ON THE CLASSIFICATION OF DIALECTS:

FROM LINGUISTIC ATLAS TO DIALECTOMETRY

Hans GOEBL [*]

Hans.Goebl@plus.ac.at

University of Salzburg

ORCID 0000-0001-5489-2858

Abstract

This article offers an exemplary presentation of the quantitative classification of dialects based on data extracted from the linguistic atlas of France ALF (Gilliéron & Edmont 1902-1910). First, the historical background of the genesis of the ALF (1897-1901) and its impact on Romance linguistic geography are described. Then some samples of the statistical and visual methods of "Salzburg dialectometry" applied to ALF data are presented.

The samples mentioned contain the following sections: preparation of the ALF data for the measurement of inter- and intra-dialectal similarity, establishment and subsequent evaluation of the calculated similarity matrix by means of the following mapping tools: similarity maps, parameter maps, isoglottic synthesis, tree analysis and correlation analysis. These mapping tools are presented in detail and the results they produce are discussed in detail. The maps presented are all coloured and all have a polygonal background map.

Keywords: dialectology, dialect classification, dialectometry

[*] Fachbereich Romanistik, Erzabt-Klotz-Strasse 1, A-5020 Salzburg.

SOBRE LA CLASSIFICIACIÓ DELS DIALECTES:

DE L'ATLES LINGÜÍSTIC A LA DIALECTOMETRIA

Resum

Aquest article ofereix una presentació exemplar de la classificació quantitativa dels dialectes a partir de dades extretes de l'*Atlas Linguistique de la France* (ALF) (Gilliéron & Edmont 1902-1910). En primer lloc, es descriuen els antecedents històrics de la gènesi de l'ALF (1897-1901) i l'impacte que tingué en la geografia lingüística romànica. A continuació, es presenten algunes mostres dels mètodes estadístics i visuals de la "dialectometria de Salzburg" aplicats a les dades ALF.

Les mostres esmentades contenen els següents apartats: preparació de les dades ALF per a la mesura de la similitud inter- i intradialectal, establiment i posterior avaluació de la matriu de similitud calculada mitjançant les eines de cartografia següents: mapes de similitud, mapes de paràmetres, síntesi isoglòtica, anàlisi d'arbre i anàlisi de correlació. Aquestes eines cartogràfiques es presenten detalladament com també els resultats obtinguts. Els mapes són tots acolorits i tots tenen un fons poligonal.

Paraules clau: dialectologia, classificació dialectal, dialectometria

SUR LA CLASSIFICATION DES DIALECTES:

DE L'ATLAS LINGUISTIQUE À LA DIALECTOMÉTRIE

Résumé

Cet article propose une présentation exemplaire de la classification quantitative des dialectes à partir de données extraites de l'atlas linguistique de la France ALF (Gilliéron & Edmont 1902-1910). Dans un premier temps sont décrites les modalités historiques de la genèse de l'ALF (1897-1901) et de son impact sur la géographie linguistique romane. Sont ensuite présentés quelques échantillons des méthodes statistiques et visuelles de la « dialectométrie salzbourgeoise » appliquées aux données de l'ALF.

Les échantillons mentionnés contiennent les sections suivantes : préparation des données de l'ALF pour la mesure de la similarité inter- et intra-dialectale, établissement et dépouillement ultérieur de la matrice de similarité calculée au moyen des outils cartographiques suivants : cartes de similarité, cartes à paramètres, synthèse isoglottique, analyse arborescente et analyse de corrélation. Ces outils cartographiques sont présentés en détail et les résultats qu'ils produisent sont discutés avec précision. Les cartes présentées sont toutes coloriées et disposent toutes d'un fond de carte polygonisé.

Mots-clefs: dialectologie, classification dialectale, dialectométrie

1. Introduction

The aim of this article is to present a general overview of the principles of the process of "dialect classification" which, in the author's opinion, are indispensable today. It deals with the following domains: a) the collection of the necessary data and

their traditional philological interpretation and evaluation, b) the subsequent *quantitative* (*polythetic*)[1] data processing and compression, and c) the final linguistic interpretation of the *quantitative* classification results.

In this respect, for point a) all the scientific standards achieved around the French linguistic atlas *ALF-Atlas linguistique de la France* (1902-1910) are considered to be exemplary and are therefore presented in more detail; for points b) and c), this concerns the dialectometrisation of about 60% of the 1,421 original maps of Series A of the ALF carried out between 1995 and 2000 within the framework of the "Regensburg-Salzburg-Dialectometry, RS-DM".[2]

Special attention should be paid to the greatest possible clarity of the terms used. This is also in consideration of the fact that - admittedly under different names - the process of "dialect classification" has been used for about two hundred years, albeit with constant evolution and differentiation of the terms and methods used. In short, the "classification of dialects" may be very old, but it also has a relatively high degree of conceptual and methodological complexity and thus also a lack of clarity, which unfortunately has often proved to be quite disruptive.

2. The history and development of dialect data collection: from "parallel texts" to "linguistic atlases"

Towards the end of the 18th century, the opinion became firmly established within and outside the circle of language amateurs and philologists that in order to study the history of various historical individual languages, not only should their respective oldest *language monuments* be consulted, but that for this purpose the

---

[1] The terminological contrast between polythetic and monothetic refers to whether many or only individual features are taken into account in the dialect classification. Clearly, only the polythetic route is taken here. See also Altmann & Lehfeldt (1973, 17-33).

[2] Since my dialectometric work was carried out at the University of Regensburg between 1973 and 1982 and at the University of Salzburg only from 1983 onwards, I will refer to it as "Regensburg-Salzburg dialectometry" (RS-DM).

respective dialects were also useful in their current form and therefore had to be researched accordingly. This view naturally also presupposed that the dialects in question were no longer regarded as worthless or even harmful "corruptions" of the individual language in question and therefore to be fought against, but as valuable witnesses or "repositories" of the past, not only of the respective individual language, but also of the entire folk culture (Droixhe 2001, Haarmann 2001 and the timetables in Pop 1950: 1179-1198) associated with it. This development ran parallel to the gradual emergence of comparative linguistics (or "Indo-European" studies) and was based on all the grammatical-linguistic findings developing within its framework.

In the process emerged the method of using "parallel texts" (such as the "Lord's Prayer" or the text of the "Parable of the Prodigal Son" (Luke 15:11-32), not only to compare several *historical languages* but also to document and subsequently compare numerous *individual dialects* of a single historical language (See the timetables in Pop, 1950: 1179-1198).[3] Classic undertakings of this kind were carried out under Emperor Napoleon by the ethnographer and administrative official Charles-Eugène Coquebert de Montbret (1755-1831) not only in France itself, but in almost all the territories then occupied by France (Pop 1950: 5 ss. and Ködl 2014).

It became necessary to collect such parallel data as *standardised* as possible, to ensure their *linguistic comparability*. This linguistic comparison developed very quickly so that not whole texts or their individual sentences were compared with each other, but also individual words and even their components (i.e., sounds, prefixes and suffixes). In doing so, mainly lexical and phonetic interests were pursued.

This led to the compilation of *questionnaires*; at the same time, however, the knowledge of the precariousness of the empirical survey of such questionnaires grew. From then on, this should no longer be done by dialect amateurs, but preferably by linguists travelling around in the field and, moreover, in direct conversation with the respondents (Pop 1950: 39 ss. and Swiggers 2010).

---

[3] Analogous initiatives were taken - among many others - for Basque by Louis-Lucien Bonaparte (1813-1891) and for Italian by Giovanni Papanti (1830-1893) on the basis of more than 700 translations of a novella (i.e. "Novella IX della Giornata I del Decameron") by G. Boccaccio (Papanti 1875). See also the comprehensive accounts by Bourguet (1988), Leclerc (1979) and also Stagl (1995).

In the various modern philologies, there were numerous pioneering undertakings in the entire 19th century, which on the one hand were geographically oriented and on the other hand represented a mix of *lexicographic* (i.e., collection of "interesting" words and idioms) and *geolinguistic* (i.e., discovery of linguistic variability in space) interests. In this respect, post-revolutionary France (Second Empire and Third Republic) and various German states (Switzerland, individual states of the German Confederation and Habsburg Austria) were leading the way (Goebl 2017).

Already in the 18th century, within the framework of contemporary "statistics",[4] such large-scale surveys had been carried out on many anthropic domains (such as economy, trade, transport, demography, etc.) and had everywhere brought to light a previously unknown variation and variability, knowledge of which proved very useful for the respective states. In the first half of the 19th century, a general awareness developed that improved knowledge also of the dialectal and ethnographic diversity of one's own state territory, adding a public interest to the scientific value.

All this happened against the background of the slowly developing general schooling and the parallel upraise of linguistic comparison between the respective language of instruction and the local dialects that inevitably took place in many school classes.[5] This contrastive linguistic comparison, however, led to highly divergent challenges and results - depending on the cultural and/or national embedding. These were more complex and conflictual in basically *dialectophobic* societies - such as those of Romania - than in traditionally *dialectophile* regions such as in Switzerland, southern Germany or in the German-speaking parts of the former Habsburg Austria.

To sum up: at the end of the 19th century, the political relevance of dialects and thus also of the emerging field of dialectology was well known in many European countries.

---

[4] In this context, this is understood to mean research into the numerical dimensions of entire *states*, which was cultivated above all in the 18th century, and not yet - as is generally the case today - a purely mathematical discipline. See also Bourguet (1988), Leclerc (1979) and Stagl (1995).
[5] It should be emphasised here that the linguistic atlas projects carried out everywhere in the 19th century could never have been realised without the solid bilingualism of the informants (consisting of the local dialect and the respective school language).

3. The genesis of the ALF (*Atlas linguistique de la France*)

The history of the ALF cannot be separated from the biography of its author - Jules Gilliéron (1854-1926) (cf. Pop 1959, passim). Gilliéron came from French-speaking Switzerland and, although he taught at the "École Pratique des Hautes Études" in Paris from 1883 onwards and was thus in contact with all the intellectual luminaries of France, he remained closely connected with the philological traditions of his homeland. Gilliéron approached linguistic atlas work in several steps: by publishing a small regional atlas for the French part of the Swiss canton of Valais (cf. Gilliéron 1880), by systematically practicing field work using a phonetic transcription specially developed for this purpose, and by publishing a journal (albeit rather short-lived) associated with this new scientific practice: "Revue des Patois Gallo-Romans" (RPGR), 1883-1893). In addition, Gilliéron maintained close contact with Jean-Pierre Rousselot (1846-1924), a very active French dialectologist and phonetician who had introduced the then emerging method of instrumental sound recording and analysis into dialectology.

During his various inquiries and their subsequent evaluation, Gilliéron was already struck in the 1880s and 1890s by the "spontaneity", not to say the "asystematicity" of the data collected in the field, which - as was increasingly claimed in linguistics at the time - was therefore considered "unclassifiable".[6] Moreover, in France, traumatised in national terms by the lost war of 1870-71, the question of the demarcation between the two main historical languages of Gallo-Romania, i.e., the *Langue d'Oïl* (or French) in the north, and the *Langue d'Oc* (or Occitan) in the south,

---

[6] Characteristic of this is Hugo Schuchardt's habilitation lecture from 1870 („Über die Klassifikation der romanischen Mundarten"), at the end of which Schuchardt speaks out against the possibility of a scientific classification of Romance dialects: „Ich habe *über* die Klassifikation der romanischen Sprachen sprechen wollen, und ich habe *gegen* sie sprechen müssen; ich glaube aber dass selbst dieses negative Resultat insofern ein Gewinn ist als dadurch die Erklärung von nicht wenigen Erscheinungen in der Geschichte dieser Mundarten erleichtert wird." (Schuchardt 1870/1900: 188) ["I wanted to speak *about* the classification of the Romance languages, and I had to speak *against* it; but I believe that even this negative result is a gain insofar as it facilitates the explanation of not a few phenomena in the history of these dialects."].

had been raging since 1876. This was between some Parisian professors (among them the philologists Gaston Paris [1839-1903] and Paul Meyer [1840-1917]) and a group of intellectuals from southern France (among them the philologist Charles de Tourtoulon [1836-1912] and the doctor and philosopher Joseph-Pierre Durand (de Gros [1826-1900].[7]

While the southern French philologists stressed (with very good arguments) the scientific delimitability of the Occitan dialects, Gaston Paris postulated (especially in the context of a major address delivered in 1888 in Paris to a large audience) three issues: (a) the inconspicuous (and thus "imperceptible") merging of French and Occitan dialects, (b) the consequent non-existence of "dialects" (understood as geographically compact idioms in the traditional sense), and (c) the need to focus, therefore, within France on collecting precise geographical inventories for the largest possible number of individual linguistic features .

Postulates a) and b) discredited completely all further typological-classificatory discussions on the dialectal fragmentation of France, while postulate c) encouraged Gilliéron in particular to intensify his field surveys. Thus, between 1888 and 1897, Gilliéron became acquainted with Edmond Edmont (1849-1826),[8] a Picardy grocer and local historian, with whom he established the guidelines for a large-scale enterprise covering the whole of France (including the French-speaking peripheral zones: Wallonia, French-speaking Switzerland and western Piedmont):

- with regard to the network density: 5-6 inquiry points (or: localities) per département;

- with regard to the questionnaire: initially 1,421 questions (for single words, but also whole sentences) in thematic order;

---

[7] See Tourtoulon (1876 and 1890), Durand de Gros (1889) and our historical overviews from 2004 (b) and 2018 (b). There is also one in Brun-Trigaud (1990). In addition, I would like to mention a brilliant handbook on "taxinomie" by J.-P. Durand (de Gros) from 1889, which unfortunately remained completely unknown among the dialectologists of the time.
[8] Edmont was the author of a dialect dictionary of the Picardian dialect of his native village of Saint-Pol-sur-Ternoise (Edmont 1897). This village became later the inquiry point 284 of ALF.

- with regard to the number of informants: for each inquiry point, one informant to be regarded by Edmont as dialectally "competent";

- with regard to the modalities of data collection: recording of the first given answer of the informant ("traduction du premier jet") by means of the phonetic transcription already used in the RPGR and also by Rousselot himself (cf. Pop 1950: 39-44).

The inquiries were carried out between 1897 and 1901 by Edmont in the course of eight geographically precisely planned campaigns in 638 localities. [9] The questionnaire of 1,421 questions (Series A), which had already been used at the beginning of the project, was used everywhere; however, there were two extensions, with additional (rounds of) data collection: a first extension of 326 questions (Series B) throughout southern France at 326 (of 638) inquiry points; a second extension of 173 questions in south-eastern France at 204 (of 638) localities (Details can be found in the ALF Notice -1902 and in the ALF monograph by Brun-Trigaud, Le Berre & Le Dû 2005).

Today, the procedural modalities of Edmont's fieldwork are well known.[10] He conducted the interviews on the basis of a handwritten version of the original questionnaire by Gilliéron and noted down - clearly in the sequence of the questions of the original questionnaire - the answers given by the informants by means of phonetic transcription on new (and thus empty) sheets of paper. He then sent the completed sheets to Gilliéron in Paris. They have been preserved in their entirety at the "Bibliothèque Nationale" in Paris.[11]

The first of the 35 fascicles of the ALF published until 1910 was printed in 1902, only one year after the end of all fieldwork.[12] In addition to Gilliéron, Georges Protat,

---

[9] In these 638 localities, 639 inquiries were carried out. This is explained by the fact that at ALF point 284, Saint-Pol-sur-Ternoise, E. Edmont's hometown, two surveys were carried out by him (in the centre of the town and on the periphery).

[10] This is due in particular to the research of Patric Sauzet (Toulouse). See: http://symila.univ-tlse2.fr/alf/questionnaire_alf_reconstitue.

[11] See also: https://data.bnf.fr/en/15078566/jules_gillieron_atlas_linguistique_de_la_france/.

[12] In 1968, a reprint of the ALF was produced by the Italian publisher Forni (Bologna). The reprint format chosen (44 by 31 cm) was about 52 by 33 cm smaller than that of the original edition. Today, the ALF can be consulted on the web via two links: 1) via the University of Innsbruck: https://diglib.uibk.ac.at/urn:nbn:at:at-ubi:2-4568 and 2) via the University of Toulouse (project: SYMILA): http://symila.univ-tlse2.fr/alf.

one of the owners of the printing house where the ALF was printed, undoubtedly contributed to this typographical masterpiece. He optimally organised the typesetting, the subsequent correction and the final cartography of the typescript data, as can be read in detail in the introductory volume of the ALF ("Notice servant à l'intelligence des cartes", Paris 1902: 9-10).

It should be noted that the publication of the dialect data collected by Gilliéron in a geographical format, which he stubbornly[13] pushed through against all odds (cf. Storost 1994), was a real innovation at that time. It was only the general availability of the new data in a geographical - and not just tabular - order that made the great scientific success of the ALF possible, which began immediately after the publication of its first fascicle.

In the last quarter of the 19th century, there were many systematic field surveys in and outside France using questionnaires and phonetic transcriptions, but they usually did not result in the production of one map per item, as in the case of ALF, but (only) of one column (with phonetic information) per item (Gartner 1882 and 1883, and Ettmayer 1902). Clearly, the presentation of *de facto* spatially staggered data in column format did not have the same heuristic and suggestive impact the dialect maps of the ALF immediately had.[14]

4. The subsequent evaluation of the published data of the ALF

4.1 The qualitative *phase*

From 1902 onwards, the reception of the individual fascicles of the ALF by the academic world began abruptly. This was not only instigated by Gilliérons's scientific

---

[13] This resistance came mainly from Paul Meyer: see the original documents published by Jürgen Storost (1994).
[14] This is true also for the "silent maps" used since the very beginning for the linguistic evaluation of the original ALF maps.

writings based on the ALF[15] and by other researchers inflamed by the new "géographie linguistique"[16] (such as Jaberg 1908), but also through the rapid diffusion of the technique of evaluating or "discussing" original maps of the ALF through the classificatory transfer of their contents to "silent maps" (*silma*).

These *silma* were blank maps that contained only the basic network of the ALF, consisting of the departmental boundaries and the numbers of the 638 inquiry points (see Goebl 2018 for more information). They were sold in large quantities (dimensions: 33 by 27.7 cm) from 1902 by the printing house Protat Frères parallel to the individual fascicles of the ALF.

The task of the user of the ALF was now to fill in one of these *silma* with suitable colour or black-and-white symbols in the most heuristically appealing way possible, after determining the linguistic objective of his analysis of an original map of the ALF. Thus, the ALF users (or linguistic geographers) of the time advanced *eo ipso* to data analysts or classifiers and to cartographers.

These new challenges and possibilities were fully accepted by Romance scholars both inside and outside France and were also used in the publication of the new knowledge gained in this way. In addition to the *silma* offered by the printing house Protat Frères, which were tailored due to their very sparse topographical design to a circle of users optimally familiar with the geography of France, another prototype was launched in Germany by the Hamburg Romance scholar Bernhard Schädel (1878-1926), which expressed the political and topographical conditions of France more clearly or in greater detail (see a facsimile of this German *silma* in Goebl 2018: 48).

---

[15] See above all the 1905 study on *scier* written with Jean Mongin, in which all the foundations of the *géographie linguistique* conceived by Gilliéron (always with a view to the ALF) are to be found. Furthermore, I refer to his monographs from 1912 (*Études de géographie linguistique*), 1918 (on the designations of the abeille ["bee"]) and 1919 (on the *faillite de l'étymologie phonétique*). The dual term *géographie linguistique*, often used by Gilliéron himself and also his followers, did not originate with Gilliéron; it was already used in 1895 by his opponent Paul Meyer.

[16] This concerns above all Karl Jaberg (1906 and 1908). Gilliéron's regular courses on *géographie linguistique* at the "École Pratique des Hautes Études" in Paris, in which practically all the leading figures in Romance studies at the time took part, were also very important; for a list of these participants see Pop & Rodica 1959: 53-63.

For purely geometrical reasons, the variations found on an original map of the ALF could be indicated on a *silma* by means of areal or linear signatures (or symbols). In both cases, the map theme to be conveyed in this way could remain on the purely qualitative level, but it was also possible - admittedly at the price of a more elaborate cartography - to present a quantitatively oriented map theme on it. Unfortunately, this only happened very rarely.

The cartography used by Gilliéron in his geolinguistic writings remained a qualitative area-based cartography, and the respective areas were even published in colour before the First World War. As a curiosity, it should be noted that most French geolinguists[17] have actually followed him to this day in their exclusive preference for area-based cartography.

In contrast, German (e.g., Fleischer 1913 or Ettmayer 1924) and Scandinavian geolinguists (e.g., Rosenqvist 1919) observed very early the phenomenon of what was later called "isogloss bundles"[18] as the result of the fact that the *border lines* ("isoglosses") of not only individual items can be drawn on the *silma*, but also, of several distribution areas,[19] in particular when they were used in a line-oriented manner. This opened the way to a quantitatively oriented evaluation of the contents of several ALF maps on one silma, admittedly only via their line-oriented application.

Unfortunately, the geometrical (and logical) problems associated with the line-oriented combination method were not fundamentally analyzed or discussed within Romance studies,[20] either then or later. This did not happen until about 80 years later in the context of dialectometry (cf. Goebl 1984). However, as early as 1898, a German dialectologist, Carl Haag (1860-1946), recognised and presented the geometrical problems associated with the drawing of such boundary lines in an astonishingly clear-sighted manner.

---

[17] Pierre Gardette (1906-1973) and the Lyonese school of dialectology founded by him were an exception: see Gardette 1941. Very instructive are also the numerous pictorial quotations in Jochnowitz (1973), where the main issue is the "isoglottic" demarcation" of Francoprovençal.

[18] This line-based phenomenon has also become familiar to Germanists in their work on the DSA data.

[19] These distribution areas correspond exactly to the "taxatorial areas" (TA) of RS-DM.

[20] This also affected - apart from Carl Haag - German studies, which concentrated primarily on the cartographic output of the DSA: cf. on this H. Girnth (2010).

Haag discovered that due to the point-based structure of his linguistic maps, the border lines (since 1892 called *isoglosses*, see below) to be drawn on them can be defined from a geometric perspective as a sequence of polygon edges (or: polygon sides). He used a simple geometric procedure for this, which today is usually referred to as "Voronoi geometry". This procedure consists of first triangulating the grid of inquiry points (called today: "Delaunay triangulation") and then constructing the respective line bisectors on each of the resulting triangle sides (see Goebl 1983 for more details). Through the subsequent extension of these bisector lines up to their reciprocal fusion, the entire area under investigation can be covered by a closed *polygonal network* (cf. Okabe, Boots & Sugihara 1992 [2000²]).

The term *isogloss*, which is used today in many languages to designate these (individual or combined) boundary lines, goes back to the German-Baltic pastor, ethnographer and linguist August Bielenstein (1826-1907), who coined it in 1892, with deliberate reference to the meteorological term *isotherm*, which already existed at that time.

Isogloss was initially to be understood in a purely cartographic sense and spread only slowly. After its general establishment in the first decades of the 20ᵗʰ century it was used more and more frequently to designate (geo)linguistic features, especially in the field of Indo-European studies, by way of a curious metonymic transfer (Goebl 2004a). Ultimately, this was due to a misunderstanding of the underlying formal facts. This confusion arose from the fact that the first experiences with the drawing of individual and combined isoglosses were made on the basis of elongated lines which traversed - and thus subdivided - the studied area in the manner of elegant "garlands" from one end to the other. In this way, arose the false impression that the linguistic features lying on either side of these continuous "garlands" (= isoglosses) were themselves "isoglosses". The fact is, however, that the shapes, sizes, and courses of the isoglossic lines serving to border feature areas are very diverse, just like these areas themselves, and therefore by no means always resemble elongated garlands.
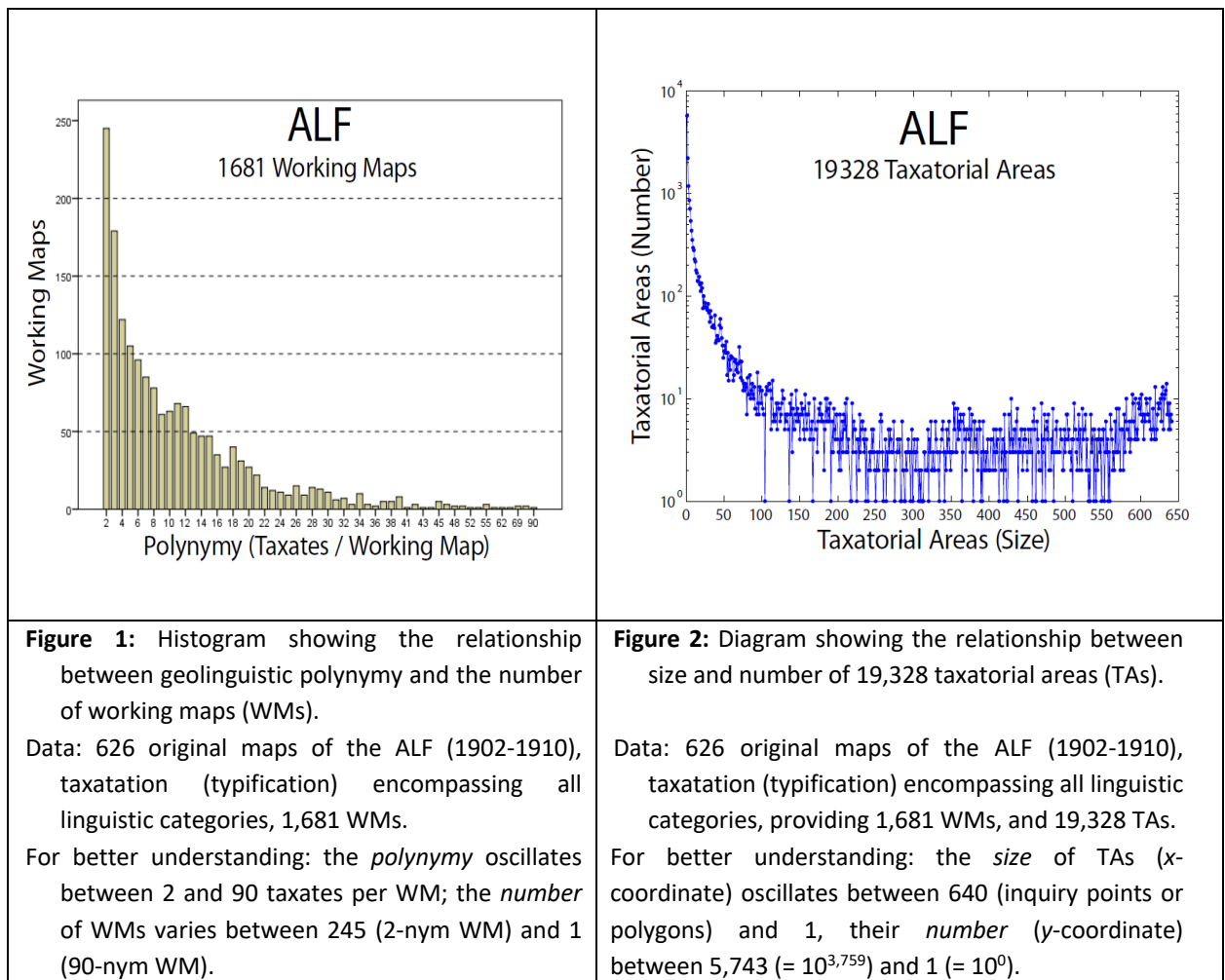
I illustrate this fact with the numerical balances (Goebl 1984, 2002, 2003, 2006, 2007, 2010 and 2013a) I obtained during the dialectometrisation of the ALF (1996-

2000) when I assessed more than 600 original ALF maps referring at the same time to the three "working maps" (WM) shown in Plate 1.

- Number of dialectometrically analyzed original maps of the ALF: 626

- Number of *working maps* (WM) obtained by linguistic analysis ("taxatation", based on phonetics, morphosyntax and lexicon): 1,681

- Number of feature areas ("taxatorial areas" - TA) occurring on these 1,681 WM: 19,328

- Range of variation in the size of these 19,328 TA: from 1 to 640 inquiry points or polygons (see Figure 2)

- Internal variability ("polynymy") of the 1,681 WM: between 2 and 90 taxates per WM (see Figure 1).

It can be seen from this - and from looking at Figures 1 and 2 - that only relatively few of the 1,681 working maps and the 19,328 taxatorial areas occurring on them are sufficiently uniform[21] or large enough to permit the drawing of elongated or garland-like isoglosses.

---

[21] In Salzburg, rather uniformly structured working maps (WM) are called *oligonymous*, their opposite *poikilonymous*.

Figure 1: Histogram showing the relationship between geolinguistic polynymy and the number of working maps (WMs).

Data: 626 original maps of the ALF (1902-1910), taxatation (typification) encompassing all linguistic categories, 1,681 WMs.

For better understanding: the *polynymy* oscillates between 2 and 90 taxates per WM; the *number* of WMs varies between 245 (2-nym WM) and 1 (90-nym WM).

Figure 2: Diagram showing the relationship between size and number of 19,328 taxatorial areas (TAs).

Data: 626 original maps of the ALF (1902-1910), taxatation (typification) encompassing all linguistic categories, providing 1,681 WMs, and 19,328 TAs.

For better understanding: the *size* of TAs (*x*-coordinate) oscillates between 640 (inquiry points or polygons) and 1, their *number* (*y*-coordinate) between 5,743 (= $10^{3,759}$) and 1 (= 10^0).

Moreover, some Romanists (Jaberg 1908: 5-6; Wartburg 1963: 22-24 (including a map)) noticed very early on, when filling and colouring "silent maps", that even in the case of categorically identical or thematically very similar original maps of the ALF, the isoglosses that could be derived from them always showed de facto divergent courses and thus by no means proved to be congruent - as would have been expected on the basis of the then prevailing theory of "regular sound change".

In the two-dimensional scheme of a dialectometric data matrix (see Figure 3), these incongruities give the impression of a "(special) entanglement" of the individual taxatorial areas (or dialect features) (Goebl & Smečka 2016b).[22]

---

[22] The "special entanglement" is in a certain manner reminiscent of the "particolar combinazione" of language features postulated in 1876 by the Italian linguist G. I. Ascoli (1829-1907) to identify and differentiate individual dialects.
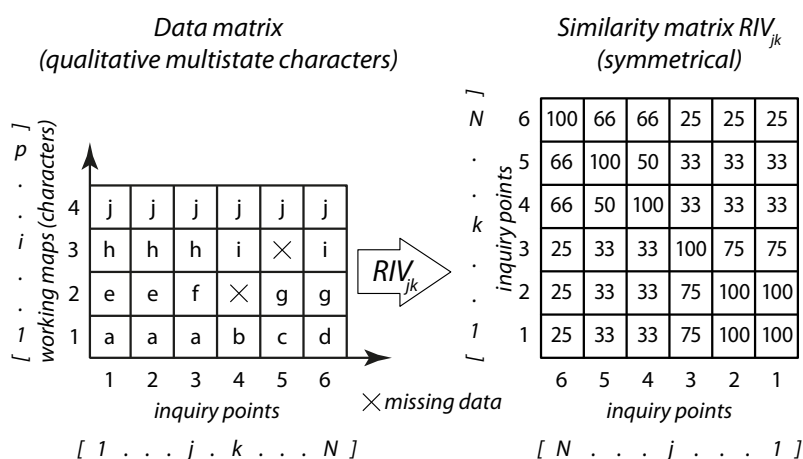
Figure 2

Figure 3. Data matrix and similarity matrix. Scheme of calculation of the interdialectal similarities via RIV$_{jk}$ (« Relative Identity Value").

For better understanding: calculation of the RIV between the inquiry points 2 and 3:

Σ of (pairwise) taxatorial co-identities (COI): 3
Σ of (pairwise) taxatorial-co-differences (COD): 1

RIV$_{2,3}$ = 100 × Σ COI / Σ COI + Σ COD = 100 (3 / 3 +1) = 75%

It should be emphasized here that this "special entanglement" of the feature areas (which initially appears aberrant from the then prevailing perspective of the supposedly regular sound laws) and the associated non-coincidence of the isoglosses surrounding these areas are *constitutive characteristics* (universals) of all linguistic atlases and therefore also of the data contained therein.[23]

In this way, however, an experience has been repeated in atlas-based linguistic geography that has been known in lexicography in principle since the times of Jakob

---

[23] The non-coincidence ("indépendance") of the feature areas was already brilliantly addressed several times around 1950 by the southern French linguist Théobald Lalanne (1880-1952): cf. Lalanne (1949-1952, 1953). In this context, however, I would also mention that this fact was already addressed very clearly in 1920 by the Spanish philologist Ramón Menéndez Pidal (1869-1968) in comparative studies of the text genre of the *Romanceros*. In this context the latter, like those mentioned above, functioned as "parallel texts".

Grimm (1785-1863), namely that "every word [has] its own history" (cf. Malkiel 1967, Christmann 1971). Based on the experiences made with the new linguistic atlases, this slogan, which refers to the vocabularies, should have been supplemented for linguistic atlases by the following: "Each feature area has its own geography". Unfortunately, this did not happen with such precision at the time.

Despite many individual observations and acute analyses of eye-catching feature areas (See the bibliographical overview in Chapter III in Iordan (1962: 171-322), the quantitative circumstances of their special entanglement remained largely unknown and unreflective until the 1980s.

However, since linguistic atlases are always representations of an underlying reality, one should ask about the (causative) reasons that generate this reality. In our opinion, these reasons lie in a particular kind of linguistic behaviour of speakers interacting linguistically (better: "basilectally") with each other over a geographical network. Since 2005, we have been using the term "basilectal management of space by its inhabitants" in this context. Ultimately, these "inhabitants" produce a special kind of "texts" during their basilectal interactions, which in turn have intrinsic regularities as known for "normal" texts from "quantitative linguistics" (see the review articles by Mehler (2005) and Altmann (1983, 2005)).

4.2 The quantitative *phase*

First of all, with regard to the left half (= data matrix) of Figure 3, it should be noted that in Romance studies since the ALF, all linguistic atlases formally correspond to a two-dimensional scheme with the following dimensions:

- horizontal (along the *x*-axis): *N* inquiry points

- vertically (along the *y*-axis): *p* atlas maps.

It is also undisputed that the efforts of the pioneers of linguistic geography were quite rightly directed towards maximizing the dimensions *N* and *p* obviously according to their own logistical and economic possibilities. With regard to the size of *N*, Gilliéron was able to use the experience that had been gained in France since 1790 with the

départment network (see also Bourguet 1988) created by the French Revolution. At the same time, he took advantage of the "drawing of a random sample" - admittedly without knowing it or wanting to. Sampling was precisely what his German geolinguistic colleague Georg Wenker (1852-1911) could not do either, who - having grown up in a completely different cultural-political tradition - made the mesh density of the survey network of his "Deutscher Sprachatlas" (DSA) dependent on the existence of a local primary school, and at that time there was one in almost every little village in Germany.

So Wenker came up with more than 50,000 inquiry points and had therefore enormous cartographic problems in the subsequent processing of this huge amount of data. These were also the main reason why the data of the DSA were never published in their original form (and thus made generally available in detail), as was the case with the ALF since the very beginning (See also our comparatively oriented contributions from 2002b, 2006b, 2007 and 2011).

Even if the dimensions of the ALF are "only" N = 638 inquiry points[24] and p = 1,421 (Series A) atlas maps, this nevertheless results in mass data whose structures - in this respect quite comparable with the oceans - have "surfaces" (= the 1,421 individual maps) that are clearly visible to the naked eye, but also have "deep structures" that are inaccessible to the naked eye and can only be detected by means of suitable procedures.

The most important task of the quantitative evaluation of the mass data stored in the linguistic atlases is to explore (and make visible) precisely these deep structures. It was exactly this impulse that Jean Séguy (1914-1973) and Henri Guiter (1909-1994), the French pioneers of dialectometry, felt in their *quantitative experiments* using the ALG ("Atlas linguistique et ethnographique de la Gascogne", with J. Séguy) and the ALPO ("Atlas linguistique des Pyrénées Orientales", with H. Guiter) (see Séguy 1971 and 1973, on the ALG, and Guiter, 1973, on the ALPO). The French term *dialectométrie*,

---

[24] It is worth recalling the difference between the number of *inquiry points* (= 638) and the number of *inquiries* (= 639) of the ALF.

which has been adopted or translated into many languages today, was deliberately "invented" and launched as a neologism by Séguy in 1973. I myself adopted it in the same year for my "pre-dialectometric" research, which had been in progress since 1971 (see Goebl & Winterleitner 1971), and used it in 1974 when presenting my first dialectometric paper at the International Congress of Romanists in Naples (Goebl 1976).

4.2.1 Two scientific amalgams: *linguistic geography + statistics* and *linguistic geography + cartography*

Since 1973, linguistic geography (or: geolinguistics) has thus found itself in a new interdisciplinary context: a) with various quantitative disciplines, which I will refer to here simplistically as "statistics",[25] and b) with very elaborate cartographic challenges (cf. Dickinson 1972 and Brunet 1987), which go far beyond filling in the *silma* mentioned above. In both cases, the methodological and theoretical links with linguistic geography, which grew between 1902 (ALF) and 1973 (Séguy), are so close that one can rightly speak of a *connubium methodicum*, i.e., a marriage-like fusion of methods.

In addition, it should not be forgotten that since the 1960s, the then brand-new (and at first still rather cumbersome) computers have spread very rapidly to the universities, so that - as in my personal case - continuously improving possibilities for the efficient processing of mass data (at first only in numerical, but later also in [carto]graphic terms) became available. This, however, posed the problem of not only adapting this new technology to the needs of geolinguistics, but also, in view of its almost inexhaustible potential, of taming it to such an extent that geolinguistics could profit from it and did not have to suffer from the onslaught of the new machines.

I recognised this benefit-damage ambiguity very early on, especially in the field of computerised statistics, and therefore always endeavoured to a) select and b) use

---

[25] This is a mathematical discipline and no longer the "science of states" cultivated in the 18th century.

the statistical instruments used exclusively according to the requirements of linguistic geography. My aim was to protect linguistic geography as a linguistic discipline and its representatives from the impact of "statistical rock falls".

Of course, given the old *anti-mathematism* endemic in the humanities, certain "frictions" with conservative colleagues were unavoidable or even predictable; but their extent could be kept within relatively narrow limits by "accompanying measures" - by which was meant a "shock-absorbing" terminology and discourse practice.[26] Unfortunately, in the last 50 years, not all adepts of the new dialectometry have made this "moderation" a primary concern, so that there was regrettably the (metaphorical) departure of some of the "rock falls" evoked above.

4.2.1.1 Linguistic geography and *statistics*

First, it must be made clear that in the present case the use of statistics can only serve to make visible (or: detect) structures hitherto hidden but already present in the data, which in turn should be plausible or easily interpretable from a linguistic point of view. However, the use of statistical methods must not lead to a situation where the geolinguistic data to be analysed are condensed into numerical artefacts which can hardly or no longer be interpreted linguistically. This dilemma corresponds to the contrast between discovering and inventing data-inherent (deep) structures, which is well known also in other empirical disciplines.

From a statistical point of view, RS-DM was modelled on the "Numerical Classification", which was already well developed in the 1970s, and which originated primarily in the field of biology and was further developed in that area (Sneath & Sokal 1973, Bock 1974, Chandon & Pinson 1981). Apart from the great methodological usefulness of the statistical methods used there also for linguistic geography, the research ideal cultivated by many biological taxonomists - namely the historical

---

[26] In general, RS-DM was very warmly received abroad and was also "exported" through the takeover of VDM. Curiously, however, this did not apply to Italy, where, for reasons I still do not understand, there are still strong reservations about DM to this day.

reconstruction of the results of biological evolution and the respective visualisations - basically also corresponded to that of the linguistic geography and dialectometry as practiced in Salzburg.

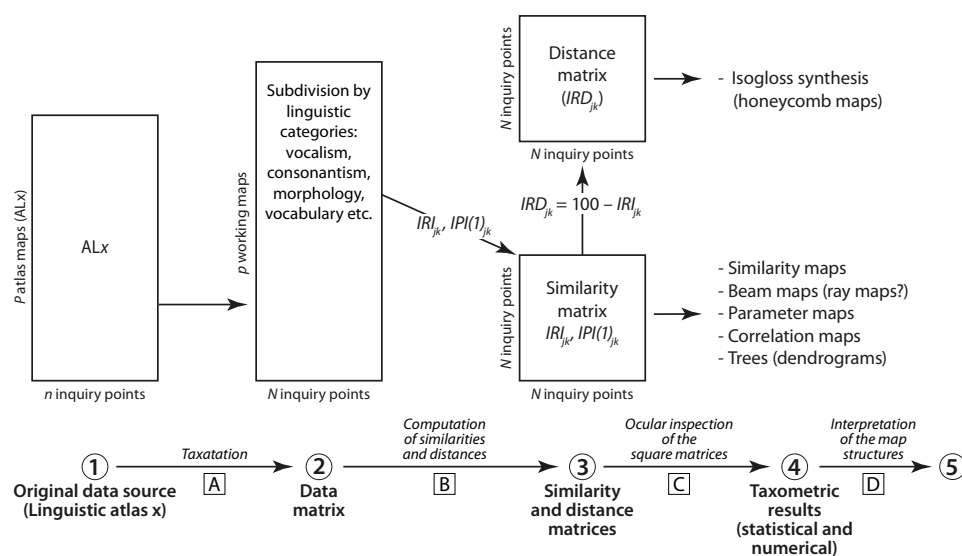The associated methodological steps can be represented graphically as shown in Figure 4.

Figure 1

Figure 4. Flow chart of the methods of the Regensburg-Salzburg-dialectometry (RS-DM)

I will make a few comments on the five steps noted in Figure 4:

- Steps 1 and 2: serve to create the data matrix. The methods applied - apart from the formal integration into the form of a (two-dimensional) data matrix - largely correspond to the philological methods used by Gilliéron and his successors. Their designation as *taxation* is based on the methodological (and terminological) proximity to the "Numerical *Taxonomy"* mentioned above.

- Steps 2 and 3: serve to measure the pairwise *similarities* between the *N* locality vectors. The selection of the *similarity measures* (or: *indexes*) used for this is delicate: it

must be done - as is the case with "Numerical Classification" (or "Taxometry") - with close attention to the "nature" of the data stored in the data matrix.[27] After many experiments, I finally decided to carry out this similarity measurement using several similarity measures,[28] mainly in order to be able to compare the results obtained with this multiplicity.

- Step 3: is used to derive a square matrix of *distances* (*dist*; between the *N* locality vectors) from the square matrix of *similarities* (*sim*) created during the similarity measurement.

- Step 4: the primary concern here is the multiple processing of the data stored in the two matrices. Only through this diversity one can obtain a sufficiently accurate representation of the depth structures existing in the atlas data. I refer to the "samples" given here in the appendix: these refer - apart from the exemplary presentation of three working maps - to the following five map types: *similarity maps* (Maps 2a, 2b and 2c), *parameter maps* (Map 3), *honeycomb maps* (Map 4), *tree graphs* (Map 5) and *correlation maps* (Map 6). Since each of these five cartographic analyses shows different aspects of the already mentioned deep structures, they should be regarded as complementary to each other.

- Step 5: all numerically calculated and graphically represented (i.e., visualised) results have to be discussed and interpreted from a primarily linguistic point of view and - whenever possible - also explained accordingly. Care is always taken to establish an organic relationship between the new (dialectometrically determined) findings and the already existing older concepts and views.

---

[27] The properties measured on the objects (= local dialects, locolects, basilects) should be relevant for the course of language history. Thus, the *joint* occurrence of the processes of *deletion, insertion* and *substitution* used in the Levenshtein measurement has never been empirically observed in the history of language. This is the main reason for my deep skepticism towards this method (which is very often used today): cf. Heeringa (2004).

[28] These are *qualitative* (such as RIV and WIV) and *quantitative* measures (such as AEM, AMM or rBP). All these indexes are contained in our DM-software VDM. (WIV [GIW] = Weighted Identity Value; AEM [DEM] = Average Euclidean Metric, AMM [DMM] = Average Manhattan Metric]. The (German) abbreviations between square brackets occur regularly in our VDM software.

4.2.1.2 Linguistic geography and *thematic cartography*

It has already been mentioned that the great (theoretical and heuristic) success of the ALF is causally related to the cartographic (and precisely non-tabular) presentation (and publication) of the data collected by Edmont. But this also included the systematic evaluation (and processing) of these original maps of ALF by means of the already mentioned *silent maps* (*silma*).

The continuous recourse to the *base map* (or: grid) of the ALF - here realised in polygonised form - is thus not only a basic feature of Romance linguistic geography per se, but also a proprium of RS-DM. It is therefore always endeavoured to visualise the calculated results not via any tables, curves or quadrants, but via the (polygonised) *base map* of the ALF (and other linguistic atlases).

In addition, care was always taken to design these visualisations according to the standards customary in international cartography (Cf. for example the manuals by Dickinson 1973 & Brunet 1987) and to always use colours and - wherever possible - no black-and-white[29] hatchings for this purpose in a precisely controlled form.[30]

5. Presentation and interpretation of Plates 1-6[31]

In the following, eight coloured plates (or map triples) are presented in sequence. Seven of them belong to five measurement moments of RS-DM (and the VDM program). It should be emphasised here that these methods not only involve a classification but also an exploration of the deep structures of the data contained in

---

[29] As a rule, *black-and-white* hatchings do not allow the clear reproduction (and recognition) of finely structured spatial gradations.

[30] The numerical and graphical standards listed in chapters 4.2.1.1. and 4.2.1.2. were all implemented in the Salzburg dialectometry software VDM ("Visual DialectoMetry"). It was created around 1996 by Edgar Haimerl, and updated and expanded several times. The VDM software is freely available to interested parties. However, its use is not self-explanatory and therefore requires a short introduction that can be given at any time in Salzburg.

[31] The results shown on maps 2-6 can be reproduced or checked interactively on the following webpage: http://dialektkarten.ch/dmviewer/alf/index.en.html.

the data matrix (see Figure 3, left half). Each of the seven maps (2-6) used to visualise certain measurement results serves to visually convey a very specific topic, which is located at the interface between *geolinguistics* and *Numerical Taxonomy*.

5.1 Presentation and interpretation of Plate 1 (three working maps)

One sees on Plate 1[32] three categorically different choropleth maps, each of which had been obtained by "taxatation"[33] from a very specific original map of the ALF:

Far left: *vocally* relevant working map (WM) to map 812 of the ALF *le marché* (< Latin MERC**Á**TU "[the] market"). This WM shows the spatial distribution of 17 different (phonetic) results of the stressed **Á** in the lat. nexus -**Á**TU. The numbers next to the 17 phonetic results are at the cardinal (or: nominal) measurement level and serve only to identify them.

---

[32] De facto, Maps 1 to 6 correspond to the *triptycha* known from the fine arts through their threefold internal structure.

[33] It should be remembered here that the process of "taxatation" was already practiced more than a hundred years ago by J. Gilliéron and or K. Jaberg in a completely analogous form, but under different names ("typification", "discussion" etc.).
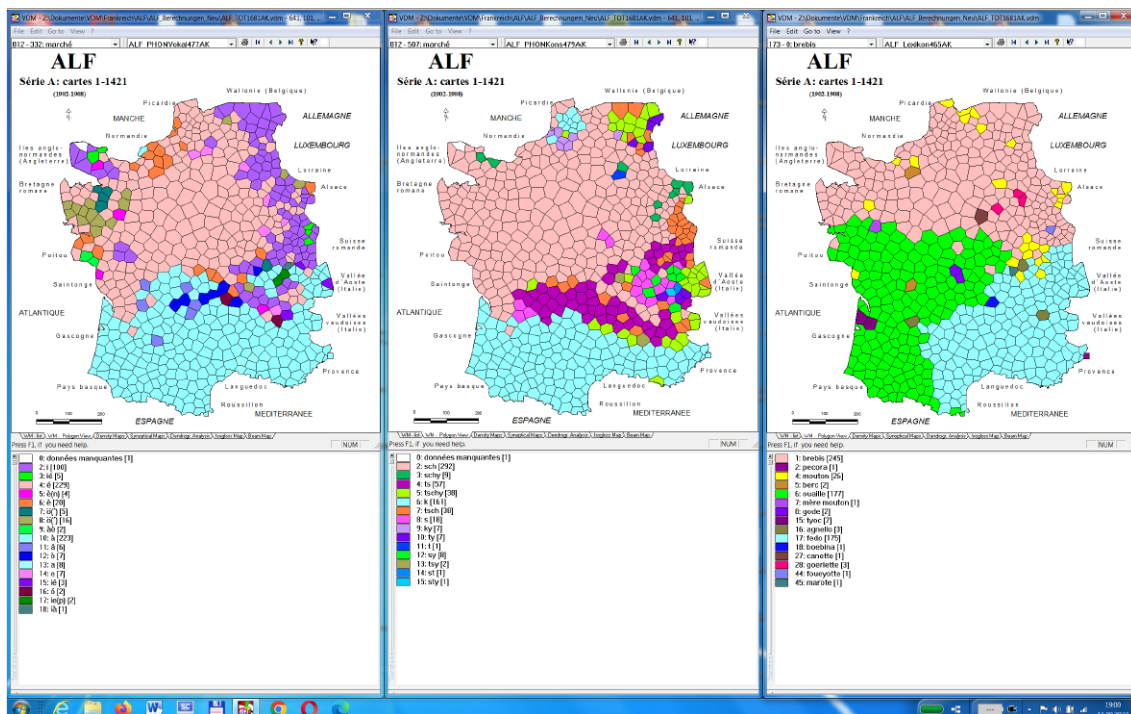
Plate 1. Three samples of working maps (WM).
  - cartographic status: *qualitative* choropleth maps.
  - for better understanding: *données manquantes* = missing data

Left: *phonetic* WM showing the spatial distribution of the Gallo-Romance results of **-ÁTU** in the Latin etymon MERC**ÁTU** "market" (following ALF 812 *le marché*)

Middle: *consonantic* WM showing the spatial distribution of the Gallo-Romance results of **C+Á** in the Latin etymon MER**CÁ**TU "market" (following ALF 812 *le marché*)

Right: *lexical* WM showing the spatial distribution of the Gallo-Romance designations of the "ewe" (following ALF 173 *la brebis*).

Between square brackets is the set of polygons (or inquiry points) covered by the respective "taxate". The two most frequent taxates have the numbers 4 (in pink, with 229 polygons) and 10 (in light blue, with 223 polygons). The ("French") taxate No. 4 (in pink), located in the northern half, is linguistically much younger than the southern ("Occitan") taxate 10 (in light blue). It spread radially at the expense of taxate 10 from around 1000 AD during the continuous expansion of the linguistic domain of Langue d'Oïl.

Our dialectometrisation of the ALF produced all in all 477 *vocally* relevant working maps (WM).

Middle: consonantally relevant WM to the original map 812 of the ALF *le marché* (< Latin MERCÁTU "(the) market"). The map shows the spatial distribution of 14 different results of the development of Latin **C** + Á (between its preservation and full palatalization) in the Latin etymon MER**C**ÁTU.

Once again, there is a clear confrontation between North (mainly via taxate 2 *sch*) and South (mainly via taxate 6 *k*). Here, too, there has been a diachronic expansion from the north (to the advantage of the Langue d'Oïl) to the south (at the expense of the Langue d'Oc).

Our dialectometrisation of the ALF produced in total 479 *consonantally* relevant WM.

Right: lexically relevant WM to map 173 of the ALF *la brebis* (< lat. VÉRVICE, "mutton"; later "ewe").

The three most important taxates are in the northeast (in pink: taxate 1), in the southwest (in light green: taxate 6: < *lat OVÍC[U]LA "sheep") and in the southeast (in light blue: taxate 17: < lat. *FÉTA "ewe"). This spatial stratification is at least a thousand years old: cf. Wartburg 1918.

Our dialectometrisation of the ALF produced on the whole 471 *lexically* relevant WM.

In cartographic terms, these three visualisations are areal-based. If one wanted to convey the same information by using line symbols or by drawing isoglosses, this would only make sense for the larger taxates and would by no means always result in elongated lines (or isoglosses) running over a large area in the manner of "garlands". It is easy to see here the cartographic (and also classificatory) precariousness of the isoline principle.

5.2 Presentation and interpretation of Plate 2a (three similarity maps) (Plate 2)

These are three *similarity profiles* to the ALF points 1 (Marcigny, Département Nièvre) [Map 2a], 285 (Ramecourt, Département Pas-de-Calais) [Map 2b] and 741 (Vaissac, Département Tarn-et-Garonne) [Map 2c], which function all as "reference points" (RP). The three RP polygons in question remain white. The remaining 640 polygons bear six colours, which visualise the range of the RIV-scores between minimum and maximum in six steps (intervals). The "Relative Identity Value" (RIV) was used as a similarity measure, which is considered the standard similarity index of RS-DM.[34]
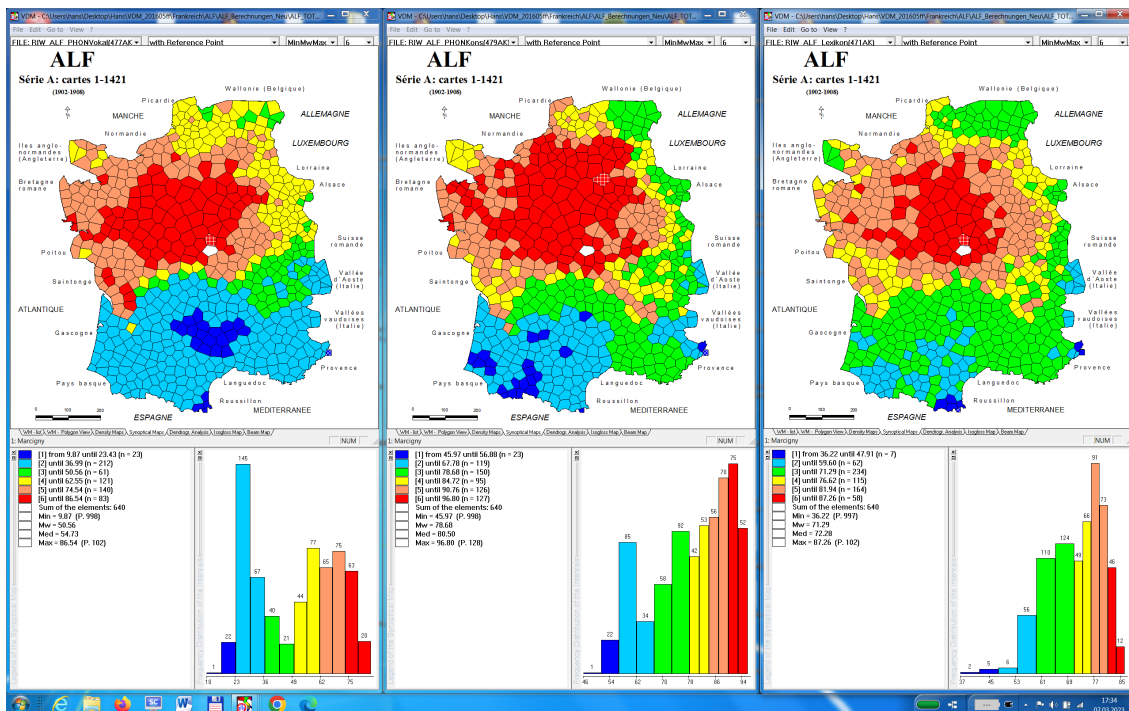

Plate 2a. Three similarity maps based on different linguistic categories referring to ALF-point 1 (Marcigny, Département Nièvre).

- similarity index: $RIV_{1,k}$

- algorithm of visualisation: MINMWMAX 6-tuple

- cartographic status: *qualitative* choropleth maps

---

[34] For a description of the RIV (in *English*: RIV: Relative Identity Value), see Goebl (1984: 74-78); in *German*: RIW: Relativer Identitätswert); Goebl (1981: 351); in *French*: IRI: Indice Relatif d'Identité) and Goebl (2010: 437).

Left: similarity map based on 477 *vocal* WM

Middle: similarity map based on 479 *consonantal* WM

Right: similarity map based on 471 *lexical* WM.

The numerical legends (bottom left in each case) show the key values of the six colour intervals, which are graded according to the logic of the solar spectrum (or rainbow). These intervals were calculated by the interval algorithm MINMWMAX. This ensures that the three "warm" marked intervals are above the arithmetic mean (4-6) of the relevant similarity distribution (consisting of 640 RIV scores), while the "cold" marked intervals (1-3) are below it. In the numerical legends, the arithmetic mean is shown as the upper value ("until...") of interval 3.

Each *similarity map* corresponds formally to one of the 641 (= N) point [or: locality] vectors of the square similarity matrix: see the right half of Figure 3.
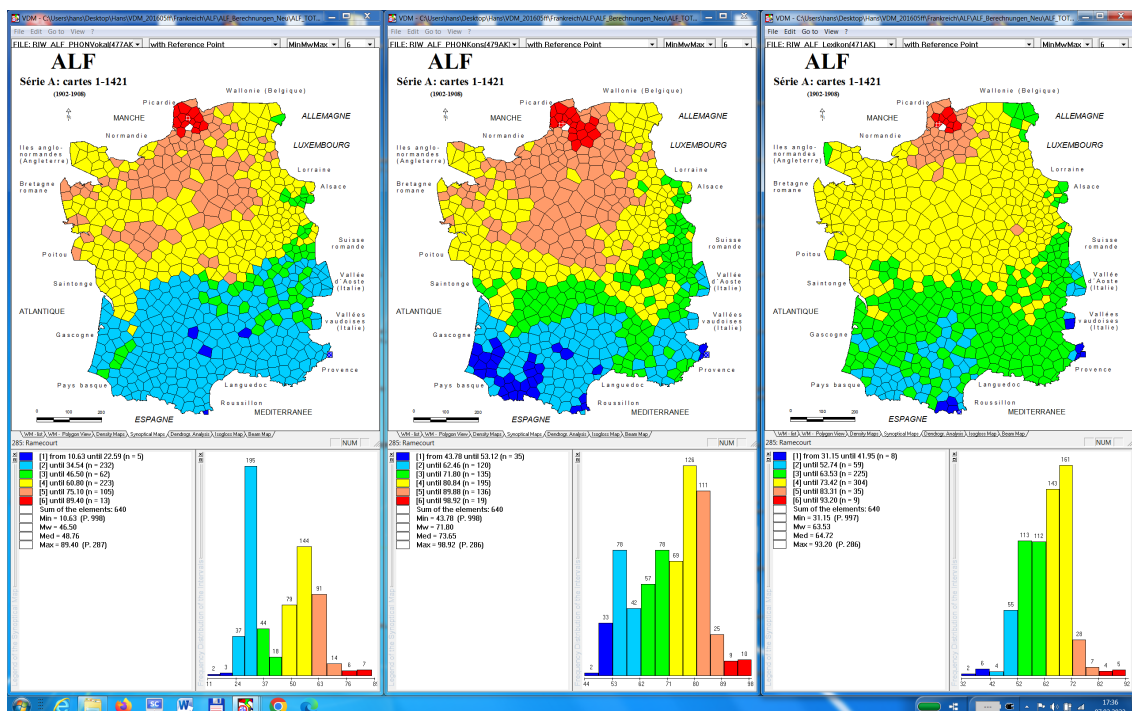

Plate 2b. Three similarity maps based on different linguistic categories referring to ALF-point 285 (Ramecourt, Département Pas-de-Calais)

- similarity index: $RIV_{285,k}$
- algorithm of visualisation: MINMWMAX 6-tuple

- cartographic status: *qualitative* choropleth maps

Left: similarity map based on 477 *vocal* WM
Middle: similarity map based on 479 *consonantal* WM
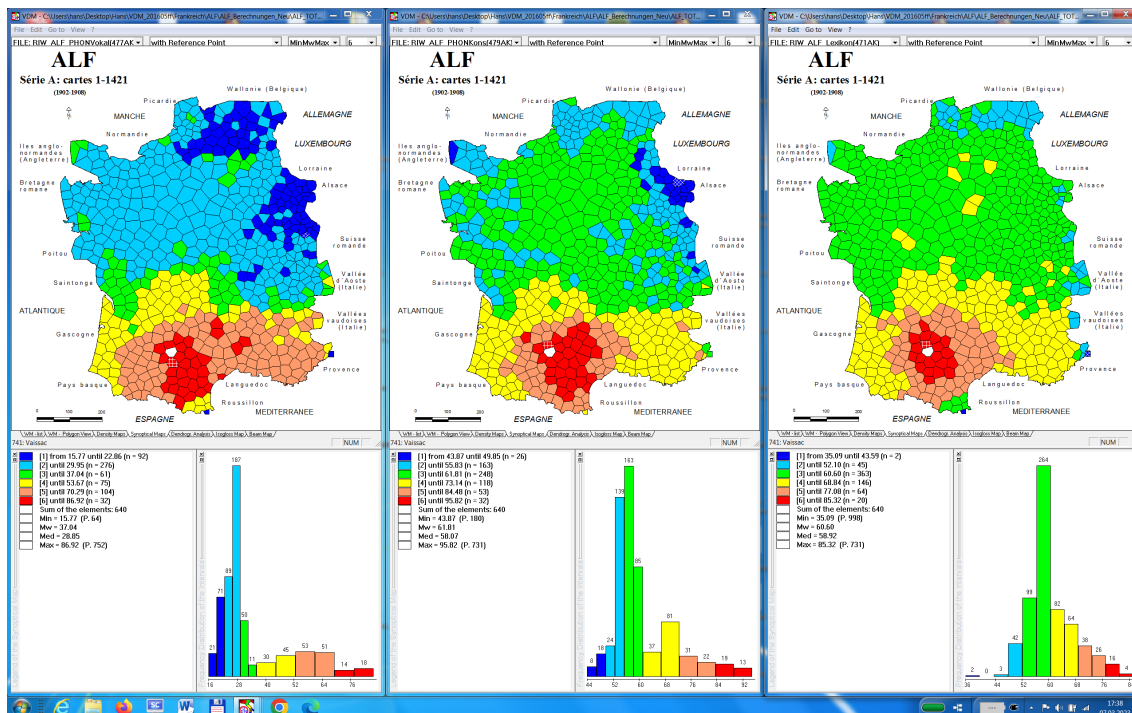Right: similarity map based on 471 *lexical* WM.

Plate 2c. Three similarity maps based on different linguistic categories referring to ALF-point 741 (Vaissac, Département Tarn-et-Garonne)
- similarity index: $RIV_{741,k}$
- algorithm of visualisation: MINMWMAX 6-tuple
- cartographic status: *qualitative* choropleth maps

Left: similarity map based on 477 vocal WM
Middle: similarity map based on 479 consonantal WM
Right: similarity map based on 471 lexical WM.

On all three maps, the concentric layering of the six colour levels clearly shows the distance-dependent decrease in linguistic similarity to the three RP (→ ALF-points 1, 285, and 741). The resulting choropleth profiles are very concisely structured in spatial terms. Since they each belong to three different linguistic categories (*vocalism,*

consonantism, lexicon), they clearly resemble each other, but are by no means identical.[35]

The "most compact" profile results from *vocalism* (left), the "most diffuse" from *lexicon* (right). In all three cases, the circular gradation of the polygons in red, orange and yellow is clearly visible.

These geographical regularities observable on the three maps (2a - 2c) have emerged in all our dialectometric projects; they also occur when using similarity measures other than the RIV.

The genuinely geolinguistic significance of the similarity maps refers to a classical dialect-classification question that has very often been asked dating from the 19th century and which can now be answered conclusively and in great detail about: namely, the issue of the "position of a local dialect within the overall framework".[36] It turns out that the spatial profiles of these "positions" are extremely variable, but always "well-ordered" in spatial terms.

The histograms (with 12 bars each) visible to the right of the numerical legends refer to the statistical "character" of the nine similarity distributions. It is easy to see that they have strongly varying shapes and are therefore by no means symmetrical. This variation can be measured by observing the corresponding characteristic values (or parameters).

Since it has been shown that the observation of some of these statistical parameters (such as *mean, maximum, skewness*, etc.) is of great linguistic interest, the parameters in question of the 641 similarity maps have each been synoptically combined and mapped. Map 3 provides an illustrative example only for the statistical parameter of *skewness*.

---

[35] Since the *lexicon* is undeniably the least "law-governed" category, the choropleth profiles it produces are generally less "precise" than those that are *vocally* or *consonantally* relevant.

[36] See the significant title of the following article by G. I. Ascoli from 1876 (b): *Del posto che spetta al ligure nel sistema dei dialetti italiani* ("On the place which belongs to Ligurian among the Italian dialects").

5.3 Presentation and interpretation of Plate 3 (parameter map: synopsis of skewness values)

The three maps are again based - from left to right – on categorically different sets of working maps: 477 vocalic WM, 479 consonantic WM, and 471 lexical WM. The respective similarity measurements have once again been done using RIV. Each of the three sub-maps shown is based on 641 skewness scores. Unlike map 2, the visualisation was realised using the interval algorithm MEDMW (instead of MINMWMAX) and over 8 (instead of 6) colour intervals. This results in a significant "tightening" of the polychrome choropleth profile, which is very useful heuristically. This "tightening" results from the fact that the same number of polygons is always sorted into each of the four intervals that are on either side of the mean. This will also be the case for the Plates 4 and 6.
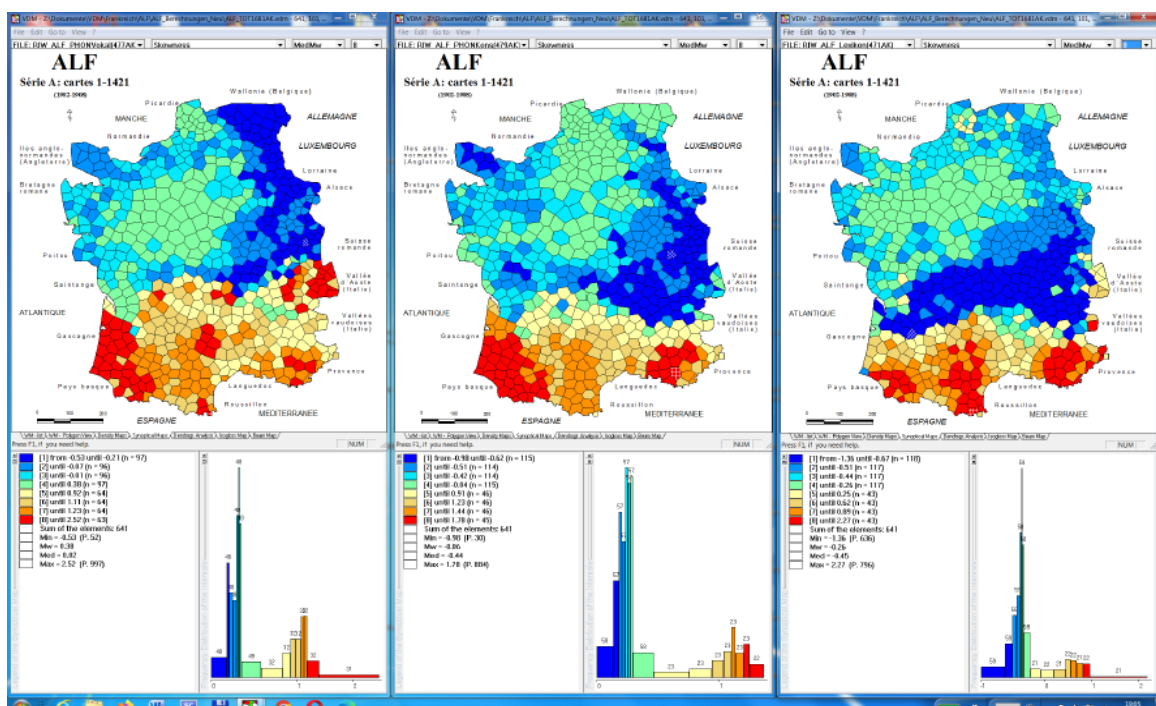

Plate 3. Three parameter maps showing the synopsis of 641 skewness scores of 641 similarity distributions (according to RIV_{jk})
- similarity index: RIV_{jk}
- algorithm of visualisation: MEDMW 8-tuple
- cartographic status: *quantitative* choropleth maps

Left: Parameter map based on 477 vocal WM
Middle: Parameter map based on 479 consonantal WM
Right: Parameter map based on 471 lexical WM.

The skewness is a measure for the symmetry of a frequency distribution: it varies theoretically between -1 and +1. With perfect symmetry of this frequency distribution, the respective mean value should be zero. However, one can clearly see from the three numerical legends that there are skewness scores above and below the theoretical value zero, and from the shape of the three histograms that none of the three skewness distributions is symmetrical.

For a deeper understanding of these three maps, one must know the linguistic meaning of the respective colours (and the underlying numerical scores): the occurrence of the colour blue means that at the polygon in question in the data matrix there is an excess of particularly large taxatorial areas (TA), all of which are the result of far reaching (and very old) diffusion processes.

On the other hand, the appearance of the colour red means the exact opposite; namely that at the respective polygon (or inquiry point) in the data matrix there is an excess of small TA, which are again the result of linguistic retreat or diluting processes.

On all three maps, the circular arrangement of blue and (dark) green hues, the majority of which is very clearly pronounced, points to the radial expansion of the linguistic type of Langue d'Oïl (= Northern French) from the Ile-de-France and Paris, respectively. This process is most likely more or less directly related to the takeover of political power by the Merovingian dynasty at the end of the 5th century AD and the continued presence of the greatest political power in this area.

In contrast, the red implants found only in the south represent linguistically defensive or even recessive zones. These are located - from west to east - mainly in Gascony, Languedoc, Provence or even in Suisse romande (in the case of vocalism and lexicon).

In any case, these three maps of plate 3 offer highly interesting diachronic insights, thus strikingly confirming the fact that deep insights into the linguistic history of a given area are possible from present-day linguistic data of the same region.

Looking at the position of the dark blue polygons, it becomes clear that in the case of vocalism the greatest expansive dynamic was found over the entire length of the eastern border of France; in contrast, the same applies to consonantism, especially in the southeast of France. In the case of the lexicon, this dynamic has again condensed into a broad belt, which extends in a very compact form in a west-east direction from the Atlantic to the Alps.

5.4 Presentation and interpretation of Plate 4 (interpoint or honeycomb map)

It is a matter of three "isoglossic syntheses", each of which is de facto based on 1,791 distance scores (*RDV*) which have been recorded between the individual inquiry points ("interpoints") (Goebl 1983) and then visualised in suitable form. The interpoint distance scores ("Relative Distance Value", *RDV*) are defined as follows: *dist = 100 - sim*. The similarities are determined by means of the already known similarity index *RIV*; the quoted formula can then also be rewritten as follows: *RDV = 100 - RIV*.

The 1,791 interpoints are given a special cartographic design in terms of thickness and colour. In this context the following cartographic principles apply: the greater the respective RD-value, the thicker and bluer the respective polygon side appears; and: the smaller the respective RD-value, the redder and thinner the respective polygon sides are visualised. By means of the Salzburg DM-program VDM, these two cartographic parameters can be freely adjusted.

Although this type of mapping corresponds in principle most closely to the drawing of "linguistic borders" (and the questions asked about them) practised in geolinguistics for more than a century, none of these three maps show spatial subdivision effects that would correspond to the linear and compact course of

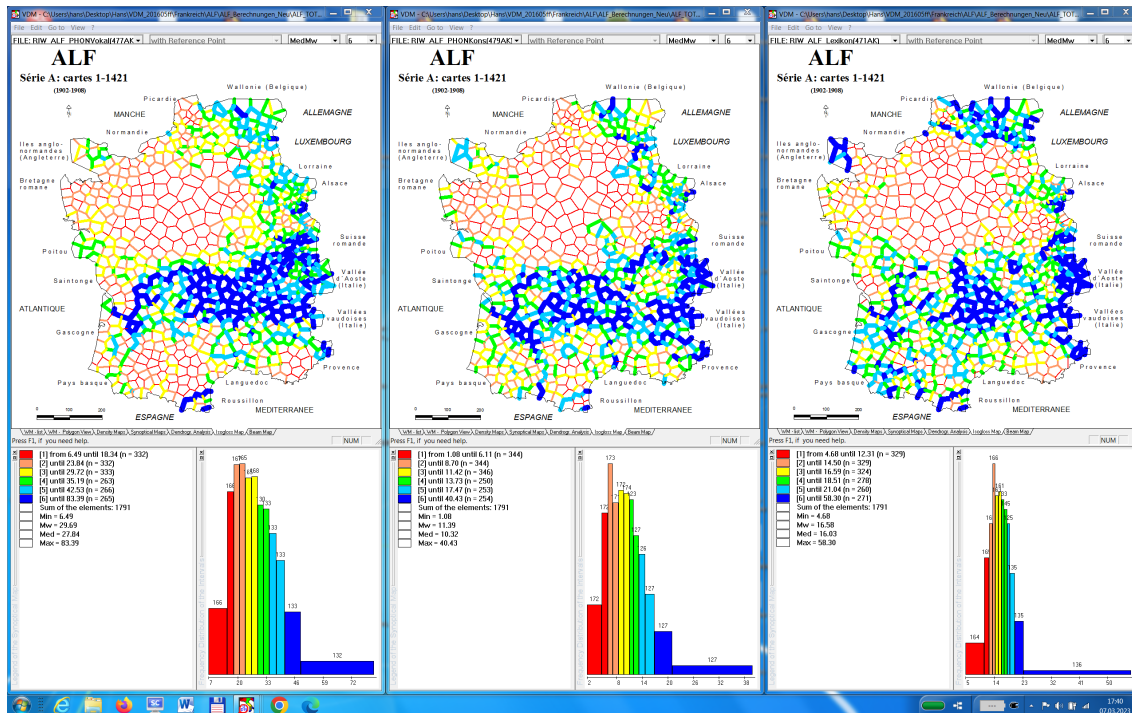"borders" - understood as precise dividing lines between states, regions or other human groupings.[37]

Plate 4. Three isogloss syntheses ("honeycomb maps") showing the synopsis of 1,791 interpoint *distance* values (according to RDV$_{jk}$)
- distance index: RDV$_{jk}$
- algorithm of visualisation: MEDMW 6-tuple
- cartographic status: *quantitative* isarithmic maps

Left: honeycomb map based on 477 vocal WM
Middle: honeycomb map based on 479 consonantal WM
Right: honeycomb map based on 471 lexical WM.

In contrast, one clearly recognises spatially broadly scattered (and yet spatially relatively compact) *bulkheading effects* that are also remotely reminiscent of the appearance of honeycombs. This is *most* the case on the left (→ vocalism) and the middle (→ consonantism) sub-map; it is least the case on the right map (→ lexicon), which shows the aforementioned compactness of the partitioning effect in only a very

---

[37] For this reason, I advise geolinguists to avoid the terms *border* or *limit*, which evokes false (because: linear) associations, and to replace it - wherever possible - with terms like *transition zone, border belt,* etc.

weakened way. But this fully corresponds to the well-known evidence of the spatio-temporal "instability" of the vocabulary.

Nevertheless, the "scraped" or "smoothed" character of the centre of northern France (Ile-de-France) and the strong internal fragmentation of the eastern and north-eastern fringes of France (Wallonia, Picardy, Lorraine) as well as of southern France proper (Occitania) can be seen on all three maps.

5.5 Presentation and interpretation of Plate 5 (three dendrograms or tree analyses)

The tree algorithms used in Numerical Taxonomy for about 70 years in similar research contexts are very much in line with the (metaphorical) ideas of the tree-like organisation of languages and dialects, which have also been anchored in Romance studies since some time (Schleicher 1863). In the present case, a "hierarchical-agglomerative" method (Sneath & Sokal 1973: 241, Bock 1974: 407-408, Chandon-Pinson 1981: 123) was used, which was developed in the 1960s by the American statistician Joe Ward Jr. (1926-2011). Its advantage is the formation of numerically very homogeneous groups (provided with a great internal "intra-group-similarity").
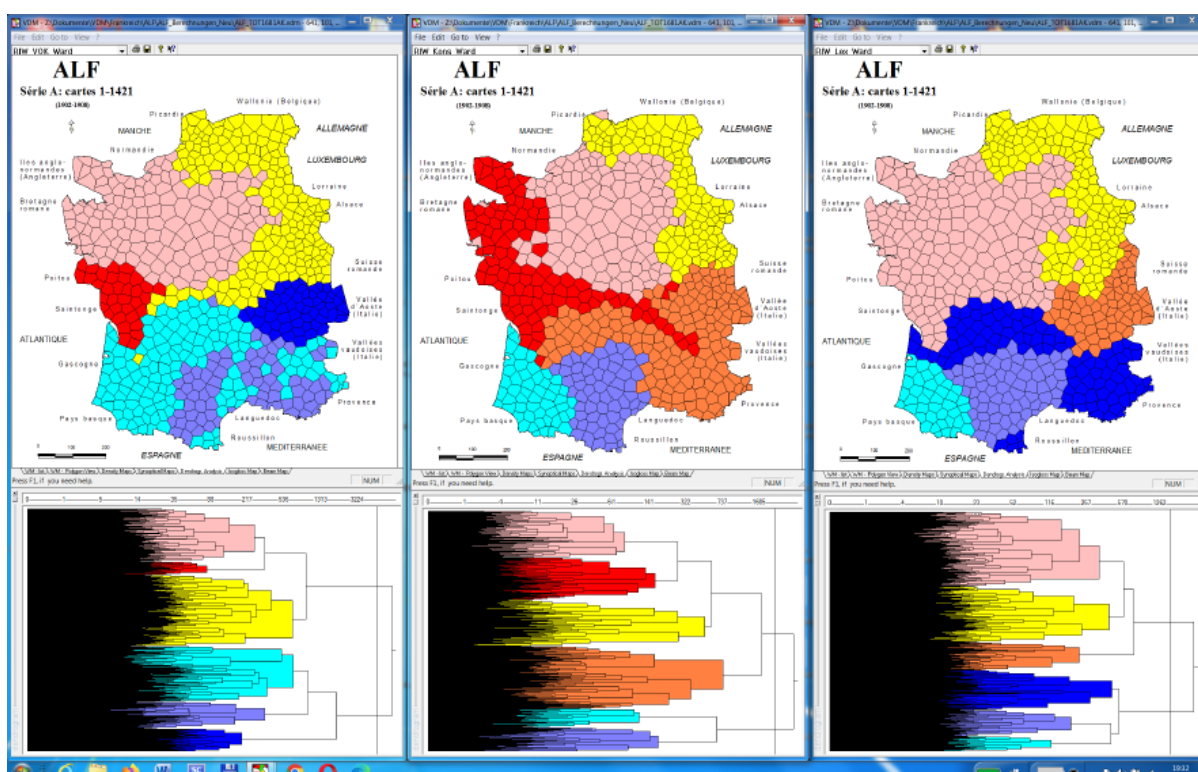
Plate 5. Three dendrographic classifications of 641 locolects (as registered in ALF) with subsequent spatial conversion of the coloured branches (*dendremes*) of the trees (below) to as many coloured *choremes* (above)

- similarity index: RIV$_{jk}$
- dendrographic algorithm: hierarchical grouping according to Joe Ward Jr.
- number of coloured *choremes* (above) and *dendremes* (below): 6
- cartographic status: trees and *quantitative* choropleth maps

Left: dendrographic classification based on 477 vocal WM
Middle: dendrographic classification based on 479 consonantal WM
Right: dendrographic classification based on 471 lexical WM.

In the *hierarchical-agglomerative* methods, *N* initially unconnected scores of a similarity matrix are merged in pairs in *N-1* successive steps using a certain algorithm (here: according to Joe Ward Jr.) to form ever larger, hierarchically structured aggregates (groups or clusters), which finally (or after the last merger) end in a "stem" or "root". The resulting (exclusively binary) dendrographic construct (or: arborescence) thus corresponds to an upside-down tree with its leaves pointing downwards and its

root pointing upwards. (On Plate 5, the "roots" of these trees have been tilted to the right).

From a linguistic point of view, such trees can be used as a "symbol" for the theoretical historical "splitting" of a given linguistic area that is assumed to be homogeneous at the very beginning. In such a diachronic perspective one must look at the tree from the root to the leaves and first ask oneself how many leaves or which parts of the respective "branches" are attached to the two uppermost (etc.) bifurcations. This question must then be renewed for each further bifurcation and answered by a corresponding colouring, admittedly only as long as the branching structure created in the process does not lose its heuristic usefulness.

In order to be able to do exactly that in a visually appealing way, in our case the spatial components hanging on the upper root branch (and thus lying in the north of France) were given warm colours, while the spatial components hanging on the lower root branch (and thus to be assigned to the south) were given cold colours.

Terminologically: The parts of the ramification of the calculated dendrogram that are to be regarded as "characteristic" are called *dendremes*, their spatial equivalents we call *choremes*. On Map 5, the dendremes and choremes always have the same colours.

For better comparability of the (quite different) dendreme and choreme structures of the three sub-maps of Map 5, only six colourings were provided in each case. In all cases, one recognises the prominent position of *Langue d'Oïl* (in pink) and Eastern France (in yellow) as well as the changing allocation of *Francoprovençal* (in blue or brown), once to the south (far left: vocalism), once to the north (far right: lexicon) and a third time (in the middle: consonantism) again to the north, although its catchment area has been extended to include large parts of the south.

As a rule, the numbered bar above the three trees can be used as a "time axis", which must be processed from right to left (or from the large to the small scores) in the case of a diachronic interpretation of the dendreme-choreme structure of the tree. It is true that as a dendreme approaches the left edge or the "leaves" of the tree, its theoretically assumed age decreases continuously. In other words: dendremes close to

the leaves are to be considered younger, dendremes close to the stems or roots are to be considered older.[38] The point of comparison is the bar of the (transversal) T that lies furthest to the right in the respective dendreme.

In this optic, on the far left (see the map to vocalism), the north has split off from the (older) south, while in the middle (consonantism) and on the right (lexicon), quite obviously the totality of the northward-oriented (French) dendremes/choremes are to be regarded as "older" than the southward-oriented (cold-coloured) dendremes/choremes.

Here, too, a dialectometric tree analysis with several algorithms (such as *Complete Linkage, Average Linkage* or *UPGMA* (Unweighted Pair Group Method with Arithmetic mean): Sneath & Sokal (1973: 230-234), etc. - and thus in a comparative manner - should be realised. This is fully guaranteed by our DM program VDM - where six dendrographic algorithms are implemented.

5.6 Presentation and interpretation of Plate 6 (three correlation maps)

Especially in the work of J. Séguy (1971, 1973), the question of the relationship between the increase in geographical distance and the decrease in linguistic similarity (to a previously defined RP) played a major role. Due to a lack of corresponding technical possibilities (such as computers), Séguy could only "think about" this problem in a very reduced way.

---

[38] Of course, the *dendremes* are also affected by this, although they are arranged *spatially* and not *dendrographically*.
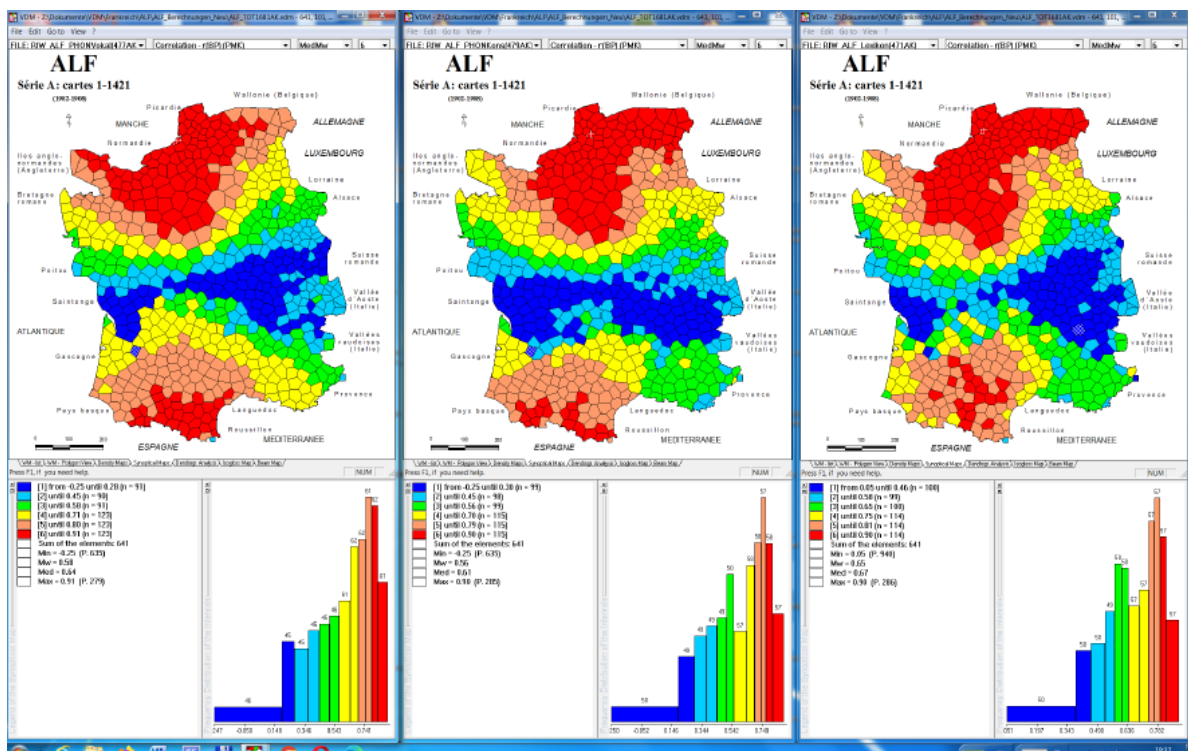
Plate 6. Three choropleth profiles based on 641 correlation scores (according to rBP$_{jk}$) between 641 *similarity* vectors (according to RIV$_{jk}$) and 641 *proximity* vectors (according to Pythagoras' theorem with subsequent conversion in proximity values [prox$_{pyth}$])
   - algorithm of visualisation: MEDMW 6-tuple
   - cartographic status: *quantitative* choropleth maps.

Left: correlation based on prox$_{pyth}$ and 477 vocal WM
Middle: correlation based on prox$_{pyth}$ and 479 consonantal WM
Right: correlation based on prox$_{pyth}$ and 471 lexical WM.

Since 2005, there has been a "correlative" add-on module to VDM (Goebl, 2005), which makes it possible to correlate different modalities of "basilectal management of space by men" - such as in *vocalic, consonantal, lexical,* or even *geometric* terms - in pairs by means of the *Bravais-Pearson product-moment coefficient* (rBP)[39] and to visualise the resulting *N* rBP- scores accordingly.

---

[39] For the formula of the rBP see: Sneath & Sokal (1973: 138); Bock (1974: 44-45, 59-61); and Chandon & Pinson (1981: 65). Here are the life dates of the two naturalists and statisticians whose names are associated with this index: Auguste Bravais (1811-1863) and Karl Pearson (1857-1936).

In this process, identically numbered vectors of two similarity matrices calculated for one and the same network (with *N* inquiry points) are pairwise correlated by means of rBP. In the case of the consideration of the geographical distance, the procedure is as follows: first, the reciprocal distances (*dist*)[40] are determined on the basis of the GIS coordinates of the *N* (= 641) inquiry points of the ALF (or of another atlas grid) by means of the well-known Pythagorean theorem ($a^2 + b^2 = c^2$), and then these distance scores (*dist*) are converted - in accordance with the logic of the similarity measures used for correlation - into proximities (*prox*) by means of a simple formula: *100 - dist = prox.*

It should be added that the correlation index rBP oscillates between -1 and +1. On our Map 6, low rBP-scores are visualised by the colour blue and high rBP-scores by the colour red. The resulting spatial profiles (see the three choropleth profiles of Map 6) can be interpreted from a linguistic point of view as follows: red zones correspond to those regions where there is a *very great (spatial) harmony* between the two correlated dimensions (here: *vocalism* and *spatial* proximity, etc.) or where the two dimensions are - metaphorically speaking - "in step".

The opposite is true for the zones coloured blue; there, dissonances and tensions of any kind are to be assumed between the correlated dimensions. It goes without saying that in such cases (*harmony* or *tension*) historical factors come into play or have to be taken into account in the interpretation.

The three maps in Plate 6 are visualisations of the correlations (using rBP) which exist on the one hand between the *geographical proximity* and with the three linguistic categories of *vocalism*, *consonantism*, and *lexicon* on the other hand.

Looking at the three very similarly structured choropleth profiles of Plate 6, one immediately recognises that the zones of highest (spatio-temporal) harmony are located in the north on the one hand and in the south and southwest of France on the other. The spatially finely graduated transition from the two poles of harmony (in the

---

[40] The distance calculations are only carried out for directly neighbouring measuring points ("interpoints"), whereby the neighborhood relations have been determined by a previously carried out triangulation of the atlas network concerned.

north and south) to the zone of "conflict", which runs through the whole middle of France from west to east in a very compact manner, is also not to be overlooked.

From our knowledge of the linguistic history of Gallo-Romance, but also from the choroplethic structure of the three maps, it can be deduced that the "harmony zone" (provided with a red core and orange-yellow border), which stretches from the English Channel in a wedge shape towards the south, is more "dynamic" and younger than its "harmony counterpart" in the south, and thus ultimately the blue "conflict zone" is the result of the pressure - which has lasted for about more than 1000 years - from the north (i. e. from the Langue d'Oïl) on the south (i. e. on the Langue d'Oc).

Given the universal applicability of VDM's correlation module, the surplus value in terms of linguistic insights and findings is considerable.

6. Final remarks

In addition to the procedures shown here, further applications to the Romance linguistic atlases ALF (France), AIS (Italy) (Goebl 2008), ALDC (Catalonia) (Goebl 2013c) and ALPI (Iberia) (Goebl 2013b) as well as to the English atlas SED ("Survey of English Dialects") (Goebl 2007) and also, to the Germanic atlas SDS ("Sprachatlas zur deutschen Schweiz") (Goebl et al. 2013) can be found on two websites:

- Romance languages and English:

http://dialektkarten.ch/dmviewer/alf/index.en.html

- German speaking Switzerland:

http://dialektkarten.ch/dmviewer/swg/index.en.html

These two websites are interactive and were designed by the Suisse IT specialist Yves Scherrer (currently in Helsinki).

Of particular interest, however, is the fact that these methods can also be applied to data extracted from medieval sources, so that a direct comparison of old and new deep structures is possible for one and the same area of investigation. So far,

this has been achieved with very good success for northern France (Goebl & Smečka 2016b) and northern Italy (Videsott 2009).

7. Acknowledgements

- Production of figures 1-4: Werner Goebl (Vienna)

- VDM: Edgar Haimerl (Seattle, USA; Blaustein, Germany)

- Webpages ("dialektkarten.ch/dmviewer"): Yves Scherrer (Helsinki)

I thank these three gentlemen very cordially for their assistance.

8. Abbreviations

ALF	*Atlas linguistique de la France*, 1902-1910
DSA	*Deutscher Sprachatlas*, 1927-1956
MEDMW	Interval algorithm
MINMWMAX	Interval algorithm
rBP	Correlation index
RDV	Relative Distance Value (*distance* index)
RIV	Relative Identity Value (*similarity* index)
RP	Reference point
RPGR	Revue des Patois Gallo-Romans
RS-DM	Regensburg-Salzburg-Dialectometry
silma	Silent map
TA	Taxatorial area
WIV	Weighted Identity Value
WM	Working map

References

ALF (1902-1910): Gilliéron, Jules & Edmond Edmont (eds.), *Atlas Linguistique de la France* (10 vols.), Paris: Champion (reprint: Bologna: Forni, 1968).

ALF-Notice (1902): *Atlas linguistique de la France, Notice servant à l'intelligence des cartes*, Paris: Champion (reprint: Bologna: Forni, 1968).

ALG (1954-1974): Séguy, Jean (ed.), *Atlas Linguistique et ethnographique de la Gascogne* (6 vols.), Paris: Centre National de la Recherche Scientifique.

ALPO (1966): Guiter, Henri (ed.), *Atlas linguistique des Pyrénées Orientales*, Paris: Centre National de la Recherche Scientifique.

ALTMANN, Gabriel (1985) "Die Entstehung diatopischer Varianten: Ein stochastisches Modell", *Zeitschrift für Sprachwissenschaft,* 4, 139-155.

ALTMANN, Gabriel (2005) "Diversification processes", in HSK, 27, 646-658.

ALTMANN, Gabriel, Werner LEHFELDT (1973) *Allgemeine Sprachtypologie. Prinzipien und Meßverfahren*, München: Fink.

ASCOLI, Graziadio Isaia (1876a) "Paul Meyer e il franco-provenzale", *Archivio glottologico italiano,* 2, 385-395.

ASCOLI, Graziadio Isaia (1876b) "Del posto che spetta al ligure nel sistema dei dialetti italiani", *Archivio glottologico italiano,* 2, 111-160.

BOCK, Hans Hermann (1974) *Automatische Klassifikation: Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten (Cluster-Analyse)*, Göttingen: Vandenhoeck & Ruprecht.

BOURGUET, Marie-Noëlle (1988) *Déchiffrer la France. La statistique départementale à l'époque napoléonienne*, Montreux: Éditions des archives contemporaines.

BRUN-TRIGAUD, Guylaine (1990) *Le croissant: le concept et le mot, contribution à l'histoire de la dialectologie française au XIXe siècle*, Lyon III (Jean Moulin): Centre d'Études Linguistiques Jacques Goudet.

BRUN-TRIGAUD, Guylaine, Yves LE BERRE, Jean LE DÛ (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 Gilliéron et Edmond Edmont augmenté de quelques cartes de l'Atlas Linguistique de la Basse- Bretagne de Pierre Leroux)*, Paris: Comité des Travaux Historiques et Scientifiques (CTHS).

Brunet, Roger (1987) *La carte. Mode d'emploi*, Paris: Fayard-Reclus.

Chandon, Jean-Louis, Suzanne Pinson (1981) *Analyse typologique: Théories et applications*, Paris: Masson.

Christmann, Hans Helmut (1971) "Lautgesetze und Wortgeschichte: Zu dem Satz *Jedes Wort hat seine eigene Geschichte"*, in Wolf-Dieter Stempel, Eugenio Coseriu (eds.), *Sprache und Geschichte: Festschrift für Harri Meier zum 65. Geburtstag,* Munich: Fink, 111-124.

Dickinson, Gordon Cawood (1973) *Statistical Mapping and the Presentation of Statistics*, London: Arnold.

Droixhe, Daniel (2000) "Les conceptions du changement et de la parenté des langues européennes aux XVIIe et XVIIIe siècles", in HSK, 18/1, 1057-1071.

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* (23 fascicles with 128 maps), Marburg / Lahn: Elwert.

Durand (de Gros), Joseph-Pierre (1889) "Notes de philologie rouergate [Note XVIII]", *Revue des Langues Romanes,* 33, 47-84; 251-287; 433-444.

Durand (de Gros), Joseph-Pierre (1899) *Aperçus de taxinomie générale*, Paris: Alcan.

Edmont, Edmond (1897) *Lexique saint-polois*, Saint-Pol: chez l'auteur / Mâcon: Protat Frères.

Ettmayer, Karl von (1902) "Lombardisch-Ladinisches aus Südtirol. Ein Beitrag zum oberitalienischen Vokalismus", *Romanische Forschungen* 13, 321-672 [republished in 1995 by Hans Goebl: *Neu herausgegeben, mit einem vorwärts und einem rückwärts alphabetischen Register der Etyma, einer kurzen geotypologischen Studie zu den neuveröffentlichten Materialien, einer Biographie und einer Bibliographie sowie einer Würdigung des wissenschaftlichen Oeuvres Karl von Ettmayers*, San Martin de Tor / St. Martin in Thurn: Istitut Cultural Ladin 'Micurà de Rü'].

Ettmayer, Karl von (1924) *Über das Wesen der Dialektbildung erläutert an den Dialekten Frankreichs,* Vienna: Denkschriften der Akademie der Wissenschaften in Wien.

Fleischer, Fritz (1913) *Studien zur Sprachgeographie der Gascogne*, Halle: Niemeyer.

Gardette, Pierre (1941) *Géographie phonétique du Forez*, Mâcon: Protat Frères.

Gartner, Theodor (1882) *Viaggi ladini, fatti e narrati dal Dr. T. G.*, Linz: Wimmer.

Gartner, Theodor (1883) *Raetoromanische Grammatik*, Heilbronn: Henniger (reprint: Vaduz, Sändig, 1992).

GILLIÉRON, Jules (1880) *Petit atlas phonétique du Valais roman (sud du Rhône),* Paris: Champion.

GILLIÉRON, Jules (1912) *Études de géographie linguistique d'après l'Atlas linguistique de la France*, Paris: Champion.

GILLIÉRON, Jules (1918) *Généalogie des mots qui désignent l'abeille d'après l'atlas linguistique de la France*, Paris: Champion.

GILLIÉRON, Jules (1919) *La faillite de l'étymologie phonétique*, Neuveville: Beerstecher.

GILLIÉRON, Jules, Jean MONGIN (1905) *Étude de géographie linguistique: scier dans la Gaule romane du Sud et de l'Est,* Paris: Champion (Italian translation by Lorenzo Massobrio, Novi Ligure: Grafica editoriale, 1990).

GIRNTH, Heiko (2010) "Mapping language data", in HSK, 30.2., 98-121.

GOEBL, Hans (1976) "La dialectométrie appliquée à l'ALF (Normandie)", in *Atti del XIV Congresso internazionale di linguistica e filologia romanza (Napoli, 15-20 aprile 1974)*, Naples / Macchiaroli / Amsterdam: Benjamins, vol. II, 165-195.

GOEBL, Hans (1983) "Parquet polygonal et treillis triangulaire: Les deux versants de la dialectométrie interponctuelle", *Revue de Linguistique Romane,* 47, 353-412.

GOEBL, Hans (1984) *Dialektometrische Studien: Anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF* (3 vols.), Tübingen: Niemeyer.

GOEBL, Hans (2002a) "Analyse dialectométrique des structures de profondeur de l'ALF", *Revue de Linguistique Romane,* 66, 5-63.

GOEBL, Hans (2002b) "Sprachatlanten: woher? womit? wozu? Einige buntgemischte Reflexionen am Gartenzaun zwischen Romanistik und Germanistik"*, in Peter Anreiter, Peter Ernst, Isolde Hausner, Helmut Kalb (eds.), *Namen, Sprachen und Kulturen. Imena, Jeziki in Kulture. Festschrift für Heinz Dieter Pohl zum 60. Geburtstag*, Wien: Edition Präsens, 257-274.

GOEBL, 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.

GOEBL, Hans (2004a) "Eine Glosse zu Isoglosse", in Thomas Krisch, Thomas Lindner, Ulrich Müller (eds.), *Analecta Homini universali dicata. Arbeiten zur Indogermanistik, Linguistik, Philologie, Politik, Musik und Dichtung. Festschrift für Oswald Panagl zum 65. Geburtstag,* Stuttgart: Heim, vol. II., 525-537.

GOEBL, Hans (2004b) "Joseph-Pierre Durand (de Gros), 1826-1900. Ein weitgehend vergessener Protagonist in der Ascoli-Meyer-Debatte über die Existenz von Dialekten", in Alberto Gil, Dietmar Osthus, Claudia Polzin-Haumann (eds.), *Romanische Sprachwissenschaft. Zeugnisse für Vielfalt und Profil eines Faches. Festschrift für Christian Schmitt zum 60. Geburtstag*, Frankfurt / Main, Berlin: Lang, vol. I, 169-192.

GOEBL, Hans (2005) "La dialectométrie corrélative: Un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme", *Revue de Linguistique Romane,* 69, 321-67.

GOEBL, Hans (2006a) "Recent Advances in Salzburg Dialectometry", *Digital Scholarship in the Humanities* (formerly *Literary and Linguistic Computing*), 21 (4), 411-435.

GOEBL, Hans (2006b) "Warum die Dialektometrie nur in einem roman(ist)ischen Forschungskontext entstehen konnte", in Wolfgang Dahmen, Günter Holtus, Johannes Kramer, Michael Metzeltin, Wolfgang Schweickard, Otto Winkelmann (eds.), *Was kann eine vergleichende romanische Sprachwissenschaft heute (noch) leisten?* Tübingen: Narr (= Romanistisches Kolloquium XX), 291-317.

GOEBL, Hans (2007) "A bunch of dialectometric flowers: A brief introduction to dialectometry", in Ute Smit, Stefan Dollinger, Julia Hüttner, Gunther Kaltenböck, Ursula Lutzky (eds.), *Tracing English through Time – Explorations in Language Variation. In Honour of Herbert Schendl on the Occasion of his 65th birthday,* Wien: Braumüller, 133-171.

GOEBL, Hans (2008) "La dialettometrizzazione integrale dell'AIS. Presentazione dei primi risultati", *Revue de linguistique romane,* 72, 25-113 (with 48 colour maps).

GOEBL, Hans (2010) "Dialectometry: Theoretical prerequisites, practical problems, and concrete applications (mainly with examples drawn from the 'Atlas Linguistique de la France', 1902-1910)", *Dialectologia*, Special Issue 1, 63-77.

GOEBL, Hans (2011) *Quo vadis, atlas linguistice? Einige wissenschaftshistorische und zeitgeistkritische Reflexionen zur atlasgestützten Geolinguistik*, in: Claudia Schlaak, Lena Busse (eds.). *Sprachkontakte, Sprachvariation und Sprachwandel. Festschrift für Thomas Stehl zum 60. Geburtstag,* Tübingen: Narr, 5-27.

GOEBL, Hans (2013a) "Dialectometry and quantitative mapping", in Alfred Lameli, Roland Kehrein, Stefan Rabanus (eds.), *Language and Space: An International Handbook of Linguistic Variation, Vol. 2. Language Mapping.* Berlin: de Gruyter, 433-457.

GOEBL, Hans (2013b) "La dialectometrización del ALPI: Rápida presentación de los resultados", in Emili Casanova Herrero, Cesáreo Calvo Rigual (eds.), *Actas del XXVI Congreso Internacional de Lingüística y de Filología Románicas* (Valencia 2010), Berlin: de Gruyter, vol. 6., 143-154.

GOEBL, Hans (2013c) "La dialectometrització dels quatre primers volums de l'ALDC", *Estudis romànics,* 35, 87-116.

GOEBL, Hans (2017) "La romanistica svizzera e la nascente ricerca geolinguistica in Europa con particolare riguardo all'opera di Jules Gilliéron", in Sandro Bianconi, Domenico De Martino, Annalisa Nesi (eds.), *La romanistica svizzera della prima metà del novecento e l'Italia.* Atti del Convegno di Firenze, Villa Medicea di Castello (9-10 novembre 2016), Firenze: Accademia della Crusca, 11-34 (with three colour maps).

GOEBL, Hans (2018a) "La face cachée de la géographie linguistique. Bref aperçu sur les « cartes muettes » produites pour l'ALF, l'AIS et le FEW", *Revue de linguistique romane,* 82, 5-63.

GOEBL, Hans (2018b) "«Des caractères et de l'extension du patois normand» von Charles Joret (1883) unter typophobem Beschuss", in Wolfgang Asholt, Ursula Bähler, Bernhard Hurch, Henning Krauss, Kai Nonnenmacher (eds.), *Engagement und Diversität. Frank-Rutger Hausmann zum 75. Geburtstag*, München: Akademische Verlagsgemeinschaft, 183-197.

GOEBL, Hans, Yves SCHERRER & Pavel SMEČKA (2013) "Kurzbericht über die Dialektometrisierung des Gesamtnetzes des "Sprachatlasses der deutschen Schweiz" (SDS)", in Karina Schneider-Wiejowski, Birte Kellermeier-Rehbein, Jakob Haselgruber (eds.), *Vielfalt, Variation und Stellung der deutschen Sprache*, Berlin / Boston: de Gruyter, 153-175.

GOEBL, Hans & Pavel SMEČKA (2016a) "The Quantitative Nature of Working Maps (WM) and Taxatorial Areas (TA): A Brief Look at two Basic Units of Salzburg Dialectometry (S-DM)*,* in Emmerich Kelih, Róisin Knight, Ján Mačutek, Andrew Wilson (eds.), *Issues in Quantitative Linguistics 4. Dedicated to Reinhard Köhler on the occasion of his 65th birthday*). Lüdenscheid: RAM-Verlag (Studies in Quantitative Linguistics, vol. 23), 113–127.

GOEBL, Hans & Pavel SMEČKA (2016b) "L'interprétation dialectométrique des atlas «scripturaux» d'Anthonij Dees", *Revue de linguistique romane,* 80, 321-368 (with 50 colour maps).

GOEBL, Hans & Norbert WINTERLEITNER (1971) "Projekt einer sprachstatistischen Auswertung von in Sprachatalanten gespeicherter linguistischer Information mit Hilfe elektronischer Rechenanlagen", *Linguistische Berichte,* 14, 60-61.

GUITER, Henri (1973) "Atlas et frontières linguistiques", *Les dialectes romans de France à la lumière des atlas régionaux* (Colloque de Strasbourg, 24-28 mai 1971), Paris: CNRS, 61-109.

HAAG, Carl (1898) *Die Mundarten des Oberen Neckar- und Donaulandes (Schwäbisch-alemannisches Grenzgebiet: Baarmundarten)*, Reutlingen: Hutzler.

HAARMANN, Harald (2000) "Die großen Sprachensammlungen vom frühen 18. bis frühen 19. Jahrhundert", in HSK, 18/1, 1081-1094.

HEERINGA, Wilbert (2004) *Measuring Dialect Pronunciation Differences using Levenshtein Distance,* Groningen: Groningen Dissertations in Linguistics, vol. 46.

HSK 18 (2000-2001-2006): Auroux, Sylvain, Koerner, Ernst Frideryk Konrad, Niederehe. Hans-Joseph, Versteegh, Kees (eds.), *History of Language Sciences, Geschichte der Sprachwissenschaften, Histoire des sciences du langage*, Berlin: de Gruyter, 3 vols.

HSK 27 (2005): Köhler, Reinhard, Altmann, Gabriel, Piotrowski, Rajmund G. (eds.), *Quantitative Linguistik, Quantitative Linguistics. Ein internationales Handbuch, An International Handbook*, Berlin / New York: de Gruyter.

HSK 30.2 (2010): Lameli, Alfred, Kehrein, Roland, Rabanus, Stefan (eds.), *Language and Space. An International Handbook of Linguistic Variation*, vol. 2: *Language Mapping*, Part 1 and 2 [Maps], Berlin / New York: de Gruyter, Mouton.

IORDAN, Iorgu (1962) *Einführung in die Geschichte und Methoden der romanischen Sprachwissenschaft,* Berlin: Akademie-Verlag.

JABERG, Karl (1906) "Zum Atlas linguistique de la France", *Zeitschrift für romanische Philologie,* 30, 512.

JABERG, Karl (1908) *Sprachgeographie: Beitrag zum Verständnis des Atlas Linguistique de la France,* Aarau: Sauerländer.

JOCHNOWITZ, George (1973) *Dialect boundaries and the question of Franco-Provençal*, The Hague / Paris: Mouton.

KÖDL, Sven (2014) *Die Enquête Coquebert de Montbret (1806-1812). Die Sprachen und Dialekte Frankreichs und die Wahrnehmung der französischen Sprachlandschaft während des Ersten Kaiserreichs*, Bamberg: Bamberg University Press.

L<small>ALANNE</small>, Théobald (1949-1952) *L'indépendance des aires linguistiques en Gascogne maritime*, 2 vols. Saint-Vincent-de-Paul: chez l'auteur (reprint: Labatut: Atelier Histoire Trait d'Union, 2018).

L<small>ALANNE</small>, Théobald (1953) "Indice de polyonymie, indice de polyphonie", *Français moderne,* 21, 263-274.

L<small>ECLERC</small>, Gérard (1979*) L'observation de l'homme. Une histoire des enquêtes sociales*, Paris: du Seuil.

LRL I/2 (2001) Holtus, Günter, Metzeltin, Michael, Schmitt, Christian (eds.), *Lexikon der romanistischen Linguistik,* Tübingen: Niemeyer, vol. I, 2.

M<small>ALKIEL</small>, Yakov (1967) "Each word has a history of its own", *Glossa,* 1, 13-149.

M<small>EHLER</small>, Alexander (2005) "Eigenschaften der textuellen Einheiten und Systeme", in HSK, 27, 325-348.

M<small>ENÉNDEZ</small> P<small>IDAL</small>, Ramón, Diego C<small>ATALÁN</small>, Álvaro G<small>ALMES</small> (1954) *Como vive un romance. Dos ensayos [1920 and 1950] sobre tradicionalidad*, Madrid: Aguirre Torre.

M<small>EYER</small>, Paul (1895) "C et G suivis d'A en provençal. Étude de géographie linguistique", *Romania,* 24, 529-575.

O<small>KABE</small>, Atsuyuki, Barry B<small>OOTS</small>, Kokichi S<small>UGIHARA</small> (1992, 2000[2]) *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, Chichester: Wiley. 2nd edition (2000) with Chiu, Sung Nok, Chichester: Wiley.

P<small>ALSKY</small>, Gilles (1996) *Des chiffres et des cartes: Naissance et développement de la cartographie quantitative française au XIXe siècle*, Paris: Comité des travaux historiques et scientifiques (CTHS).

P<small>APANTI</small>, Giovanni (1875) *I parlari italiani in Certaldo alla festa del V centenario di Messer Giovanni Boccacci*, Livorno: Vigo (reprint Bologna: Forni, 1972).

P<small>ARIS</small>, Gaston (1888) "Les parlers de France"*,* in Gaston Paris, *Mélanges Linguistique,* Paris: Champion, 1909, 432-448.

P<small>OP</small>, Sever (1950) *La dialectologie: Aperçu historique et méthodes d'enquêtes Linguistiques* (2 vols.), Gembloux: Duculot.

P<small>OP</small>, Sever, P<small>OP</small>, Doina R<small>ODICA</small> (1959) *Jules Gilliéron: Vie, enseignement, élèves, œuvres, souvenirs*, Louvain: Centre International de Dialectologie Générale.

R<small>OSENQVIST</small>, Arvid (1919) "Limites administratives et division dialectale de la France", *Neuphilologische Mitteilungen,* 20, 87-119.

SCHLEICHER, August (1863) *Die Darwinsche Theorie und die Sprachwissenschaft. Offenes Sendschreiben an Herrn Dr. Ernst Häckel, a. o. Professor der Zoologie und Director des zoologischen Museums an der Universität Jena,* Weimar: Böhlau.

SCHUCHARDT, Hugo (1870/1900/1928) *Über die Klassifikation der romanischen Mundarten [Probevorlesung, gehalten zu Leipzig am 30. April 1870]*, Graz: Styria, 1900 (first reprint in: Spitzer, Leo (ed.), *Hugo Schuchardt-Brevier. Ein Vademecum der allgemeinen Sprachwissenschaft*, Halle, Niemeyer, 1928, 166–188; second reprint: Darmstadt, Wissenschaftliche Buchgesellschaft, 1976).

SÉGUY, Jean (1971) "La relation entre la distance spatiale et la distance lexicale", *Revue de Linguistique Romane 35, 335–57.*

SÉGUY, Jean (1973) "La dialectométrie dans l'Atlas Linguistique de la Gascogne", *Revue de Linguistique Romane,* 37, 1–24.

SNEATH, Peter Henry Andrews, Robert Reuven SOKAL (1973) *Numerical Taxonomy: The Principles and Practice of Numerical Classification,* San Francisco: CA, Freeman.

STAGL, Justin (1995) *A History of Curiosity. The Theory of Travel 1550-1800*, Chur: Harwood Academic Publisher.

STOROST, Jürgen (1994) "Noch einmal zur Diez-Stiftung", *Romanistisches Jahrbuch,* 45, 74-84.

SWIGGERS, Pierre (2010) "Mapping the Romance Languages of Europe", in HSK 30.2., 269–300.

TOURTOULON, Charles, de (1890) "Communication […] sur la classification des dialectes", *Revue des Langues Romanes,* 34, 130-175.

TOURTOULON, Charles de, Octavien BRINGUIER (1876) "Rapport sur la limite géographique de la langue d'oc et de la langue d'oïl", *Archives des Missions scientifiques et littéraires,* III/3, 544-606, 1 map.

VIDESOTT, Paul (2009) *Padania scrittologica. Analisi scrittologiche e scrittometriche di testi in italiano settentrionale antico dalle origini al 1525*, Tübingen: Niemeyer (Beihefte zur Zeitschrift für romanische Philologie, vol. 343).

WARTBURG, Walther von (1918) *Zur Benennung des Schafes in den romanischen Sprachen. Ein Beitrag zur Frage der provinziellen Differenzierung des spätern Lateins*, Berlin: Reimer (Abhandlungen der Preußischen Akademie der Wissenschaften, phil.-hist. Klasse, vol. 10).

WARTBURG, Walther von (19632) *Problèmes et méthodes de la linguistique*, Paris: Presses Universitaires de France.

WEINHOLD, Norbert, WOLF, Lothar (2001) *Areallinguistik / Géolinguistique*, *LRL* I/2, 823-835.

WINKELMANN, Otto (2001) *Romanische Sprachatlanten / Les atlas linguistiques des langues romanes*, *LRL* I/2, 1004-1068.