and features and words are less freely combinable in the search function than in the Nordic Dialect Corpus map. To sum up: user-defined maps generated on the fly from databases are still in the early stages of development, and in cartographic terms they are very simple. However, this type of mapping will become very important in the future.

8. Conclusions

In order to offer an overview of the state of the art in linguistic cartography, this article has presented recent advances in the field. Whereas traditional dialectology was the main field of application for linguistic cartography in the twentieth century, the new cartographic forms presented here are a result of the introduction of methods and approaches from fields such as theoretical linguistics (e.g., generative grammar), sociolinguistics, statistics and geoinformatics. The combination of data and methods from various disciplines has proved fruitful, usually for all of the disciplines involved in the exchange. It is important to point out that new cartographic methods allow new perspectives even on old data, as was exemplified here, particularly with aggregate and dialectometric mappings (section 3) and the *DiWA* system (section 7). One of the aims of my contribution to the Nordic Dialectology Meeting has been to extend the invitation to apply these new methods to the impressive amount of data

collected by generations of dialectologists, so as to reaffirm for these data the scientific value they continue to deserve, today and in the future.¹

Cited atlases

- ADDU = Thun, Harald & Adolfo Elizaincín, 2000: Atlas lingüístico Diatópico y Diastrático del Uruguay. I, 1–2: Consonantismo y vocalismo del español. Kiel.
- ALF = Gilliéron, Jules & Edmond Edmont, 1902–1915: Atlas Linguistique de la France. 10 vols. Paris.
- ALI = Jaberg, Karl & Jakob Jud, 1928–1940: Sprach- und Sachatlas Italiens und der Südschweiz. 8 vols. Zofingen.
- DiWA = Schmidt, Jürgen Erich & Joachim Herrgen (eds.), 2001–2009: Digitaler Wenker-Atlas. Compiled and prepared by Alfred Lameli, Tanja Giessler, Roland Kehrein, Alexandra Lenz, Karl-Heinz Müller, Jost Nickel, Christoph Purschke & Stefan Rabanus. Online publication. Available from <<http://www.diwa.info>>. Marburg: Forschungszentrum Deutscher Sprachatlas.
- DynaSAND = Barbiers, Sjef et al., 2006: Dynamische Syntactische Atlas van de Nederlandse Dialecten. Online publication. Available from <<http://www.meertens.knaw.nl/sand>>. Amsterdam: P. J. Meertens Instituut.
- Klaproth, Julius, 1823: Asia polyglotta nebst Sprachatlas. Paris.
- LANE = Kurath, Hans, 1939–1943: Linguistic Atlas of New England. In collaboration with Miles L. Hanley, Bernard Bloch, Guy S. Lowman Jr. & Marcus L. Hansen. 3 vols. Providence, Rhode Island.
- MRhSA = Bellmann, Günter, Joachim Herrgen & Jürgen Erich Schmidt, 1994–2002: Mittelrheinischer Sprachatlas. 5 vols. Tübingen.
- SAND = Barbiers, Sjef, Hans Bennis, Johann van der Auwera, Eeftje Boef, Gunther de Vogelaer & Margreet van der Ham, 2005–2008: Syntactische Atlas van de Nederlandse Dialecten. 2 vols. Amsterdam.
- Schweizer, Bruno, 1954: Zimbrischer Sprachatlas. Dießen/Ammersee. Manuscript held in the archives of the Forschungszentrum Deutscher Sprachatlas, Marburg.
- SMF = Munske, Horst Haider & Alfred Klepsch (eds.), 2003–: Sprachatlas von Mittelfranken. Heidelberg.
- SOB = Eichinger, Ludwig M. (ed.), 2008–: Sprachatlas von Oberbayern. Heidelberg.
- TSA = Kühebacher, Egon, 1965–1971: Tirolischer Sprachatlas. 3 vols. Marburg.
- VALTS = Gabriel, Eugen (ed.), 1985–: Vorarlberger Sprachatlas. Mit Einschluss des Fürstentums Lichtenstein, Westtirols und des Allgäus. Bregenz.
- Wenker, Georg, 1889–1923: Sprachatlas des Deutschen Reichs. Hand-drafted by Emil Maurmann, Georg Wenker and Ferdinand Wrede. Manuscript held in the archives of the Forschungszentrum Deutscher Sprachatlas, Marburg.

Literature

- Auer, Peter & Jürgen Erich Schmidt (eds.), 2010: Language and Space. An International Handbook of Linguistic Variation. Vol. 1: Theories and Methods. Berlin/New York.

¹ I want to thank Mark Pennay (Forschungszentrum Deutscher Sprachatlas, Marburg) for correcting the English manuscript.

- Bucheli Berger, Claudia & Elvira Glaser, 2004: Zur Morphologie des (ko)prädikativen Adjektivs und Partizips II im Alemannischen und Bairischen. In: Franz Patocka & Peter Wiesinger (eds.), *Morphologie und Syntax deutscher Dialekte und Historische Dialektologie des Deutschen*. Beiträge zum 1. Kongress der Internationalen Gesellschaft für Dialektologie des Deutschen, Marburg/Lahn, 5.–8. März 2003. Wien. Pp. 189–226.
- Ellis, Alexander J., 1889: *The Existing Phonology of English Dialects, Compared with that of West Saxon Speech*. London.
- Fuchs, August, 1849: *Die romanischen Sprachen in ihrem Verhältnisse zum Lateinischen*. Nebst einer Karte des romanischen Sprachgebiets in Europa. Halle.
- Gilles, Peter & Beat Siebenhaar, 2010: Areal variation in prosody. In: Auer & Schmidt. Pp. 786–804.
- Gilliéron, Jules & Edmond Edmont, 1902: *Atlas linguistique de la France*. Notice servant a l'intelligence des cartes. Paris. [Authors not indicated on the title page.]
- Girnth, Heiko, 2010: Mapping language data. In: Lameli, Kehrein & Rabanus. Pp. 98–121.
- Goebel, Hans, 2010: Dialectometry and quantitative mapping. In: Lameli, Kehrein & Rabanus. Pp. 433–457.
- Göschel, Joachim, 1992: *Das Forschungsinstitut für deutsche Sprache "Deutscher Sprachatlas"*. Wissenschaftlicher Bericht. Marburg.
- Hempl, George, 1896: Grease and Greasy. In: *Dialect Notes* 1. Pp. 438–444.
- Johannessen, Janne Bondi, Joel Priestley, Kristin Hagen, Tor Anders Åfarli, & Øystein Alexander Vangsnes, 2009: The Nordic Dialect Corpus—an Advanced Research Tool. In: Kristiina Jokinen & Eckhard Bick (eds.), *Proceedings of the 17th Nordic Conference of Computational Linguistics NODALIDA 2009*. Tartu. (NEALT Proceedings Series 4.) Pp. 73–80.
- Kortmann, Bernd, 2010: Areal variation in syntax. In: Auer & Schmidt. Pp. 837–864.
- Kurath, Hans, 1939: *Handbook of the linguistic geography of New England*. In collaboration with Marcus L. Hansen, Julia Bloch & Bernard Bloch. Providence, Rhode Island.
- Lameli, Alfred, 2010: Linguistic atlases—traditional and modern. In: Auer & Schmidt. Pp. 567–592.
- Lameli, Alfred, Roland Kehrein & Stefan Rabanus (eds.), 2010: *Language and Space*. An International Handbook of Linguistic Variation. Vol. 2: Language Mapping. Berlin/New York.
- Nerbonne, John, 2010: Mapping aggregate variation. In: Lameli, Kehrein & Rabanus. Pp. 476–495.
- Peters, Jörg, 2006: *Intonation deutscher Regionalsprachen*. Berlin/New York.
- Rabanus, Stefan, Roland Kehrein & Alfred Lameli, 2010: Creating digital editions of historical maps. In: Lameli, Kehrein & Rabanus. Pp. 375–385.
- Šafařík, Pavel Josef, 1842: *Slovanský zeměvid*. Prague.
- Thun, Harald, 2010: Pluridimensional Cartography. In: Lameli, Kehrein & Rabanus. Pp. 506–524.
- Wagener, Peter, 1985: Zur Dialektgeographie Südniedersachsens. In: *Niederdeutsches Jahrbuch* 108. Pp. 147–163
- Wattel, Evert & Pieter van Reenen, 2010: Probabilistic maps. In: Lameli, Kehrein & Rabanus. Pp. 495–505.
- Wikle, Thomas A. & Guy Bailey, 2010: Mapping North American English. In: Lameli, Kehrein & Rabanus. Pp. 253–268.