Discovering and mapping syntactic areas: old and new methods

Elvira Glaser and Gabriela Bart
Univ. Zurich, Institute of German studies
eglaser@ds.uzh.ch, gabriela.bart@ds.uzh.ch

Keywords: Dialect syntax, cartography, Swiss German

The starting point of the following exposition is the hypothesis that syntactic variation basically behaves as other linguistic variation in phonetics, morphology and lexis with respect to geographical distribution. On the basis of several traditional as well as recent investigations, some of which will be referred to in the following, it is possible to oppose to some authors¹ questioning the existence of specific syntactic rules in dialect grammar by considering syntactic peculiarities in dialect speech to be a general marker of spoken language. Thus, if dialect specific syntactic rules can be taken for granted, we can conclude that there is an areal distribution.²

As a consequence, we consider the representation of dialect-syntactical patterns on maps as crucial. Only maps allow for a visual perception of the feature distribution in the geographical space. From the beginning, dialectologists recognized this necessity and, as time went on, the cartographic methods concerning dialectological maps improved. According to Haas (2004) who discusses several types of German dialect atlases, visualization should follow the principles of accuracy and clarity.³ Against this backdrop, our paper deals with the cartographic representation of dialect syntactic variation and the methods of visualization in order to provide evidence for the clustering of syntactic variants in geographical space. We therefore compare several types of mapping techniques with respect to their visualisation of the spatial distribution, whenever possible using syntactic maps. At the end, we give a forward look on the application of geostatistical analyses in order to corroborate some hypotheses on the distribution of syntactical variants in space.⁴

In the beginning of German dialect geography, interpretative maps with isoglosses indicating the distribution of the predominant variants were the preferred type of visualization. As an example we take a syntactic map from Sperschneiders early work on Thuringian dialects (Sperschneider 1959):

² For a recent discussion on this issue cf. Bucheli Berger/Glaser/Seiler (forthcoming).
³ We will not go further into the general principles of cartographic representation, in turn referring to the seminal work of Bertin (1973).
⁴ These analyses were carried out within a collaborative project between linguists and geographers of the University of Zurich. The syntactic data was collected from 2000 to 2004 within the project “Dialect syntax of Swiss German dialects” (http://www.ds.uzh.ch/dialektsyntax/pro_beschrieb.html). A master student in GIScience (Pius Sibler, under the supervision of Robert Weibel) used this data in order to test adequate geostatistical methods for detecting relevant distributions.
The line divides a southern area with inflected conjunctions (in the second person singular) – the four graphic symbols show the occurrence of four different inflected conjunctions – from a northern area without inflected conjunctions, as in Standard German (1b):

1. wenns du mitgehst ...
   if.2SG you come.2SG-along
   ‘if you come along’

2. wenn du mitgehst ...
   if you come.2SG-along
   ‘if you come along’

On his map, Sperschneider also indicates some cases where a variant appears on the “wrong” side, the dash in the northern area and the circle in the southern area. The map can be considered quite an accurate visualisation even if we do not see all the measuring points where the investigation was carried out. This kind of maps highlights the areas as such. In general, the isogloss maps are an attempt to give an overall view on the predominant variants and they often suggest homogeneity. The irregular occurrences, though recorded, remain in the background. The principle of clarity dominates accuracy. In the present case, however, the exact specification of the inflected conjunctions occurring at a certain measuring point renders the map somehow uneven. The recording of the specific
conjunction, mostly wī ‘as’, wën ‘if’ and ob ‘whether’, however, makes the map quite accurate.

There was much critique against this type of maps especially because of the isoglosses suggesting an absolute dividing line between two variants and so the suggestion of homogeneity, which, of course, does not even apply to phonological variants and much less to the lexicon. The question was raised whether it is justified at all to assume linguistic areas if there is variation within.

As far as syntax is concerned, it is, however, not yet clear to which type of area formation it belongs. There are too few syntactical studies so far taking into consideration the geographical distribution. This is why it has not been possible to undertake a proper comparison with other linguistic levels.

A further attempt to cope with both principles of accuracy and clarity was the mapping technique elaborated by Rudolf Hotzenköcherle in the dialect atlas of Swiss German (SDS 1962-1998), after having discussed various earlier methods (Hotzenköcherle 1962: 139-142). The cartographers of the SDS followed the principle of accuracy by putting a symbol for every type of answer at the exact place of its occurrence. If there were several variants at a place, they consequently mapped several symbols. Isoglosses were not sketched in. However, in trying to find an adequate symbolization which helps to detect different areas of predominating variants and variation zones, they also followed the principle of clarity. The following map shows an example out of the few syntactic maps of the SDS:

MAP 2. Serialisation within the Swiss German verbal group in subordinate clauses (SDS III: 261)

The map shows serialisation differences within the sentence final verbal group in a subordinate clause (example in Standard German):
The obligatorily analytical expression of past tense in Swiss German by an auxiliary and the past participle of the lexical verb ‘to be’ allows at least two options, symbolized by a dash and circle, the ordering 2-1 and 1-2 respectively. As it is our intention to discuss the general lines of the mapping technique and not the construction as such, we will concentrate on the two main variants.\(^5\) Whereas a large area exhibits the inverted 2-1 serialisation (dash) with the auxiliary at the last place as in standard German, *gsii bii* (‘been have’, lit. ‘been am’), the 1-2 order (circle), *bii gsii* (‘have been’, lit. ‘am been’) appears especially in the Bernese and Valais region in the south-western part of Swiss German. Hotzenköcherle refrains from drawing an isogloss which would not be impossible – apart from the north-west and some eastern areas, cf. map 2 – as it is shown by the following illustration.

\(^5\) A further differentiation could be made within the straight word order variant considering the position of the nominal predicate *ein kleines Mädchen* ‘a little girl’ with respect to the verbal group as it is indicated on the SDS map by graphically varying the circle.
of Swiss German) and another compact one where the straight serialisation appears in between. But there are also small transition zones where the areas border on each other, and there are some places showing variation within the inverted 2-1 serialisation, above all in the south-east. Taking the map’s message as presented in the atlas, we can see an example of a quite clear geographical distribution of the two main variants. This distribution is in line with other examples of Swiss German variant distribution where we recurrently find a similar west-eastern division (Hotzenköcherle 1986). The dividing line varies, however, sometimes lying near the political border between the cantons Aargau and Zurich, sometimes lying farther to the west as in the case at hand. Summing up, this kind of syntactic map shows the distribution of syntactic variants in a quite reliable way, so that everybody interested in the syntactic construction can look for the co-occurrence of variants, draw isoglosses and even prepare a different categorization and symbolization out of the data presented.

From that time on, such point symbol maps are used in the overwhelming majority of dialect atlases in the German-speaking area. If there are not too many variants, these maps show clear distributions and they allow the identification of the exact place of occurrence of a variant. The SDS maps represent good examples of the point symbol technique. There is, however, a lot of other atlases using the same technique in a less sophisticated way, so that sometimes areas may be difficult to discover. In the following, we present some examples of syntactic maps taken from the Sprachatlas von Niederbayern, containing dialect data from Lower Bavaria.

MAP 4. Serialisation within the verbal group in subordinate clauses
(Eroms & Spannbauer-Pollmann 2006: 235)
The map shows serialisation differences in spontaneous speech within the sentence-final verbal group (present perfect) in a subordinate clause (represented here by the Standard German equivalent):

(3) ...weil er nicht hat arbeiten müssen

‘...because he did not have to work’

Instead of a black and white geographic background, the authors chose a coloured relief of the area concerned. Concerning the symbolization technique the major principles remained. The spontaneous answers are classified, coded by a symbol and put on the map at the exact place of elicitation. The authors chose three types of symbols, squares, dots and triangles with respect to the position of the auxiliary hat (‘has’) at the first, second or third element within the modal perfect construction (i.e. the ordering 1-3-2, 3-1-2 and 3-2-1). As the 3-1-2 variant symbolized by the dot is predominant everywhere in the region investigated and there is only a slight concentration of 3-2-1 order in the north-eastern margin, the map is not very expressive with respect to areal distribution.

Further differences (with respect to the lexical choice) are left out here being not relevant for the discussion at hand.
This smaller map in black and white without relief gives additional information on the spontaneous or suggested character of the predominant 3-1-2 variant. Again, the map does not show a relevant geographic distribution. It makes clear, however, that at most places with a different symbol on map 4 the 3-1-2-variant was accepted, too. The comparison of the two maps in the original atlas is, however, not easy, because they are placed on the front and on the back of the leaf, respectively.

The Sprachatlas von Niederbayern (Eroms & Spannbauer-Pollmann 2006) contains also – relatively few – maps with clearer feature distribution, as for example in the case of prepositional dative marking:

Map 6 shows the concentration of the spontaneous use and acceptance of the suggested dative marker α in the southern part of Lower Bavaria in the following sentence (represented by the Standard German equivalent):

(4) Gib es der Kathi
    Give it the.DAT Kathi
    ‘Give it to Kathy!’

**LEGEND:**

- Δ *gibs α der Kathi* – akzeptiert
- • *gibs α der Kathi* – spontan
- ● *gibs der Kathi* – akzeptiert

Map 6. Prepositional dative marking with a feminine proper noun
(Eroms & Spannbauer-Pollmann 2006: 235)
The map focuses on the occurrence of the prepositional construction \( \alpha da Kathi \) ‘to the.DAT
Kathi’, leaving out the predominant variant with the exception of the places where the
standard variant was accepted along with the prepositional construction. Thus, the map
shows a nice geographical distribution at the expense of completeness. Though, the
situation is explained in the commentary.

Judging by the great amount of syntactic maps in the *Sprachatlas von Niederbayern*
(Eroms & Spannbauer-Pollmann 2006), one can get the impression that, for the most part,
syntactic variants are dispersed all over the region and do not show any areal clustering
within the region under investigation. On the background of comparable projects with
opposite results, the reason for this divergence is not clear. It could be an effect of the
homogeneity of Middle Bavarian in this respect, so that the areas at issue are larger than the
area investigated or an outcome of the mapping technique.

As far as we see, it has only been after the year 2000 that instead of black and white
symbol maps coloured symbol maps came up. As for syntax, the maps of the *Syntactic Atlas
of the Dutch Dialects* (SAND 2005/2008) provide a good example.

![MAP 7. Infinitival complementiser in purposive clauses in Dutch dialects (SAND 2005: 18)](image)

The map shows the geographical distribution of the two variants of the complementiser
introducing infinitival purposive clauses: *om* and *voor* ‘in order to’. The mapping technique
relies on the colour and the relative position of the squares given in the legend together
with the total number of the attestations. The distribution of the dark green squares shows
the distribution of the southern complementiser type *voor* which is in most areas used
together with the type *om*, whereas *om* is the only variant in most places especially in the
northern part of the Dutch area. Obviously, this visualisation technique also allows the
perception of syntactic areas, that means, an area with predominant use of *om*, several
small areas with predominant or even exclusive use of *voor* – as in the west – and a larger
southern area with variation of the two types. Just as Hotzenköcherle (SDS III) and Eroms & Spannbauer-Pollmann (2006), the authors of the SAND do not use isoglosses either. Yet with two variants, the syntactic distribution is easy to recognize. If there are more than two variants at a location, the perception of the areal distribution sometimes turns out to be much more difficult, especially because of the problems of recognizing the difference between the various colours (cf. Hoekstra 2007).

MAP 8. Strong forms of subject pronouns 3PL in Dutch dialects (SAND 2005: 47)

In the Zurich dialect syntax project (SADS) we have used similar techniques as the authors of the SDS and the SAND in order to visually represent our Swiss German syntactic data. We use black and white symbols on relief maps, sometimes together with coloured symbols or coloured squares as on the SAND maps illustrated above. As a rule, we do not integrate topics lacking a significant geographic distribution of the variants. On the other hand, we focus on visualizing the extant geographic distribution of syntactic variants, leaving out distracting ones, if necessary. Of course, all these decisions are explained in the commentary. Unlike the other atlases discussed so far, for the most part, we integrate information on quantity into our maps. It is, however, a simple type of quantity indication, distinguishing between variants occurring only once at a place and variants occurring more than once, as is demonstrated by the following maps.

---

7 Cf. on further details Bucheli Berger (2008).
MAP 9. Infinitival complementiser in purposive clauses in Swiss German dialects (SADS)

This map shows the variants concerning a syntactic variable similar to the one in the SAND discussed above, i.e. the two complementisers zum and für ‘in order to’ introducing a purposive infinitival clause, equivalent to the Standard German complementiser um ... zu:

(5) um ein Billett zu lösen
in-order a ticket to buy
‘in order to buy a ticket’

There are two symbols for the two variants, a dot and a dash. We use smaller and larger symbols in order to differentiate between singularly occurring answers and multiple answers. Yet the corresponding form of the smaller and larger symbols allows perceiving at least two areas on the map: a western area where the dot is the only variant and a large zone with both symbols in the east. If we have a closer look at the north-east of Swiss German we note that there are less and less locations showing the für-variant (dot) further to the east, and in addition the für-variant mostly occurs as a singular answer (small dot).
MAP 10. Infinitival complementiser in purposive clauses in north-east Swiss German (section of map 9)

The continuous decrease of the für-variant in the east appears perhaps even better on a coloured symbol map including information on quantity:

MAP 11. Quantitative distribution of the infinitival complementiser für in Swiss German (SADS)
The red dots decrease and get smaller (mentioned once at a place) further to the east, and the dark coloured dots, symbolising the majority at a location, increase to the west. This situation made Guido Seiler (2005) suggest the model of an inclined plane instead of clear cut syntactic areas. This model is backed up by the comparison of the use of the für-variant in different syntactic contexts, as illustrated on the next map.

MAP 12. Complementiser für in two types of infinitival purposive clauses

The map shows the use of für in an infinitival clause containing a direct object (red dot: ‘to buy a ticket’) and in an infinitival clause containing only the verb (black triangle: ‘to get asleep’). For both contexts it holds true that the farther to the west the use of für increases. If we compare the two clauses we can also say that the farther to the west the more we find für in different syntactic contexts.

To sum up, we have seen that the notion of a syntactic area is not a straightforward one. Problems appear with the concept of an area as a clearly defined or homogeneous space. From the perspective of a single variant, we can define it as the range of a variant, no matter whether there are also other variants around or whether the one at issue is the predominant one. Thus, there can be overlapping areas with an even distribution of several variants along with areas of a predominant or sole variant. With respect to the variation in an area defined in such a way, it is obvious that a quantitative analysis of the data will be interesting. As the number of our informants vary strongly at a place so that at some places the absolute number can be quite low (< 5) we refrain from giving relative values at single places.
Nevertheless, a quantitative overall treatment of the data can be helpful. This will perhaps enhance getting a more general idea of how syntactic variants are distributed and comparing their distribution in a quantitative way. Thus, dialectometrical methods are concerned as developed in the last decades for example by the teams of Goebl and Nerbonne (cf. Goebl 2006 and Nerbonne 2010). Their research focuses on the aggregation of different variants, whereas recently a team from Augsburg and Ulm started work on variant-based dialectometry (cf. Rumpf/Pickl/Elspass/König/Schmidt 2009, 2010), an approach compatible with our definition of areas as the range of a variant. Yet up to now, their work has been focussed on lexical variation. They take the predominant variants at a location and use kernel density estimation (KDE) as a statistical method to get from the individual locations to coloured area class maps as the following one:

MAP 13. Thiessen polygons of predominant lexical variants for ‘dry loppings’ in the south-west of Bavaria (Rumpf et al. 2009: 307)

Within the framework of collaboration between the GIScience Center of the Department of Geography and the German Department of the University of Zurich, Sibler (2011) applied the variant based approach on our dialect syntactic data. The method is based on the estimation of a value for the locations without a documented form on the basis of the surrounding values, an approach which is also perfectly applicable to syntactic variants as nominal data. Up to now, the rare dialectometrical work on dialect syntax was focussed on the aggregation of distance measures, so that distance maps were created (Spruit 2008),8 showing areas

---

8 Finding association rules between syntactic variables is another topic in Spruit (2008).
based on syntactic differences. This kind of work has to tackle the problem of measuring the differences between syntactic variants.

In the following, the cartographic results of the application of KDE on Swiss German syntactic data are presented. The maps 14 to 16 correspond to the point symbol maps on purposive clause linkage, shown above (maps 9-12). Map 14 represents an interpolated colour map of the für-variant, map 15 of the zum-variant, and map 16 is a combination map showing the intensity of the respective dominant variant (based on polygons of measuring points).

MAP 14. Interpolated intensity of the für-complementiser in infinitival purposive clauses (based on the data of map 9) (Sibler 2011)

MAP 15. Interpolated intensity of the zum-complementiser in infinitival purposive clauses (based on the data of map 9) (Sibler 2011)
MAP 16. Interpolated combination map of the *zum-* and *für*-complementiser in infinitival purposive clauses (based on the data of map 9) (Sibler 2011)

Map 16 seems to underline the idea of two syntactic areas, because the coexisting variants are not presented besides the predominant variants. Their existence is only to be inferred by the intensity of the colour.

Sibler (2011) applied KDE to various phenomena: In addition to several types of purposive clauses, we decided together to pick out the linkage of the standard of comparison in comparative clauses (6) and the position of the indefinite article in adjective phrases (7) because of different linguistic hypotheses on their distribution. The examples are given in Standard German:

(6) Sie ist grösser als ich
    ‘She is bigger than me’

(7) Susi wäre eine ganz liebe Frau für Markus
    ‘Susy would be a very nice woman for Mark’

Map 17 shows a phenomenon (complementizer in comparative clauses) with four variants as linking element, one – *als* – dominating over the whole area, the others – *wan, wie* and *weder* – prevailing only in very restricted zones, so that KDE interpolation sweeps them away. Such blended maps obviously do not reflect our intuition about existing variation. They do not correspond to the principle of accuracy mentioned above. Nevertheless they
reflect quantitative relations in a reliable way and thus are suitable for questions of a more abstract character and, of course, for the determination of the quantitative situation. Map 18 shows three variants with respect to the position of the indefinite article together with an adjective: preponed, as in Standard German, postponed, as e.g. ganz å liebi Frau ‘very a nice woman’ – and doubled, as e.g. å ganz å liebi Frau ‘a very a nice woman’. Previous work (Steiner 2005; Steiner forthcoming) had provided evidence for the claim that the two main variants, the postponed article and the doubled article, largely do not show a relevant geographic distribution. It is quite astonishing that the interpolation reveals a kind of core area of the doubling variant reaching diagonally (NW to SE) from Bâle to Grisons. In this case, the dialectometrical approach helps finding a syntactic area where the doubling of the indefinite article is quite common. Conversely, the traditional point symbol maps present a rather chaotic distribution which is difficult being interpreted.

The first attempts to use dialectometrical methods for the visualisation and interpretation of our Swiss German syntactic data make us believe that it could be a good idea to integrate these methods in future dialect syntactic research along with point symbol maps giving an accurate picture of the variants’ occurrence. It is, however, necessary to evaluate the usefulness of interpolated area class maps in every single case. In a next step, the area class maps could be characterised with respect to structural features such as complexity, compactness and homogeneity (cf. Rumpf et al. 2009). The quantitative evaluation of these structural features must be compared to the dialectologist’s intuition about complex or homogeneous maps based on the visual perception and cognition. Thus, these methods can help evaluating large amounts of maps, which a dialectologist would not be able to memorize and compare. Comparing the created syntactic area class maps to the maps of Rumpf et al., we cannot yet draw any conclusion, since the base of our calculation is still too small. It is, however, interesting that among our small sample there is one map (comparative linkage) seemingly completely homogeneous after the interpolation procedure, whereas among Rumpf et al.’s more than 80 maps no one shows such behaviour.

Besides creating area class maps, the quantitative treatment of our syntactic data implied applying further geostatistical methods which we will discuss elsewhere (Bart et al. in preparation).

The syntactic phenomena in question behave differently with respect to areal distribution. Up to now, we can at least distinguish between three different types: variants with large zones of complementary areal distribution, variants without any significant distribution, and variants forming hot spots within a larger area with a dominating variant. This distinction, already resulting from the informal comparison of the point symbol maps, is thus corroborated statistically.

9 For an evaluation of the different types of maps with respect to the demands of an atlas, cf. Bucheli Berger/Glaser/Seiler (forthcoming).
10 For a reflection on the future benefit of GIS methods in dialectology cf. the more general considerations in Hoch & Hayes (2010).
MAP 17. Linking element in comparative constructions in Swiss German (Sibler 2011): interpolation of predominant variants

MAP 18. Position of the indefinite article in an adjective phrase in Swiss German (Sibler 2011): interpolation of predominant variants
References


Steiner, Janine (2005) “Also d’Susi wär e ganz e liebi Frau für de Markus! Zur Verdoppelung des indefiniten Artikels in der adverbiell erweiterten Nominalphrase im Schweizerdeutschen”. In Helen Christen (ed.) Dialektologie an der Jahrtausendwende. *Linguistik online* 24, 3. On-line: [http://www.linguistik-online.de/24_05/steiner.html](http://www.linguistik-online.de/24_05/steiner.html) (last checked on 21/3/2012)