21-9-2014

Hans Goebl (University of Salzburg)

hans.goebl@sbg.ac.at

https://www.sbg.ac.at/rom/people/prof/goebl/goebl.htm

Text of the paper, read in this form in Amsterdam, Meertens Instituute, the 17th of september 2014

Slide 1 (Titeldia)

Hans Goebl

Language in space and language on maps:
some methodological and historical considerations

Dear colleagues! The aim of my talk is to present to you the close relationship which existed and still exists in the field of Romance geolinguistics between empirical research and theoretical thought. For me as for many other Romanists in the past and the present, the mapping of geolinguistic facts had and has always two functions: 1) the verification of already existing assumptions and expectations, and 2) the completely unexpected discovery of new insights on maps which had been drawn before – so to speak – in complete scientific innocence: you might know that this kind of discovery is often called „serendipity".

You might also know that the founding father of Romance geolinguistics is Jules Gilliéron, and that our geolinguistic „bible" is his „Atlas linguistique de la France" (ALF), which I consider one of he greatest achievements of linguistic research in general. When creating the ALF, Gilliéron could avail himself – despite his undeniable personal merits – of a specific geographic and cartographic substratum which had developed continuously in France from the XVIIth century. The focal point of this evolution was the desire of the royal administration to know as many details as possible of the life of the entire body of France, beginning with the *size* of the territory and the *number* of its inhabitants, and ending with very particular aspects of *economic*, *financial* and *social* life. This geodetic tradition was substantially accelerated during the French Revolution and the subsequent Napoleonic administration.

One of the most important measures in this regard was the establishment of the famous departmental grid which contributed enormously to the standardization and promotion of statistical investigations of every sort. So, a top-ranking officer of the Napoleonic administration, Charles-Eugène Coquebert de Montbret, organized, between 1806 and 1813, the first nationwide inquiry on linguistic differences by collecting a great quantity of dialect translations of the well-known „Parable of the Prodigal Son".

The empirical and cartographical zeal of nationwide inquiries continued, indeed increased after Napoleon. So, we can observe the establishment, in France, of an excellent *theory* and *practice* of geographic and cartographic research.

Slide 2 (Literacy in France: Charles Dupin 1827)

This is an excellent example of these activities which very often led to brillant mappings. Here you see a map of the spatial distribution of *literacy* in France. The clear shape of the map triggered a lot of acrimonious political and intellectual debates, aiming to *diminish* or even to *deny* the cultural differences between „la France illuminée“ in the North-East and „la France obscure“ in the South-West.

Slide 3 (Crimes in France: Adolphe d'Angeville 1836)

The following two slides are socially relevant. The choroplethic map drawn by Adolphe d'Angeville shows the spatial distribution of *criminal energy* in France, whereas…

Slide 4 (Regional origin of Parisian prostitutes: Alexandre Parent-Duchâtelet 1836)

…the Parisian physician Alexandre Parent-Duchâtelet analyzed the regional origins of the prostitutes of Paris, and their numerical presence in the 48 quarters of Paris.

Note that these activities never stopped during the 19th century and permeated the whole political, administrative and intellectual landscape of France. Obviously there were also vast numbers of coloured maps, and even a kind of competition between different empirical and cartographic „schools“.

Jules Gilliéron was born and educated in French-speaking Switzerland and came to Paris in 1876, when he was 22 years old. Ten years later, he took French citizenship. But already from 1882 onwards he was entrusted at the famous „École Pratique des Hautes Études“ with the teaching of *Gallo-Romance dialectology*. In the last two decades of the 19th century he underwent a rapid intellectual evolution as the result of close contact with the upper echelons of Parisian academic life. The span of this evolution becames apparent by the comparison of the internal structure of his first linguistic atlas published in 1881, and that of his later masterpiece, the ALF, published from 1902 onwards.

Slide 5 (Map 3 of the first linguistic atlas compiled by Jules Gilliéron (1881)

Here you see one of the 30 maps of his „Petit atlas linguistique du Valais romand“, where the data presentation is still very rudimentary. Note that the colouring of the map is done by hand and that the coding principle which has been adopted blocks our direct insight into the collected data.

In contrast the principles of data presentation applied in the ALF – compiled at the end of the 19th century and published from the beginning of the 20th century – is much better.

Slide 6 (Map 1 *abeille* („bee“) of the „Atlas linguistique de la France“ (ALF), compiled and published by Jules Gilliéron and Edmond Edmont (1902 sqq.))

Here you see the northern part of the first map of the ALF showing more than 600 names of the *bee* in fully-fledged phonetic transcription and distributed in space according to their real geographic location. This kind of mapping, which later became prevalent in Romance linguistics, is called a „full text map“. Considered individually, such a mapping is not very innovative or even revolutionary.

Slide 7 (Blank copy of a „silent map“ for the (geo)linguistic interpretation of ALF maps)

It gets however its very *innovative impact* by the parallel use of so-called „silent maps“, which have been distributed at low costs and in great quantities, and already have been since the beginning of the publication of the ALF. When filling in a silent map, the researcher had to isolate first some specific aspects of the original raw data and then transpose them graphically onto the silent map. This activity could only be executed if based on clear *theoretical* and *analytical* grounds. Very early, Romanists called this activity *discussion*, *typisation* or *classification* of ALF data. The *discussed* (or *classified*) silent maps could be used only for personal or academic purposes, although some were published, which happened, s long ago as before World War I, very frequently in colour.

Slide 8 („City block“ colour map, drawn by Jules Gilliéron (1905): the names for „cock“ (*coq*) in Gascony)

This is a classical mapping published by Gilliéron himself in 1905, showing the names of the cock in Gascony which, amongst others, served to explain the onomasiological fragmentation of Gascony. Gilliéron described it as a consequence of the refusal of the local speakers to accept the same denomination – namely *gat* – for the cock and the cat as well. He called this kind of linguistic blocking „homonymie fâcheuse“ (unfortunate/problematic homonymy).

Slide 9 („Free style“ colour map, drawn by Guylaine Brun-Trigaud (2005): the names for „cock“ (*coq*) in Gallo-Romance)

Obviously there are many other ways of visualizing the linguistic complexity of ALF maps. This is a more recent mapping published in 2005, already using electronic devices, but ignoring formal constraints such as Voronoi polygonization or frequency counting.

Slide 10 („City block“ map, drawn by Jules Gilliéron (1918): the names for „bee“ (*abeille*) in Gallo-Romance)

This is another classic map drawn by Gilliéron himself, published but in black-and-white for economic reasons. Note the persistence of the use of *city block technique* in cartography!

Slide 11 („Free style“ colour map, drawn by Guylaine Brun-Trigaud (2005): the names for „bee“ (*abeille*) in Gallo-Romance)

This slide is based on the same original data but has been produced by neglecting some less frequently-attested denominations of the bee. Note that when filling in silent maps, one can proceed by *lumping together* or by *splitting* the raw data!

Slide 12 (Polygonized working map of VDM: the names for „bee“ (*abeille*) in Gallo-Romance)

Splitting – which means taking into account all the details of the original map – was the guideline when I established this working map during the dialectometrization of a huge number of original maps from the ALF.

Slide 13 (Traditional isogloss synthesis drawn by Karl Jaberg (1908) showing the variable intensity of the conservation of Latin C+A in Gallo-Romance)

Cartographically speaking one can develop silent maps in two directions: 1) for the creation of *line* maps or 2) for the creation of *area* maps. In both cases there were early attempts to create synthetic mapping in order to visualize the variable intensity of a given geolinguistic feature. Here you see a map drawn by Karl Jaberg in 1908, in his famous introduction to the ALF, which shows the changing degree of preservation of original Latin k, derived from the Latin nexus C+A in the North and in the South of France. Obviously the heuristic value of *line*-based synthesis is rather limited. Analogous attempts with *area*-based maps gave much better results.

I'll show you now some handmade examples, which have been created in the same style by my students on silent maps of ALF, enhanced by Voronoi polygonization. First I'll show you four typifications related to the already mentioned etymological context: the development of Latin C before A. Note that the colour blue marks the preservation of Latin k.

Slide 14 (Hand-drawn taxation of ALF 224 *chambre* „room" (< CÁMERA): Gallo-Romance outcomes of Latin C+A)

Slide 15 (Hand-drawn taxation of ALF 225 *champ* „field" (< CÁMPU): Gallo-Romance outcomes of Latin C+A)

Note also that despite the identical etymological context the different typifications - which in dialectometry I call „taxation" – diverge slightly from each other.

Slide 16 (Hand-drawn taxation of ALF 229 *chandelle* „candle" (< CAN+*DÍLLA): Gallo-Romance outcomes of Latin C+A)

Slide 17 (Hand-drawn taxation of ALF 233 *chanter* „to sing" (< CANTÁRE): Gallo-Romance outcomes of Latin C+A)

Slide 18 (Hand-drawn isogloss synthesis showing the variable intensity of the preservation of Latin C+A in Gallo-Romance (based on ALF maps 224, 225, 229 and 233)

This is again a *line*-based synthesis of the k-preservation observed on our four classified maps from the ALF. No doubt that you will agree when I suggest that its heuristic expressivity is rather *low*…

Slide 19 (Hand-drawn choroplethic map showing the variable intensity of the preservation of Latin C+A in Gallo-Romance (based on ALF maps 224, 225, 229 and 233))

.. whereas the message of an *area*-based synthesis is *much better* and *clearer*.

Slide 20 (Density map, drawn by Guylaine Brun-Trigaud (2005) showing the variable intensity of the preservation of Latin C+A in Gallo-Romance (based on 47 ALF maps))

This is a similar synthesis, based on 47 maps from the ALF, which was produced in 2005 by Guylaine Brun-Trigaud, using a computer program but in a merely graphical way, without any numerical definition of the changing degree of k-preservation. Nevertheless the expressivity of this map is quite good.

Now I'll show you some examples of *line*-based attempts at data synthesis which were all produced immediately after World War I by Austrian and Swedish Romanists. It's very strange to see that *line*-based isogloss synthesis was not produced by French researchers before 1955. It seems that their strong *space-dividing power* was for a long time too dangerous for French eyes.

Slide 21 (Traditional isogloss synthesis, drawn by Karl von Ettmayer (1924) showing the general dialectal structure of Northern Gallo-Romance (based on 97 ALF maps))

This slide and the next one were produced in 1924 by the Austrian Romanist Karl von Ettmayer in order to illustrate his seminal book „Über das Wesen der Dialektbildung" (On the nature of dialect formation), which I consider one of the best dialectological texts ever written, and not only in the field of Romance linguistics. This is a general isogloss synthesis whereas the following slide …

Slide 22 (Composite isoplethic map, drawn by Karl von Ettmayer (1924) showing the presence of geolinguistic incursions (*Einbrüche*) and relicts (*Schollen*))

… serves to illustrate two *geodynamic phenomena* which occur on many of our atlas maps: namely geolinguistic *expansion* and geolinguistic *withdrawal*.

Slide 23 (Traditional isogloss synthesis, drawn by Arvid Rosenqvist (1919) showing the general dialectal structure of Gallo-Romance (based on 68 ALF maps))

This is the best Gallo-Romance isogloss synthesis realized by *traditional* means, which has often been reproduced for propaedeutic purposes. Its author is the Swedish Romanist Arvid Rosenqvist.

Slide 24 (Dialectometric isogloss synthesis of Gallo-Romance (based on 1681 ALF working maps and produced by VDM))

When using our computer program VDM for the production of isogloss synthesis one gets this result, whose cartographic appearance can be easily changed concerning the *number* and the *distribution* of colours, and the *thickness* of polygon sides as well. This drawing can be considered as the computer-assisted endpoint of the whole *interpoint technique*, which was introduced in geolinguistics, as early as 1898, by the German linguist Carl Haag.

My personal route to this type of mapping was not so easy. Before the general availability of plotters and good graphics software, we had no choice but to produce such maps by hand.

Slide 25 (Draft of a „naive" interpoint synthesis, drawn by Hans Goebl (1981) (based on 696 AIS working maps))

Here you see an early attempt going back to the middle of the seventies, when I did not yet have any access to Voronoi polygonization and plotters. Nevertheless note the presence of a general triangulation giving each interpoint its exact place between two neighbouring atlas points, but also the graphic *rudimentarity* of the intensity signatures in black-and-white technique.

On the contrary, for the publication of my habilitation thesis in 1984, I benefitted already from the availability of polygonization programs and powerful plotters.

Slide 26 (Dialectometric isogloss synthesis of Northern Italy, drawn by Hans Goebl (1984) (based on 696 AIS working maps): linear growth of the thickness of polygon sides)

This is a monochrome visualization where the thickness of polygon sides changes in a linear manner: see the *wedge-shaped* legend below. To my eye, nowadays, the heuristic impact of this visualization is fairly *small*.

Slide 27 (Dialectometric isogloss synthesis of Northern Italy, drawn by Hans Goebl (1984) (based on 696 AIS working maps): algorithmic swelling and colouring of the polygon sides)

In the late 70s, I discovered that the use of algorithmically-defined intervals and rainbow colours could improve considerably the heuristic quality of the maps.

Slide 28 (Dialectometric isogloss synthesis of Northern Italy (based on 696 AIS working maps and produced by VDM))

Here you see the ultimate solution in VDM where the range and the distribution of *colours* as well as the *thickness* of the polygon sides can be easily modified, as and when required.

Let me now go back to the *area*-based mapping of linguistic features of all kinds. I guess that since the time of Jules Gilliéron, Romanists have produced and discussed, in the many branches of Romance linguistics, several thousand silent maps, of which only a small minority has been published. One of the great discoveries made in this period was the complicated and polymorphic behaviour of single-feature areas. Later on, I'll call this property „specific entanglement of areas".

Gilliéron himself developed a sophisticated method of discussing the past and present state of such areas, which he called *stratigraphie linguistique [linguistic stratigraphy]*. Furthermore, he claimed that every area is the product of particular human behaviour, driven by specific psycholinguistic principles and necessities.

Slide 29 („Free style" colour map, drawn by Guylaine Brun-Trigaud (2005): the names for „to saw" (*scier*) in Gallo-Romance)

This map refers to a classic paper of Gilliéron where he described the long lasting struggle between reflexes of the original Latin verb SECÁRE (Fr. *scier* „to saw") and some younger competitors in terms of processes of the *birth* and *death* of words. During this long period and on the basis of many experiments, we learned that the *shape* and the *frequency* of the different areas could vary enormously on the silent maps in use, and that, in practice, each *area* has its own history, just as – since the period of Jakob Grimm – each *word* is meant to have its own history – in Gilliéron's famous formulation, „chaque mot a son histoire".

Slide 30 („Free-style" colour map, drawn by Guylaine Brun-Trigaud (2005): the names for „garden" (*jardin*) in Gallo-Romance)

In the course of these studies there arose a close link between Romance *geolinguistics*, Romance *historical lexicology* and Romance *diachronic linguistics*, witness the beginning of the great etymological dictionary of French – „Französisches etymologisches Wörterbuch"

(FEW) promoted, compiled and published by Walther von Wartburg – whose first volume bears a double dedication to Jules Gilliéron, the founding father of Romance geolinguistics, and to Wilhelm Meyer-Lübke, the greatest Romance specialist of historical grammar, which he practised largely following the principles of the Neogrammarians of Leipzig.

Obviously the majority of such mappings have been published in black-and-white as demonstrated by this slide…

Slide 31 (Free style" line map, drawn by Gaston Tuaillon (1976): the names for „pig" (*cochon*) in Gallo-Romance)

… drawn from a modern French review dedicated mainly to problems of linguistic geography.

As already demonstrated in some previous slides, the presentation of synthetic data remained somewhat problematic because of the *absence of colours*. It's obvious that the linguistic and visual expressivity of the two following slides could only …

Slide 32 (Density map, drawn by Guylaine Brun-Trigaud (2005) showing the variable diffusion of *-o* (< Latin -ÉLLU) in Gallo-Romance (based on 6 ALF maps))

…be achieved using *colours*. This slide furnishes a synthetic view of the spread of a typical Parisian phonetic feature. Note the *parachuted*- and *tentacular*-like diffusion of the Parisian sound *-o* in words such as *bateau* (boat), *beau* (beautiful), *château* (castle) and so on.

Slide 33 (Density map, drawn by Guylaine Brun-Trigaud (2005) showing the variable diffusion of the diphthongs *-wɛ* and *-wa* (< stressed Latin -É[) in Gallo-Romance (based on 10 ALF maps))

This slide exhibits the same spatial dynamics, albeit in more intense form for the spread of the diphthong *wɛ̀* und *wa* from the Ile-de-France, the heart and linguistic centre of Northern French.

Slide 34 (Density map, drawn by Hans Goebl (1973) in „naive" city block technique, showing the variable diffusion of the enunciative particle *ké* in Gascony (based on 12 maps of the „Atlas linguistique de la Gascogne", vols. I-II))

In this respect, I produced, myself, in my early years, some very *naive* cartographic „exploits", due to lack of colours and of appropriate polygonization.

The following slide shows you another attempt realized in the same style, but in city block technique. Linguistically, it refers to medieval data from Normandy.

Slide 35 (Density maps, drawn by Hans Goebl (1979) in city block technique, showing the variable spatial diffusion of the preservation of Latin C+A in medieval documents and modern dialects of Normandy)

Let me now comment on another aspect of mapping in geolinguistics. As I have already mentioned, each of the 1421 maps of the series A of ALF embraces the whole territory of France or – in other words – the *totality* of the *object under investigation*. This synoptic effect was and is very fruitful. Let me quote just one historical detail for the sake of clarity. In 1901. i.e. one year before the publication of the ALF, the French Ministry of Internal Affairs –

which hitherto had supported financially the ALF enterprise – asked for an expert report on the publication of the ALF from an elderly, but still highly renowned, French philologist. His name: Paul Meyer. In the 70s and 80s of the 19th century, he denied vigorously the *existence of dialects,* claiming that only the cartographic study of isolated linguistic features is worthwhile. In his report, Paul Meyer condemned the Gilliéronian idea of publishing his data in synoptic maps and argued instead for the publication of the ALF data in *tabular form*.

Fortunately, Gilliéron succeeded in enforcing his personal view! You can easily imagine the damage to geolinguistics, if Paul Meyer had won the argument!

Data publication in tabular form is not lacking in Romance geolinguistics. But it is very rare: see the following example drawn from the „Tableaux phonétiques des patois de la Suisse romande“, published in 1925.

Slide 36 (Publication of geolinguistic data in (coherent) tables: „Tableaux phonétiques des patois suisses romands“ (TPPSR) (1925))

The dimensions of this pseudo-atlas are as follows: 62 inquiry points and about 400 „maps“ presented in tabular form. Very probably the publication of the „Tableaux phonétiques des patois de la Suisse romande“ in tabular form is due to economic reasons.

As a Romanist, I was always amazed by the fact that 50 years after the publication of the ALF and the glorious flourishing of French (and Romance) *géographie linguistique*, the data of the linguistic atlas of England – called the „Survey of English Dialects“ (SED) – have not been published in maps or at least in coherent tables like the „Tableaux phonétiques des patois de la Suisse romande“. In reality the data of the SED have been published in a very intricate tabular form…

Slide 37 (Publication of geolinguistic data in (dispersed) tables: „Survey of English Dialects“ (SED) (Basic Material, 1962-1971)

…so that the equivalent of one atlas map must be sought in several different volumes of the „Basic material“ of the SED. Here you see the beginning of the data of the „map“ referring to the names for *bilberries*. You see the answers for only 6 of the historical 40 counties of England, beginning with Northumberland and Cumberland and, in addition, with the Isle of Man. In my mind, it’s a real miracle that nevertheless several atlas-like discussions of these data have been compiled manually or by computer, and been published later on.

Slide 38 (Map of the 16 sub-atlases of the series „Reeks Nederlandse Dialectatlassen” (RND) (1925-1982))

I am wondering whether the solution – adopted for the geolinguistic inquiry in Flanders and Holland – to create a linguistic atlas for such a great domain as the sum of several regional atlases – was altogether efficient. When you are eager to check the 638 Gallo-Romance variants for *tree*, you only have to look at map 51 of the ALF; when however you are eager to do so for the 360 variants of *tree* in Flanders and Holland, you must first get 16 different books, set them out on a big table and finally open them at the same page. Obviously, this is not the best way to promote geolinguistic thinking!

In this context allow me a critical remark concerning the general scientific context in the past: the rather problematic English and Dutch-Flemish experiences both took place some decades after the brillant French exploit of the ALF and the development of the related discipline of *géographie linguistique*. In retrospect, this prompts some serious doubts about the quality of inter-*national* and inter-*disciplinary* cooperation as practised at this time.

After the presentation of the principles and techniques of Gilliéronian geolinguistics, let me now discuss the respective output since the emergence of dialectometry. As dialectometry is based on counting, we became rapidly aware of the numerical dimensions which exist inside the numerous coloured „taxations" (or working maps) of original atlas maps, and also of the relationship which exists between working maps which show different geolinguistic fragmentation.

Slide 39 (Histogram showing the relationship between *map polynymy* (2-90) and the *number* of corresponding working maps. Corpus: ALF, 1681 working maps)

This histogram refers to our dialectometrization of about the half of the original maps of the ALF. It is based on 1681 working maps where you can find almost 19 000 taxatorial areas of various size, shape and location within the grid of the ALF. This curve shows you the ratio which exists between *roughly* structured working maps – provided thus with little *polynymy* – and *finely* structured working maps – provided therefore with a great *polynymy*. On the x-axis, the polynymy goes from 2 to 90. The y-axis shows you that there are about 250 *di*-nymic working maps and also that with increasing polynymy of the working maps their respective number decreases continously.

The shape of the curve is very similar to that of a Zipf-Mandelbrot-distribution, well known in lexicometry. The same regularity could be observed in all our dialectometric analyses which embrace – beside France – Italy, Spain, Portugal, Catalonia and England as well. As early as 1985, the German quantitative linguist Gabriel Altmann demonstrated that these regularities are the natural property of synergetic systems based – metaphorically speaking – on the alternation of *birth*, *florescence* and *death* of their components.

However, the real components of these synergetic systems are the *polymorphic areas* whose discovery had been rendered possible by the mass application of the famous *silent maps*.

(Just a little incidental remark: note that by jumping over the data matrix and thus by avoiding the whole areal analysis – as it occurs in Levenshtein dialectometry – you can never detect what I am now talking about.)

As to the structural properties of the areas which I have just mentioned, the dialectometric analysis of the data matrix led to the following insights:

1. The taxatorial areas, even when they are etymologically similar, never coincide precisely.
2. The spatial co-existence of the taxatorial areas is very complicated. It resembles the *interwovenness* of a complex tissue. I call this effect „special entanglement of areas". This phenomenon, which has never been seriously studied before, seems to be the very secret of the function of geolinguistic networks.

3. As a consequence of the special entanglement of areas, each locality of a given geolinguistic grid is characterized by the existence of a specific set of areas of different size and shape, which obviously reflects the communicative position of the given locality within the respective grid.

The next slide will illustrate this fact. It refers to data drawn from the Italian linguistic atlas AIS.

Slide 40 (one pair of taxatorial density maps for AIS-P: 1 (Breil/Brigels) and AIS-P. 376 (Venice))

The „pillar“ on the left demonstrates the potential of small and greater areas of a *peripheral* locality amidst the AIS grid, whereas the „pillar“ on the right belongs to Venice, whose interactional position in the AIS grid is very *central*. Each pillar consists of ten horizontal bars of variable length. Each of them indicates the number of the respective taxatorial areas. At the bottom, you see the number of *small* areas, at the top the number of *big* areas. The average size of the big areas is near the size of the whole AIS grid.

At first glance you can see that the *areological potential* of the peripheral locality (Breil/Brigels) is quite different from that of Venice, and that it is characterized by a *greater* number of *small* taxatorial areas, and by a *smaller* number of *greater* taxatorial areas, whereas in Venice you find the opposite ratio. Obviously this divergence is a consequence of the different linguistic histories of the two localities.

The areological circumstances visualized in this slide can be mapped in a very impressive way by summing up, in ten steps, the respective areas in the form of ten „density maps“.

Slide 41 (ten pairs of taxatorial density maps for AIS-P. 1 (Breil/Brigels) and AIS-P. 376 (Venice))

The visualization of all density maps is done by six colour classes. The red zones show the highest areological coincidence. On the left, this refers to the western part of Surselva (in Grisons), on the right, to Venice itself. The yellow zones symbolize a weaker degree of coincidences indicating zones of a looser degree of „special entanglement“ with the AIS atlas points 1 or 376. Attention: one should not confuse *area entanglement* with *linguistic similarity*.

In a few seconds, I'll start the dynamic picture sequence which shows, in ten steps, the evolution of the special entanglement by steady amplification of the size of the taxatorial areas. Each step will last about three seconds. You will note on the choroplethic maps, here presented in pairs, a progressive loss of areal *specificity* and a parallel increase of areal „*banality*“, which will become overwhelming from the seventh step onwards.

Start

I emphasize that this demonstration is only based on the *data matrix* of our AIS-project and *excludes* therefore both the *measurement of similarity* and the use of the *similarity matrix*. Obviously such demonstrations could be done with all our dialectometric corpora.

For many years, I have been pointing out that linguistic atlases reflect a specific *linguistic behaviour* of man in space which is not chaotic but governed by *special laws* or *regularities* which are comparable to many other linguistic laws which are already known.

I'll show hereafter another „hidden order", whose discovery is intimately connected with the mapping of taxatorial areas.

Slide 51 (Histogram showing the relationship between *map polynymy* (2-90) and the *number* of corresponding working maps. Three polynymic sub-corpora: 2-5, 6-12, 13-90 taxates/working map)

You already know the shape of this histogram. In its new form it reflects the partition of our ALF corpus into three sub-corpora. Each of them has an equal number of working maps. The sub-corpus A contains *clearly* shaped working maps which are often reproduced in elementary text books and general introductions for undergraduate students. The sub-corpus B refers to *tightly* shaped working maps whose geolinguistic interpretation requires already a good command of Romance linguistics. The sub-corpus C however refers to working maps whose inner areological structures are very *complex* or even *chaotic*. Normally such maps served only as a source for purely *etymological* research and have never been used for genuine *geolinguistic* investigations.

At first glance, one would assume that only the two first corpora will contain well-ordered deep structures. However, evidence shows that this is not true.

Slide 52 (Comparison between three dialectometric analyses based on different polynymic data sets. Visualized variable: skewness (= linguistic compromise))

On this slide, the three dialectometric profiles based respectively on the corpora A, B and C, furnish practically the same result.

Obviously you can create different dialectometric data matrices also by randomly changing their quantitative volume.

Slide 53 (Comparison between three dialectometric analyses based on data sets with different size: 100, 338, 843 working maps. Visualized variable: skewness (= linguistic compromise))

Here you can see again three dialectometric profiles which were calculated on the base of three corpora which have been created by random extraction from a pool of data containing the total corpus of 1681 working maps.

The lesson of this experiment is that the deep structures hidden in the data matrix are very robust and have therefore a very high degree of inner redundancy.

Let me conclude my talk with some demonstrations referring to the phenomenological difference between *maps* and *formulae*. The former belong to the world of *images*, the latter to that of *numbers*. The problem is that in geolinguistics – which is a map-driven discipline – *maps* often depend on *formulae*.

Slide 54 (Comparison between three similarity maps calculated by the „Weighted Identity Value“ [WIV(x)] with three different weights: x= 1, 1.5 and 2. Reference point: ALF-P. 760, Léguevin)

This slide shows you three similarity maps with a given reference point from the ALF, generated by three different states of the same *similarity formulae*. Their choroplethic profiles differ significantly, mainly by the frequency of dark blue polygons. Whereas the difference between the three formulae states is absolutely clear from the *statistical* point of view, such is not the case for the differences between the *linguistic relevance* of the three choroplethic maps.

Slide 55 (Comparison between three similarity maps calculated by three different similarity indexes: Average Euclidean Metric, Average Manhattan Metric, Bravais-Pearson Correlation Coefficient. Reference point: ALF-P. 760, Léguevin. Corpus: ALF, 1681 working maps)

Here you have three other similarity profiles belonging to the same reference point, but created by three other similarity measures. Statistically speaking, they are completely different. In an overall perspective this is true also for the three choroplethic maps, which differ considerably from each other. When choosing one of these three similarity indexes, it was completely impossible to foresee, on the one hand, the structure of the future choroplethic map and, on the other hand, the difference between the three visualizations.

A few moments ago, I evoked the fundamental difference between the worlds of *numbers* and that of *maps,* which are linked by the tie of visualization. Unfortunately we forget too often that when passing from the world of *numbers* to the world of *maps* we change automatically the status of data and of information that we rely on. So, what may be *significant* in terms of *numbers*, can be *irrelevant* in terms of *maps* or *pictorial structures*. And obviously *vice versa*.

In this respect, I will show you now three fading sequences, where you can see, on the left, *maps* and, on the right, the respective histograms, which clearly reflect the world of *numbers*.

Slide 56 (Dynamic cartography: sequence of 31 similarity maps in rectilinear order)

We will complete three typo-diagnostic journeys from Gascony to Wallonia. The first journey will be done in a very rough perspective, using only *two*-colour-maps. The second trip will run along the same path, but with *four*-colour-profiles, whereas the third journey will be realized with *six*-colour-maps. Each journey consists of 31 similarity profiles, arranged following a straight line in black which leads from Gascony to Wallonia. The black line contains some small dashes which predict deeper changes of the current similarity profile.

After the start in Gascony, we touch a linguistic island with a French (and not Occitan) dialect, enter an ambiguous transitional zone between Southern and Northern France, and arrive finally in the domain of genuine Northern French dialects.

Note that your eyes have to cope with two *intellectual dimensions* or *challenges* and two kinds of *changes*: on the left with *maps* and therefore with *linguistics*, on the right with histograms and therefore with statistics. It’s up to you whether to concentrate on the left or on the right, thus to have a look as a *linguist* or as a *statistician*.

Each journey will last one minute.

Slide 57 (Start: 2-tuple)

Slide 88 (Start: 4-tuple)

Slide 120 (start: 6-tuple)

As you could see, both the choroplethic profiles and the histograms continuously changed their structures and shapes. As a *linguist*, I am mainly interested in the change of the *choroplethic profiles* and much less in changes of the height and the location of the *histogram bars*. You may decide for yourself whether you share my emphasis on *maps* or whether you prefer the focus on *numerical histograms*. Obviously this is also a *philosophical* problem which cannot be resolved within the context of this talk.

Slide 152: Schlußdia