On the geolinguistic change in Northern France between 1300 and 1900: a dialectometrical inquiry

Hans Goebl
Salzburg University
Department of Romance Philology
Akademiestrasse 24
A-5020 Salzburg
hans.goebl@sbg.ac.at

Abstract

With the supply of 8 closely interpreted dialectometrical maps, this paper analyses the linguistic change of the geolinguistic deep structures in Northern France (Domaine d'Oïl) between 1300 and 1900. As a matter of fact, the result will show – with one exception – the great stability of these deep structures.

1 Introduction to the issue

Through the comparison of two data sets of 1300 and of 1900, the present contribution discusses, if and in which way the basic geolinguistic structure of Northern France (Domaine d'Oïl) changed in the course of this period. In this investigation, a number of different methods of dialectometry (DM) will be applied. DM is a subdiscipline of quantitative linguistics which concentrates on the exploration of the actual deep geolinguistic structures of a given space, using as data source linguistic atlases or similarly structured data collections (consisting of N inquiry points and p atlas or working maps). Of course, it has to be assumed that these deep structures were generated by a genuine specific activity of man (i.e of of the *homo loquens*), that is to say: the « linguistic (or dialectal) management of space by the *homo loquens* ». Insofar as man has obviously many *other* opportunities of managing a given natural space besides the *linguistic* management, there result many opportunities for interdisciplinary cooperation with DM.

The Salzburg-based DM (Goebl 2006a) pursues the genuine principles of traditional (Romance) linguistic geography with quantitative means. It therefore defines its main aim in the empowering of the diagnostic virtue of traditional linguistic geography by introducing global or synthetic (quantitative) methods.

2 Data basis

It consists of two machine-readable data matrices, the first resuming the period around 1300, the other one resuming the period around 1900.

2.1 Corpus 1300 (drawn from Dees 1980)

The medieval corpus was borrowed from the scripta-atlas (1980) of the Amsterdam Romance linguist A. Dees. This atlas is based on the comprehensive interpretation of 3300 original charters of Northern France of the second half of the 13th century, which were analysed in that instance according to a list of ca. 300 written (or scripta-) attributes. These scripta-attributes are mainly of phonetic relevance, most of them referring to vocalism (189 attributes), but also to consonantism (87 attributes), and some of them even to morphology (22 attributes). As a result, the data matrix holds 298 attributes and 85 « inquiry points ». The latter correspond actually to *scripta-centres* (scriptoria, chanceries) which are distributed as evenly as possible all over the Domaine d'Oïl. For the measuring of the graphic variation in the 3300 charters, A. Dees developped a specific method. As a result, he was able to determine – for each single attribute – its relative occurrence (in percentage) in the charters of the 85

Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology, pages 75–83, Prague, June 2007.

scripta-centres. The content of the data matrix lies therefore on a *metrical* scale.

In the ninetees, A. Dees and his collaborator Piet von Reenen handed me over this data matrix, as a basis which allowed me to realize many dialectometrical experiments. Its only disadvantage is that the machine-readable matrix holds less attributes (268) than the printed atlas (298). Nevertheless, by applying the « Average Euclidean Metric » (AEM), the « Average Manhattan Metric » (AMM) and the « Bravais-Pearson correlation coefficent » [r(BP)], the dialectometrical results are very profitable (see Goebl 2006b). The scripta-atlas published by Dees in 1980 shows quantitative visualisations of the spatial distribution of the 298 attributes, but does not encompass global data interpretation with dialectometrical (or similar) methods.

2.1.1 The Dees-data: one illustrative example

In his scripta-atlas (1980: carte 87, p. 93), A. Dees also investigated the regional variation in the spelling of the French possessive pronoun: *leur, leurs, leurz* etc. which are all derived from the Latin etymon ILLÓRU. Most probably they were created under the influence of a specific regional dialect pronunciation. At the end of the 13th century, the geographic contrast between these *eu*-spellings and the older equivalent forms *lor, lors, lors* etc. was quite sharp in the Domaine d'Oïl. Hence, Dees checked the number of all occurrences of *eu*-spellings (belonging to the possessive pronoun) in the 3300 above mentioned charters and listed, for each of the 85 scripta-centres of his atlas, the percentages of those charters which show at least one occurence of the spelling *-eu-*. As a result, 81 out of the 93 charters of the scripta-region 26 « Somme, Pas-de-Calais » (located in the medieval Artois: see the top of the figures 1, 3, 5 and 7) showed a considerable amount of *eu*-spellings, unlike the remaining 12 charters. In the 105 charters of the scripta-region 1 « Charente, Charente-Maritime » (South-western corner of the Domaine d'Oïl), no occurrences of the *eu*-spellings were found. Obviously, the different spellings of the possessive pronoun in that region were still on *-o-*. Thus, Dees registered the value 87% (= 81 : 93) for the scripta-region 26 in the North and the value 0% for the scripta-region 1 in the South-west.

As Dees analysed 298 scripta-features in the same way, he succeeded in covering the whole range of the stressed and unstressed vocalism and consonantism of Old French.

2.2 Corpus 1900 (drawn from ALF)

The second corpus, referring to 1900, was drawn from the data of the French linguistic atlas ALF, precisely: from a data matrix which had been established in the process of dialectometrization of the total ALF grid. The dimensions of this data matrix are: N = 641 inquiry points (distributed all over France), p = 1687 working maps, 1117 referring to phonetics (612 to vocalism, and 505 to consonantism), 417 referring to vocabulary, and 99 to morphology. 347 inquiry points (out of the 641 points on the total ALF grid) are located in Northern France : they represent therefore the Domaine d'Oïl. Among these 347 inquiry points, 85 points were selected in geographic correspondance to the 85 scripta-centres of the Dees-atlas, and subsequently reunited to a new grid (see the right halves of Maps 1-8).

Among the 1687 workings maps mentioned above, we took only into consideration those of phonetic relevance, thus: 1117 maps. They derived from 247 original maps of the ALF by phonetic typization, which is a common procedure in Romance linguistics. The units of this ALF data matrix are upon the *nominal* scale. With the supply of the « Weighted Identity Value (with the weight 1) » [WIV (1)], the dialectometrical interpretation of this data matrix proved to be very successful (see Goebl 1984, I: 83-86, and 2006a: 418-419).

2.2.1 The ALF-data : two illustrative examples

An example for two characteristic phonetic features is given in Map 812 of the ALF *le marché* « the market ». The 85 occurrences in the Domaine d'Oïl all derive from the Latin etymon MERCÁTU. The different dialectal followers of the stressed Á which is considered in this instance show the following results: a) pronunciation with *-i* (19 ALF-points), b) with (closed) *-é* (60 ALF-points), c) with (open) *-è* (1 ALF-point), d) with *-ö* (4 ALF-points), e) with (neutral) *-e* (1 ALF-point). From the metrological point of view, these five phonetic types represent what is called « (nominal) multistate characters ». As the corresponding working map contains five different (phonetic) types (or « taxates » in Salzburg terminology), it is called as well a « 5-nymic working map ».

Nevertheless, the data of the same ALF-map can also be analysed according to consonantal principles, which is realized by listing the dialectal results of the postconsonantal C before stressed Á in MERCÁTU. The results are as follows: a) *š* (72 ALF-points), b) *šy* (2), c) *ts* (1), d) *tšy* (3), e) *k* (2), f) *tš* (3), g) *ky* (1), h) *ty* (1). On the map, these eight consonantal types show a geographic distribution which is far from being similar to the former one of the five vocalic types. Actually, this experience is also valid for the great majority of our ALF-working maps.

The reduced data matrix drawn from the integral ALF-grid (with 1687 working maps) consists of 914 working maps: it starts with 2-nymic maps and has up to 23-nymic maps, embracing a total of 4263 phonetic types or « taxates ».

3 Establishement of the dialectometrical maps

DM is a map-based discipline: It visualises systematically all its results by using previously defined cartographic standards and by a very handsome computer program called VDM (« Visual DialectoMetry »), which supports and resets these visualisations perfectly. With VDM, choropleth maps and isarithmic maps, as well as trees can be generated. The results are always mapped in colours that are ranged according to the solar (or rainbow) spectrum, the warm colours lying above the arithmetic means of the respective frequency distribution, and the cold colours below it. The trees are all « spatialized » in principle, which means that their structural information is projected directly from the tree on the map.

The comparison between the medieval versus the modern data occurs basically in visual form, a methodically correct procedure, as the two corresponding iconic patterns are established according to the same cartographic norms. Further, the respective frequency distributions may also be correlated in order to gain a correlation map. For reasons of space, this procedure will not be demonstrated in this paper.

All the maps shown in section 4 are taken from two square similarity matrices (N x N) consisting of 85 items (N = 85), calculated by means of special similarity indexes – AEM and WIV(1) – on the basis of two data matrices (N = 85 ; p_{1300} = 268 metrical attributes, p_{1900} = 1117 nominal attributes). Hence, this demonstration includes two similarity maps, two parameter maps, two interpoint maps and two trees (with the respective spatializations). These four comparison planes are actually of special relevance, by allowing a global comparison which is also precise to the last detail of the medieval versus the modern data.

4 Four comparison planes between 1300 and 1900

4.1 Comparison plane 1 : two similarity maps

The most important instrument of DM is the similarity map. Each similarity map consists of a reference point and N-1 similarity values distributed in space, which values decrease proportionally with their geographical distance from the reference point. The geographic pattern of the progressive drop of these measurement values is clearly shown with the cartographic means of DM. In Maps 1 and 2, the reference point is located in the Poitou (South-west). The visual comparison of the two choropleth profiles shows their great similarity. The same effect occurs also from the remaining 84 reference points. This means that the linguistic management of the Domaine d'Oïl was very similar in the Middle Ages (through the linguistic activity of the scribes) and in modern times (through the linguistic activity of the dialect speakers). It must be added that, generally speaking, medieval non-Latin charters of the 13th and the 14th centuries (mainly) had a strong dialectal colouring, a phenomenon which was noted not in France only, they showed therefore a great number of local and/or regional written attributes. In the 19th century already, it was assumed that this graph(et)ic variation was generated or at least partly caused by the oral variation of the different medieval dialects. In Northern France, this regional colouring of the charters decreases rapidly after ca. 1400, and vanishes after 1450.

4.2 Comparison plane 2 : two parameter-maps : synopsis of the skewness values

Maps 3 and 4 reveal an entirely different question. The synopsis (or combination) of the N skewness values of a given similarity matrix indicates the degree of variation between different regions in regard to the so-called « linguistic compromise or exchange ». This phenomenon is defined as the degree of the intermixing of geolinguistic attributes with (respectively) regionally varying extension and/or intensity. Our DM-classification distinguishes therefore zones of high linguistic compromise (here : clear shadings) and zones of weak linguistic compromise (here : dark shadings). Where this linguistic exchange is high or great, a strong linguistic intermixing is prevailing. Where it is weak, the linguistic interaction is also low: these areas went on keeping a strong linguistic autonomy and were not yet seized by the general intermixing.

In Map 3 (left), the zones of high linguistic compromise or exchange form a kind of cross: they are located in the centre of the Domaine d'Oïl, whereas on its peripherical boarders the areas of different historical provinces (such as: Normandy, Picardy, Lorraine, etc.) are found. In Map 4 (right), the clear shaded zone occupies now the main part of the grid of the Domaine d'Oïl: in comparison with the left map it has virtually « exploded » (note the black circle), as a consequence of the continuous expansion of the language type of the Ile-de-France, which had been strongly supported by the French kings and after 1789 also by the Republic. Only on the Eastern peripherical boarders, some provinces (Picardy, the Walloons, Lorraine, etc.) could elude the general language compromise and thus the general linguistic intermixing.

Both maps consist of respectively 85 skewness values which were gained by respectively 85 similarity distributions. Since almost 20 years, it is well-known that the skewness value is an excellent instrument for measuring language compromise or exchange; in many instances, evidence of this fact has been given with different data sets (see Goebl 1984, I: 150-153, and 2006a: 419-420).

4.3 Comparison plane 3 : two interpoint or honeycomb maps

Actually, Maps 5 and 6 represent two honeycomb maps, each of them consisting of 225 polygon sides which vary according to thickness and darkness. Every one of these polygon sides lies between (= *inter*) contiguous inquiry points (hence the name *inter*point map), and indicates virtually the relative dialectal differences. Instead of the linguistic *similarities* (*sim*), the potential linguistic *differences* or *distances* (*dist*) were mapped. In quantitative regard, they are interrelated according to the formula: $dist + sim = 100$. Thus, the distance related counterpart of the above mentioned similarity index WIV(1) is the WDV (1) (« Weighted Distance Value (with the weight 1)) ».

The cartographic message of the two maps largely corresponds to the evidence of the traditional isogloss syntheses which were commonly established during the 20th century in Romance, German and English linguistics. The thick (and dark) polygon sides represent the so called « linguistic boundaries», a linguistic term which is rather colloquial and imprecise. One clearly recognizes that in Map 5 (left) in the North (Picardy) and the South-west (Poitou, Saintonge) there are very prominent and distinct « boundaries ». But it also shows very clearly in Map 6 (right) that in the period between 1300 and 1900 these « boundaries » were moved to the North (and East) as well as to the utmost borders of the South by an « invisible force » and that a zone with only very weak interpunctual demarcations emerged in the middle of the Domaine d'Oïl. Our knowledge of the history of the French language allows us to identify this « invisible force »: it is the irradiation of the linguistic type of the Ile-de-France, pushed by the politics.

4.4 Comparison plane 4: two dendrographic analyses (following Ward's method)

Moreover, the two similarity matrices can first be processed by dendrographic methods, in a next step, the two trees are compared. In this procedure, one has to pay attention to those bifurcations of the tree which are located near the trunk (or the root). Among the relevant « hierarchic agglomerative methods » applied for the generation of trees, Ward's method has proved to be most appropriate. In Maps 7 and 8, the tree and the map were drawn and visualised, isolating thereby (respectively) three distinct cartographic clusters. These clusters are called « dendremes » in the tree, and their correspondences on the map « choremes ». The heuristic comparison of Maps 7 and 8 concentrates on the position of the dendremes in the tree and simultaneously on the position of the choremes on the map. First, the perfect spatial coherence of all choremes is striking. Further, it clearly results that the three dendremes (No. 1-3) at the top seize the East, the North and the Centre (including the West) of the Domaine d'Oïl, though in such a way that the central dendreme-choreme (No. 1) expanded in the course of the six centuries between 1300 and 1900 at the expense of the Eastern (No. 3) and the Northern (No. 2) choreme-dendreme. Again, this is a consequence of the irradiation of the dialect of the Ile-de-France, supported by the French royal dynasty and the Republic.

5 Final remarks

By the visual comparison of four pairs of maps established with dialectometrical methods, evidence was given that the geolinguistic deep structures of the Domaine d'Oïl (Northern France) – in the period between 1300 and 1900 – maintained a large stability, that is to say : remained mostly identical in regard to their phonetics. Hence the question arises on determining the chronological development and elaboration *before 1300* of these phonetic deep structures. Nevertheless, the present investigation revealed the actual expansion of the linguistic type of the Ile-de-France between 1300 and 1900 which represents the typological basis for standard French. The dialectometrical techniques, which were again applied in this contribution, have proven many times their great diagnostic value in the last three decades.

References

ALF: Jules Gilliéron and Edmond Edmont. 1902-1910. *Atlas linguistique de la France*, 10 vol., Paris, Champion.

Anthonij Dees. 1980. *Atlas des formes et des constructions des chartes françaises du XIII[e] siècle*, Tübingen, Niemeyer.

Hans Goebl. 1984. *Dialektometrische Studien. Anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*, Tübingen, Niemeyer.

Hans Goebl 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.

Hans Goebl. 2006a. Recent Advances in Salzburg Dialectometry. *Literary and Linguistic Computing* 21 (4): 411-133.

Hans Goebl. 2006b. Sur le changement macrolinguistique survenu entre 1300 et 1900 dans le domaine d'Oïl. Une étude diachronique d'inspiration dialectométrique. *Linguistica* 46: 3-43.

Frequently used abbreviations (also in the legends of the Figures 1-8)

AEM: Average Euclidean Metric: see chapter 2.1.

ALF: Atlas linguistique de la France: see also the References

AMM: Average Manhattan Metric: see chapter 2.1.

DM: Dialectometry

r(BP): Bravais-Pearson correlation coefficent: see chapter 2.1.

VDM: Visual DialectoMetry

WDV(1): Weighted Distance Value (with the weight 1): see chapter 4.3.

WIV(1): Weighted Identity Value (with the weight 1): see chapter 2.2.

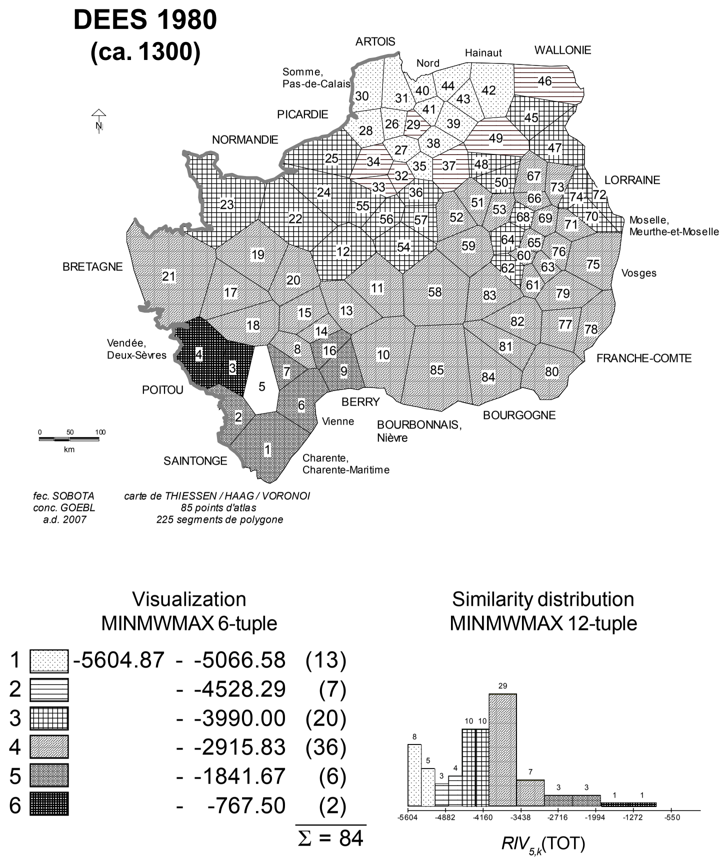

Figure 1: A similarity profile of the medieval Domaine d'Oïl: similarity map to the scripta-region 5 (Deux-Sèvres)
Similarity index: $AEM_{5,k}$
Corpus: 268 quantitative maps (from Dees 1980)
Algorithm of visualization: MINMWMAX (6-tuple)

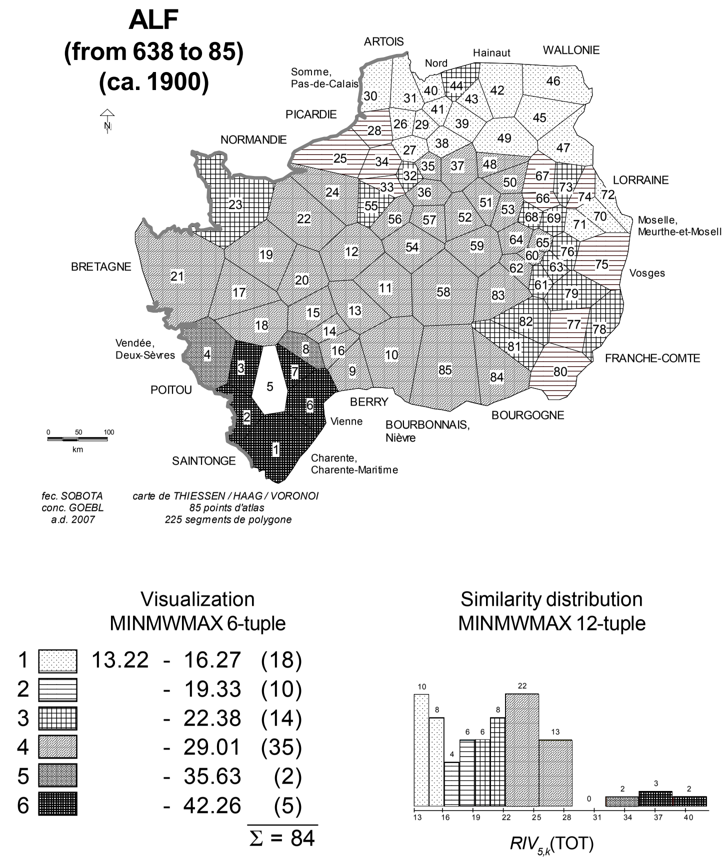

Figure 2: A similarity profile of the modern Domaine d'Oïl: similarity map to the ALF-point 510 (Echiré, Département Deux-Sèvres)
Similarity index: $WIV(1)_{510,k}$
Corpus: 914 phonetic working maps (from ALF)
Algorithm of visualization: MINMWMAX (6-tuple)

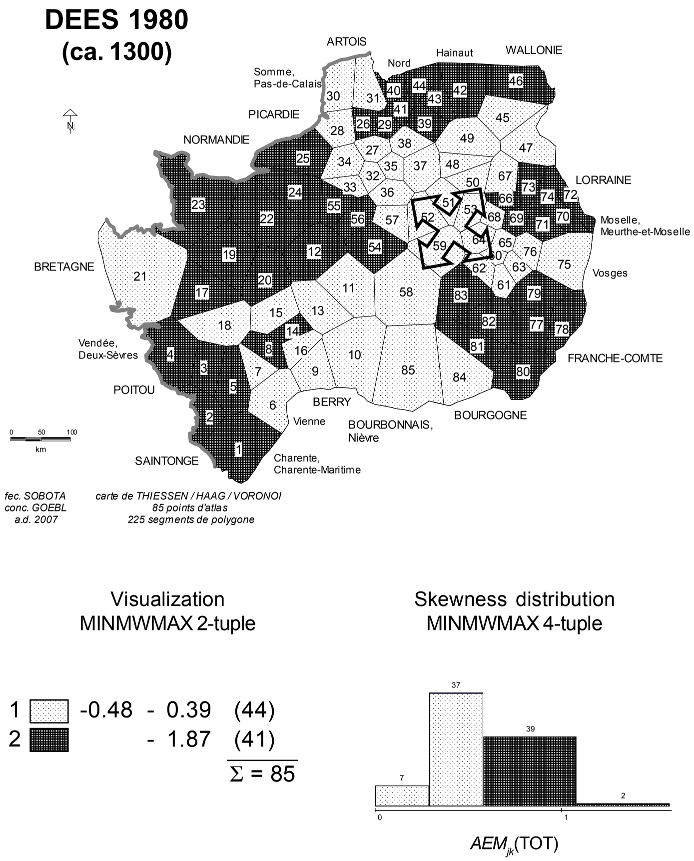

Figure 3: Choropleth map of the medieval Domaine d'Oïl: the synopsis of the skewness values of 85 similarity distributions
Similarity index: AEM_{jk}
Corpus: 268 quantitative maps (from Dees 1980)
Algorithm of visualization: MINMWMAX (2-tuple)

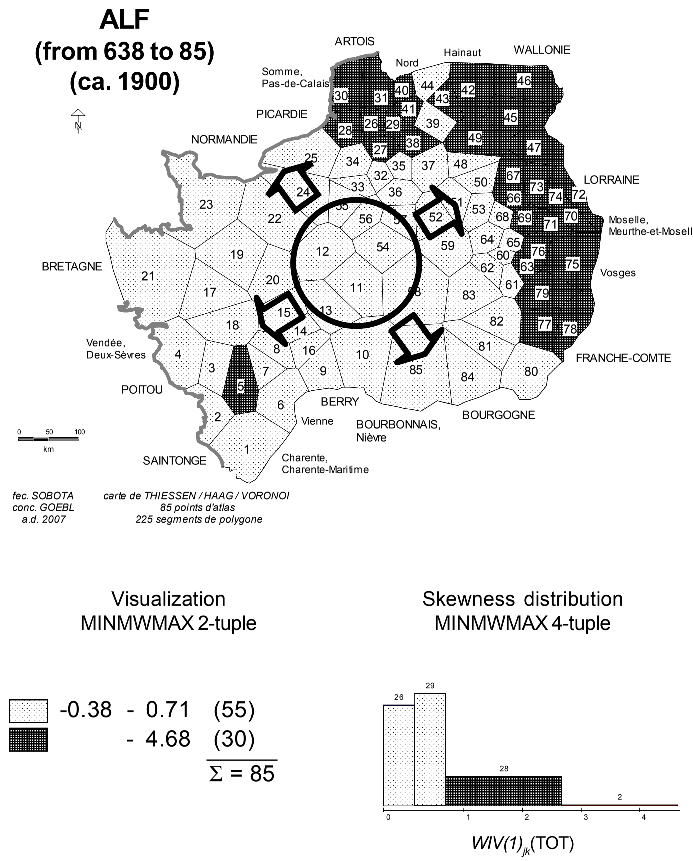

Figure 4: Choropleth map of the modern Domaine d'Oïl: the synopsis of the skewness values of 85 similarity distributions
Similarity index: $WIV(1)_{jk}$
Corpus: 914 phonetic working maps (from ALF)
Algorithm of visualization: MINMWMAX (2-tuple)

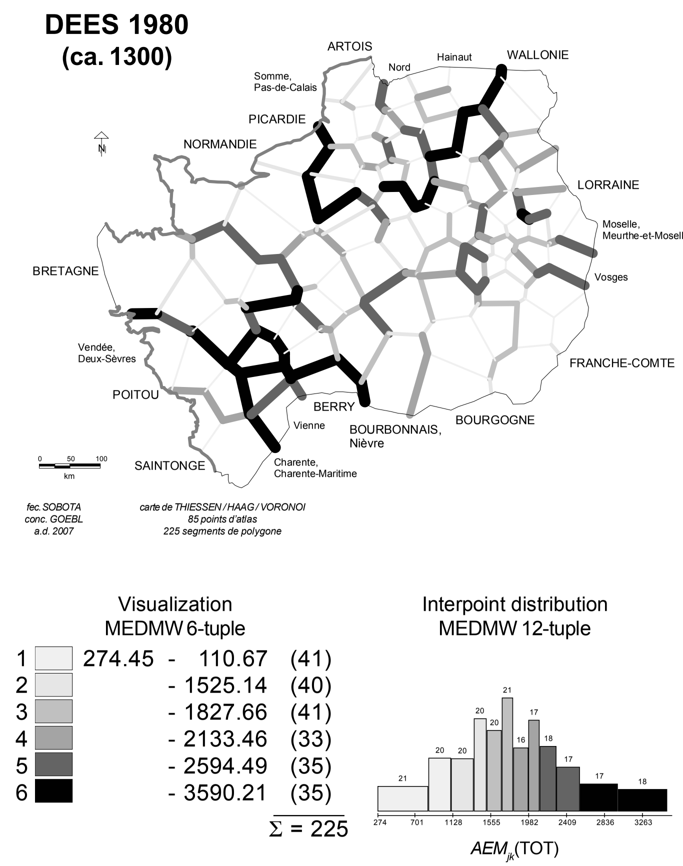

Figure 5: Honeycomb map of the medieval Domaine d'Oïl showing a synopsis of 225 interpoint distance values
Distance index: AEM_{jk}
Corpus: 268 quantitative maps (from Dees 1980)
Algorithm of visualization: MEDMW (6-tuple)

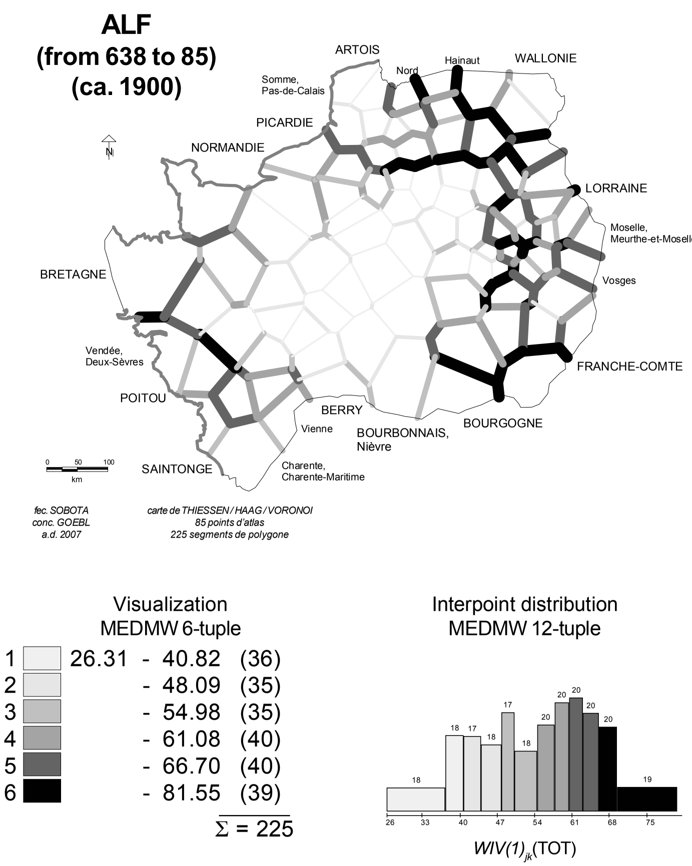

Figure 6: Honeycomb map of the modern Domaine d'Oïl showing a synopsis of 225 interpoint distance values
Distance index: $WDV(1)_{jk}$
Corpus: 914 phonetic working maps (from ALF)
Algorithm of visualization: MEDMW (6-tuple)

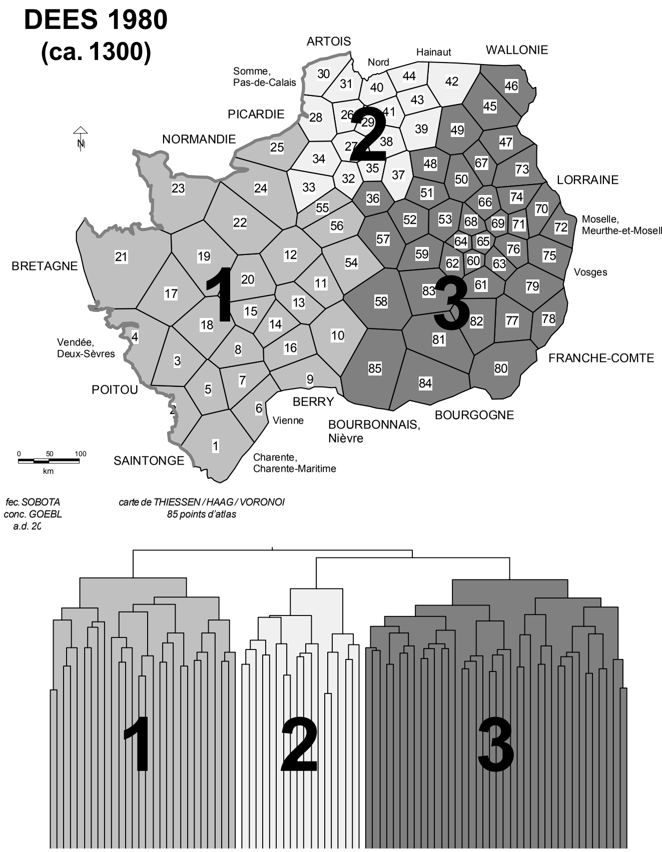

Figure 7: Dendrographic classification (and corresponding spatialization) of the medieval Domaine d'Oïl (85 scripta-regions according to Dees 1980)
Similarity index: AEM_{jk}
Dendrographic algorithm: hierarchical grouping method of Ward
Number of marked dendremes resp. choremes: 3

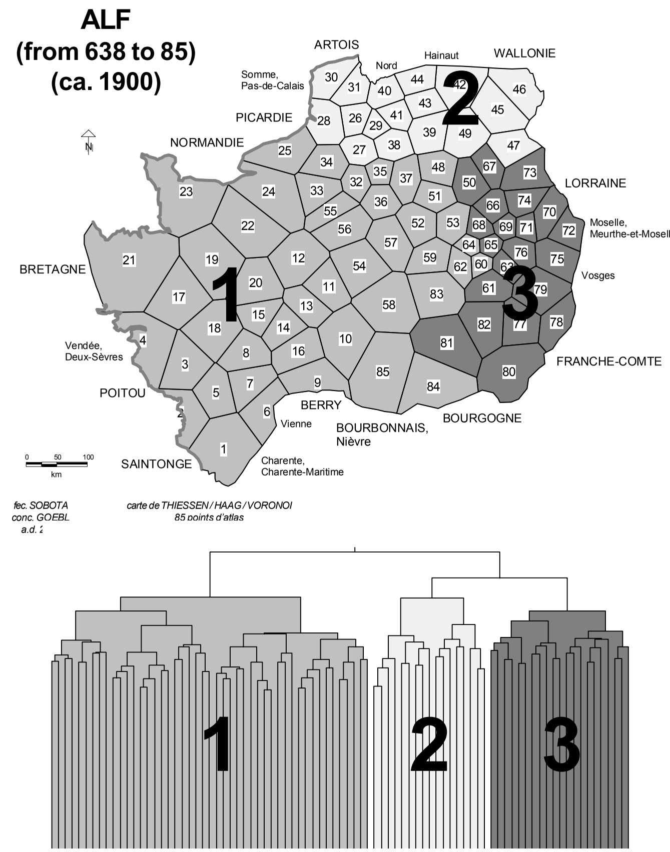

Figure 8: Dendrographic classification (and corresponding spatialization) of the modern Domaine d'Oïl (85 ALF-points)
Similarity index: $WIV(1)_{jk}$
Dendrographic algorithm: hierarchical grouping method of Ward
Number of marked dendremes resp. choremes: 3