

Language and Space

An International Handbook of
Linguistic Variation

Volume 2: Language Mapping

Edited by

Alfred Lameli

Roland Kehrein

Stefan Rabanus

Offprint

Handbücher zur Sprach- und Kommunikationswissenschaft [HSK] 30.2.

Berlin: De Gruyter Mouton, 2010 (pp. 433-457, 2201-2212)

De Gruyter Mouton

Nile, Richard and Christian Clerk

1996 *Cultural Atlas of Australia, New Zealand, and the South Pacific*. New York: Facts on File.

Peterson, Michael P.

1993 Interactive cartographic animation. *Cartography and Geographic Information Systems* 20: 40–44.

Thrower, Norman J. W.

1961 Animated cartography in the United States. In: Eduard Imhof (ed.), *Internationales Jahrbuch für Kartographie*, 20–30. Gütersloh: Bertelsmann.

Vath, Nuria, Marcus Hasselhorn and Gerd Lüer

2001 *Multimedia-Produkte für das Internet – Psychologische Gestaltungsgrundlagen*. Munich: Oldenbourg.

Yattaw, Nancy J.

1999 Conceptualizing space and time: A classification of geographic movement. *Cartography and Geographic Information Systems* 26: 85–98.

Zedi, Roger

2000 Interaktive und animierte 3D-Szenen in VRML. In: Gerd Buziek, Doris Dransch and Wolf-Dieter Rase (eds.), *Dynamische Visualisierung. Grundlagen und Anwendungsbeispiele für kartographische Animationen*, 83–96. Berlin/Heidelberg/New York: Springer.

Christian Hanewinkel, Leipzig (Germany)

Eric Losang, Leipzig (Germany)

22. Dialectometry and quantitative mapping

1. Introduction
2. From the atlas data to the data matrix
3. From the data matrix to the similarity and distance matrices
4. Similarity maps: Computation, cartographic description and geolinguistic meaning
5. The parameter maps
6. The interpoint maps: Honeycomb and beam maps
7. Correlative dialectometry
8. Dendrographic dialectometry
9. Review
10. Frequently used abbreviations
11. Atlases
12. References

1. Introduction

Dialectometry, as it is introduced in this contribution, is a global and also quantitative response to a problem which arose more than one hundred years ago, with the mapping of single linguistic features. Today, it is widely recognized that the prominent German and French linguistic atlases (*DSA* and *ALF*), both of which have set standards in geolinguistics, do not map complete languages, but rather single linguistic features. Further-

more, the interpretation of the mapping in the *DSA* and the *ALF* has clearly shown that the spatial profiles of the different linguistic features diverge markedly from feature to feature with respect to their areas and the surrounding boundary lines (generally called *isoglosses*). Many linguists wondered why this also happened in those instances in which the etymological cognacy of two linguistic features led one to expect almost identical areal or isoglottic structures. In German linguistics, this phenomenon, crystallized in the so called “Rhenish fan” (*Rheinischer Fächer*), has even become emblematic: see Putschke (2001). Strangely enough, such an emblematic crystallization did not occur in Romance linguistics, although the *ALF* data suggest an analogue in the Rhône Valley (located between Lyon and Marseille).

Traditional geolinguistics has coped with this notorious non-coincidence of single areas and their surrounding isoglosses in a rather one-sided manner:

- The heuristic-cartographic response: if possible, the diverging isoglosses were superposed on the maps. As a result, isogloss bundles of various thickness would emerge, representing the basis for spatial dialect classifications. But, as a matter of fact, German and Romance linguists have rarely applied this technique of graphic superposition or cumulation to the corresponding *areas*. [The cartographic result of the superposition of areas are the so-called “density maps” (*Dichtekarten, cartes de synthèse*). There are but very few specimens of such density maps. Recently, this technique has been applied to a larger extent to the data of the *ALF*: see Brun-Trigaud, Le Berre and Le Dù 2005 (passim).]
- The quantitative-statistical response: unfortunately, as a consequence of the relatively small number of areal superpositions, quantitative questions related to cumulative areal classifications have not been treated seriously. Hence, heuristic-cartographic, quantitative-statistical and systematic-classificatory aspects have not been questioned. In view of this methodical shortfall in *areal* dialectal cartography, Carl Haag’s theoretical fundamentals of *isoglottic* dialectal cartography of 1898 are highly valuable.

Historically speaking, a scientific retrospective reveals (given the complex spatial structure of linguistic atlas mappings) that German and Romance linguists’ endeavors concentrated on concrete explanations or “excuses” for the “genuinely unnatural” non-coincidence of determinate isoglosses. These external interpretations drew on political, demographic, economic and communicative influences on the tensions or fluctuations in space, which determined how the diverse *isoglosses*, or the shape of the *areas* they surrounded, had developed in space. Even if this interpretation did not eliminate the fundamental error concerning the “theoretical” assumption about the bundling of isoglosses, it created a critical consciousness of the relevant internal and external tensions and relations in space amongst geolinguists of all minds. With this new sensibility for space, the interdisciplinary dialogue with geographers and historians could begin, as well as with those scholars and scientists in the humanities and life sciences who dealt with space-related data: see the whole range of interdisciplinary work done in the period between 1926 (Aubin, Frings and Müller) and 2005 (Scapoli et al.). Unfortunately, quantitative-statistical problems in connection with the synthesis of geolinguistic data were hardly ever discussed in German and Romance linguistics. The discussion restricted itself to some timid approaches in Germany (see Rosenkranz 1938) and France (see Lalanne 1953). In France in the decades before and after the publication of the *ALF* (1902–1910), the almost absurd postulate (from a modern point of view) arose that it was not

“scientific” to pay attention to the actual existence of “dialects” in space, as the noblest task was the scientific search for the geographic distribution of *single* linguistic features and their description. [Nonetheless, the general meaning of *dialect* was always related to a space-based entity of typological relevance.] This extremely *atomistic* and moreover *anti-systematic* doctrine (see Paris 1888) was based on the almost religious belief that France was as a matter of principle indivisible and could therefore not be subdivided into different dialect regions. In addition, this doctrine represented an enormous obstacle in French geolinguistics to any further development of dialect classification and dialect data synthesis. Until 1950, not a single isogloss synthesis of *ALF* data was produced by French researchers.

Nevertheless (and curiously enough), the concept of *dialectométrie* developed in a French research environment, namely in the context of the regional linguistic atlas of Gascoigne (*ALG*; see Séguy 1973: 1). In Séguy’s lifetime (1914–1973), the Gallo-Romance dialectal variety had already been registered and described not only in the *ALF*, but also in other regional atlases, empirically confirming the dilemma mentioned above. Séguy’s historical merit consisted in his radical methodical *volte-face*: whereas previous *qualitative* methods had been unable to overcome the apparently uncontrollable chaos of the data from the linguistic atlas, Séguy began to explore data *quantitatively*. Unfortunately, his premature death in 1973 put an end to his auspicious activities, so that his actual contributions were confined to only a promising name (*dialectométrie*) and a few tentative statistical and quantitative-cartographic approaches. [In this context, one should not forget the dialectometrical work of Henri Guiter (1909–1994): see, above all, his ground-breaking paper of 1973.]

By the early seventies though, the following technical and methodical resources had become available:

- an extremely elaborate numerical classification based on research in disciplines such as biology, psychology and economics, etc. (see the handbooks of Sneath and Sokal 1973 [in English], Bock 1974 [in German] and Chandon and Pinson 1981 [in French]);
- the renewed interest of human geography for regional differences which began to attune itself to quantitative approaches (see Haggett 1973);
- an increasingly efficient and accessible information technology, which managed to liberate the actual numerical and cartographic problems and challenges from the constraints of manual work (see the reader published by W. Putschke in 1977).

Thanks to the propitious research climate at Regensburg University in the 1970s, these three opportunities were used extensively to gain a satisfactory numerical and cartographic global interpretation of linguistic atlas data (see Goebel 1981, 1982 and 1984, which are representative for this period).

The developments in dialectometry in Regensburg and those fostered in Salzburg since 1982 will be the subject of this contribution. It must also be mentioned that – in addition to the “Salzburg school of dialectometry” (SDM) – dialectometrical research has diversified considerably in the last twenty years, mainly in the field of German (cf. Hummel 1993; Schiltz 1997) and Romance linguistics (see *inter alia* Melis, Verlinde and Derynck 1989; Verlinde 1988; Viaplana 1999). Nowadays there are two further dialectometrical research centers worth mentioning: the “Groningen school” (see Heeringa 2004; Heeringa and Nerbonne 2001; Nerbonne and Heeringa 2001), and the “Athens [Georgia, USA] school” (see Nerbonne and Kretzschmar 2003, 2006; Kretzschmar and Schneider 1996).

The wide variety of activities of the Salzburg school can be characterized in its theoretical and methodical principles as follows:

(a) Theoretical prerequisites of the linguistic atlas work

- The effects of the dialectal and basilectal management of geographic space by *Homo loquens* that is executed according to determinant communicative, social and other similar principles.
- The assumption that linguistic atlases (in their traditional form) represent a reliable and effective basis for the investigation and documentation of this complex linguistic management of space by *Homo loquens*.
- The linguistic management of space by humans can be analyzed through both *direct* (i.e., by considering individual linguistic atlas maps) and *indirect* (i.e., by statistical analysis of a large number of linguistic atlas maps) observation.
- The discovery that underlying deep structures of the linguistic management of space by humans show well-formed spatial patterns: they are actually the spatial counterpart to the famous “sound laws” (*Lautgesetze*) discovered by the Neogrammarians at the end of the nineteenth century and which developed along the diachronic axis.

(b) Methodical principles

- In continuation of the tradition of *qualitative* geolinguistics via *quantitative* means, the Salzburg school attaches great importance to up-to-date visualization and the continuous mapping of quantitative results.
- In essence, the Salzburg school practices a compound discipline: it makes use of selected methods of descriptive statistics, numerical classification, and computational cartography, with the objective of promoting state-of-the-art Romance geolinguistics.
- Despite its genuinely *geolinguistic* anchorage, the Salzburg school is always open to *transdisciplinary cooperation*.

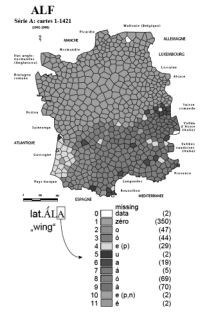
All the maps presented and discussed in this contribution are drawn from a substantial research project based on the French linguistic atlas *ALF* (and financed through the Austrian Research Fund FWF [*Fonds zur Förderung der wissenschaftlichen Forschung in Österreich*] via two grants: no. 12 414 and 13 349), in which almost half of its original data were classified (or *taxated*) into types and then subjected to dialectometrical analysis (see Goebel 2000, 2002 and 2003 for the scientific results of this project). The bibliographical notes give further references for the dialectometrical processing, “dialectometrization”, of other Romance (and also non-Romance) linguistic atlases. It should be mentioned that, parallel to this research project (1997–2000), our former project assistant Edgar Haimerl created a special purpose-built software for the Salzburg school called “Visual DialectoMetry” (VDM), which has since become an important element of our effective dialectometrical work. All of the color maps referred to in this chapter were generated with the help of VDM (boosted by the geographical software MapInfo). Incidentally, VDM is available free to all interested researchers (further information is available from <<http://ald.sbg.ac.at>>).

2. From the atlas data to the data matrix

In a first step, all the data from a given linguistic atlas must be converted into machine-readable form. At present, two methods are available:

- The Groningen approach: the original transcriptions from a particular linguistic atlas are rendered into machine-readable form, then the quantitative differences between the respective lococolects (or “localities”) are automatically measured, pairwise (e.g., by applying the “Levenshtein algorithm”).
- The Salzburg approach: the original data from a given linguistic atlas are analyzed (viz. typized or taxated via “taxatation”) by trained geolinguists according to traditional linguistic criteria (phonetics, morphology, vocabulary, etc.). This results in a great number of disjunctive (qualitative) units called *taxates*, each associated with a well-defined geographical area. In a further step, the quantitative similarities between the different localities are computed using similarity indexes suitable for processing qualitative data.

Taxatation captures the quality of the linguistic attributes on a nominal measurement scale, see Figure 22.1. In German and Romance linguistics, the principle of taxating qualitative features in diverse linguistic atlas maps (always according to given linguistic criteria) has a long and rich tradition that is continued by the Salzburg School: compare the examples given in Map 2201 (*vocalic taxatation*) and Map 2202 (*lexical taxatation*) in the separate maps volume.



Map 2201

Map 2201 was generated on the basis of *ALF* Map 18 *l'aile* ‘the wing’. It shows eleven different outputs for final Latin -A, all derived from the Latin etymon *ÁLA*. This etymon occurs at all 641 *ALF* points. The most frequent result (at 350 *ALF* points) is the item “zéro” (i.e., final -A van-

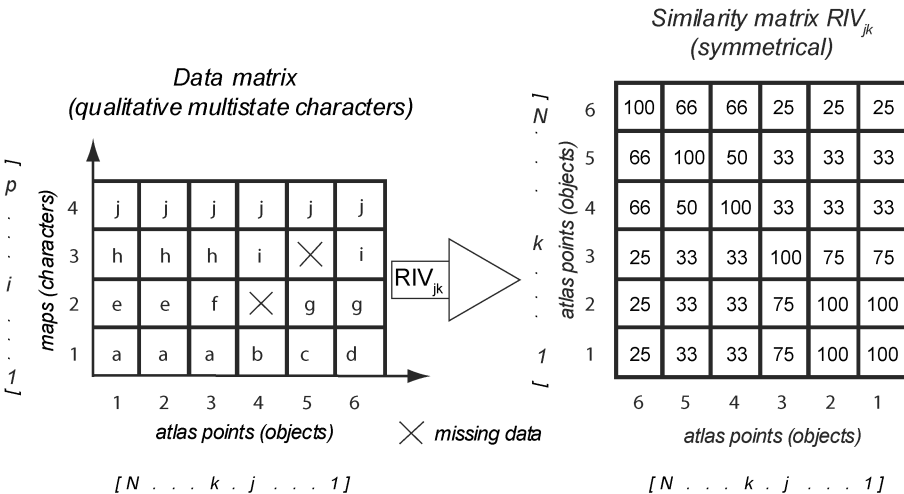
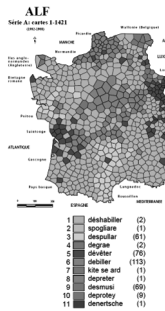


Fig. 22.1: From the data matrix to the similarity matrix: Schematic representation of the measurement and processing of linguistic atlas data

ishes completely). The other outputs are: *-à* (at 70 *ALF* points), *-ò* (at 69 *ALF* points), etc.: see the legend to Map 2201.



Map 2202

The geographic shape of Map 2201 clearly shows the north–south division of Gallo-Romania. It suggests that in more recent times the northern taxate 1 (zero) has dominated at the expense of other vocalic variants. By including diachronic information derived from the different parts of Gallo-Romania, the linguistic interpretation of the map profile can be further refined and extended.

Map 2202 (with a completely different choropleth profile) shows, on the basis of *ALF* Map 394 *déshabiller* ‘to undress’, the spatial distribution of the Gallo-Romance denominations for the concept. The dominant spatial distribution (at 371 *ALF* points) of the omnipresent taxate 1 (*déshabiller*, the standard French term) provides evidence that the French form was able to prevail over a great number of regional variants across almost the entire Gallo-Romance area. In addition, Map 2202 shows the old homelands of the supplanted regional terms to be nevertheless still recognizable.

Similar mappings can actually be derived from any existing linguistic atlas and then cast in a data matrix. In the symbolic language of *numerical taxonomy* (also called *numerical classification* or *taxometry*) this kind of data matrix consists of N objects or elements (here, *locolects*, *localities*, *inquiry points*, or their vectors in the matrix) and p attributes or features (here, *atlas maps*, or the *working maps* derived from the original atlas maps via taxation).

For both statistical and heuristic reasons, the number of *locolects* (N) to be analyzed and the number of taxed working maps (p) to be synthesized should be as high as possible (several hundred). A comprehensive – theoretical and technical – description of the different steps in the dialectometrization of a linguistic atlas has been given in an earlier publication (Goebel 1984).

Moreover, it is of utmost importance for any geolinguist undertaking the dialectometrization of a linguistic atlas to keep in mind the following facts and their consequences for further research. Through the global synthesis and subsequent analysis of mass data, he/she moves from the level of the *particular* to the *general*, which is an *inductive* approach. As a consequence, the final results of a dialectometrical synthesis have a different epistemological status to the many individual maps fed into the data synthesis (see, for example, Maps 2201 and 2202). Experience shows that this epistemological metamorphosis has not always been adequately interpreted by linguists.

3. From the data matrix to the similarity and distance matrices

Once the data matrix is established, the next procedure can be compared with the processing of a complex signal, whose genuine content or meaning is still to be decoded. Therefore, it is factually correct to consider the sequence of dialectometrical methods as a flow chart (see Figure 22.2).

Using the global and inductive approach, one can either compare, in pairs, the N atlas point vectors or p working map vectors with the data matrix. As traditional geolinguistics aims to determine the similarities (or cognacies) between atlas points, only the

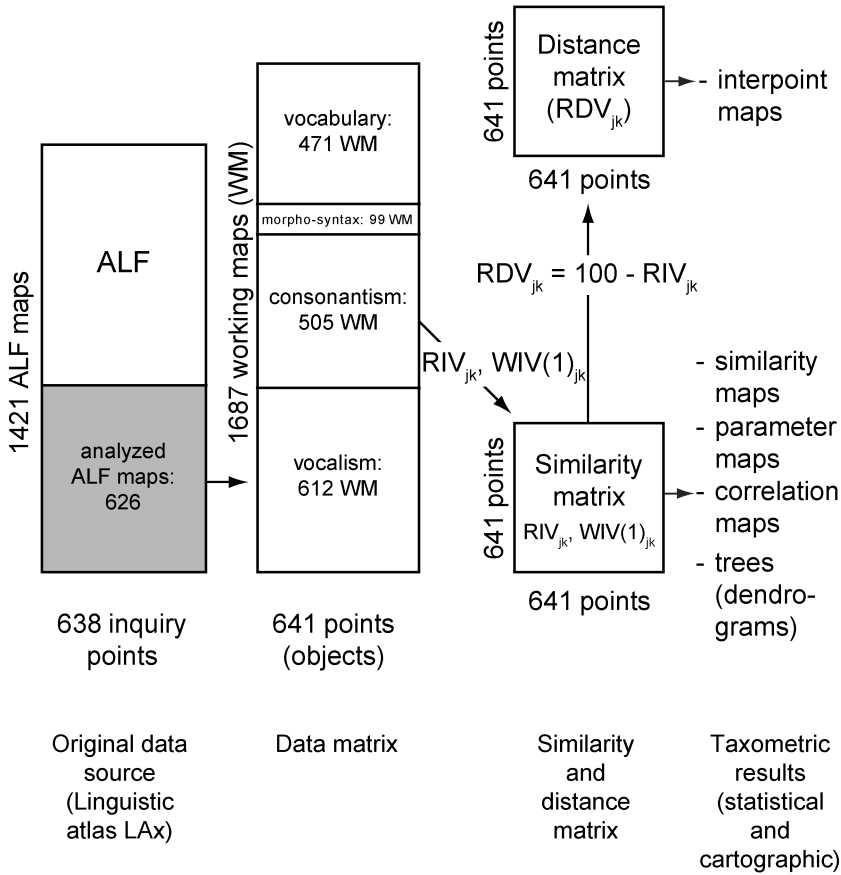


Fig. 22.2: Flow chart of the dialectometrical methods used by the Salzburg school of dialectometry

first of these two possibilities will be adopted by the dialectometrician. For the pairwise comparison of object vectors, modern numerical taxonomy offers us many similarity measures (also known as *similarity indexes* or *metrics*). The selection of an appropriate measure therefore depends on the dialectometrician’s theoretical ideas about interdialectal similarity. In this instance, the “Relative Identity Value” (RIV_{jk}) has proven extremely successful. The RIV_{jk} is a similarity index which is calculated from the number of pairwise matchings (also called co-identities, COIs) of taxates and the number of pairwise mis-matchings (co-differences, CODs). The RIV_{jk} values oscillate between zero and one hundred percent, according to the following formula (the meaning of the symbols is explained in Table 22.1):

$$RIV_{jk} = 100 \frac{\sum COI(i)_{jk}}{\sum COI(i)_{jk} + \sum COD(i)_{jk}}$$

In principle, it is possible to use different similarity measures: actual dialectometrical practice has demonstrated the proficiency of similarity indexes other than RIV_{jk}, namely

Tab. 22.1: Key to the symbols in the RIV_{jk} formula

Symbol	Meaning
$COD(i)_{jk}$	co-difference between two taxates (on the map i and for the atlas points j and k)
$COI(i)_{jk}$	co-identity between two taxates (on the map i and for the atlas points j and k)
i	one of p atlas or working maps
j	reference atlas point
k	atlas point to be compared with the reference point j
p	total number of maps in the data matrix
N	total number of atlas points in the data matrix
RIV_{jk}	Relative Identity Value (between the attribute vectors of the atlas points j and k)

all those indexes which weigh the pairwise taxate identities (co-identities) according to the extension of the areas of the corresponding taxates.

In the whole procedure, the “usefulness” of a particular similarity index is crucial for the geolinguist’s choice and for his/her further typo-diagnostic work. Normally, the relevant selection criteria should always be linguistic, and not borrowed from other sciences. The geolinguistic usefulness of a similarity measure reveals itself only after the measurements taken have been mapped; that is, after the *visual* inspection of the shapes of the respective quantitative maps. In fact, a similar dilemma also occurs in all disciplines that seek to uncover and diagnostically analyze structural patterns in their data that are invisible at first sight.

Between geolinguistic similarity (*sim*) and geolinguistic distance (*dist*), the following relationship is assumed: $sim + dist = 100$. After calculating the square ($N \times N$) similarity matrix and deriving the appropriate distance matrix (see the formula above), the dialect-

Tab. 22.2: Key features of the *ALF* dialectometrization used here

Number of <i>ALF</i> maps analyzed		626 (44 % of the 1,421 original <i>ALF</i> maps)		
<u>Number of derived working maps</u>				
total corpus	<i>phonetically</i> relevant	<i>lexically</i> relevant	<i>morphosyntactically</i> relevant	
1,687	1,117	471	99	
<u>Number of taxates (areas) analyzed^a</u>				
total corpus	<i>phonetics</i>	<i>vocabulary</i>	<i>morphosyntax</i>	
20,043	12,023	7,031	989	
Number of <i>ALF</i> inquiry points analyzed		638 + 3 artificial points (see section 4.2.)		

^a Note that, in the final instance, the spatial complexity of the interwovenness of such a great number of discrete units (or areas) has provoked most of the misunderstandings related to the rare coincidence of isoglosses that have arisen since the end of the nineteenth century. Also note that to each taxate belongs a more-or-less extended area which, theoretically, can vary between 1 and 641 (= N) atlas points.

tometrician has to visualize and to map the quantitative space-related information from the two matrices in a (more or less) comprehensive way, in line with the research objectives.

The following heuristic tools have proven their usefulness over the past few decades: similarity maps, parameter maps, interpoint maps, correlation maps, and trees. (This list could of course be expanded). All these mappings were implemented with the VDM software. They can be directly computed and visualized in variable forms on the screen. Table 22.2 shows the key features of the *ALF* dialectometrization used in the examples in this article.

4. Similarity maps: Computation, cartographic description and geolinguistic meaning

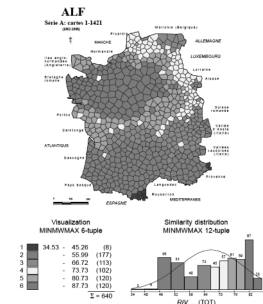
4.1. Computation of the similarity maps

Each of the N atlas point vectors of a similarity matrix contains N similarity values. Along the diagonal of this similarity matrix, the values all equal 100 percent (i.e., 1), with the subsequent reciprocal symmetry of the two halves of any similarity matrix: see Figure 22.1 (right) and Figure 22.3 (in section 7; middle).

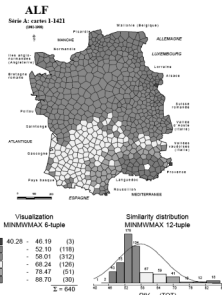
The N atlas point vectors can now be mapped separately. Their $N - 1$ measurement values oscillate between zero and one hundred percent. One measurement value (for the reflexive measurement RIV_{jj}) is always equal to one hundred percent. Customarily, this reflexive value is not colored on the maps and remains white. All the other $N - 1$ values are mapped. The cartographic challenge is how to represent the *numerical* variation of the $N - 1$ measurement scores with a corresponding variation of *colors* (usually with four to eight different colors). This iconic representation must always take into account the physiological capabilities of the human eye. The transfer from (quantitative) *numerical* into (quantitative) *iconic* information is extremely important in dialectometry (and many other pattern-oriented sciences as well). In dialectometry, it has always been assumed that this transfer is *algorithmic* (and therefore not authoritative) and *variable* (meaning that the selection and number of color tones is free). In order to realize these transfers, all available quantitative cartographic experience (see Dickinson 1973) was incorporated into VDM.

4.2. The configuration of the choropleth maps

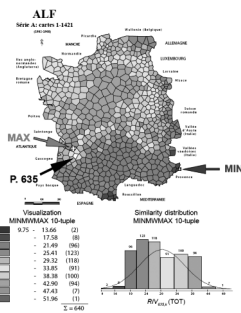
Cartographically, Maps 2203 to 2206 belong to the choropleth class of maps. In this case, the base map has 641 inquiry points (or localities), consisting of the original 638 inquiry points of the *ALF*, plus three artificial points (corresponding to the standard languages: French = P. 999, Italian = P. 998, and Catalan = P. 997). The insertion of these artificial points is of special geolinguistic interest and relevance, as it reveals the influence of the sociolinguistically higher-ranking standard languages on the different traditional lococlects.



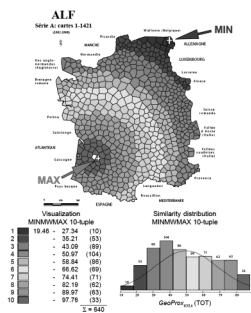
Map 2203



Map 2204



Map 2205



Map 2206

Next, the basic grid of 641 inquiry points is triangulated and then polygonized in line with Delaunay-Voronoi geometric principles (see Okabe, Boots and Sugihara 1992: *passim*). The great advantage of the resulting tessellation structure is a regular partitioning amongst the 641 inquiry points across the whole area. It also allows simultaneous *visual* comparison of a great number of choropleth maps.

The 640 measurement values of less than one hundred percent are visualized on the choropleth maps for this article using six or ten color intervals via two interval algorithms, known within the research project as “MINMWMAX” and “MEDMW”. For both, the measurement values are equally spread above and below the arithmetic mean (MW, from Ger. *Mittelwert* ‘arithmetic mean, average’).

The algorithm MINMWMAX divides by three the range of the measurement scores between the minimal value (MIN) and the arithmetic mean (MW), or, alternatively, between the maximal value (MAX) and the mean (see Maps 2203–2206). As for the interval algorithm, MEDMW, those measurement values below the arithmetic mean (or their polygons) have to be split up into subsets as evenly as possible: on Maps 2207, 2208 and 2211 this results in five groups. The same procedure is applied to establish the number of measurement values (and hence polygons) lying between the arithmetic mean and the maximum. The main purpose of the interval algorithm MEDMW is to generate highly suggestive and clear-cut choropleth profiles, whereas the interval algorithm MINMWMAX generates more evenly structured patterns.

The quantitative maps (Maps 2203 to 2211) all consist of three parts (plus the caption):

- the map itself (choropleth: Maps 2203–2208, 2211–2212; isarithmic: Maps 2209 and 2210);
- a numerical legend (bottom left); and
- a histogram (bottom right).

The numerical legend details the division into six or ten color steps and the quantitative thresholds used. The corresponding arithmetic mean is the upper threshold of class 3 (on Maps 2203 and 2204) or of class 5 (on Maps 2205 and 2206). The histograms (at the lower right) convey information (relevant for the statistician rather than the geolinguist) on the statistical nature of the underlying frequency distribution of the map. The number of bars on the histogram corresponds exactly to the number of color steps in the numerical legend. For statistical comparisons, the silhouette of the theoretical bell curve is once again superimposed. This bell curve is calculated from the characteristic statistical parameters of the empirical similarity distribution; for further discussion cf. Goebel (1984: I, 87–98).

4.3. The geolinguistic meaning of similarity maps

In geolinguistics, similarity maps can be analyzed in many different ways. In the first instance, they show the position of a lococlect within the investigated atlas grid, which is interpreted as a relational framework. Map 2203 shows the (relational) position of the linguistic potential of a Norman lococlect (corresponding to *ALF* point 347, Dompierre, Département Orne), while Map 2204 demonstrates the same with a Provençal lococlect (corresponding to *ALF* point 885, Gréoux, Département Basses-Alpes).

Both maps show a progressive and regular decrease in dialectal similarities across space. The same phenomenon can be observed on all the remaining 639 similarity maps of the same similarity matrix. Thus, the similarity maps give evidence that, besides the well-known sound laws located on the *time* axis, there exist other linguistic laws, which unfold within the two dimensions of *space*. Thus, strictly speaking, they should be considered as (geolinguistic) “spatial laws”.

Note that when reading or interpreting the visual pattern of a similarity map, one should always begin with the *warm*-colored polygons and then proceed to the *cold*-colored polygons.

Each choropleth profile of a given similarity map – with its specific spatial pattern – has a determinate linguistic meaning. A visual comparison of Maps 2203 and 2204 suggests that the dialectality of *ALF* point 347 (Map 2203) is better integrated into the whole *ALF* grid than *ALF* point 885 (Map 2204). This is indicated by the different shape and extent of the corresponding red zone. Comparing the numerical legends confirms the fact: on Map 2203, 342 (102 + 120 + 120) of the 640 measurement values are above the arithmetic mean, on Map 2204 it is only 207 (126 + 51 + 30).

Supposing that the 641 lococlects of the *ALF* grid were members of a more-or-less well-connected group of people, one could posit – from an analogical point of view – that “person” 347 (from Normandy) of this community is more communicative than the average, with a score of 53.4% (342 : 640), whereas “person” 885 (from Provence) is below average with only 32.3% (207 : 640). In this analogy, the Provençal member of the *ALF* “group” is at a clear disadvantage in comparison to the Norman member.

Similar communicative considerations related to the quantitative structure of similarity maps provide the basis of the parameter maps discussed in section 5.

The algebraic logic of the similarity index RIV_{jk} corresponds equally well to the electrical signals of phone calls: any single co-identity (COI) could represent contact via telephone, and any co-difference (COD) the opposite (no contact). Hence, the two RIV_{jk} -based similarity maps might be interpreted by analogy in terms of the spatial telephone bills of “network customers” 347 (Normandy) and 885 (Provence). From this perspective, “telephone connection” 347 (Map 2203) has been phoning much more often and over greater distances than “connection” 885 (Map 2204).

For the Romance linguist, the stratification of the rainbow colors on Maps 2203 and 2204 is *per se* characteristic: on Map 2203, the aggregate of warm-colored zones corresponds (with the exception of Wallonia in the north) to the influence area of the Northern French linguistic type (generally called *Langue d’Oil* or *Domaine d’Oil*), whereas on Map 2204, the warm-colored polygons cover the *Domaine d’Oc* (i.e., the *Occitanian* or *Langue d’Oc* area of influence).

The dimensions and location of the two surfaces in red-orange-yellow would change little or not at all if the reference point were displaced at random within the *Domaine*

d'Oil or the Domaine d'Oc. With VDM, this effect is almost perfectly illustrated in a visualization with only two classes (red and blue) and a changing reference point within the red zone (through clicking rapidly on different polygons).

The visible decrease in the linguistic similarities in all directions on Maps 2203 and 2204 (and also Map 2205) is not a natural result of the individual communication experiences of the *locolect* speakers at the *ALF* points 347 and 885. This decrease can in fact be explained by the synergy of the forces of all those language laws which determine the diachronic and synchronic nature of all 641 *locolects* of the *ALF* grid.

4.4. Similarity measurement and linguistic islands

A very interesting application of the similarity measurement is the analysis of the dialectometrical position of a linguistic island within a larger inquiry grid: see Map 2205. The reference point for this map is *ALF* point 635, Andraut (located in the Département Gironde). The locality Andraut is part of the French linguistic island of Petite Gavacherie. It was founded in the last quarter of the fifteenth century in northern Gascoigne (which linguistically belongs to the Domaine d'Oc) by settlers that migrated from the Province Saintonge (and therefore from the Domaine d'Oil). Determining (see the red arrow labeled MAX) the locality (or polygon) which according to RIV_{jk} shows the greatest similarity to reference point 635 (Andraut) should indicate the dialectal "homeland" of this reference point.

In fact, the maximal value of the similarity distribution (which ranges between 9.75 and 51.96 percent) is found at *ALF* point 632 (Abzac). [The minimal value is found at the artificial atlas point 998 (Standard Italian); see the blue arrow labelled MIN on Map 2205.] This inquiry point lies within the historical province of Saintonge: according to historical sources, the settlers did in fact emigrate from its southern border. The clear "oïlitan" (i.e., Northern French, Langue d'Oil) character of the *locolect* of *ALF* point 635 is identified by the position and the shape of all warm colored polygons in northern Gallo-Romania.

It may be assumed that for the identification of the linguistic homeland of a linguistic island, small and smallest-sized linguistic features and their areas – the geographic extension of a single linguistic feature can vary between 1 and 641 atlas points – are of particular interest. Map 2205 was therefore established with a weighted similarity index, namely $WIV(1)_{jk}$, the "Weighted Identity Value (with the weight 1)", which responds ideally to these prerequisites. For a detailed description of the Weighted Identity Value (in German *Gewichtender Identitätswert*) see Goebel (1984: I, 83–86).

Moreover, the question of the origin of *ALF* point 635 may also be tested in two separate linguistic categories, namely through two corpora of either *phonetically* or *lexically* relevant working maps. This procedure reveals that the *locolect* of *ALF* point 635 has been able to completely preserve its Northern French (oïlitan) *phonetic* character, but was nevertheless forced to make considerable *lexical* concessions to the Gascon environment.

4.5. Proximity maps and the Euclidean management of space

While the distribution of linguistic similarities in space (caused by the linguistic activity of *Homo loquens*) can be visualized by means of *similarity* maps, *proximity* maps have a

completely different objective: they show how *Euclidean* proximities are distributed in space. Euclidean proximity (*prox*) is defined as the complement of Euclidean *distance* (*dist*). Again, the formula $prox (= sim) + dist = 100$ is applied.

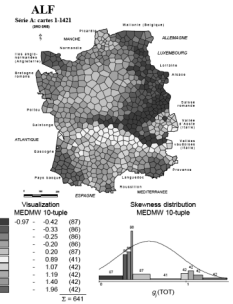
Euclidean distances are the exact straight-line distances from any single locality within the *ALF* grid to any other. As the *x/y*-coordinates of all *ALF* points are already stored in the VDM software, it is easy to calculate the straight-line distances between individual *ALF* points with the well-known Pythagorean formula ($a^2 + b^2 = c^2$).

Map 2206 shows the spatial stratification of the Euclidean proximities with respect to *ALF* point 635. The locations of the maximal (MAX) and minimal values (MIN) of the corresponding proximity distribution depend entirely on the geographical distance from reference point 635: see the two arrows labelled MAX and MIN on Map 2206. A visual comparison of the choropleth profiles of Maps 2205 and 2206 shows that the two are quite different. A statistical correlation (using the Bravais-Pearson correlation coefficient) of the 641 value-pairs (consisting of the *linguistic* similarities on the one hand and of the respective *Euclidean* proximities on the other) produces a comparable result: $r(BP) = -0.019$.

This score is the minimal value of the corresponding correlation map (see Map 2211), which implies that the position of the dialect of the linguistic island of Andraut (*ALF* point 635) within the *ALF* grid has almost nothing in common with Euclidean realities, unlike other *locolects* of the same area; for more information cf. section 7 and Goebel (2005a).

5. The parameter maps

In section 4.3 it was mentioned that the characteristic statistical parameters of the similarity distributions have a special linguistic meaning. In principle, parameter maps investigate the linguistic meaning of different characteristic statistical parameters of the computed similarity distributions (such as minimum, maximum, arithmetic mean, median, standard deviation, skewness, etc.). In this procedure, one statistical parameter is taken from each of the *N* similarity distributions of a given similarity matrix. The resultant set of all the *N* parameter values are subsequently conveyed in a cartographic synopsis and mapped in the usual way. In this article, this procedure is illustrated by the synopsis of 641 *skewness* values (see Map 2207) and the synopsis of 641 *maxima* (see Map 2208). Both synopses are of great linguistic interest and have often proved their usefulness in the analysis of diverse linguistic atlas data.



Map 2207

5.1. Combining the skewness values

Skewness is one of the characteristic parameters of a given similarity distribution, measuring how pronounced its symmetry or asymmetry is; for more information (formula) cf. Goebel (1984: I, 150–153). In an absolutely symmetrical frequency distribution, there

are the same number of scores on both sides of the arithmetic mean. If that is the case, the skewness value is zero. When, however, most of the measurement scores are concentrated *below* the arithmetic mean, the skewness values are *positive*. In the opposite case (the majority of the measurement scores are *above* the arithmetic mean), they are *negative*. To date, a large number of dialectometrical analyses have shown that the synopsis of the skewness values is extremely useful for the detection of a phenomenon typical of all linguistic atlases; viz., for the uncovering of so-called “linguistic compromise or exchange” (in German *Sprachausgleich*). In essence, “linguistic compromise” is defined as the specific intermixture of small, middle and large-scale features (with their respective areas) in any locolect of a given atlas grid; for further discussion cf. Goebel (2003: 81–82).

Any locolect showing many *large-scale* and only a few small or middle-scale attributes is *strongly* involved in the general dynamics of linguistic exchange. If, however, it consists of many *small* or middle-scale attributes and only a few broad-scale traits, it participates to a *lesser extent* in the general linguistic compromise prevailing in the respective atlas grid.

Actually, the phrase and concept of “linguistic compromise or exchange” was originally introduced into German linguistics to describe geolinguistic intermixture phenomena of many different kinds: see the seminal work of Besch (1967).

On Map 2207, with its typical shape, the colors *blue* (corresponding to the *negative* skewness values) and *red* (the *positive* skewness values) thus have the following meaning:

- *blue*: zones of great linguistic compromise, that is, very strong intermixture with many large-scale linguistic attributes (and their areas);
- *red*: zones of little linguistic compromise, that is, with a high percentage of small and middle-scale linguistic attributes (and their areas).

Note that the dark blue polygons form two circular, or pincer-shaped patterns with clear spatial distribution: the big dark blue circle in the east surrounds the entire *Domaine d’Oil*, while the *Francoprovençal* region (on the southeastern periphery) is located between the two jaws of a pincer whose pivot (metaphorically speaking) lies to the west of Lyon.

These two dark blue configurations are, on the one hand (the *Domaine d’Oil*), the visible result and consequence of the secular expansion and propagation of the linguistic type of the *Langued’Oil*, and, on the other hand (the *Francoprovençal*), the visible result and consequence of the retreat of the old Latin linguistic heritage of *Lugdunum/Lyon* with all of its concomitant linguistic contacts and conflicts, squeezed on two sides by the *Langued’Oil* in the north (especially the dialects of Burgundy and the *Franche-Comté*) and by the *Langued’Oc* in the south (especially the dialects of the *Dauphiné*).

The red and orange colored polygon aggregates in the south (*Gascoigne*, *Languedoc* and *Provence*) are, on the contrary, real zones of linguistic isolation; between them, there are some weak flows of linguistic compromise (see the polygons of the classes 6–8). As the choropleth structure of Map 2207 – which most clearly indicates the dynamic propagation in the north and the punctual (block-like) resistance in the south – is of great importance for the history of the whole of Gallo-Romania, it also has to be compared with the diachronic evidence in individual cases. [For similar situations in northern Italy see Goebel (1981: 394–405) and (1984: I, 150–154), in England see Goebel and Schiltz (1997) and Goebel (2007a: 141–142), and in the whole of Italy see Goebel (2007b: 215–217, 2008: 46–49).

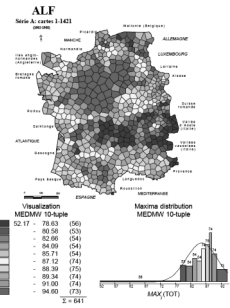
5.2. Combining the maxima

The comparative analysis of numerous similarity maps and the parameters of the corresponding similarity distributions has often shown that the maxima vary in a manner clearly dependent upon the geotypological position of the reference point. In geotypologically compact areas, the RIV maxima are high, whereas in blended areas they are low, despite the RIV maximum on each similarity map being, in most cases, in close proximity to the reference point. Map 2208 gives clear evidence that the synoptic mapping of the RIV maxima enables us to register a phenomenon associated with the concept of “dialect kernel” in traditional qualitative geolinguistics; for more specific details see Goebel (1984: I, 140–143).

In this procedure, it is assumed that linguistically highly similar locolects build strongly cross-linked dialect kernels (comparable to the peaks of a mountain landscape), whereas dialects that are linguistically less integrated in their surroundings (metaphorically speaking, they would have but very few “best friends”) form transitional zones, thus representing the valleys in the same fictional (geo)typological landscape. In fact, the message of Map 2208 fits perfectly with this metaphor.

The compact zone of red polygons in the center of the *Domaine d’Oil*, with a smaller and set back dialect kernel (also in red) on the northern border (Picardy), is separated from another, smaller compact area in the south (with two red-orange colored kernels) by a large blue-green transition zone. The overall structure of the map is an authentic landscape formed by mountains and valleys. Similar transition zones (in blue and green) are also found on the western periphery (Poitou, Anglo-Norman islands), on the northern border (Wallonia) and on the eastern periphery (Lorraine, French-speaking Switzerland and the Aosta Valley).

Note that this spatial distribution would be even more apparent in a stereographic visualization: see Map 14 in Goebel (2003: 111).



Map 2208

6. The interpoint maps: Honeycomb and beam maps

The label “interpoint map” can be traced back to Romance geolinguistics (Lalanne 1953), but it was the German linguist Carl Haag (1898) who established the basic directives for the cartographic generation of these maps (for a historical overview see Goebel 1983). According to the interpoint principle, the triangulation of the atlas grid has to be established first. In the present case, 1,791 triangle sides which are always located between two contiguous inquiry points emerge. As a result, each *triangle side* can be used to capture the potential of the linguistic *similarity* (using RIV_{jk}) of two neighboring atlas points. For the assessment (and the subsequent visualization) of the linguistic *difference* between two such neighboring points, however, one should use the relevant *polygon side* and a distance index (RDV_{jk}). The cartographic generation of the *honeycomb* map therefore relies upon the *polygon sides*, whereas the *beam map* relies on the *triangle sides*.

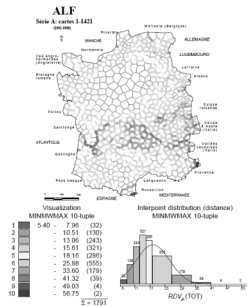
In both instances, as the mappings are obviously no longer based on areas (choropleth mapping technique), but rather on lines (isarithmic mapping technique), there are

new challenges to cope with in the process of visualization. The polygon and triangle sides are encoded in two ways on the maps, through different colors and by varying the thickness of the line. On the honeycomb map, the greatest interpoint differences are colored in *blue* (underlying metaphor: TO GIVE SOMEONE THE COLD SHOULDER), whereas the beam map uses the color *red* to display the greatest interpoint similarities (in metaphorical terms, WARM AFFECTION).

As for the layout of the entire map, one should note its global cartographic significance, which relies on the iconic interplay of 1,971 linear signatures (isarithms). This problem represents a still greater cartographic challenge than the generation of choropleth maps. [The VDM software is capable of readily solving the numerous problems associated with the creation of easily readable interpoint maps. It allows users to combine three interval algorithms and ten sets of different map granulations (varying between 2 and 20 intervals).] However, using the powerful VDM software, a perfect solution can be rapidly set up. In the present case, the visualization emphasized the “frontier” phenomenon on the *honeycomb* map and the “interconnectedness” phenomenon on the *beam* map.

6.1. The honeycomb map (Map 2209)

The iconic message relies on ten classes marked by variation in the colorings and the thickness of the polygon sides. With regard to the spatial concentration of thick polygon sides (also marked by the color steps 7 and 8), it clearly appears that there are line-like phenomena at the following four spots on the map: between Roussillon and the Languedoc (in the south), between the Domaines d’Oc and d’Oil (see the curved line in the middle of the map), on the northern border of the Francoprovençal area (at the eastern edge of the map) and between the Picardian and Wallonian zones (at the northern limit of the map). In contrast, there are zones of little compartmentalization in the central areas of the Domaines d’Oc and d’Oil. It is also worth mentioning that the segregation between the different *ALF* points is in general rather more gradual than sharp-edged.



Map 2209

A visual comparison of the most extensive isogloss syntheses in older French studies (see Rosenqvist 1919 and Ettmayer 1924) so far confirms that the general cartographic message of those first maps (still done by hand) coincides in principle with the actual interpoint message of Map 2209.

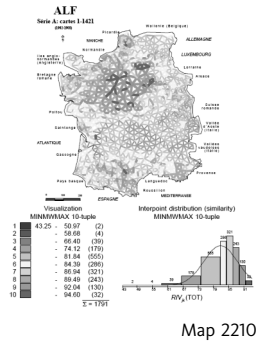
Nevertheless, it must not be forgotten that the general classificatory status of interpoint maps in general and of honeycomb maps in particular is rather limited, although they are well shaped and therefore highly suggestive and self-explanatory. Consider the following example. The distance or similarity matrices used to generate Maps 2209 and 2210 have [according to the formula $N/2(N-1)$; $N = 641$] all in all 205,120 distance or similarity values. However, for the establishment of these two maps, only 1,791 values were recorded, which corresponds to 0.87 percent of the full amount of the distance or similarity values of the two matrices. One must therefore be aware of the fact that the taxometric relevance of interpoint analysis is rather small.

6.2. The beam map (Map 2210)

The cartographic basis of Map 2210 largely overlaps with that of Map 2209. It should be mentioned that the two histograms are absolutely symmetrical and that adding the minimum value in the numerical legend of Map 2209 (5.40) to the maximum in the legend of Map 2210 (94.50) gives the sum of 100, due to the familiar formula $sim + dist = 100$, which adjusts the quantitative connection between similarity (sim) and distance ($dist$).

Map 2210 relies upon a triangulation of the atlas grid and therefore – metaphorically speaking – indicates the variable intensity of the communication and interaction currents between two contiguous atlas points. The shape and position of a powerful interaction landscape in the north (Domaine d’Oil) and the occurrence of several small and much less prominent interaction kernels in the south (Roussillon, Languedoc, Provence) can be clearly recognized. Thin green and blue triangle sides marking the transitional zones in the center of the picture and on the eastern periphery correspond in position and shape exactly to those zones on Map 2209 where stronger segregations are found.

Curiously enough, from an historical point of view, although they are the cartographic inversion of the well-known honeycomb maps (since Haag 1898) and reveal very interesting linguistic phenomena, beam maps only appeared with the emergence of dialectometry (see Goebel 1983).



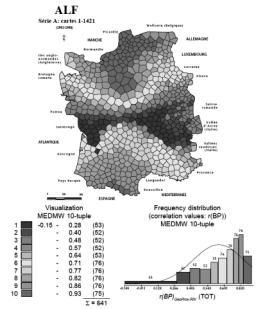
Map 2210

7. Correlative dialectometry

Correlative dialectometry is a fairly recent instrument of the Salzburg School of dialectometry: it was only integrated into the VDM software in 2004 (see Goebel 2004, 2005a, 2005b, 2006, 2007a, 2007b and 2008 for previous applications of correlative dialectometry). The basic idea was to develop a cartographic procedure which enables the visualization of the highly fluctuating correlation of two spatially prominent variables. It was also assumed that the spatial variation of these correlations would lead to highly elaborated and meaningful cartographic results.

Figure 22.3 illustrates the necessary formal requirements. For the correlation of the two variables, two similarity or distance matrices with exactly the same dimensions are needed. The measurement values of these two matrices must be of relatively similar range. Further, the actual correlation is calculated by correlating by pairs the N atlas point vectors of the two matrices with a previously selected correlation measurement (the examples make use of the the Bravais-Pearson correlation coefficient). As a result, we obtain $N r(BP)$ values, which are subsequently mapped in the usual way.

Map 2211 illustrates the correlation between geolinguistic similarities (according to RIV_{jk}) and Euclidean proximities (according to the Pythagorean formula). Maps 2205 and 2206 have already illustrated the spatial stratification of these two dimensions for ALF point 635 (Andraut).



Map 2211

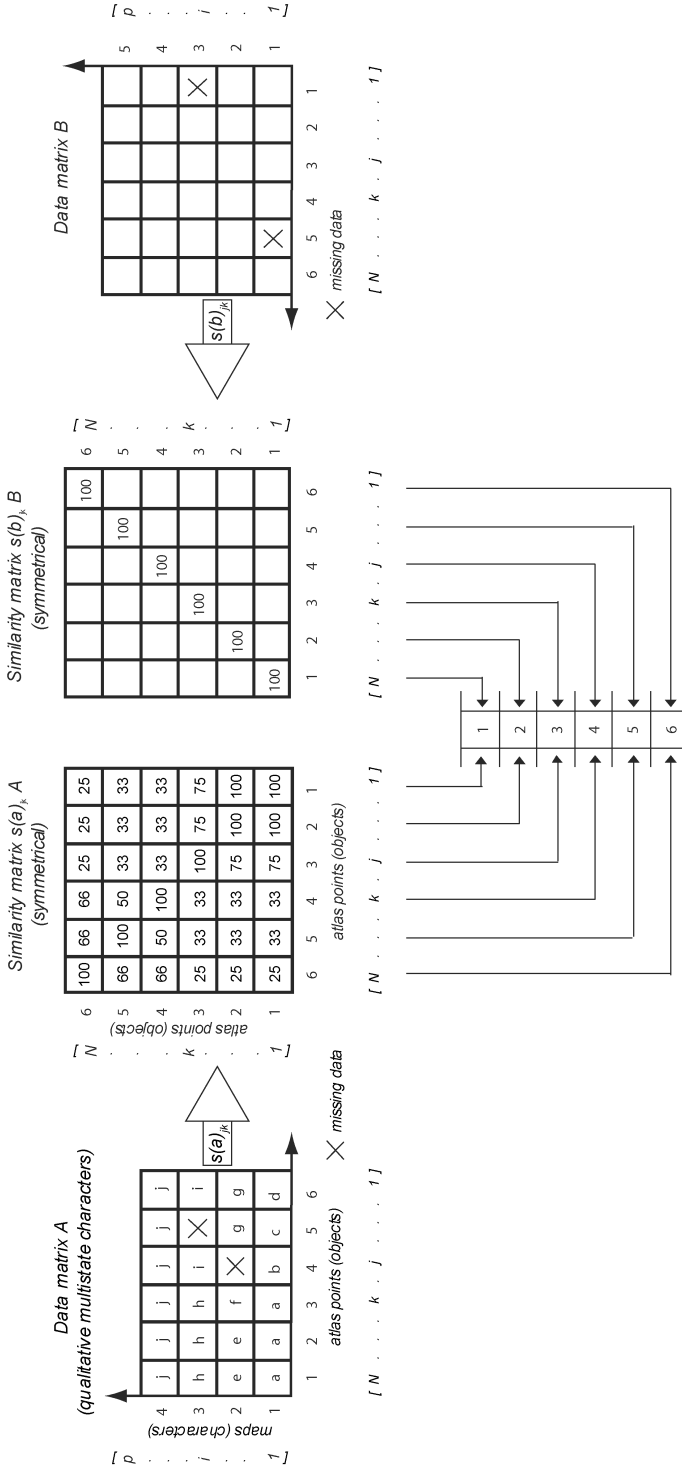


Fig. 22.3: Schematic representation of the calculation of the correlation maps [Note the symmetry of the two similarity matrices (A and B), and that the content of the data and similarity matrix A corresponds to that of Figure 1]

There are three possibilities for decreasing geolinguistic similarity across space: it can be slower than the Euclidean model predicts, it can closely match the prediction, or it can be faster. In the first instance, the geolinguistic “energy” of a lococlect (metaphorically speaking) is conveyed and increased by an extra-linguistic force; in the latter situation, the opposite occurs, and the result is a braking action, or a weakening of a lococlect’s geolinguistic “energy”. In both instances, historical, political, social or economical factors are involved.

The correlation coefficient $r(\text{BP})$ oscillates between -1 (inverse correlation between the two variables), 0 (no correlation), and $+1$ (perfect correlation). It measures, statistically speaking, the linear correlation between two numerical variables.

Map 2211, based on 641 $r(\text{BP})$ values, shows a highly regular shape, which is of course not the product of mere chance. The question is how to interpret the different colors from a linguistic point of view. Red indicates a high correlation between the spatial stratification of the geolinguistic similarities and the Euclidean proximities: it signifies a kind of “primitive harmony” between the “energies” of language and space. Blue, in contrast, refers to those areas where this “primitive harmony” has been to some degree disturbed, or no longer exists.

The perfectly regular spatial pattern of Map 2211 suggests that in Gallo-Romania there existed (and still exist) two prominent and distinct correlations between language and space that meet in the middle of the map. However, this antagonism is not abrupt, but gradually developed in a progressive “phasing out” still visible today. In accordance with the history of Gallo-Romania, the compact stratification of the northern red zone and the more graduated structure of the southern red zone further suggest that the “northern force” was and is poised to displace the “southern force” beyond the hyperbola-shaped blue ditch or to exert a capillary influence on it. The result in the blue zone is a kind of dislocation in the primitive spatial convergence between the linguistic and the Euclidean management of space.

The relevance of this “dislocation theory” can be convincingly illustrated by the classification of *ALF* point 635 (Andraut). This inquiry point – which has the minimum $r(\text{BP})$ value of -0.15 – corresponds to a linguistic island transplanted from the dialectal space of the Saintonge to that of Gascoigne in the fifteenth century (cf. section 4.5). Actually, the process of the foundation of a linguistic island represents, in principle, a total disengagement or decoupling of all “natural” or “primitive” relations between the dialectal and Euclidean management of space.

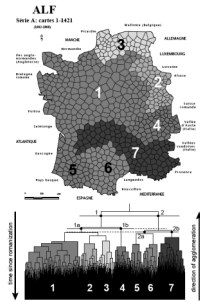
Correlative dialectometry can also be successfully employed to compare other varieties of the human management of space, and thus to compare different intralinguistically relevant categories, such as *phonetics* and *vocabulary*, *vocalism* and *consonantism*, etc. It also allows for the comparison of taxometrically different modalities (such as similarity measurements with RIV_{jk} and $\text{WIV}(1)_{jk}$ on the one hand, and similarity measurements with *many* working maps and with *less* workings maps on the other).

8. Dendrographic dialectometry

Classification trees, which are used in other disciplines within the humanities and life sciences, are likewise employed in geolinguistics. Nevertheless, their heuristic usefulness depends on their appropriate and careful application. More-or-less recent developments

in the field of numerical classification have provided relevant algorithms, which belong to the class of “hierarchic-agglomerative methods”. They generate trees consisting of a hierarchy of disjunctive clusters with a binary structure. The inner binary organization of these trees is indeed of great relevance for classification and pattern recognition.

These dendrographic algorithms presuppose the establishment of a similarity or distance matrix with the dimensions $N \times N$. Through the successive fusion (or agglomeration) of two highly “similar” yet originally unconnected objects (i.e., dialects or locolects), they generate a tree-like hierarchy, which becomes increasingly narrow and ends ultimately in the stem (or root): see the arrow at the right-hand side of the tree on Map 2212. (In this regard, the definition of “similar” or “similarity” depends on the applied dendrographic algorithm. It is therefore extremely important that the dialectometrician choose the algorithm in accordance with his/her own theoretical conception of “[linguistic] similarity”.)



Map 2212

In principle, classification trees consist of $N - 1$ bifurcations. The biggest clusters hanging on the bifurcations near the stem (or the root) of the tree also have the greatest inner heterogeneity; the smallest clusters (with the greatest internal homogeneity) are located near the leaves. For the dialectometrician, the analysis of the structure of the tree is of less interest than the projection in space of the resulting dendrographic classification (the so-called “spatialization”). In this process, specially marked clusters from the tree (or “dendremes”) are linked to analogous parts of the map (the “choremes”).

Map 2212 illustrates this principle. Thus, the “northern” dendremes 1–4 (on the left), and the “southern” dendremes 5–7 (on the right) hang from two ramifications near the root of the tree. The map clearly shows the position of the corresponding choremes in space. In all instances, the seven dendremes/choremes correspond to several well-known dialect domains of Gallo-Romania. Moreover, their perfect spatial coherence is striking.

From a linguistic point of view, these classification trees can always be interpreted both *diachronically* and *synchronically*.

The *diachronic* interpretation analyzes the structure of the tree by proceeding from the root to the leaves: see the arrow to the left of the tree on Map 2212. It is assumed that the progressive fragmentation of the ramifications of the tree simulates the progressive dialectal fragmentation (or “dialectalization”) of Gallo-Romania, which would have represented a linguistically homogeneous area at the outset. The space of this area would have first separated into the Domaines d’Oc and d’Oil; subsequently, the Franco-provençal would have separated from the rest of the Domaine d’Oil (dendreme/choreme 4); then the Domaine d’Oc would have split into the dendremes/choremes 5–6 and the dendreme/choreme 7, and so on. Moreover, similar views are also shared in lexicostatistics: for more information see Goebel (1993) and (2003: 84–89). The *synchronic* interpretation concentrates on the relative position of the dendremes within the tree, and interprets their reciprocal relatedness. The results show that compared with dendreme/choreme 1 (Central and Western France), dendremes/choremes 2 (Eastern France) and 3 (Picardy and Wallonia) have more reciprocal closeness. The same results are found in the south, when the dendremes/choremes 5 and 6 (Gascoigne and Languedoc) are contrasted with dendreme/choreme 7 (Limousin, Auvergne, Provence), etc.

According to the classification goals of geolinguistics, the following hierarchic-agglomerative methods have proven their practical usefulness: the Ward algorithm (see

Map 2212), Complete Linkage, Average Linkage (UPGMA), and the Simple Average Linkage algorithm.

The dendrographic algorithms give access to structures that are deeply hidden or engraved in the similarity matrices and which can now be visualized. In such cases, however, the linguistic interpretation of the trees should always also take into account their quantitative properties, which is why a sound knowledge of their mathematical fundamentals is a *sine qua non* for the dialectometrician. All necessary information is provided in the relevant handbooks: see Sneath and Sokal (1973: 204–245), Bock (1974: 383–411) and Chandon and Pinson (1981: 122–131).

9. Review

The dialectometrical procedures introduced in this chapter have an exclusively diagnostic and exploratory character. In a strict sense, they can only be usefully employed for scientific explanations if they occur in an adequately comprehensive research context that encompasses many different factors. Moreover, dialectometrical methods as a whole represent a hybrid compound which includes elements of linguistics, statistics and cartography, without being genuinely “at home” in any of these sciences. It is all the more important to place emphasis on the specific scientific scope of this compound of methods, which is the enlargement of our knowledge concerning the structure and function of geolinguistic networks through quantitative means.

The last thirty years have shown that dialectometrical outlooks are still met with reservation and comprehension difficulties in the humanities in general and in linguistics in particular. These hindrances are connected with the use of mathematical and statistical methods. The same is true of the highly visual and map-centered character of dialectometry, which is often not easily accessible for many linguists. In the future, these reservations will hopefully disappear or at least recede against the background of the increasingly extensive use of IT-based quantitative methods and IT-supported visualization techniques.

Dialectometry, in the way it has been documented in this paper, defines itself programmatically as the *quantitative* arm of classical, atlas-based and *quality*-oriented linguistic geography. With its support, it could in the end be demonstrated that, in the data of many linguistic atlases, many hitherto hidden and unexpectedly complex spatial patterns are to be found, which suggest the existence and agency of genuine “spatial laws”. As far as they are concerned, these patterns are connected to our previous postulate concerning the basilectal management of space by *Homo loquens*, which is defined as one of the numerous semiotic activities of humankind.

In conclusion, it would be highly regrettable if, in the current scientific contest between *geo-* and *sociolinguistic* perspectives, traditional interest in the “linguistic atlas” as a classical source of data should decline or even subside in the long term.

Acknowledgment

The work reported on here was enabled by two research grants (12414 and 13349) financed by the Fonds zur Förderung der wissenschaftlichen Forschung in Österreich

(FWF; Austrian Science Foundation) in Vienna. The work on extracting 1,687 working maps from 626 original *ALF* maps by taxation was carried out by Barbara Aigner, Ingrid Dautermann, Hildegund Eder, Susanne Oleinek and Annette Ida Schatzmann, all former students at Salzburg University. Edgar Haimerl, Seattle (USA), formerly Blaustein (Baden-Württemberg, Germany), designed, implemented and has supported the dialectometric software VDM (Visual DialectoMetry). Slawomir Sobota from Salzburg University created the maps and figures for this paper, and Gabriele Blaikner-Hohenwart, Salzburg University, translated the original German text into English, which was then edited by Benjamin Wright. My most cordial thanks to the FWF and to all my abovementioned students, assistants and collaborators as well.

10. Frequently used abbreviations

MEDMW	Interval algorithm (see section 4.2.)
MINMWMAX	Interval algorithm (see section 4.2.)
RDV_{jk}	Relative Distance Value (see section 6.1.)
RIV_{jk}	Relative Identity Value (see section 4.1.)
$WIV(1)_{jk}$	Weighted Identity Value with the weight of 1 (see section 4.4. and Map 2205)

11. Atlases

AIS • ALF • ALG • CLAE • DASB • DSA

12. References

- Aubin, Hermann, Theodor Frings and Josef Müller
 1926 [1966] *Kulturströmungen und Kulturprovinzen in den Rheinlanden: Geschichte, Sprache, Volkskunde*. Bonn: Röhrscheid. [Reprint: Darmstadt: Wissenschaftliche Buchgesellschaft.]
- Besch, Werner
 1967 *Sprachlandschaften und Sprachausgleich im 15. Jahrhundert. Studien zur Erforschung der spätmittelhochdeutschen Schreibdialekte und zur Entstehung der neuhochdeutschen Schriftsprache*. Munich: Francke.
- Bock, Hans Hermann
 1974 *Automatische Klassifikation. Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten (Cluster-Analyse)*. Göttingen: Vandenhoeck & Ruprecht.
- Brun-Trigaud, Guylaine, Yves Le Berre and Jean Le Dù
 2005 *Du temps dans l'espace. Explication linguistique de la France. Essai d'interprétation des cartes de l'Atlas linguistique de la France de Jules Gillieron et Edmond Edmont augmenté de quelques cartes de l'Atlas linguistique de la Basse-Bretagne de Pierre Leroux*. Paris: CTHS.
- Chandon, Jean-Louis and Suzanne Pinson
 1981 *Analyse typologique. Théories et applications*. Paris/New York/Barcelona/Milan: Masson.

- Dickinson, Gordon Cawood
1973 *Statistical Mapping and the Presentation of Statistics*. London: Arnold.
- Ettmayer, Karl von
1924 *Über das Wesen der Dialektbildung erläutert an den Dialekten Frankreichs*, 1–56. (Denkschriften der Akademie der Wissenschaften in Wien, phil.-hist. Klasse 66.) Vienna/Leipzig: Hölder-Pichler-Tempsky.
- Goebel, Hans
1981 *Eléments d'analyse dialectométrique (avec application à l'AIS)*. *Revue de linguistique romane* 45: 349–420.
- Goebel, Hans
1982 *Dialektometrie. Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialektgeographie*, 1–123. (Denkschriften der Österreichischen Akademie der Wissenschaften, phil.-hist. Klasse 157.) Vienna: Verlag der Österreichischen Akademie der Wissenschaften.
- Goebel, Hans
1983 *Parquet polygonal et treillis triangulaire. Les deux versants de la dialectométrie interponctuelle*. *Revue de linguistique romane* 47: 353–412.
- Goebel, Hans
1984 *Dialektometrische Studien. Anhand italoromanischer, rätomanischer und galloromanischer Sprachmaterialien aus AIS und ALF*. 3 vols. Tübingen: Niemeyer.
- Goebel, Hans
1993 *Eine hierarchisch-agglomerative Klassifikation von Daten des Sprachatlases AIS*. In: Jürgen Schmidt-Radefeldt and Andreas Harder (eds.), *Sprachwandel und Sprachgeschichte. Festschrift für Helmut Lüdtke*, 63–78. Tübingen: Narr.
- Goebel, Hans
2000 *La dialectométrisation de l'ALF: présentation des premiers résultats*. *Linguistica* 40: 209–236.
- Goebel, Hans
2002 *Analyse dialectométrique des structures de profondeur de l'ALF*. *Revue de linguistique romane* 66: 5–63.
- Goebel, Hans
2003 *Regards dialectométriques sur les données de l'Atlas linguistique de la France (ALF): relations quantitatives et structures de profondeur*. *Estudis Romànics* 25: 59–120.
- Goebel, Hans
2004 *Sprache, Sprecher und Raum. Eine kurze Darstellung der Dialektometrie. Das Fallbeispiel Frankreich*. *Mitteilungen der Österreichischen Geographischen Gesellschaft* 146: 247–286.
- Goebel, Hans
2005a *La dialectométrie corrélative: un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme*. *Revue de linguistique romane* 69: 321–367.
- Goebel, Hans
2005b *Dialekte und Familiennamen in Frankreich. Ein interdisziplinärer Vergleich mit den Mitteln der Dialektometrie*. In: Günter Hauska (ed.), *Gene, Sprachen und ihre Evolution. Wie verwandt sind die Menschen – wie verwandt sind ihre Sprachen?*, 68–99. Regensburg: Universitätsverlag Regensburg.
- Goebel, Hans
2006 *Recent advances in Salzburg dialectometry*. *Literary and Linguistic Computing* 21/4: 411–435.
- Goebel, Hans
2007a *A bunch of dialectometric flowers: A brief introduction to dialectometry*. In: Ute Smit, Stefan Dollinger, Julia Hüttner, Gunther Kaltenböck and Ursula Lutzky (eds.), *Tracing*

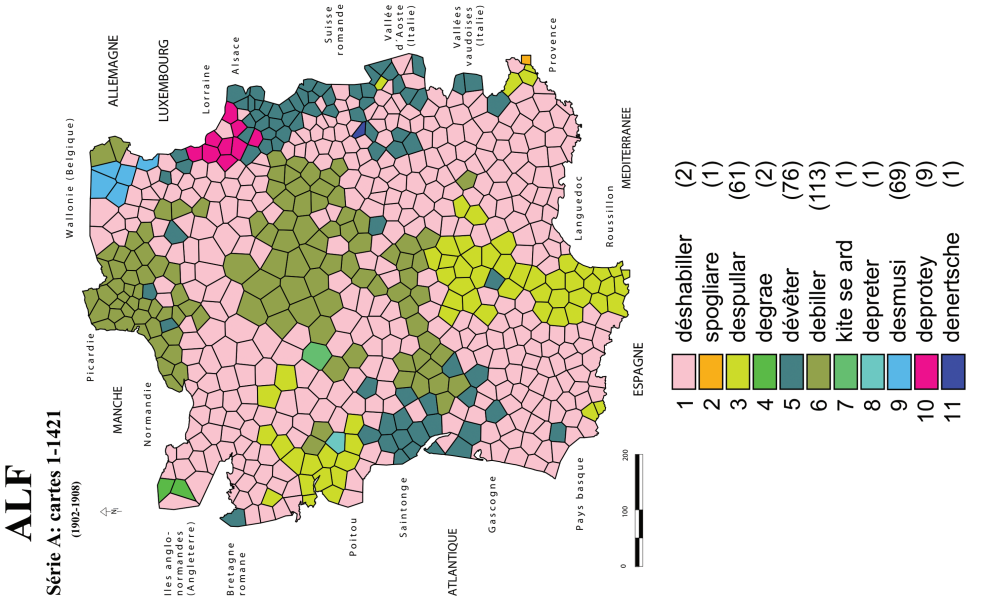
- English Through Time. Explorations in Language Variation. In Honour of Herbert Schendl on the Occasion of his 65th Birthday*, 133–171. Vienna: Braumüller.
- Goebel, Hans
2007b Dialektometrische Streifzüge durch das Netz des Sprachatlases AIS. *Ladina* XXXI: 187–271.
- Goebel, Hans
2008 La dialettometrizzazione integrale dell’AIS. Presentazione dei primi risultati. *Revue de linguistique romane* 72: 25–113.
- Goebel, Hans and Guillaume Schiltz
1997 A dialectometrical compilation of CLAE I and CLAE II. Isoglosses and dialect integration. In: Wolfgang Viereck and Heinrich Ramisch (eds.), *Computer Developed Linguistic Atlas of England CLAE*, vol. 2, 13–21. Tübingen: Niemeyer.
- Guiter, Henri
1973 Atlas et frontières linguistiques. In: Georges Straka and Pierre Gardette (eds.), *Les dialectes romans de France à la lumière des atlas régionaux*, 61–109. Paris: Centre National de la Recherche Scientifique.
- Haag, Carl
1898 *Die Mundarten des oberen Neckar- und Donaulandes (Schwäbisch-alemannisches Grenzgebiet: Baarmundarten)*. (Beilage zum Programm der königlichen Realanstalt). Reutlingen: Hutzler.
- Haggett, Peter
1973 [1965] *Einführung in die kultur- und sozialgeographische Regionalanalyse*. Berlin/New York: Walter de Gruyter. [Originally published as *Locational Analysis in Human Geography*. London: Edward Arnold.]
- Heeringa, Wilbert
2004 *Measuring Dialect Pronunciation Differences using Levenshtein Distance*. (Groningen Dissertations in Linguistics 46.) Groningen: University Library Groningen.
- Heeringa, Wilbert and John Nerbonne
2001 Dialect areas and dialect continua. *Language Variation and Change* 13: 375–400.
- Hummel, Lutz
1993 *Dialektometrische Analysen zum Kleinen Deutschen Sprachatlas (KDSA). Experimentelle Untersuchungen zu taxometrischen Ordnungsstrukturen als dialektaler Gliederung des deutschen Sprachraums*. 2 vols. Tübingen: Niemeyer.
- Kretschmar, William A. and Edgar W. Schneider (eds.)
1996 *Introduction to Quantitative Analysis of Linguistic Survey Data. An Atlas by Numbers*. Thousand Oaks/London/New Delhi: Sage.
- Lalanne, Théodore
1953 Indice de polyonymie. Indice de polyphonie. *Français Moderne* 21: 263–274.
- Melis, Ludo, Serge Verlinde and Patricia Derynck
1989 La notion de dialecte supralocal: essai de définition dialectométrique. *Orbis* 33: 70–132.
- Nerbonne, John and Heeringa, Wilbert
2001 Computational comparison and classification of dialects. *Dialectologia et Geolinguistica* 9: 69–83.
- Nerbonne, John and William Kretschmar
2003 Introducing computational techniques in dialectometry. *Computers and the Humanities* 37: 245–255.
- Nerbonne, John and William Kretschmar, Jr.
2006 Progress in dialectometry: Toward explanation. *Literary and Linguistic Computing* 21(4): 387–397.
- Okabe, Atsuyuki, Barry Boots and Kokichi Sugihara
1992 *Spatial Tessellations. Concepts and Applications of Voronoi Diagrams*. Chichester: Wiley.

- Paris, Gaston
1888 [1909] Les parlers de France. [In: Gaston Paris, *Mélanges linguistiques*, 432–448. Paris: Champion.]
- Putschke, Wolfgang (ed.)
1977 Automatische Sprachkartographie. Vorträge des Internationalen Kolloquiums zur automatischen Sprachkartographie in Marburg (1977). *Germanistische Linguistik* 3–4: 1–395.
- Putschke, Wolfgang
2001 Die Dialektologie, ihr Beitrag zur historischen Sprachwissenschaft im 19. Jahrhundert und ihre Kritik am junggrammatischen Programm. In: Auroux, Sylvain, E. F. K. Koerner, Hans-Josef Niederehe and Kees Versteegh (eds.), *History of the Language Sciences. An International Handbook on the Evolution of the Study of Language from the Beginnings to the Present / Geschichte der Sprachwissenschaften. Ein internationales Handbuch zur Entwicklung der Sprachforschung von den Anfängen bis zur Gegenwart / Histoire des sciences du langage. Manuel international sur l'histoire de l'étude du langage des origines à nos jours*, vol. 2: 1498–1513. (Handbücher der Sprach- und Kommunikationswissenschaft 18.2.) Berlin/New York: de Gruyter.
- Rosenkranz, Heinz
1938 *Mundart und Siedlung im Gebiet der oberen Saale und des nördlichen Frankenwaldes*. Jena: Fischer.
- Rosenqvist, Arvid
1919 Limites administratives et division dialectale de la France. *Neuphilologische Mitteilungen* 20: 87–119.
- Scapoli, C., H. Goebel, S. Sobota, E. Mamolini, A. Rodriguez-Larralde and I. Barraï
2005 Surnames and dialects in France. Population structure and cultural evolution. *Journal of Theoretical Biology* 237: 75–86.
- Séguy, Jean
1973 La dialectométrie dans l'Atlas linguistique de la Gascogne. *Revue de linguistique romane* 37: 1–24.
- Sneath, Peter H. A. and Robert R. Sokal
1973 *Numerical Taxonomy. The Principles and Practice of Numerical Classification*. San Francisco: Freeman.
- Verlinde, Serge
1988 La dialectométrie et la détection des zones dialectales: l'architecture dialectale de l'Est de la Belgique romane. *Revue de linguistique romane* 52: 151–181.
- Viaplana, Joaquim
1999 *Entre la dialectologia i la lingüística. La distància lingüística entre les varietats del català nord-occidental*. Barcelona: Publicacions de l'Abadia de Montserrat.

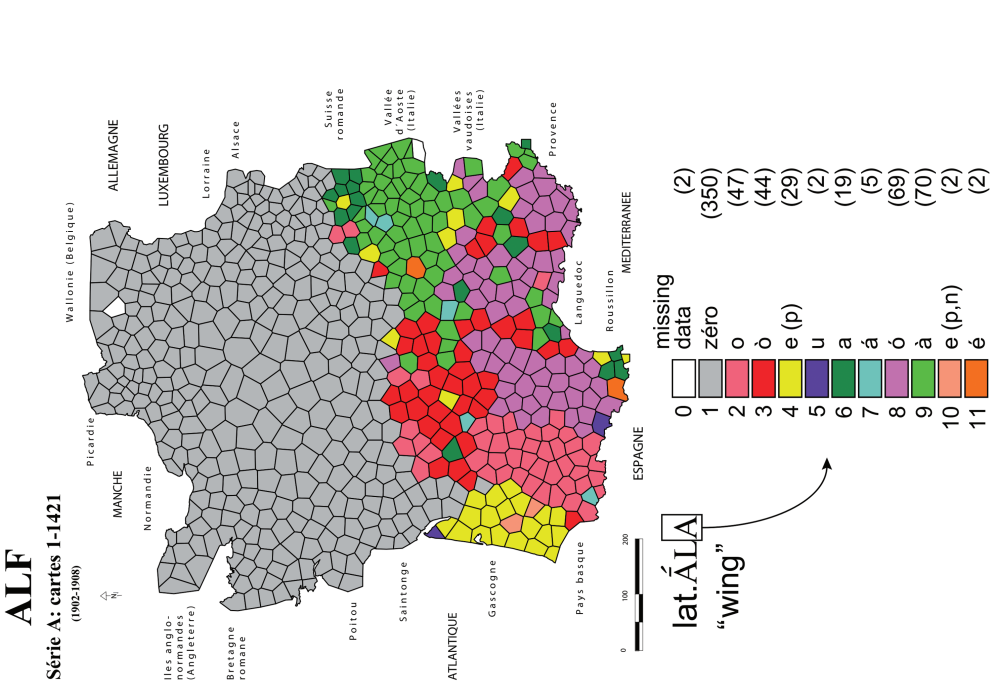
Hans Goebel, Salzburg (Austria)

Linguistic atlases:

- AIS 1928-194 Jaberg, Karl and Jud, Jakob. *Sprach- und Sachatlas Italiens und der Südschweiz*. Zofingen: Ringier, 8 vols. (reprint: Nendeln: Kraus, 1971).
- ALF 1902-1910 Gilliéron, Jules and Edmont, Edmond. *Atlas linguistique de la France*. Paris: Champion, 10 vols. (reprint: Bologna: Forni, 1968).
- ALG 1954-1974 Séguy, Jean. *Atlas linguistique et ethnographique de la Gascogne*. Paris: Centre National de la Recherche Scientifique, 6 vols.
- CLAE 1991-1997 Viereck, Wolfgang and Ramisch, Heinrich (eds.). *Computer Developed Linguistic Atlas of England*. Tübingen: Niemeyer, 2 vols.
- DASB 1996 Schiltz, Guillaume. *Der Dialektometrische Atlas von Südwest-Baden (DASB). Konzepte eines dialektometrischen Informationssystems*. Marburg: Elwert, 4 vols.
- DSA 1927-1956 *Deutscher Sprachatlas aufgrund des von Georg Wenker begründeten Sprachatlas des Deutschen Reiches in vereinfachter Form begonnen von Ferdinand Wrede, fortgesetzt von Walther Mitzka und Bernhard Martin*. Marburg/Lahn: Elwert, 23 fascicles with 128 maps.



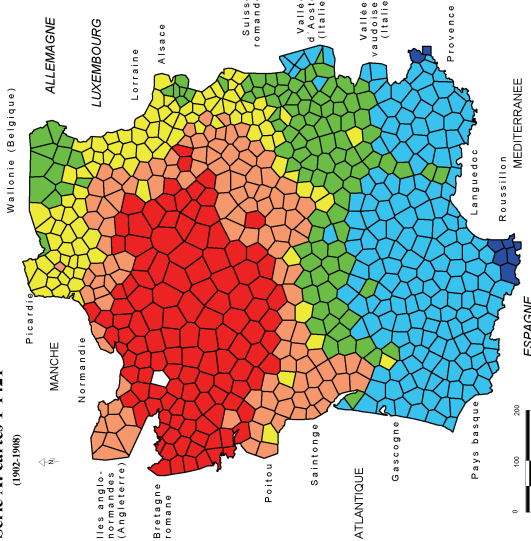
Map 2202: Sample of a lexical working map showing the spatial distribution of the Gallo-Romance designations of 'to take off (clothes)' (following ALF 394 *déshabiller*)



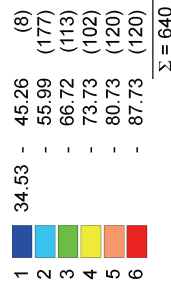
Map 2201: Sample of a phonetic working map showing the spatial distribution of the Gallo-Romance results of final -A in the Latin etymon ÁLA 'wing' (following ALF 204 *aile*)

ALF

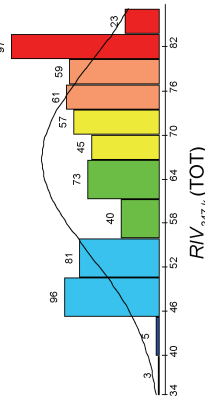
Série A: cartes 1-1421
(1902-1908)



Visualization
MINMWMAX 6-tuple



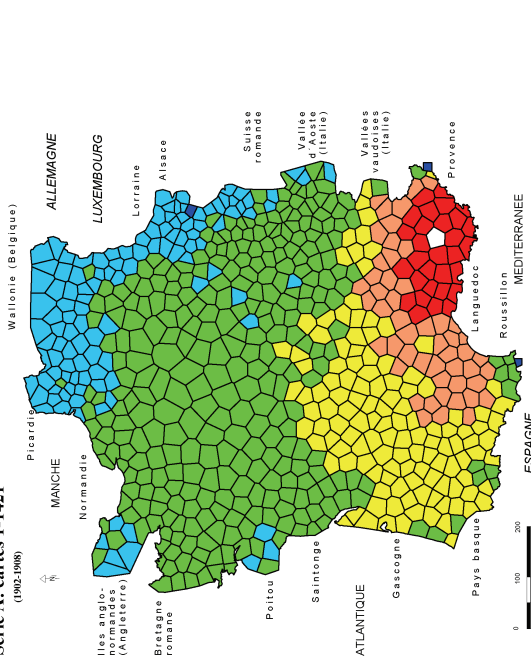
Similarity distribution
MINMWMAX 12-tuple



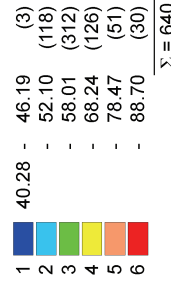
Map 2203: Sample of a similarity map showing the spatial distribution of the similarity values (according to RIV_{jk}) for ALF point 347 (Dompierre, Département Orne). Similarity index: $RIV_{347,k}$; corpus: 1687 working maps (total corpus); algorithm of visualization: MINMWMAX 6-tuple.

ALF

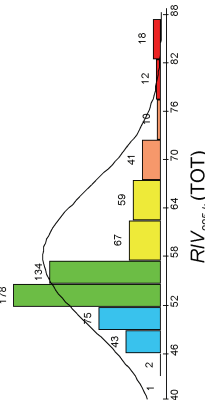
Série A: cartes 1-1421
(1902-1908)



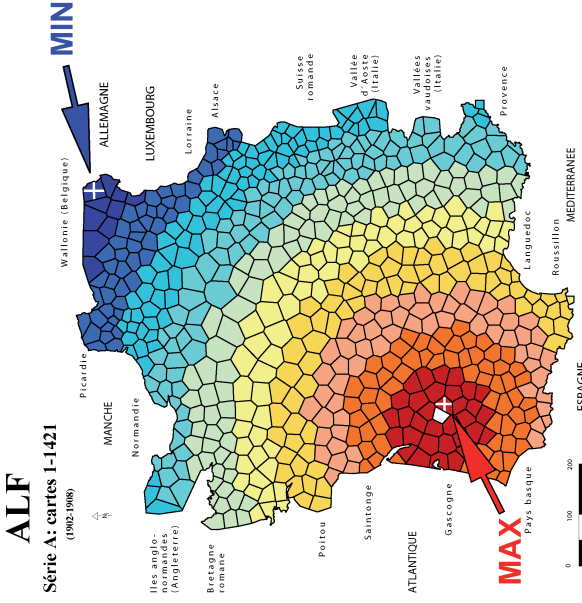
Visualization
MINMWMAX 6-tuple



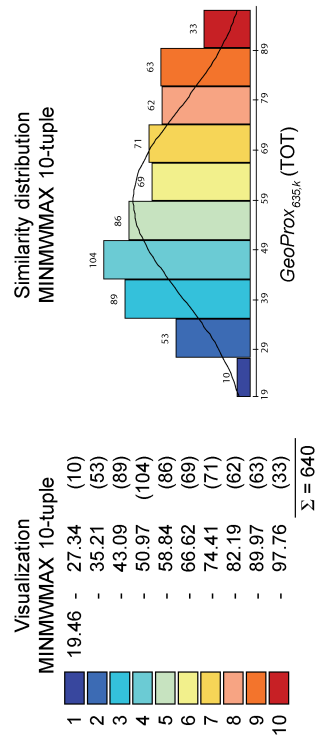
Similarity distribution
MINMWMAX 12-tuple



Map 2204: Sample of a similarity map showing the spatial distribution of the similarity values (according to RIV_{jk}) for ALF point 885 (Gréoux, Département Basses-Alpes). Similarity index: $RIV_{885,k}$; corpus: 1687 working maps (total corpus); algorithm of visualization: MINMWMAX 6-tuple.



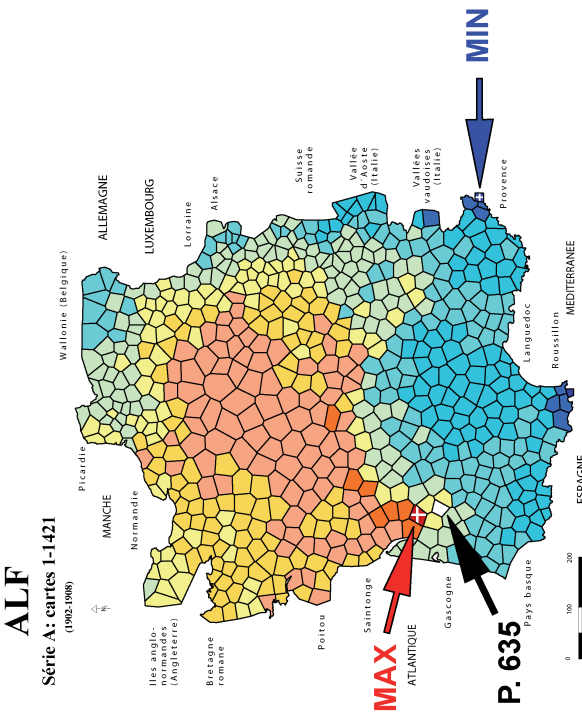
ALF
Série A: cartes 1-1421
(1902-1986)



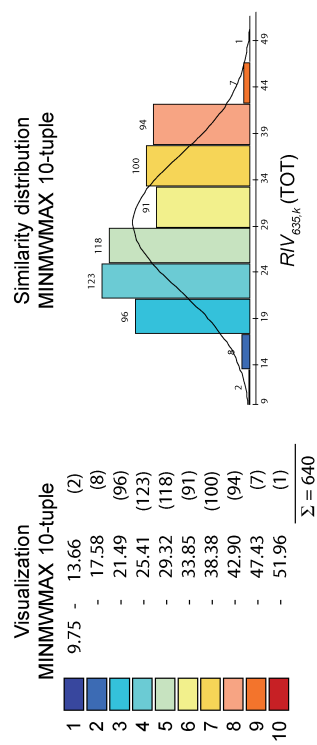
Visualization
MINMWMAX 10-tuple

Similarity distribution
MINMWMAX 10-tuple

Map 2206: Sample of a proximity map showing the spatial distribution of the proximity values for ALF point 635 (Andraut, Département Gironde). Proximity index (*prox*): = $100 - \text{dist}_{\text{Eucl}}$ (Euclidean distance according to Pythagoras); algorithm of visualization: MINMWMAX 6-tuple.



ALF
Série A: cartes 1-1421
(1902-1986)



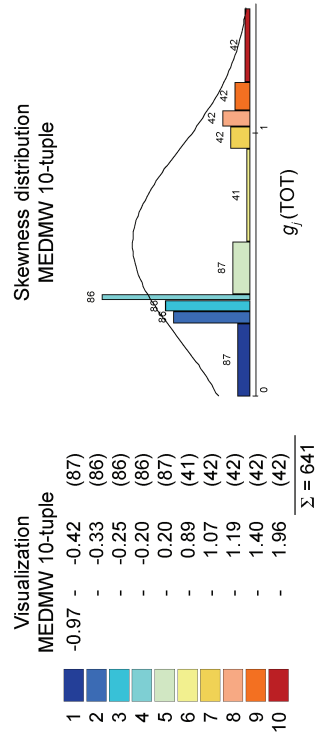
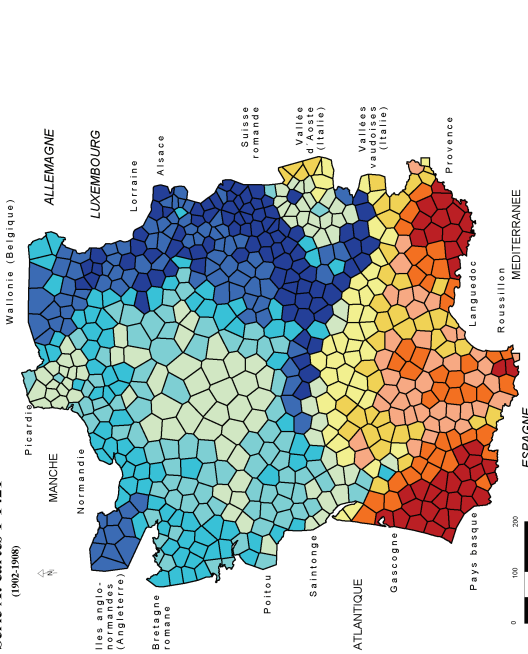
Visualization
MINMWMAX 10-tuple

Similarity distribution
MINMWMAX 10-tuple

Map 2205: Sample of a similarity map showing the spatial distribution of the similarity values (according to $WIV[i]_{jk}$) for ALF point 635 (Andraut, Département Gironde). Similarity index: $WIV(1)_{635,k}$; corpus: 1687 working maps (total corpus); algorithm of visualization: MINMWMAX 10-tuple.

ALF

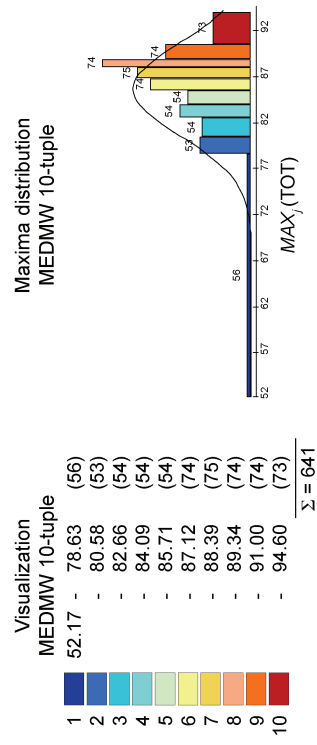
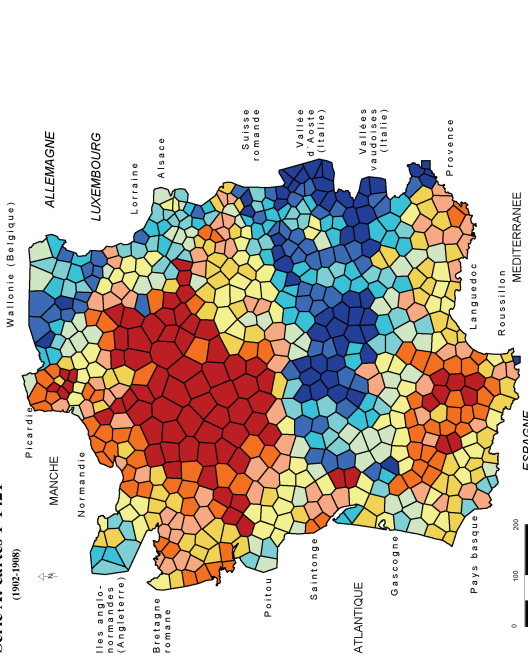
Série A: cartes 1-1421
(1902-1986)



Map 2207: Choropleth map of the synopsis of the skewness values of 641 similarity distributions. Similarity index: RIV_{jk} ; corpus: 1687 working maps (total corpus); algorithm of visualization: MEDMW 10-tuple.

ALF

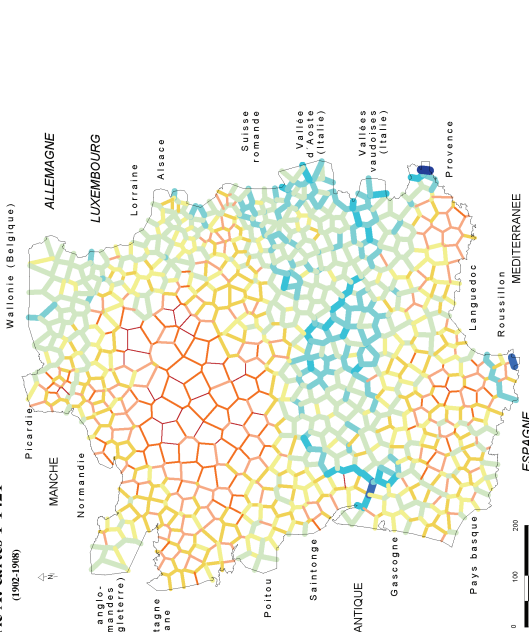
Série A: cartes 1-1421
(1902-1986)



Map 2208: Choropleth map of the synopsis of the maximal values of 641 similarity distributions. Similarity index: RIV_{jk} ; corpus: 1687 working maps (total corpus); algorithm of visualization: MEDMW 10-tuple.

ALF

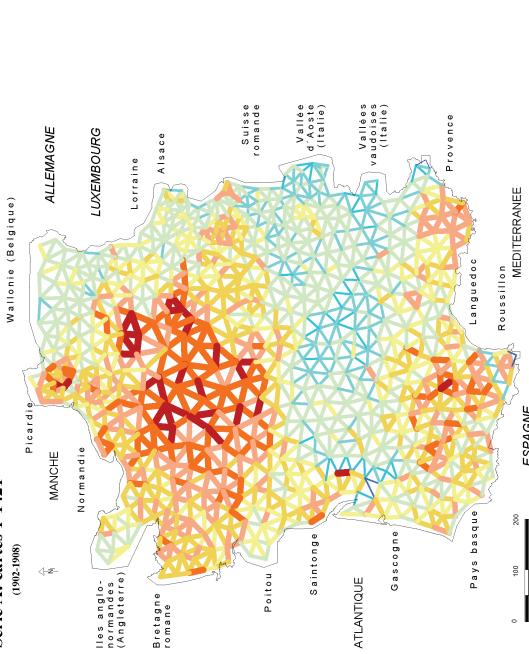
Série A: cartes 1-1421
(1902-1986)



Map 2209: Honeycomb map showing a synopsis of 1791 interpoint distance values. Distance index: RDV(1)_{jk}; corpus: 1687 working maps (total corpus); algorithm of visualization: MINMWMAX 10-tuple.

ALF

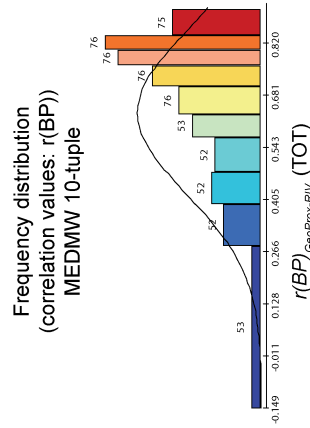
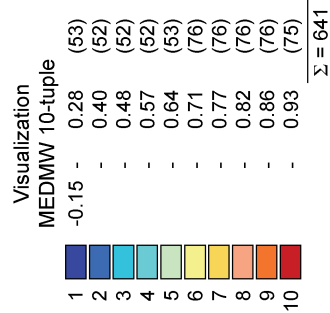
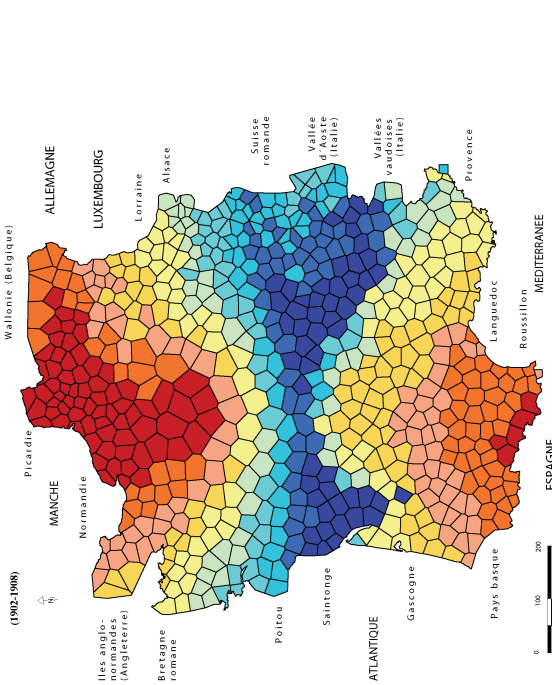
Série A: cartes 1-1421
(1902-1986)



Map 2210: Beam map showing a synopsis of 1791 interpoint similarity values. Similarity index: RIV(1)_{jk}; corpus: 1687 working maps (total corpus); algorithm of visualization: MINMWMAX 10-tuple.

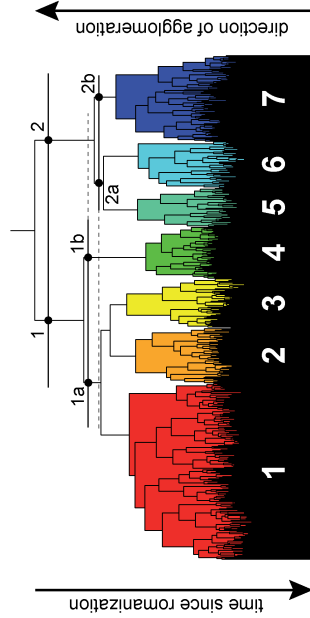
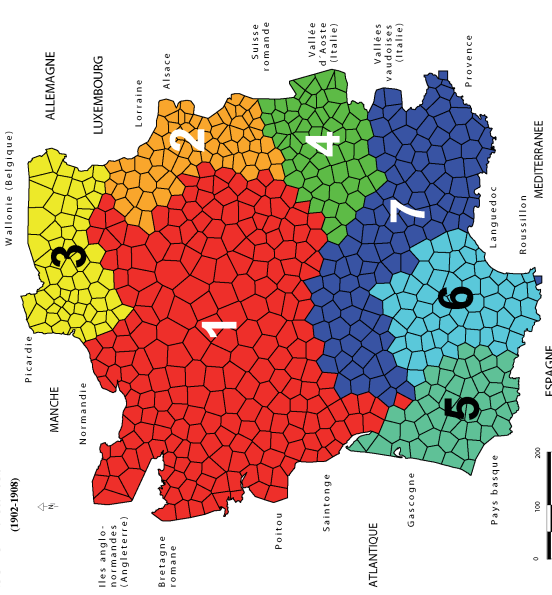
ALF

Série A: cartes 1-1421
(1902-1908)



ALF

Série A: cartes 1-1421
(1902-1908)



Map 2211: Choropleth map of the correlation values according to $r(\text{BP})_{jk}$ between 641×641 similarity values (according to RIV_{jk}) and 641×641 proximity values (according to Euclidean proximity). Corpus (of the similarity measurement): 1687 working maps (total corpus); algorithm of visualization: MEDMW 10-tuple.

Map 2212: Dendrographic classification of 641 locollects (as registered in the ALF) and spatial conversion of the colored branches (dendrogram) of the tree. Similarity index: RIV_{jk} ; dendrographic algorithm: hierarchical grouping method of Joe Ward, Jr.; number of colored choremes (above): 7; number of colored dendrems (below): 7.