

Coup d'œil dialectométrique sur les Tableaux phonétiques des patois suisses romands (TPPSR)

Autor(en): **Goebl, Hans**

Objektyp: **Article**

Zeitschrift: **Vox Romanica**

Band (Jahr): **44 (1985)**

PDF erstellt am: **10.07.2024**

Persistenter Link: <https://doi.org/10.5169/seals-34619>

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

Coup d'œil dialectométrique sur les *Tableaux phonétiques des patois suisses romands (TPPSR)*¹

0. But de l'article

Les *Tableaux phonétiques des patois suisses romands (TPPSR)* constituent, comme chacun sait, un atlas linguistique *avant la lettre* étant donné le fait que les matériaux y sont présentés non pas sous la forme de *cartes géographiques* à proprement parler mais sous la forme de *colonnes verticales*. Mis à part ce détail d'ordre plutôt typographique, les *TPPSR* disposent de tous les atouts d'un atlas linguistique conçu dans la tradition gillieronienne. Les relevés des *TPPSR* ont été faits simultanément, entre 1904 et 1907, par deux enquêteurs de l'équipe du *Glossaire des patois de la Suisse romande (GPSR)*². Les notations de l'enquêteur permanent, J. Jeanjaquet, ont servi plus tard à l'établissement du texte imprimé des *TPPSR*. Les notations divergentes des autres enquêteurs sont indiquées, dans l'ouvrage imprimé, en bas de page. Sans aucun doute, les *TPPSR* représentent, mis à part le fait qu'ils ont été élaborés en vue de remplacer l'*Atlas linguistique de la Suisse romande* jamais publié³, un précieux outil géolinguistique dont, hélas, il a été rarement fait usage pour des analyses dialectales plus poussées.

Comme, d'une part, nous voulions aussi expérimenter la validité de l'analyse dialectométrique à l'aide de données *phonétiques* (et ceci dans la perspective de dialectométrer un jour les données de l'atlas ibérique *ALPI*), et que, de l'autre, l'agencement typographique des *TPPSR* offre certains avantages pratiques pour le codage manuel des données brutes, nous avons décidé de soumettre les colonnes 1-249 des *TPPSR* à une analyse dialectométrique selon les méthodes mises au point par nous. L'échantillon retenu constitue la moitié des matériaux des *TPPSR* qui, sur 62 «patois-types», totalisent 480 «colonnes» ou – en d'autres termes – cartes d'atlas.

¹ Les travaux présentés par la suite font partie d'un programme de recherche subventionné par les organismes suivants:

- Fonds zur Förderung der wissenschaftlichen Forschung in Österreich - FWF (Vienne).
- Deutsche Forschungsgemeinschaft - DFG (Bonn).

Toilette stylistique: F. GENTON (Ratisbonne) et PH. REGÉRAT (Salzbourg).

J'adresse ici ma profonde gratitude aux MM. GENTON, REGÉRAT, SELBERHERR, RASE et PUDLATZ – les trois derniers ayant droit au titre de co-auteurs – ainsi qu'aux organismes de recherche (FWF et DFG) mentionnés ci-dessus.

Les cartes choroplèthes (Fig. 9-18) ont été réalisées par W.-D. Rase à l'aide d'un traceur à jet d'encre (angl. ink-jet-plotter).

² Cf. *TPPSR* V-VII.

³ Pour l'histoire de cet atlas cf. GAUCHAT/JEANJAQUET 1920, 17 (nos. 1099 et 1100).

0.1. Conditions méthodiques et méthodologiques de l'analyse dialectométrique

L'utilisation de la dialectométrie présuppose – à l'instar de toute analyse mathématique en sciences humaines – la maîtrise de deux volets épistémologiques bien distincts: celui de la *méthode* et celui de la *méthodologie*. En matière de méthode, il s'agit du manie-ment adéquat d'un instrument mathématique plus ou moins complexe alors qu'en matière de méthodologie il s'agit de trouver une réponse appropriée à la question visant le COMMENT et le POURQUOI de l'approche métrisante choisie. Quant à la dialectométrie – issue de l'interaction de la *géolinguistique traditionnelle* et de la *classification automatique* (analyse des données, analyse typologique, taxonomie numérique, taxométrie etc.⁴) – il importe de ne jamais oublier que son application se fait toujours en fonction d'un objectif scientifique précis et que – par conséquent – le *choix* des procédés taxométriques utilisés doit toujours se faire en fonction de *l'objectif préalablement fixé*. La maîtrise des procédés taxométriques appliqués relève du domaine de la *méthode*, alors que la fixation de l'objectif scientifique et le problème très épineux du choix des méthodes à utiliser et de leur adaptation au contexte de recherche donné, relèvent du domaine de la *méthodologie*.

Comme, tant en méthode qu'en méthodologie, le nombre des approches possibles est théoriquement illimité, les résultats dialectométriques présentés par la suite ne constituent qu'une fraction infime de ce que l'on pourrait tirer des données-TPPSR par l'application de *l'analyse numérique des données*.

Nous ne répétons que pour mémoire les jalons les plus importants de toute analyse dialectométrique⁵:

- a) choix des données empiriques à classifier.
- b) choix du principe métrologique (c.-à-d. de mesurage) permettant de mesurer les données empiriques à classifier.
- c) choix de l'indice de similarité (ou de distance).
- d) choix des procédés statistiques pour exploiter la matrice de similarité soit partiellement soit en entier.
- e) choix de la méthode de visualisation des résultats numériques de l'analyse dialectométrique.

1. Le codage (la taxation) des données originales: des TPPSR à la matrice des données

La matrice des données constitue le point de départ de toute analyse taxométrique. C'est un tableau à double entrée qui groupe N *éléments* (objets, patois-types, parlers, points d'atlas etc.) et p *attributs* (caractéristiques, traits linguistiques etc.). Voir la Fig. 1.

⁴ Manuels de taxométrie: SNEATH/SOKAL 1973, BOCK 1974, SODEUR 1974 et CHANDON/PINSON 1981.

⁵ Cf. à cela Chandon/Pinson 1981, 27-31.

**Matrice des données
(échelle nominale)**

p	5	j	j	j	j	j	j
i	4	h	i	h	X	i	h
i	3	e	f	e	e	f	g
i	2	c	c	X	c	d	d
1	1	a	a	b	a	a	b
		1	2	3	4	5	6

Objets (points d'atlas)
[1 . . . j . k . . . N]
X case vide

Fig. 1: Matrice des données constituée de 6 objets (points d'atlas) et de 5 attributs nominaux (cartes d'atlas voire «de travail»).
En gris: l'équivalent matriciel d'une carte d'atlas (voire «de travail»).

La matrice des données n'est rien d'autre – en dernière analyse – qu'une «photographie» – prise sous un angle bien défini – des données originales. Or, comme chacun sait, toute image constitue une réduction, voire une diminution de la réalité originale reproduite. Cet effet réductionniste est d'ailleurs caractéristique de toute analyse classificatoire. Il a l'avantage cependant de compenser la perte d'informations atomistes sur le plan du *particulier*, par un gain d'informations typologiques sur le plan du *général*. Tout classificateur se sert donc d'un *métabolisme transformationnel* placé entre la considération d'un *grand nombre de détails phénoménologiques* et la *reconnaissance de structures typologiques majeures*, et ceci toujours en vue de reconnaître – pour le dire avec des mots très simples – la «forêt» derrière une multiplicité d'arbres.

Les données brutes à classifier ont été tirées des *TPPSR*. En voici les coordonnées essentielles:

- a) données dialectales brutes (taxandum): *TPPSR* colonnes 1-249.
- b) objets (colonnes-*TPPSR*, points d'atlas, parlers): 62 points-*TPPSR* + 1 point artificiel (P. 99).
- c) P. 99: équivalent du français standard tel qu'il figure dans les en-têtes des colonnes des *TPPSR*.
- d) critère discriminatoire pour l'isolement des unités taxatoires (taxats): différences phonétiques *majeures* (par voie de *macro-taxation*) et *mineures* (par voie de *micro-taxation*) (cf. ci-dessous).
- e) attributs (cartes «analysées» ou cartes «de travail» obtenues par la taxation des données des *TPPSR*): 1069.
- f) nombre des unités taxatoires (taxats) consignées dans les lignes de la matrice des données: 4567 taxats-tokens nominaux⁶.

⁶ Le chiffre de 4567 indique le nombre de toutes les occurrences (= tokens) des 90 taxats (= types) répertoriés sur la Fig. 2.

- g) nombre des unités taxatoires (taxats) utilisées pour l'analyse taxatoire des données-*TPPSR*: 90 taxats-*types* nominaux (voir aussi la Fig. 2).
- h) nombre de taxats (-tokens)/carte «de travail»: 4,27.
- i) nombre des cartes «de travail» d'ordre vocalique⁷: 586.
- j) nombre des cartes «de travail» d'ordre consonantique⁸: 415.
- k) nombre des cartes «de travail» d'ordre tant vocalique que consonantique («nexus phonétiques mixtes»)⁹: 71.

Des 480 colonnes des *TPPSR* nous n'avons utilisé que la moitié. Ce choix est purement arbitraire¹⁰. Du reste, nous avons écarté toutes les colonnes comportant des cases vides, si bien que le corpus des données codé est complètement exempt de cases vides.

Quant au principe de codage (de taxation) nous avons fait nôtre l'hypothèse suivante: il doit être possible d'utiliser la fragmentation spatiale de la Suisse romande telle qu'elle ressort de l'analyse comparative des caractères du système de transcription des *TPPSR*, pour constituer une «photographie» métrologiquement valable des données brutes-*TPPSR*. L'analyse comparative mentionnée ci-dessus a été faite à deux niveaux:

- a) au niveau de la *macro*-taxation (attributs phonétiques majeurs ou additionnels comme p. ex. la diphtongaison, l'oralité, la nasalité etc.).
- b) au niveau de la *micro*-taxation (attributs phonétiques dont les modalités correspondent exactement aux différents caractères de la «transcription des sons» présentée à la page 1 des *TPPSR*).

Pour une liste exhaustive des unités taxatoires utilisées pour la macro-taxation et la micro-taxation voir la Fig. 2.

Il va de soi que quelques-unes des 90 unités taxatoires (taxats-*types*) n'ont été utilisées que très rarement, p. ex. les taxats no. 2, 8, 12, 13, 19, 21, 28 etc.

Le taxat no. 1, Zéro, mérite une mention particulière. C'est un taxat «passe-partout» qui doit être utilisé là où l'analyse taxatoire demande l'emploi d'un attribut nominal *non marqué* par rapport à des attributs *marqués*. Voir, dans la Fig. 3, la ligne 0411 de la matrice des données-*TPPSR*, où les continueurs valaisans de G latin (premier élément du nexus latin initial GL-) sont taxés respectivement par les taxats [g] et θ «Zéro»¹¹. La même remarque vaut pour la ligne 0416 où il s'agit de coder les succédanés de -A latin final. Les succédanés valaisans en [è] engendrent une identité nominale entre les points-*TPPSR* 18, 21, 22 et 24 alors que les points-*TPPSR* 19, 20 et 23 n'ont rien d'autre en

⁷ P. ex. succédanés de A dans lat. FACIT (*TPPSR* 1).

⁸ P. ex. succédanés de S dans lat. SATIONE (*TPPSR* 9).

⁹ P. ex. succédanés de -ARIU dans lat. FEBRARIU (*TPPSR* 38).

¹⁰ Fort de nos expériences antérieures, nous pouvons dire que la validité du corpus ainsi constitué est très satisfaisante tant par la quantité que par la qualité des attributs relevés.

¹¹ L'assignation respective des succédanés romands aux différents éléments d'un étymon latin complexe obéit à une convention préalablement établie. On aurait pu tout aussi bien coder les succédanés de lat. GL- en [z, y] ou [d] dans la ligne 0411 (voir la Fig. 3) quitte d'ailleurs à réserver la ligne 0412 à l'utilisation du taxat θ (Zéro).

NIX	0	
ZERO	1	
ANAPTYXE	2	
MONOPHTHONG	3	
DIPHTHONG	4	
TRIPHTHONG	5	
KURZ	6	
MITTEL	7	
LANG	8	
ORAL	9	
HALBNASAL	10	
VOLLNASAL	11	
KLEIN	12	
GROSZ	13	
a	14	
ā	15	
ã	16	
á	17	
à	18	
ä	19	
b	20	
c	21	
ç	22	
e	23	
ë	24	
ê	25	
d	26	
đ	27	
ḏ	28	
e	29	
é	30	
é	31	
è	32	
è	33	
ë	34	
f	35	
g	36	
h	37	
ì	38	
ì	39	
î	40	
j	41	
ĵ	42	
k	43	
k ^t	44	
k	45	
k ^t	46	
l	47	
l	48	
l ^t	49	
l	50	
t	51	
m	52	
n	53	
ñ	54	
ñ	55	
ñ	56	
o	57	
o	58	
ó	59	
ò	60	
ò	61	
œ	62	
œ	63	
œ	64	
œ	65	
p	66	
r	67	
r	68	
ř	69	
s	70	
š	71	
t	72	
t	73	
t	74	
u	75	
u	76	
ü	77	
ü	78	
u	79	
u	80	
ù	81	
ú	82	
v	83	
w	84	
w	85	
w	86	
y	87	
ÿ	88	
z	89	
z	90	

Fig. 2: Liste des 90 taxats nominaux utilisés dans notre analyse dialectométrique des *TPPSR*.
 Taxat 0: case vide.
 Taxats 1-13: unités nominales relevant de la macro-taxation.
 Taxats 14-90: unités nominales relevant de la micro-taxation (correspondant aux caractères de la transcription phonétique des *TPPSR*).

commun que de ne pas avoir conservé de continuateurs de -A latin final: d'où le codage de leur identité nominale par le taxat θ «Zéro». Pour l'utilisation du concept de «Zéro» en linguistique cf. Schifko 1973.

Quant au procédé de codage appliqué par nous, il importe de souligner le principe suivant: comme le but de tout codage est d'obtenir une représentation tant soit peu homomorphe des données originales, on doit coder (et ceci veut dire: représenter), par l'attribution de taxats identiques, des *identités taxatoires* (entre deux ou plusieurs attributs d'objet) là où les données originales (les *TPPSR* en l'occurrence) offrent, par l'intermédiaire de caractères identiques dans la notation phonétique utilisée, des *identités phonétiques*. Ceci signifie que l'on néglige, au cours de l'analyse taxatoire, le degré et la qualité des *différences* phonétiques entre les attributs d'objet pour ne s'en tenir qu'aux *similarités* (= identités, ressemblances etc.). Il est bien évident que la considération exclusive de la similarité phonétique au cours du codage ne permet – quant au calcul de la matrice de similarité à partir de la matrice des données – que l'utilisation d'un indice de *similarité* à l'exclusion de tous les indices de *distance* ou de dissimilarité. *Similarité* et *distance linguistique* se révèlent, une fois de plus, comme étant deux choses bien distinctes.

Par l'application consécutive de la macro-taxation et de la micro-taxation aux données d'une seule colonne des *TPPSR*, il est possible d'en tirer plusieurs cartes «de travail». Voilà comment nous avons pu tirer plus de mille cartes «de travail» de quelques 250 colonnes-*TPPSR*.

Comme l'ensemble des unités taxatoires disponibles (taxats-types) est limité et ne dépasse jamais le nombre de 90, on est en droit de qualifier ce genre de taxation de «taxation fermée». Une taxation «ouverte» en revanche, opère d'entrée de jeu avec une liste de taxats non limitative. Une taxation ouverte a été pratiquée dans nos analyses dialectométriques de l'*ALF* et de l'*AIS*, faites d'ailleurs sur la base de critères discriminatoires *lexicaux* et *morphosyntaxiques*: cf. nos contributions entre 1976 et 1984. Pour un exemple de taxation mixte (macro-taxation et micro-taxation) voir la Fig. 3.

Il est bon de comparer la Fig. 3 avec les données originales des *TPPSR* (colonne 41, nos. 17-24). Retenons pour mémoire que les données brutes susceptibles d'être soumises à l'analyse taxatoire, sont appelées «taxandum», et que c'est de la décomposition, sur échelle nominale¹², du taxandum que résultent les unités de la macro-taxation ainsi que celles de la micro-taxation (polynymie taxatoire à proprement parler).

Repassons à la Fig. 1. L'attribut 1 (ou la carte d'atlas 1) dispose de deux taxats (a et b) appelés aussi *modalités* de l'attribut, alors que les attributs 3 et 5 disposent respectivement de 3 taxats (e, f et g) ou d'un seul taxat (j). Le morcellement taxatoire d'une matrice des données sur échelle nominale (non binaire) est donc fort variable. Théoriquement il peut varier entre 1 et N (nombre total des objets ou points d'atlas consi-

¹² Par *échelle nominale*, on entend, en statistique et en taxométrie, une échelle de mesure qui ne retient – à l'intérieur d'une classe d'objets – que l'*équivalence qualitative des objets*, c.-à-d. sans en retenir d'éventuelles modalités quantitatives: cf. CHANDON/PINSON 1981, 33-35.

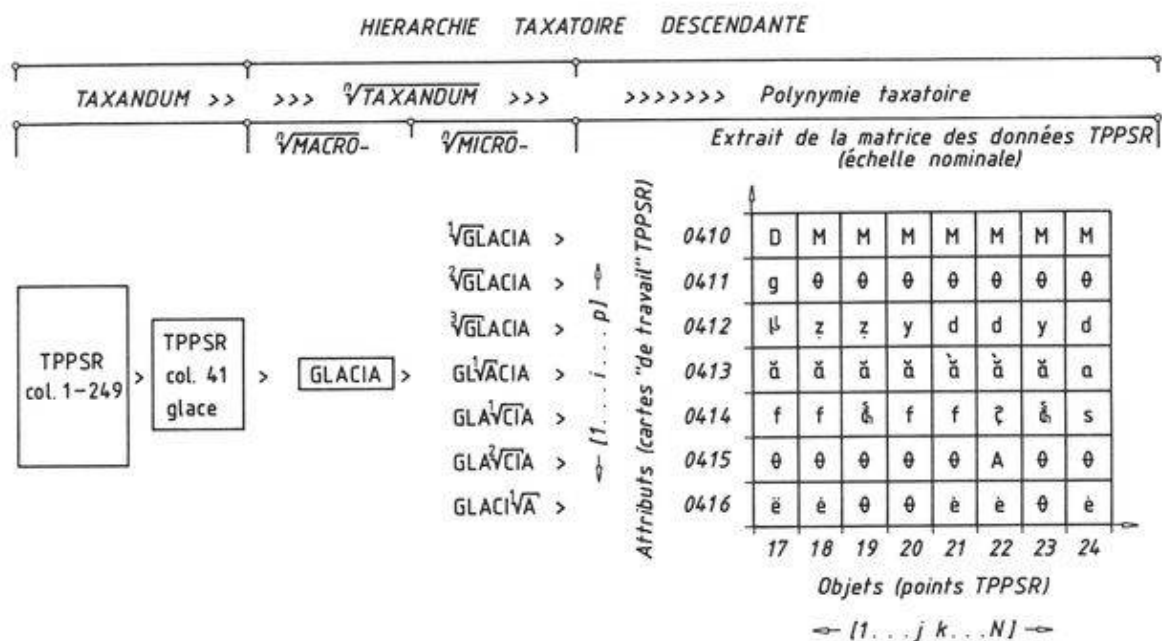


Fig. 3: Schéma de taxation à l'exemple de la colonne 41 des TPPSR «glace» et des points-TPPSR 17-24.

De la colonne 41 «glace» (carte d'atlas) nous avons tiré, par voie taxatoire, 7 attributs nominaux (cartes «de travail»).

D diphtongue (consonantique).

A anaptyxe (consonantique).

M monophthongue (consonantique).

θ taxat «Zéro».

Phon-Suisse N-TYP	Liste III-TOT N-KART	%-1069
1	109	10.20
2	338	31.62
3	185	17.31
4	77	7.20
5	59	5.52
6	64	5.99
7	47	4.40
8	61	5.71
9	36	3.37
10	28	2.62
11	27	2.53
12	16	1.50
13	5	.47
14	4	.37
15	5	.47
16	5	.47
17	1	.09
18	1	.09
19	1	.09

Fig. 4: Nombre de taxats par carte «de travail».

La première ligne de la liste est à interpréter comme suit:

«De 1069 cartes «de travail» (en-tête de la colonne %-1069), 109 (colonne N-KART, en haut), ou 10,20% (colonne %-1069, en haut), ne disposent que d'un seul taxat (colonne N-TYP, en haut) et sont donc mononymes.»

dérés). Pour un examen analogue du morcellement taxatoire de la matrice des données tirée des *TPPSR* et utilisée dans nos travaux dialectométriques relatifs à la Suisse romande, voir la Fig. 4.

La ressemblance numérique des valeurs (voire de leur *décalage* respectif) de la Fig. 4 avec celles des Fig. 2.3 et 2.26 dans Goebel 1984 I (41 et 59) est évidente. Il semble bien que ces régularités numériques soient l'expression d'une loi mathématique supérieure dont il ne peut être question ici.

2. De la matrice des données à la matrice de similarité: la mesure de la similarité

Pour pouvoir mesurer la similarité (ressemblance, identité, proximité etc.) entre deux vecteurs d'objet, il faut que le dialectométricien ait arrêté d'avance ses positions autour du concept théorique de «similarité géolinguistique mesurée par des moyens dialectométriques». L'élaboration de tels concepts théoriques fait partie du travail *methodologique* mentionné au paragraphe 0.1, travail dont nul dialectométricien ne peut se passer. A ce propos, nous renvoyons au passage suivant, qui malgré son ancienneté, présente assez clairement une excellente conception théorique de la ressemblance (similarité) et de la dissemblance (distance) de deux parlers locaux: «Et maintenant, qu'est-ce qui constitue le degré de ressemblance qui rapproche deux langues entre elles, et le degré de dissemblance qui les éloigne l'une de l'autre? La ressemblance se mesure à la proportion des caractères communs, la dissemblance à la proportion des caractères particuliers.» (Durand 1889, 63).

Il importe avant tout de ne pas confondre la notion de *similarité* (ressemblance, proximité, identité etc.) avec celle de *distance* (différence, divergence etc.). Pour la constitution (tant diachronique que synchronique) de l'espace géolinguistique tel qu'il se présente à nos yeux par le jeu capricieux des aires dialectales voire des isoglosses les entourant, l'élément constitutif est sans nul doute la similarité linguistique entre deux parlers plus ou moins éloignés ou - en d'autres termes - la co-présence de telle modalité d'attribut linguistique en deux ou plusieurs points d'un réseau d'atlas. De telles réflexions nous amènent à utiliser l'*Indice Relatif d'Identité* (IRI_{jk}) pour l'analyse taxométrique de l'espace dialectal. En voici la formule:

$$IRI_{jk} = 100 \cdot \frac{\sum_{i=1}^{\hat{p}} (COI_{jk})_i}{\sum_{i=1}^{\hat{p}} (COI_{jk})_i + \sum_{i=1}^{\hat{p}} (COD_{jk})_i} \quad (1).$$

Dans cette formule,
 IRI_{jk} est l'Indice Relatif d'Identité entre les vecteurs d'objet (parlers) j et k .
 j est l'indicatif du point de référence.
 k est l'indicatif du point comparé.
 \bar{p} est le nombre des attributs présents et dans le vecteur d'objet j et dans le vecteur d'objet k .
 i est l'indicatif d'un attribut.
 $(COI_{jk})_i$ est une co-identité entre les points (objets, parlers) j et k à l'emplacement de l'attribut i .
 $(COD_{jk})_i$ est une co-différence entre les points (objets, parlers) j et k à l'emplacement de l'attribut i .

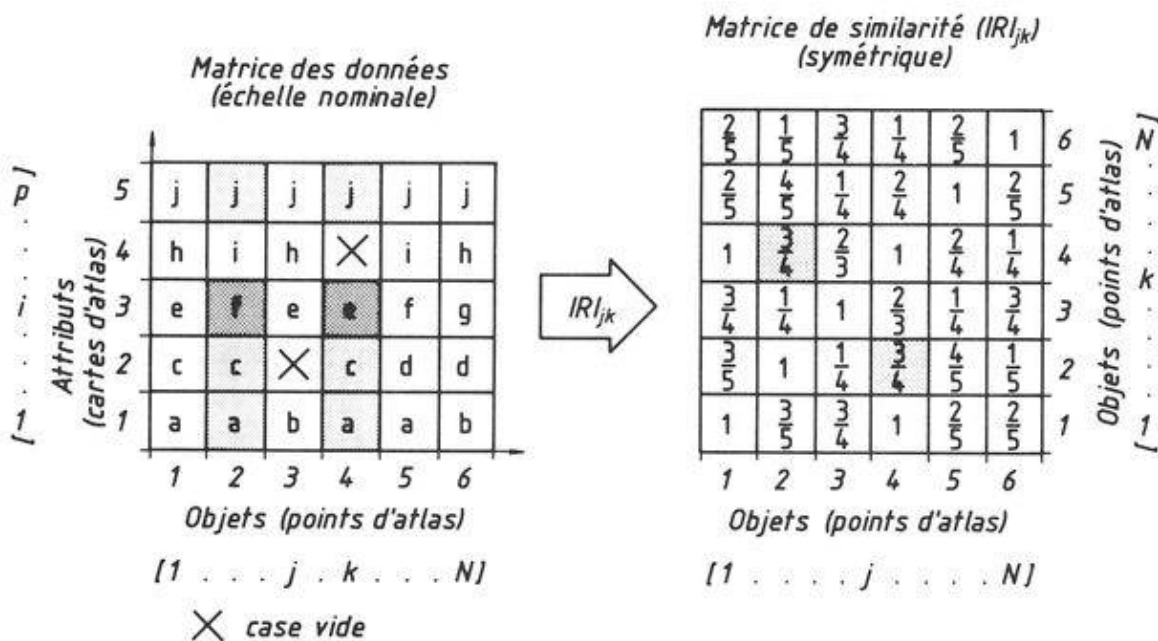


Fig. 5: Génération de la matrice de similarité à partir de la matrice des données par l'application de l'Indice Relatif d'Identité (IRI_{jk}).

A gauche (matrice des données):
 grisé clair: co-identités (COI) pour $i = 1$ (a, a), pour $i = 2$ (c, c) et pour $i = 5$ (j, j).
 grisé sombre: co-différence (COD) pour $i = 3$ (f, e).
A droite (matrice de similarité):
 en gris: $IRI_{2,4} = IRI_{4,2} = \frac{3}{4}$ (= 0,75).
 (Valeurs sans mise en pourcentage).

Vérification de l' IRI_{jk} pour $j = 2$ et $k = 4$ à l'aide de la Fig. 5:

- Remarquer que, dans la matrice des données, la présence d'une case vide dans $i = 4$ et $k = 4$ cause la diminution de \bar{p} de 5 à 4.
- Nombre des co-identités entre les points 2 (j) et 4 (k):
 $\sum_{i=1}^4 (COI_{2,4})_i = 3.$

- Nombre des co-différences entre les points 2 (*j*) et 4 (*k*):

$$\sum_{i=1}^4 (\text{COD}_{2,4})_i = 1.$$

- Calcul de l'IRI_{2,4} selon la formule (1):

$$\text{IRI}_{2,4} = 100 \cdot \frac{3}{3 + 1} = \frac{3}{4}.$$

La logique de la comparaison de tous les vecteurs d'objet par paires moyennant l'application d'un indice de similarité présuppose, pour *N* objets (parlers), *N*² comparaisons. Or, ce sont précisément les dimensions de la matrice de similarité (voir la moitié droite de la Fig. 5). Comme tous les scores calculés sont symétriques ($\text{IRI}_{jk} = \text{IRI}_{kj}$) et que les *N* valeurs répertoriées le long de la diagonale de la matrice de similarité équivalent toutes à 1 (ou à 100%) (IRI_{ij}), l'on obtient la formule suivante pour calculer le nombre des scores de similarité utilisables à des classifications typologiques ultérieures:

$$\frac{N^2 - N}{2} = \frac{N}{2} (N - 1) \quad (2).$$

N est le nombre des objets (points d'atlas, parlers) de la matrice des données.

La définition de l'indice de similarité et le calcul de la matrice de similarité relèvent du domaine de la méthode. L'exploitation taxométrique de la matrice de similarité, par contre, relève du domaine de la méthodologie. Car parmi les innombrables procédés de l'analyse multivariée pouvant être utilisés à partir de la matrice de similarité, il faut opérer bon gré mal gré un choix judicieux. Selon nous, ce choix doit être fait en fonction de la charpente théorique de la géolinguistique traditionnelle, quitte d'ailleurs à la faire *progresser en douceur*. Il nous paraît d'ailleurs complètement contre-indiqué de faire éclater l'horizon théorique de la dialectologie traditionnelle par l'utilisation inconsidérée de tout ce qu'il y a «de meilleur» (c.-à-d. de plus sophistiqué) dans la panoplie de l'analyse des données.

3. Notices servant à l'intelligence des cartes choroplèthes et des cartes numériques

3.1. Notice servant à l'intelligence des cartes choroplèthes

Voir les Fig. 9-18 (annexe).

Le fond des cartes choroplèthes a été soumis à un traitement cartographique et partant géométrique particulier. Les 62 points d'enquête du réseau-*TPPSR*¹³ ont été entourés d'un réseau polygonal afin de substituer aux implantations ponctuelles des données

¹³ Le point artificiel 99 (équivalent au français standard) a été localisé en dehors du réseau-*TPPSR* à proprement parler.

originales (représentation discrète) un agencement disjonctif de surfaces de polygones (représentation quasi-continue ou pavage). Ces polygones sont appelés «polygones de Thiessen (ou de Dirichlet ou de Voronoï)». La construction en est fort simple et à la portée de collégiens de 12 à 13 ans: voir la Fig. 6.

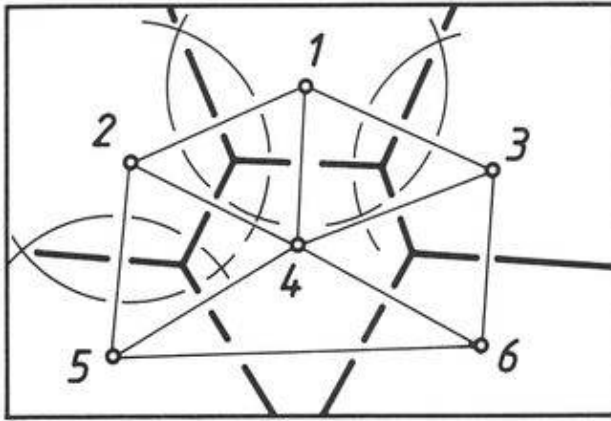


Fig. 6: Construction de polygones de Thiessen (de Dirichlet, de Voronoï) dans un réseau de 6 points d'atlas.

1. Triangulation des points d'atlas du réseau examiné. Les côtés des triangles inscrits entre les points du réseau examiné doivent être aussi courts que possible.
2. Les côtés des triangles (segments triangulaires) reliant les points d'atlas entre eux sont pourvus de leurs perpendiculaires (médiatrices de segments).
3. Fusion des perpendiculaires dans les points de concours. Ces points de concours constituent les angles des polygones de Thiessen. Les angles des polygones de Thiessen sont les centres de cercles circonscrits aux différents triangles inscrits dans le réseau d'atlas soumis à la polygonation.
4. Déplacement, après l'établissement du pavage polygonal, des numéros d'ordre vers le centre des polygones respectifs pour éviter, le cas échéant, des enchevêtrements entre le tracé du numéro et celui des segments de polygone.

Au sujet des polygones de Thiessen (de Dirichlet ou de Voronoï) cf. Goebel 1984 I, 90-92; 1983a, 358-359 et 1981, 363-364 ainsi que Thiessen 1911 et Haggett 1973, 277.

Les polygones de Thiessen serviront plus tard de supports cartographiques pour l'étalement des figurés hachurés en plages. Le pavage de Thiessen a été fait d'après la carte originale reproduite à la page IV des *TPPSR* sur une table à digitaliser. L'emplacement (quelque peu arbitraire) du point 99 (français standard) s'explique par le fait que le foyer d'irradiation du français standard, c.-à-d. de Paris, se trouve au nord-ouest de la Suisse romande.

Pour la construction des histogrammes (en bas de page, à droite) cf. 3.2.

3.2. Notice servant à l'intelligence des cartes numériques

Voir les Fig. 19-28 (annexe).

3.2.1. Les en-têtes des cartes numériques

Voir les Fig. 19-28 (annexe).

Chaque carte numérique dispose, en haut de page à gauche, d'un en-tête comprenant trois lignes de longueur inégale d'abord, et de deux blocs de trois lignes chacun ensuite.

Explications relatives aux trois lignes de longueur inégale (voir la Fig. 19):

PUNKT: 62 point de référence 62 (P. 62).
 MW moyenne arithmétique (all. *Mittelwert*).
 S écart type (all. *Standardabweichung*).
 G coefficient d'asymétrie (de Fisher) (all. *Schiefe*).
 N-KART. nombre des attributs disponibles dans le vecteur du point de référence (ici: P. 62).

Le bloc supérieur se compose de trois lignes indexées respectivement MINMWMAX, MEDMW et MED. Ce sont les noms de trois algorithmes d'intervallisation. Algorithme d'intervallisation? Qu'est-ce que c'est? Revenons à la Fig. 5. La matrice de similarité représentée est un schéma de $N \times N$ cases. Or, N est le nombre des points de l'atlas considéré. Comme un des buts de la dialectométrie telle que nous la pratiquons depuis deux lustres, est de faire parler, d'une façon quantitative, le réseau de l'atlas linguistique examiné, il est nécessaire de trouver un algorithme adéquat pour

Algorithme d'intervallisation MINMWMAX (à 6 intervalles)				
1	2		3	4
Intervalle	Variable: $IRI_{62,k}$		largeur d'intervalle	nombre de points d'atlas par intervalle
	de	à		
1	49,111 ^{a)}	52,117	3,0056	4 + 4 = 8
2	> 52,117	55,122	3,0056	6 + 14 = 20
3	> 55,122	58,128 ^{b)}	3,0056	9 + 5 = 14
4	> 58,128	67,533	9,4046	9 + 4 = 13
5	> 67,533	76,938	9,4046	3 + 1 = 4
6	> 76,938	86,342 ^{c)}	9,4046	1 + 2 = 3
62				
a) minimum (valeur discrète)				
b) moyenne arithmétique (valeur continue)				
c) maximum (valeur discrète)				

Fig. 7: Calcul des intervalles de visualisation à l'aide de l'algorithme d'intervallisation MINMWMAX (à 6 intervalles).

Pour les données numériques voir l'en-tête de la Fig. 19.

visualiser N (ou $N - 1$) valeurs tirées d'une façon ou d'une autre de la matrice de similarité. Un des algorithmes les plus appropriés parmi ceux que nous ayons expérimentés au cours de nos recherches dialectométriques, est l'algorithme MINMWMAX. Son fonctionnement peut être aisément démontré par la considération simultanée des Fig. 7 et 19.

D'un point de vue cartographique et partant iconique, il s'agit de représenter la variation de N (ou de $N - 1$) scores numériques par une variation analogue des contrastes de 6-8 teintes hachurées (ou couleurs) implantées zonalement, c.-à-d. étalées sur les polygones de Thiessen. Pour ce faire, l'algorithme MINMWMAX se sert du *minimum*, de la *moyenne arithmétique* et du *maximum* de la distribution de fréquence à visualiser. L'écart entre la moyenne et le minimum est divisé par 3. C'est ainsi que l'on calcule aisément tant la *largeur* que les *seuils* des intervalles 1-3. Même processus pour le maximum et la moyenne arithmétique. La Fig. 7 permet, dans les colonnes 1-3, de vérifier l'élaboration mathématique de la première ligne (MINMWMAX) du grand bloc à trois lignes de l'en-tête de la Fig. 19, alors que la colonne 4 de la Fig. 7 se réfère à la première ligne du petit bloc (également à trois lignes) situé sur la Fig. 19 au-dessous du grand bloc en question.

Comme, dans les Fig. 9-18, nous n'utiliserons que l'algorithme MINMWMAX, nous laisserons de côté le fonctionnement mathématique des deux autres algorithmes (MEDMW et MED¹⁴). En matière de cartographie statistique la représentation d'une *variation numérique* par une *variation iconique analogue* constitue un problème standard: cf. à ce sujet p. ex. Dickinson 1973, 83 ss. ainsi que Jenks/Coulson 1963.

Quant aux histogrammes figurant en bas de page sur les Fig. 9-18, leur mode de construction repose également sur l'algorithme d'intervallisation MINMWMAX avec, cette fois-ci, non plus 6 intervalles mais 12 intervalles. Les dimensions des 12 colonnes d'histogramme se calculent comme suit:

largeur:

moitié de la largeur d'intervalle (voir la colonne 3, Fig. 7).

hauteur:

quotient (p_j/d_j) entre la fréquence relative des points d'atlas affectés à l'intervalle considéré (p_j) et la largeur de cet intervalle (d_j).

Exemple: voir les colonnes 3 et 4 de la Fig. 7:

- $j = 1$ (= intervalle 1, sur 12 intervalles)
- $p_1 = 4/62$
- $d_1 = 1,50283$ (= 3,0056: 2)
- $p_1/d_1 = 0,0429$ unités factices de hauteur.

3.2.2. La lecture du corps des cartes numériques

Voir les Fig. 19-28 (annexe).

Toute carte numérique contient un en-tête et les coordonnées, répandues dans

¹⁴ Pour les algorithmes d'intervallisation MEDMW et MED cf. GOEBL 1984 I, 95-97.

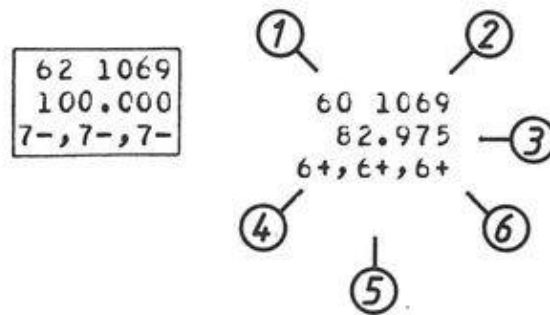


Fig. 8: Données numériques relatives à un point d'atlas.

Exemple tiré de la Fig. 19 (en haut, au centre).

A gauche: point de référence (j) (encadré).

A droite: point comparé (k).

1 indicatif du point d'atlas (ici: P. 60)

2 nombre des co-présences (= co-identités + co-différences) entre les vecteurs d'objet j (ici: P. 62) et k

(ici: P. 60) (équivalent au nombre total des attributs de notre matrice des données).

3 IRI_{jk} (ici: $IRI_{62,60}$).

4 classement par intervalle selon MINMWMAX (ici: intervalle 6 selon MINMWMAX à 6 intervalles; intervalle 6 + (= intervalle 12) selon MINMWMAX à 12 intervalles).

5 classement par intervalle selon MEDMW (ici: intervalle 6 selon MEDMW à 6 intervalles; intervalle 6 + (= intervalle 12) selon MEDMW à 12 intervalles).

6 classement par intervalle selon MED (ici: intervalle 6 selon MED à 6 intervalles; intervalle 6 + (= intervalle 12) selon MED à 12 intervalles).

l'espace et imprimées en carré, des 63 points de la matrice des données. Chacun de ces carrés obéit au même schéma typographique. Voir la Fig. 8.

La Fig. 8 permet en outre de contrôler la mesure de la similarité selon IRI_{jk} entre le point de référence 62 (encadré) et le point comparé 60: $IRI_{62,60} = 82,975$. Les attributions d'intervalle pour le point de référence 62 (7-, 7-, 7-) sont purement factices et n'ont aucune signification taxométrique.

4. Huit cartes de similarité

4.1. Comment interpréter une carte de similarité?

Toute carte de similarité dispose, pour N points d'atlas, de $N - 1$ scores de similarité et d'un point de référence (j) qui en constitue le pivot central. Les $N - 1$ valeurs de similarité (s) se réfèrent toutes à ce point de référence et constituent, du point de vue statistique voire mathématique, une *distribution de fréquence* susceptible d'être analysée en tant que telle suivant les principes de la *statistique descriptive*. En nous fondant sur de nombreuses expériences entreprises depuis le début des années 70 (et publiées à partir de 1976) nous pouvons dire que toute carte de similarité (et par conséquent toute distribution de similarité) tirée de la matrice de similarité, possède une originalité propre tant sur le plan numérique que – après transformation de la variabilité numérique en une

variabilité visuelle – sur le plan iconique. Cette originalité propre aux différentes distributions de similarité repose sur le fait que chacune des *N* distributions de similarité calculables à partir d'un réseau de *N* points d'atlas, constitue l'*actualisation* d'un *archétype sous-jacent*. Or, le nombre de ces archétypes sous-jacents étant inférieur à *N*, reste à résoudre le problème du décompte exact de ces archétypes. Comme ce problème ne peut être résolu qu'à l'aide de procédés statistiques multivariés fort compliqués, nous nous contenterons ici d'en signaler l'existence.

Les huit cartes de similarité présentées par la suite correspondent à cinq archétypes spatiaux iconiquement bien différenciés l'un de l'autre. Il n'est pas du tout fortuit que ces archétypes coïncident très souvent avec des régions géographiquement et historiquement bien définies.

Les cartes de similarité permettent d'ailleurs – et ce sont là des réflexions d'ordre méthodologique relatives à un « mariage heureux » entre dialectométrie et géolinguistique voire dialectologie traditionnelles – de mieux cerner les trois problèmes suivants:

a) Quelle est la *position linguistique*¹⁵ de tel *dialecte* (c.-à-d. de tel parler, de tel point d'enquête etc.) à l'intérieur d'un *réseau d'observation donné*? Depuis leur genèse au siècle dernier, la dialectologie et la géolinguistique traditionnelles n'ont cessé de poser cette question ou bien ouvertement ou bien d'une façon plus ou moins cachée. En guise d'exemple et de justification voici quelques titres fort éloquentes:

Del posto che spetta al ligure nel sistema dei dialetti italiani (Ascoli 1876).

The lexical affiliations of Vegliote (Fisher 1976).

La situació del català entre les llengües romàniques (Güter 1970).

Die Stellung des Katalanischen in der romanischen Sprachenfamilie (Kuen 1973).

Der Dialekt von Bergün und seine Stellung innerhalb der rätoromanischen Mundarten Graubündens (Lutta 1923).

La posizione del dalmatico nella Romania (Muljačić 1965).

La place des parlers nissarts dans la famille des langues romanes (Palermo 1977).

La questione del posto da assegnare al gallurese e al sassarese (Wagner 1943).

L'instrument méthodique de la carte de similarité permet dorénavant de mieux cerner ce problème et de le résoudre d'une façon beaucoup plus détaillée¹⁶.

b) Quel est le degré d'*interaction linguistique* de tel parler local avec son entourage plus ou moins immédiat? Qui dit *interaction* dit aussi *diffusion* voire *évolution par interaction dans le temps*. Aussi pourrait-on formuler cette question de la manière suivante: Quel est le degré de diffusion (ou d'enchevêtrement) de la dialecticité de tel parler local (défini par un vecteur d'un grand nombre d'attributs linguistiques) au sein d'un champ ou réseau d'observation donné (défini, lui, par un nombre plus ou moins

¹⁵ Il convient cependant de se rendre pleinement compte de toutes les implications méthodiques de l'utilisation du concept de « position ». A condition que l'on entende par là un *endroit* situé au sein d'un *espace tridimensionnel*, il faut en préciser les trois coordonnées de localisation (x, y et z) ainsi que l'étendue de l'espace respectif.

¹⁶ Pour plus de détails cf. GOEBL 1984 I, 100-104.

grand de parlers locaux)¹⁷? La carte de similarité donne, à cette question, des réponses précises.

- c) Quel est le degré d'*insertion* de tel parler local dans l'ensemble des autres parlers locaux qui l'entourent? Ou – en des termes plus adéquats: quelle est la «centralité» de tel parler local au sein d'une hiérarchie de parlers locaux plus ou moins «centraux»? L'instrument méthodique de la carte de similarité révèle que la centralité des points d'un réseau d'atlas est très variable et que, partant, tous les indices de centralité représentent une hiérarchie finement articulée¹⁸.

C'est ainsi que le chercheur dialectométricien dégage des perspectives tout à fait nouvelles (pour ne pas dire révolutionnaires) sur la structuration de l'espace géolinguistique. Inutile de dire qu'il faut à tout prix éviter de vouloir forcer les possibilités méthodiques et méthodologiques de la dialectométrie. Pour mieux pouvoir en évaluer les perspectives méthodiques et les problèmes méthodologiques, il est recommandé, au stade actuel de la dialectométrie, de suivre de près l'évolution tant méthodique que méthodologique des sciences suivantes: géographie quantitative (géographie humaine à orientation quantitative), sociologie quantitative/sociométrie, économie/économétrie, aménagement du territoire, biologie/biométrie. L'activité langagière de l'homme étalée dans l'espace ne constitue, somme toute, qu'une infime partie de tout ce qui, parmi ses activités sémiologiques, est susceptible de se manifester dans des structures spatiales¹⁹.

4.2. *Regard jurassien sur la Romandie linguistique des TPPSR*

Voir les Fig. 9 et 19 (annexe).

L'agencement spatial des six teintes hachurées en plages polygonales crée l'impression optique d'un relief montagneux dont les «sommets» se trouvent aux alentours du point de référence (P. 62) alors que les «vallées les plus profondes» occupent la partie méridionale (Valais: PP. 26, 30, 22 etc.) du champ d'observation. Remarquer la décroissance non linéaire de la *similarité linguistique* au fur et à mesure de l'agrandissement de la *distance kilométrique* entre le point de référence et le point comparé. Le relief de la Fig. 9 est l'actualisation d'un prototype dialectométrique jurassien. Une carte de similarité relative aux points de référence 60, 61 ou 59 aurait un profil très similaire. Noter en outre les observations suivantes:

que les points-*TPPSR* du Jura-Nord (correspondant à l'actuel canton du Jura) constituent un bloc homogène caractérisé par l'intervalle 6 (PP. 61, 60, 59 + point de référence 62).

¹⁷ Pour plus ample information cf. GOEBL 1984 I, 108–109 ainsi que HAGGETT 1973, 110–115 et 199–204.

¹⁸ Cf. aussi GOEBL 1984 I, 107–108 et HAGGETT 1973, 137–145.

¹⁹ Cf. à cela aussi GOEBL 1984 I, 100–113.

que les voisins linguistiques les plus proches de l'entourage immédiat du point de référence (classés dans l'intervalle 5) se trouvent tant dans le Jura du Sud (partie francophone de l'actuel canton de Berne) que dans le canton de Neuchâtel (P. 53). que la position du français standard (P. 99) par rapport au Jura est relativement bonne (intervalle 3).

que les antipodes linguistiques du point de référence (et, avec lui, du Jura tout entier) se situent dans le Valais.

4.3. *Regard vaudois sur la Romandie linguistique des TPPSR*

Voir les Fig. 10 et 20 (annexe).

Le type iconique représenté par la Fig. 10 est tout à fait différent de celui de la Fig. 9. Le déplacement du point de référence du point-*TPPSR* 62 au point-*TPPSR* 1 nous a fait «pénétrer dans un autre monde». Remarquer la vaste étendue des plages choroplèthes des intervalles 6 et 5 qui occupent toute la partie centrale de la Suisse romande (Fribourg et Vaud). Noter en outre que les antipodes linguistiques du point de référence (P. 1) ne se trouvent plus dans le Valais, mais bien plutôt dans le Jura.

4.4. *Regard fribourgeois sur la Romandie linguistique des TPPSR*

Voir les Fig. 11 et 21 (annexe).

Le type iconique sous-jacent à la Fig. 11 correspond de très près à celui de la Fig. 10: mêmes antipodes (Jura), mêmes voisins les plus proches (intervalles 6 et 5: Fribourg, Vaud), même position par rapport à la force assimilatrice du français standard (P. 99: intervalle 1) qui se situe, de concert avec le Jura-Nord, aux antipodes de la dialecticité du point de référence fribourgeois. De tout ceci il appert que les points de référence 38 et 1 appartiennent au même sous-ensemble classificatoire (type) spatial.

4.5. *Regard valaisan sur la Romandie linguistique des TPPSR*

Voir les Fig. 12 et 22 (annexe).

Si l'on considère les antipodes linguistiques (intervalle 1) et la position approximative des polygones choroplèthes dans les intervalles 2, 3 et 4, il y a de fortes ressemblances entre le type de la Fig. 11 et celui de la Fig. 12. Les plus fortes divergences entre les deux types iconiques se trouvent par contre aux alentours du point de référence même (P. 28). Remarquer que le nombre des polygones classés dans les intervalles 6 et 5 (5 polygones) est très restreint par rapport à celui des Fig. 11 (13 polygones) et 10 (17 polygones). Cette divergence est un signe de la forte individualité linguistique (et partant taxatoire) des points-*TPPSR* du Bas-Valais oriental.

4.6. *Regards genevois sur la Romandie linguistique des TPPSR*

Voir les Fig. 13–15 (cartes choroplèthes) et les Fig. 23–25 (cartes numériques) (annexe).

Les Fig. 13, 14 et 15 représentent toutes un seul archétype iconique, genevois en l'occurrence. Comme les profils choroplèthes des trois Figures se ressemblent beaucoup, il est possible d'en faire une courte séquence cinématographique (en trois images)²⁰. Les changements du décalage choroplèthe qui interviennent lors d'un déplacement de la référence du point 32 (Fig. 13) au point 33 (Fig. 14) et du point 33 (Fig. 14) au point 34 (Fig. 15), ne renversent pas, bien qu'ils affectent toutes les parties de la surface choroplèthe, l'architecture du type iconique. On a l'impression, en passant successivement en revue les Fig. 13, 14 et 15, de percevoir, malgré une légère fluctuation dans les détails, toujours la même macrostructure iconique.

Aux antipodes linguistiques des dialecticités genevoises se situent le Jura (intervalle 1) et le Valais (intervalle 2: avec des provignements en territoire fribourgeois et vaudois). Remarquer du reste que le français standard (P. 99), lui aussi, fait pour une fois partie des antipodes secondaires (Fig. 14: intervalle 2).

La constance avec laquelle les points-*TPPSR* du Jura-Nord (PP. 59, 60, 61 et 62) occupent les antipodes absolus du réseau-*TPPSR* est d'ailleurs fort remarquable. Cette particularité vaut pour tous les points de référence romands situés en dehors du Jura. Cette constatation cadre d'ailleurs fort bien avec la conception géolinguistique traditionnelle selon laquelle l'espace dialectal romand est partagé entre le type dialectal *francoprovençal* et le type dialectal *franc-comtois* voire carrément *oilique* (c.-à-d. relatif au type linguistique d'oïl).

4.7. Regard français sur la Romandie linguistique des *TPPSR*

Voir les Fig. 16 et 26 (annexe).

Le point de référence (P. 99) est un «point d'atlas» factice. L'insertion de ce point artificiel permet de mesurer (et partant de faire voir ensuite) l'état de francisation des parlers romands tels qu'ils ont été enregistrés par les enquêteurs des *TPPSR*. Comme le P. 99 ne fait pas partie du réseau-*TPPSR* et que, par conséquent, sa dialecticité n'a jamais pu influencer directement – c'est-à-dire par l'effet de la contiguïté spatiale – sur celle de la Romandie, on pourrait s'attendre à ce que le profil choroplèthe de la carte de similarité respective fût plus ou moins chaotique. En vérité, il n'en est rien. L'agencement des plages choroplèthes revêt en revanche des apparences fort logiques. Les zones les plus francisées de la Romandie (intervalles 6 et 5) correspondent à des régions majoritairement protestantes (Neuchâtel, Berne, Jura-Sud) alors que les zones les plus réfractaires à la francisation (intervalles 1 et 2) coïncident – à deux exceptions près (PP. 5 et 6) – avec des régions à forte majorité catholique (Bas-Valais, Fribourg). Comparer à ce sujet les cartes 26 (confessions) et 28 (langues II)²¹ de l'«Atlas de la Suisse (Atlas der

²⁰ Il s'agit là d'une application pratique de la cartographie dynamique; pour plus de détails cf. GOEBL 1984 I, 104-105 et 1982, 38-39.

²¹ Voir aussi la carte «Vitalité du patois en 1966» dans l'*Atlas de la Suisse* (carte 28 II) ainsi que les cartes thématiques en appendice à l'ouvrage de SCHULER/NEF 1983.

Schweiz)» qui, elles, confirment pleinement la covariance *religiosité-francisation* démontrée ci-dessus.

Le canton de Genève qui, du point de vue religieux, est de nos jours mi-protestant mi-catholique, occupe, quant à la francisation, une position plutôt privilégiée (intervalles 4 et 5).

En résumé, on peut dire que l'empreinte du *facteur d'innovation* «français standard» apparaît, dans le tissu dialectal de la Romandie telle qu'elle se présente dans les relevés géolinguistiques des *TPPSR*, le plus souvent là où le tissu religieux de la Romandie se trouve marqué par le *facteur d'innovation* «protestantisme». C'est par la mise en lumière de telles covariances intrasystémiques que la dialectométrie ouvre des perspectives indubitablement interdisciplinaires et sans nul doute fort prometteuses.

5. Au-delà de la mesure de similarité: la synopsis de deux paramètres caractéristiques des distributions de similarité

Il est bien évident que le rang taxométrique de la méthode dialectométrique utilisée en détermine la valeur méthodologique. Or, les cartes de similarité sont bien loin d'épuiser la panoplie dialectométrique tout entière. Bien au contraire: elles n'en constituent que des pièces d'artillerie légère, très utiles, certes, mais de portée réduite. Nous présentons par la suite deux méthodes dialectométriques d'une portée méthodologique plus grande et dont la complexité mathématique reste néanmoins fort limitée.

5.1. Synopsis des maxima des distributions de similarité

Voir les Fig. 17 et 27 (annexe).

Toute distribution de similarité dispose d'un *minimum*, d'une *moyenne arithmétique* et d'un *maximum*. Les maxima des distributions de similarité des Fig. 19 (ou 9), 20 (ou 10) et 21 (ou 11) sont respectivement: 86,342; 83,536; 88,494. Voir les en-têtes des cartes numériques (Fig. 19, 20 et 21). Comme il y a, pour N points d'atlas, N distributions de similarité différentes, il est possible de constituer une distribution de fréquence composée de N valeurs maximales. Sur la carte numérique ainsi constituée, les différentes valeurs maximales sont imprimées à l'endroit du point de référence de la distribution de similarité respective.

Cette synthèse taxométrique permet de cerner de plus près le problème géolinguistique du «noyau dialectal». Malheureusement, la plupart des linguistes ont utilisé ce concept sans trop se soucier d'en donner une définition exacte. Toujours est-il qu'un grand nombre d'entre eux entendait par «noyau dialectal» une sorte de «centre de gravité» d'un ensemble dialectal (ou sous-dialectal) donné capable de propager des flux d'innovations linguistiques et/ou métalinguistiques, et qui devait assumer par là le rôle d'une place centrale entourée de satellites en position moins centrale. Or, c'est exactement cette configuration spatiale que la méthode de la synthèse des maxima permet de

découvrir. Sans trop nous perdre dans des détails méthodiques nous passons à l'interprétation de la Fig. 17 qui permettra avant tout d'examiner les ondulations du profil choroplèthe et d'y repérer les différents sommets. Un sommet se définit comme tel par la présence d'élévations moins hautes dans le voisinage. On peut repérer ainsi, sur la Fig. 17, les sommets ou noyaux dialectaux suivants:

- noyau jurassien: PP. 62, 61.
- noyau neuchâtelois: PP. 47, 46, 51.
- noyau fribourgeois: PP. 39, 38.
- noyau vaudois: PP. 10, 12.
- noyau genevois: PP. 33, 35.
- noyau valaisan: PP. 28, 31.

Remarquer en outre que tous ces noyaux sont séparés l'un de l'autre par des dépressions de terrain plus ou moins accusées, et que le passage d'un noyau à l'autre nécessite toujours, de la part d'un voyageur imaginaire, une descente et une remontée consécutive. Ainsi le profil des noyaux dialectaux prend des apparences pour ainsi dire harmoniques voire esthétiques.

La faculté discriminatoire de cette méthode dialectométrique est d'ailleurs remarquable: aucune des individualités géolinguistiques de la Romandie (qui coïncident normalement avec les domaines cantonaux) n'a été omise. Pour d'autres applications non moins pertinentes de cette méthode cf. Goebel 1981, 382-388; 1982, 42-43; 1983b, 29-40 et 1984 I, 140-143.

5.2. *Synopse des coefficients d'asymétrie (de Fisher) des distributions de similarité*

Voir les Fig. 18 et 28 (annexe).

Pour expliquer les assises méthodologiques de cette méthode, nous avons recours à une analogie tirée de la réalité sociale. Soit un groupe de N individus. Or, il est bien connu que tous les groupes humains se composent, d'un côté, d'individus plus ouverts et dont le goût d'interaction sociale est plus marqué, et, de l'autre, d'individus plus fermés et repliés sur eux-mêmes. Aussi les membres de tout groupe social peuvent-ils être classés selon ce critère différenciateur (ouverture-fermeture). Il en va de même en géographie linguistique où les concepts de «résidu géolinguistique» (ou de «zone conservatrice») et de «zone de transition» font écho aux concepts sociologiques d'ouverture et de fermeture sociales. Bien que ces deux concepts géolinguistiques, à l'instar de celui de noyau dialectal, ne soient pas encore nettement définis, il semble indiqué de se mettre à la recherche d'un corrélat dialectométrique.

Définitions approximatives des notions de «résidu géolinguistique» et de «zone de transition»:

Résidu géolinguistique (zone conservatrice):

Interaction voire affinité linguistique à très courte distance, conservation de particularités linguistiques qui n'existent nulle part ailleurs, étendue spatiale assez limitée.

Zone de transition:

Prépondérance de traits linguistiques plus ou moins «omniprésents» et dont la diffusion spatiale est grande, manque de particularités linguistiques trop marquées, étendue spatiale assez considérable.

Les particularités mentionnées ci-dessus nous conduisent à l'étude d'une des propriétés les plus saillantes de toute distribution de fréquence, à savoir de celle de sa *symétrie* ou *asymétrie*. En termes dialectométriques, une distribution de similarité asymétrique vers la *gauche* (voir l'histogramme de la Fig. 9) serait celle d'un *parler résiduaire*, alors qu'une distribution asymétrique vers la *droite* (voir l'histogramme de la Fig. 16) correspondrait à celle d'un *parler de transition*. Les notions de «résidu» et de «transition» sont d'ailleurs purement relationnelles. Un parler dit résiduaire ou de transition ne peut l'être que par rapport à un ensemble d'autres parlers préalablement réunis en groupe.

En statistique descriptive, la meilleure mesure d'asymétrie est le troisième moment de la distribution de fréquence $((x_i - \bar{x})^3)$ divisé par le cube de l'écart type (s_j^3). En voici la formule²²:

$$g_j = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n \cdot s_j^3} \quad (3).$$

Dans cette formule,

g est le coefficient d'asymétrie (de Fisher) d'une distribution de similarité relative au point de référence j .

j est l'indicatif du point de référence d'une distribution de similarité.

i est l'indicatif (allant de 1 à 62) des différents scores de similarité (calculés selon IRI_{jk}).

n est le total (= 62) des scores de similarité ($n = N-1$; pour $N = 63$ points-*TPPSR*).

x_i est un des 62 scores de similarité (calculés selon IRI_{jk}).

\bar{x} est la moyenne arithmétique d'une distribution de similarité ($\bar{x} = MW$; cf. le paragraphe 7.).

s_j est l'écart type (all. Standardabweichung) d'une distribution de similarité relative au point de référence j (s_j étant calculé pour 62 scores de similarité).

Pour plus ample information cf. Goebel 1981, 394-401; 1982, 44-46; 1983 a, 432-436 et 1984 I, 150-154.

Le coefficient g_j prend la valeur 0 quand la distribution de similarité est parfaitement symétrique, des valeurs positives quand la distribution est asymétrique vers la gauche (type du parler «résiduaire») et des valeurs négatives quand la distribution de similarité est asymétrique vers la droite (type du parler «de transition»). Soit dit en passant, l'utilité taxométrique du coefficient de Fisher est double: elle repose tant sur la taille du coefficient elle-même que sur la modalité du signe opérationnel (+ ou -).

²² Pour la formule du coefficient de Fisher cf. BAHRENBERG/GIESE 1975, 54.

Quelle est la signification *linguistique* voire *géolinguistique* du degré d'asymétrie d'une distribution de similarité? Quand une distribution de similarité est asymétrique vers la gauche (voir l'histogramme respectif de la Fig. 9), la masse des différents scores de similarité qui la constituent, est accumulée vers la gauche et se trouve ainsi concentrée au-dessous de la moyenne arithmétique. Or, cette accumulation des scores de similarité au-dessous de la moyenne arithmétique signifie, du point de vue géolinguistique, que le vecteur du point d'atlas (parler) considéré est mal inséré dans l'ensemble des autres points d'atlas et que ses capacités d'interaction avec le reste du réseau sont plus ou moins réduites ou carrément mauvaises. En comparant le vecteur d'attribut d'un point d'un réseau d'atlas avec celui d'un membre d'un groupe humain, on peut dire qu'il s'agit alors du vecteur d'une personne socialement peu intégrée, peu acceptée des autres et montrant une attitude plutôt hostile aux autres.

L'accumulation des scores de similarité au-dessus de la moyenne arithmétique signifie l'inverse: c'est-à-dire que le point d'atlas respectif est bien intégré, qu'il fait partie d'échanges linguistiques fréquents et intenses, bref qu'il se situe à un carrefour important de la fluctuation interactive à l'intérieur du réseau considéré. Le profil choroplèthe de la Fig. 18 montre donc très bien que, par rapport à l'ensemble du réseau des *TPPSR*, le Jura (du Nord et du Sud), le canton de Genève et le Valais oriental sont les pôles réfractaires du champ d'observation alors que la superficie des cantons de Vaud, de Fribourg et du Valais occidental constituent une sorte de glacis interférentiel avec, au centre, une zone d'interaction maximale (intervalle 1) le long de l'axe est-ouest du Lac Léman. Remarquer que le français standard (P. 99) représente, dans cette optique, un parler non pas réfractaire mais déjà tout à fait transitionnel (intervalle 2). Ceci signifie que le français standard – tout en étant une variété parachutée de loin (point non contigu et artificiel en plus) et non «de souche» – semble être intégré dans le réseau-*TPPSR* à l'instar des parlers autochtones du Vaud et du Bas-Valais occidental.

Noter aussi que l'agencement des plages choroplèthes est tout autre que brusque et qu'on a ainsi l'impression d'un relief montagneux doucement ondulé. Pour le même effet voir les profils analogues dans Goebel 1984 I, 153 (profil lissé) et 169 (profil non lissé) exécutés sur d'autres données géolinguistiques.

6. Epilogue et perspectives de recherche

Les procédés dialectométriques décrits ci-dessus constituent une branche de *l'analyse typologique* ou *typo-diagnostique*. Ce genre de recherche se situe aux antipodes de ce que l'on appelait autrefois (et, peut-être, appelle encore) le «positivisme atomiste», et n'a jamais connu de grandes heures dans le domaine de la dialectologie. Nous avons essayé à plusieurs reprises (Goebel 1982, 57–59; 1983a, 383–385 et 1984 I, 13–15, 98–105) de tracer l'historique de la pensée classificatoire en matière de géographie linguistique et d'en faire voir les orientations épistémologiques profondes, orientations qui, à partir

de la négation stricte du concept de « dialecte » par P. Meyer et G. Paris dans le dernier quart du siècle passé, ont fini par décourager, en géolinguistique, tout effort de synthèse et synopse, bref de typo-diagnose. C'est en 1973 seulement que Jean Séguy, hanté depuis longtemps par la conception *typophobe* de l'indomptabilité du jeu interférentiel des isoglosses, a fini par se convertir à la *typophilie*, tout en jetant les bases de ce qu'il appelait lui-même (Séguy 1973, 1) la « dialectométrie ». Désormais, la dialectométrie permet non seulement de combler le vide créé par la prépondérance de conceptions typophobes mais aussi de découvrir des aspects jusqu'alors inconnus de la stratification spatiale de l'activité langagière de l'homme. Elle rejoint par là l'efficacité exploratoire de toutes les métriques appliquées. Il est cependant évident que la nouvelle approche métrisant requiert la mise en place de nouveaux instruments méthodiques et aussi – ce qui est très important – une réorientation méthodologique. Cette réorientation méthodologique passe par la maîtrise scientifique du concept de *quantité*. Or, la quantité n'est nullement – comme d'aucuns le croient – l'ennemi irréconciliable de la *qualité*. Bien au contraire: il y a, entre ces deux notions, un rapport de complémentarité et de réciprocité ineffaçable et très fécond qu'il convient de ne jamais perdre de vue.

7. Abréviations et glossaire (termes techniques et allemands)

G	Coefficient d'asymétrie (de Fisher): voir les Fig. 19–28 (cartes numériques) ainsi que le paragraphe 5.2.
Häufigkeiten	fréquences (absolues): nombre des points d'atlas par intervalle: voir les Fig. 9–18 (cartes choroplèthes).
MAX	valeur maximale d'une distribution de similarité: voir les Fig. 17 et 27 ainsi que le paragraphe 5.1.
MED	algorithme d'intervallisation (pas traité dans cet article): voir les Fig. 19–28 (cartes numériques).
MEDMW	algorithme d'intervallisation (pas traité dans cet article): voir les Fig. 19–28 (cartes numériques).
MINMWMAX	algorithme d'intervallisation: voir les Fig. 9–28 (cartes choroplèthes et cartes numériques) ainsi que le paragraphe 3.2.1.
MW	moyenne arithmétique (all. <i>Mittelwert</i>), calculée à partir de 62 scores dans les Fig. 19–26 et à partir de 63 scores dans les Fig. 27 et 28: voir les Fig. 19–28 (cartes numériques).
N-KART.	nombre des attributs nominaux disponibles dans le vecteur d'objet du point de référence: voir les Fig. 19–26 (cartes numériques de similarité).
PUNKT: 62	point d'atlas 62: voir les Fig. 19–26 (cartes numériques de similarité).
S	écart type (all. <i>Standardabweichung</i>), calculé à partir de 62 scores dans les Fig. 19–26 et à partir de 63 scores dans les Fig. 27 et 28: voir les Fig. 19–28 (cartes numériques).
SCHIEFE	coefficient d'asymétrie (de Fisher): voir les Fig. 18 et 28 ainsi que le paragraphe 5.2.

8. Bibliographie

- AIS: Sprach- und Sachatlas Italiens und der Südschweiz*, ed. K. JABERG/J. JUD, vol. I-VIII, Zofingen 1928-1940.
- ALF: Atlas linguistique de la France*, ed. J. GILLIÉRON/E. EDMONT, vol. I-IX, Paris 1902-1910.
- ALPI: Atlas lingüístico de la península ibérica*, vol. I: *Fonética*, 1, Madrid 1962.
- ASCOLI G. I. «Del posto che spetta al ligure nel sistema dei dialetti italiani», *AGI2* (1876), 111-160.
- BAHRENBERG G./E. GIESE, *Statistische Methoden und ihre Anwendung in der Geographie*, Stuttgart 1975.
- BOCK H. H., *Automatische Klassifikation. Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten (Cluster-Analyse)*, Göttingen 1974.
- CHANDON J.-L./S. PINSON, *Analyse typologique. Théories et applications*, Paris - New York - Barcelone - Milan 1981.
- DICKINSON G. C., *Statistical mapping and the presentation of statistics*, Londres 1973.
- DURAND J.-P., «Notes de philologie rouergate (suite)», *RLaR* 33 (1889), 47-84.
- FISHER J., *The lexical affiliations of Vegliote*, Rutherford (New Jersey) 1976.
- GAUCHAT L./J. JEANJAQUET, *Bibliographie linguistique de la Suisse romande*, Neuchâtel 1920, 2 vol.
- GOEBL H., «*La dialectométrie appliquée à l'ALF (Normandie)*», in: *XIV Congresso internazionale di linguistica e filologia romanza. Atti*, ed. A. VÄRVARO, Naples - Amsterdam 1976, vol. II, p. 165-195.
- GOEBL H., «*Éléments d'analyse dialectométrique (avec application à l' AIS)*», *RLiR* 45 (1981a), 349-420.
- GOEBL H., *Dialektometrie. Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialektgeographie*, Vienne 1982 (*Denkschriften der Österreichischen Akademie der Wissenschaften, phil.-hist. Klasse*, vol. 157).
- GOEBL H., «*Parquet polygonal et treillis triangulaire. Les deux versants de la dialectométrie interponctuelle*», *RLiR* 47 (1983), 353-412 (a).
- GOEBL H., «*Stammbaum und Welle. Vergleichende Betrachtungen aus numerisch-taxonomischer Sicht*», *Zeitschrift für Sprachwissenschaft* 2 (1983), 3-44 (b).
- GOEBL H., *Dialektometrische Studien. Anhand italo-romanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*, Tübingen 1984, 3 vol.
- GPSR: Glossaire des patois de la Suisse romande*, Neuchâtel 1924ss.
- GUITER H., «*La situació del català entre les llengües romàniques*», *Miscellanea barcinonensia* 25 (1970), 35-45.
- HAGGETT P., *L'analyse spatiale en géographie humaine*, Paris 1973.
- JENKS G. F./M. C. COULSON, «*Class intervalls for statistical maps*», *Internationales Jahrbuch für Kartographie* 3 (1963), 119-134.
- KUEN H., «*Die Stellung des Katalanischen in der romanischen Sprachfamilie*», in: *Studia iberica. Festschrift für Hans Flasche*, ed. K.-H. KÖRNER/K. RÜHL, Berne - Munich 1973, p. 331-352.
- LUTTA, C. M., *Der Dialekt von Bergün und seine Stellung innerhalb der rätoromanischen Mundarten Graubündens*, Halle 1923.
- MULJAČIĆ Ž., «*La posizione del dalmatico nella Romania*», in: *Actes du X^e Congrès international de linguistique et philologie romanes*, Paris 1965, vol. III, p. 1185-1194.
- PALERMO J., «*La place des parlers nissarts dans la famille des langues romanes*», *RLiR* 41 (1977), 339-347.

- SCHIFKO P., «Zero in der Allgemeinen und Romanischen Sprachwissenschaft», *ZRPh.* 89 (1973), 1-51.
- SCHULER M./R. NEF, *Räumliche Typologien des schweizerischen Zentren-Peripherien-Musters*. Arbeitsberichte des Nationalen Forschungsprogrammes «Regionalprobleme in der Schweiz», no. 35, Berne-Bossière (VD) 1983.
- SÉGUY J., «La dialectométrie dans *l'Atlas linguistique de la Gascogne*», *RLiR* 37 (1973), 1-24.
- SNEATH P. H. A./R. R. SOKAL, *Numerical taxonomy. Principles and practice of numerical classification*, San Francisco 1973.
- SODEUR W., *Empirische Verfahren zur Klassifikation*, Stuttgart 1974.
- THIESSEN A. H., «Precipitation averages for large areas», *Monthly weather review* 39 (1911), 1082-1084.
- TPPSR: Tableaux phonétiques des patois suisses romands. Relevés comparatifs d'environ 500 mots dans 62 patois-types*, ed. L. GAUCHAT/J. JEANJAQUET/E. TAPPOLET, Neuchâtel 1925.
- WAGNER M. L., «La questione del posto da assegnare al gallurese e al sassarese», *CN* 3 (1943), 243-267.

Salzbourg

Hans Goebel

Avec le concours de:

Siegfried Selberherr (calculs électroniques)

Wolf-Dieter Rase (cartographie automatique)

Hilmar Pudlatz (pavage de Thiessen)

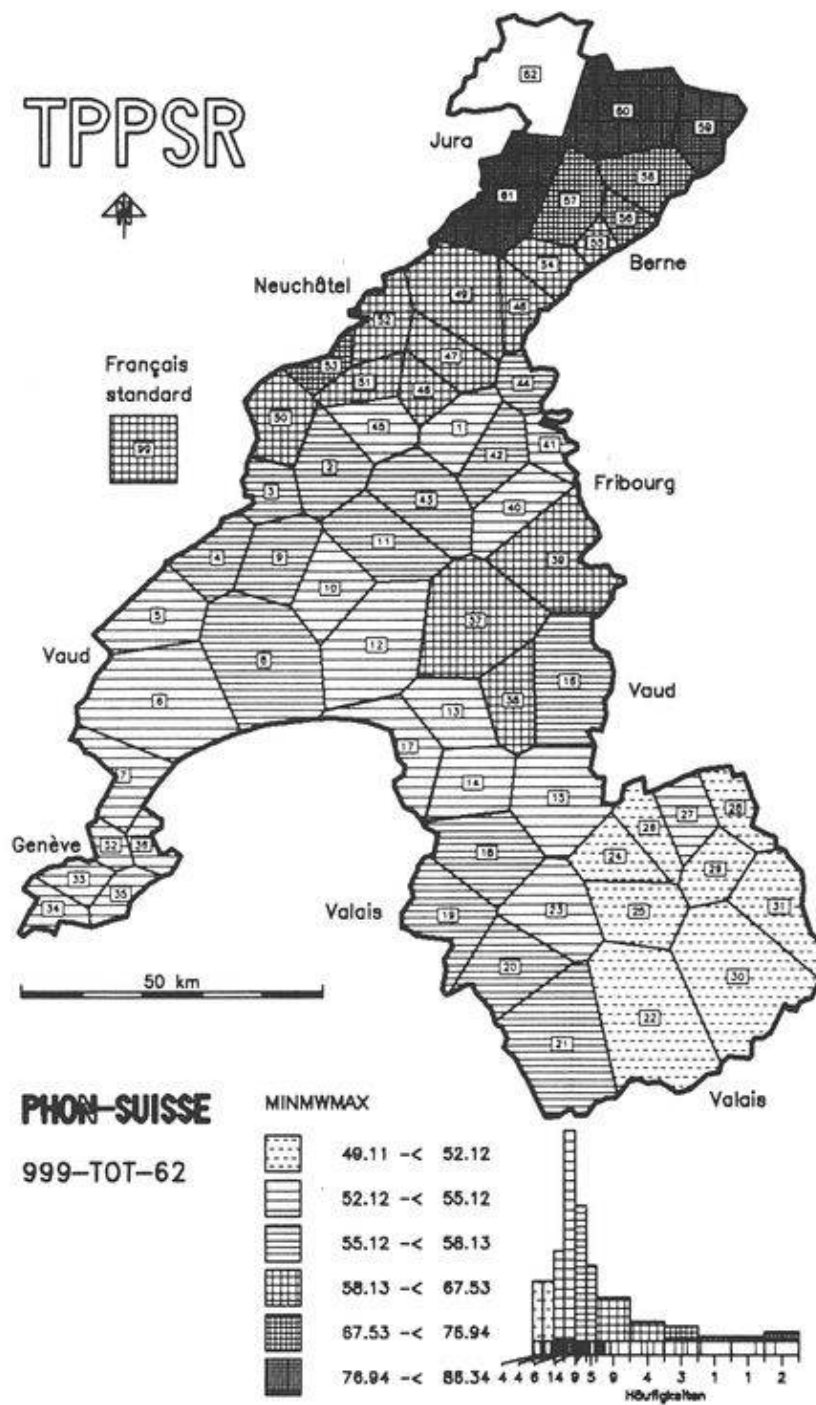


Fig. 9: Carte choroplèthe de la distribution de similarité (selon IRI_{jk}) relative au point de référence 62 (Courtedoux, Jura). Algorithme d'intervallisation: MINMWMAX à 6 intervalles. Voir aussi la Fig. 19.

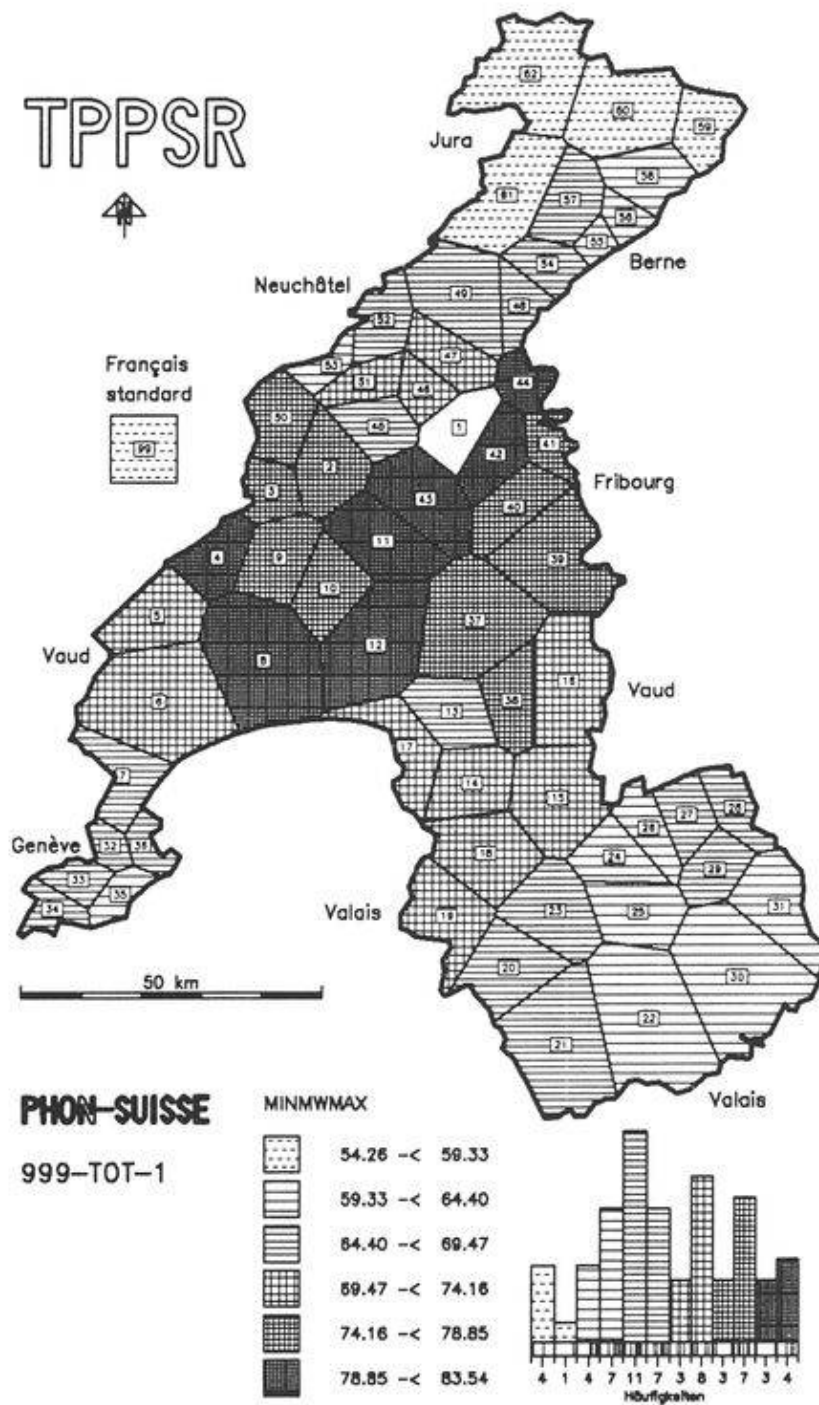


Fig. 10: Carte choroplèthe de la distribution de similarité (selon IRI_{jk}) relative au point de référence 1 (Chevroux, Vaud). Algorithme d'intervallisation: MINMWMAX à 6 intervalles. Voir aussi la Fig. 20.

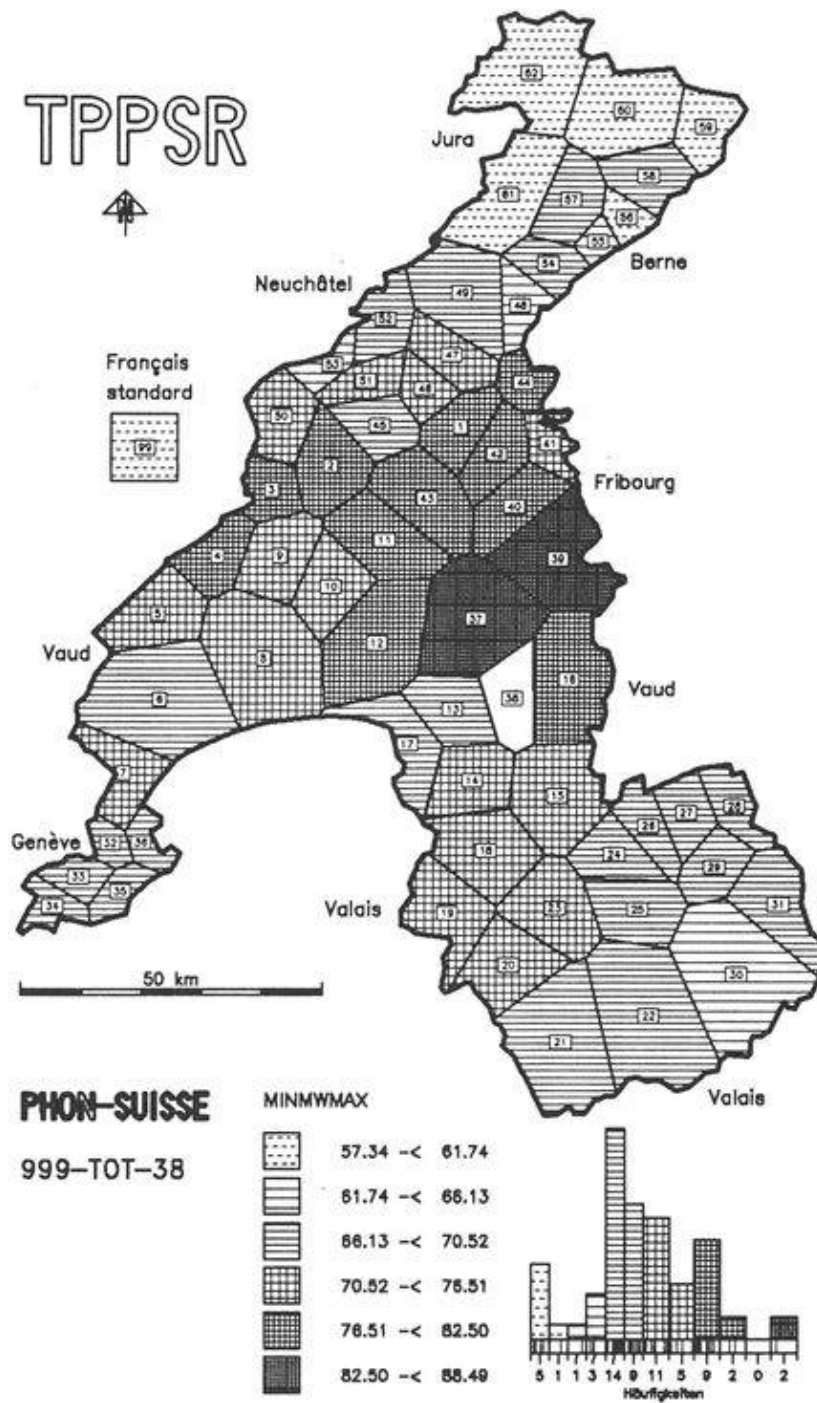


Fig. 11: Carte choroplèthe de la distribution de similarité (selon IRI_{jk}) relative au point de référence 38 (Montbovon, Fribourg). Algorithme d'intervallisation: MINMWMAX à 6 intervalles. Voir aussi la Fig. 21.

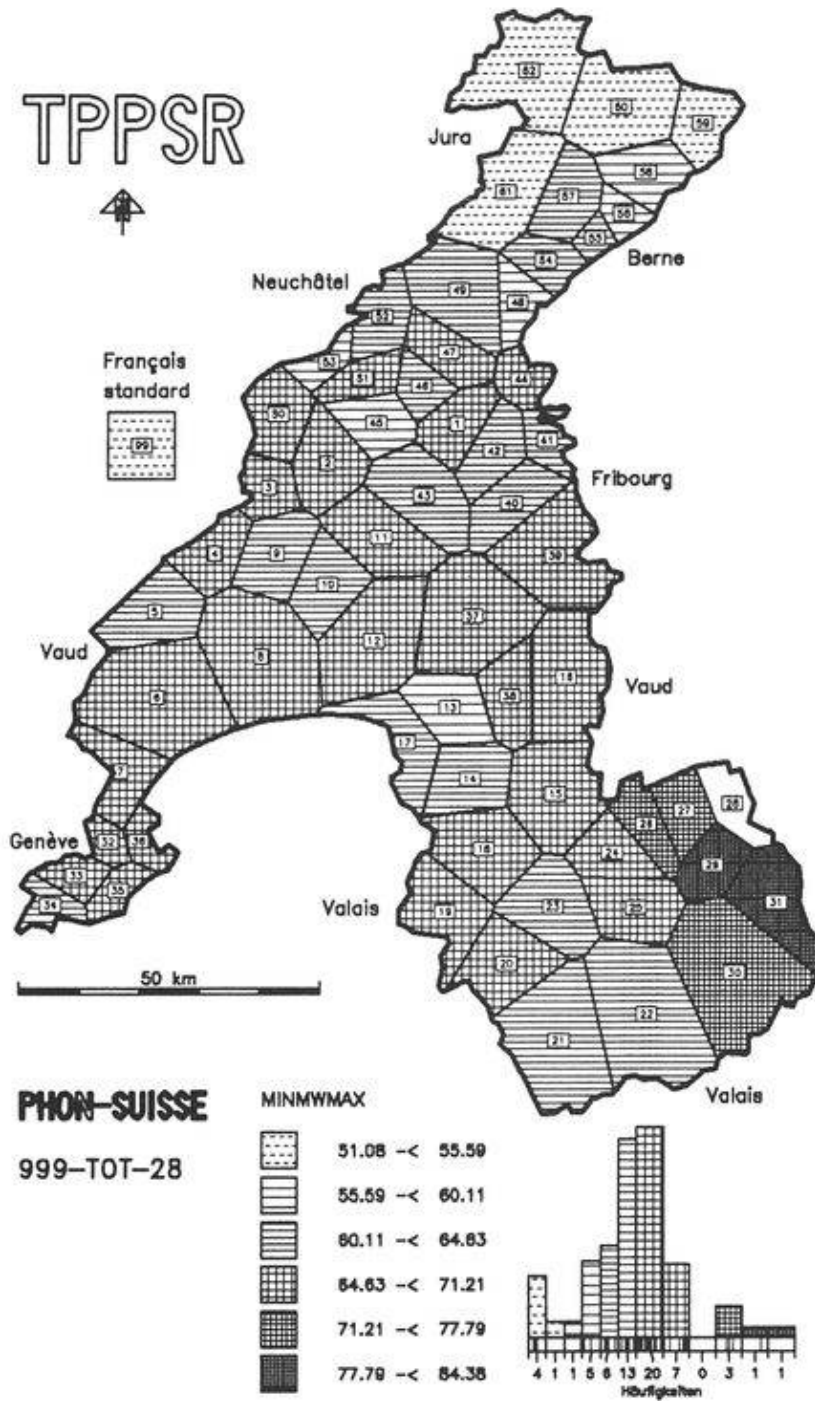


Fig. 12: Carte choroplèthe de la distribution de similarité (selon IRI_{jk}) relative au point de référence 28 (Miège, Valais). Algorithme d'intervallisation: MINMWMAX à 6 intervalles. Voir aussi la Fig. 22.

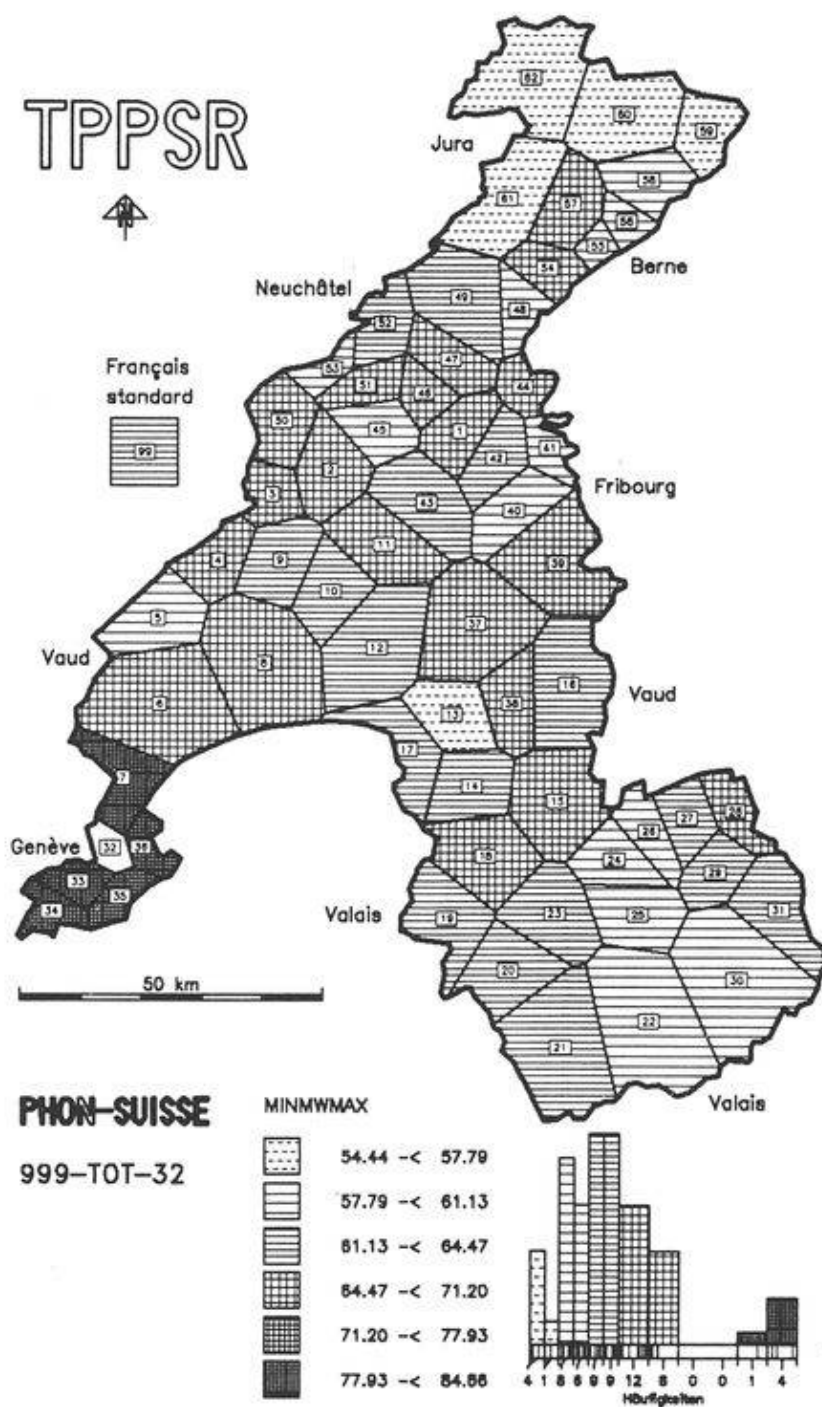


Fig. 13: Carte choroplèthe de la distribution de similarité (selon IRI_{jk}) relative au point de référence 32 (Collex, Genève). Algorithme d'intervallisation: MINMWMAX à 6 intervalles. Voir aussi la Fig. 23.

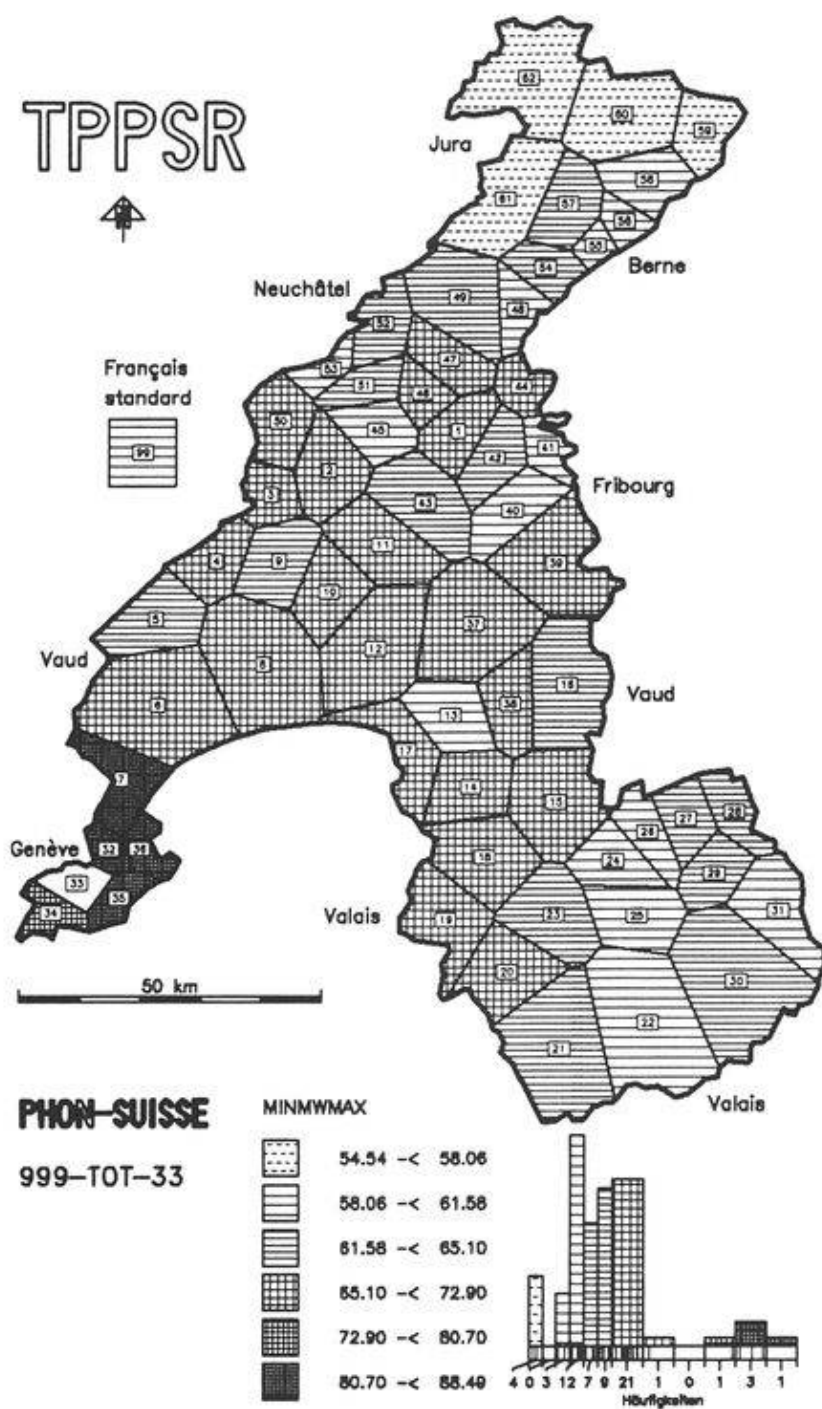


Fig. 14. Carte choroplèthe de la distribution de similarité (selon IRI_{jk}) relative au point de référence 33 (Vernier, Genève). Algorithme d'intervallisation: MINMWMAX à 6 intervalles. Voir aussi la Fig. 24.

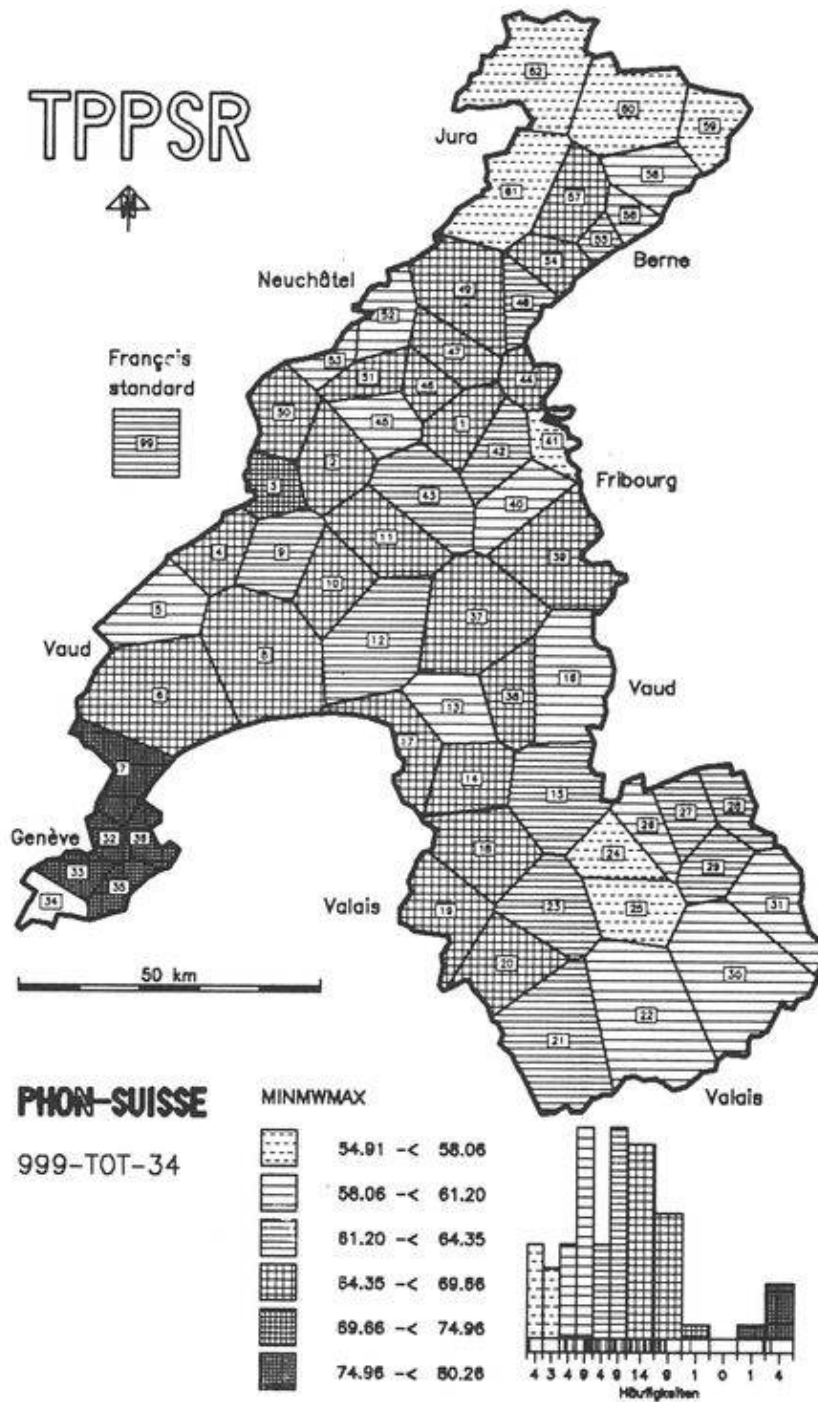


Fig. 15: Carte choroplèthe de la distribution de similarité (selon IRI_{jk}) relative au point de référence 34 (Laconnex, Genève). Algorithme d'intervallisation: MINMWMAX à 6 intervalles. Voir aussi la Fig. 25.

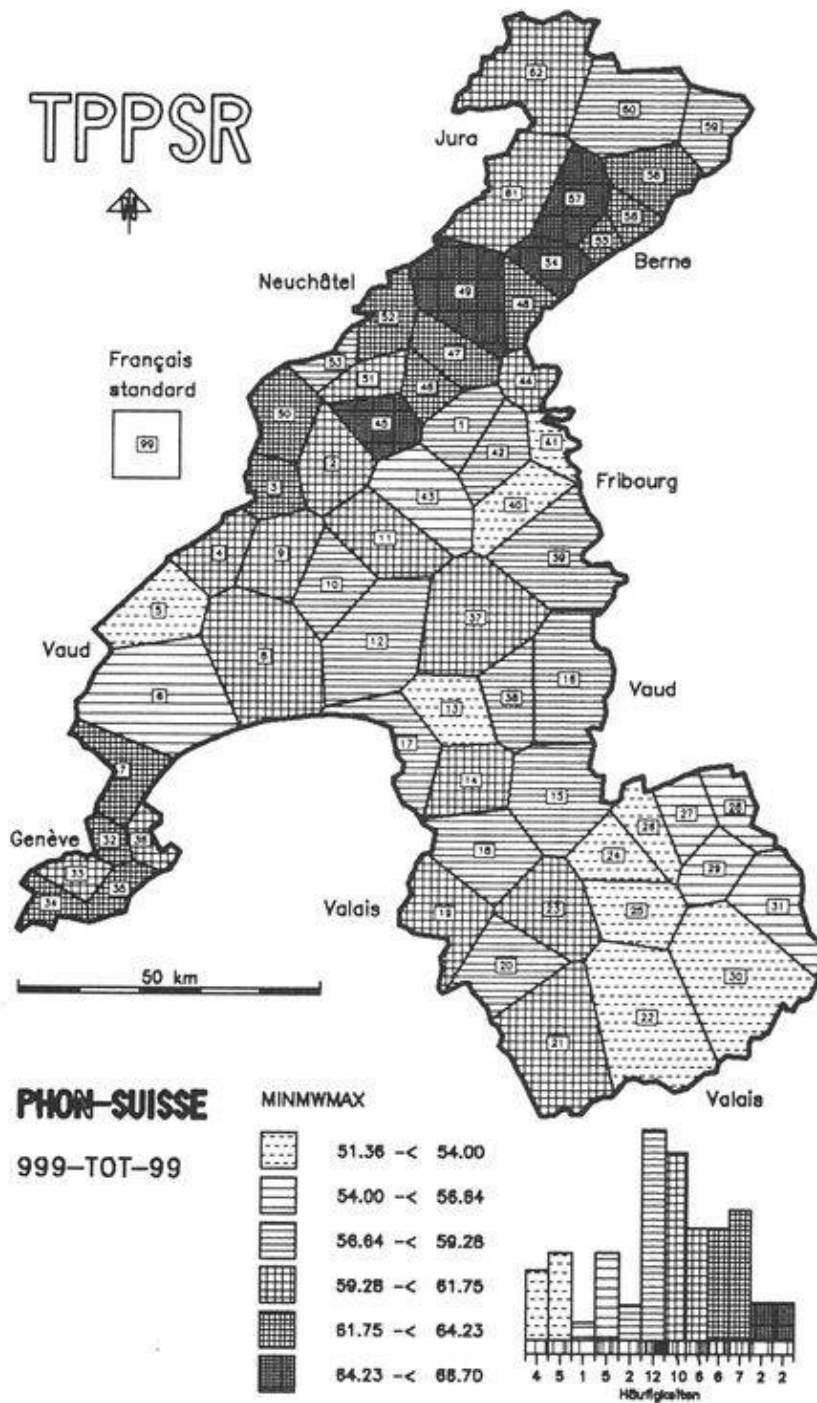


Fig. 16: Carte choroplèthe de la distribution de similarité (selon IRI_{jk}) relative au point de référence 99 (français standard). Algorithme d'intervallisation: MINMWMAX à 6 intervalles. Voir aussi la Fig. 26.

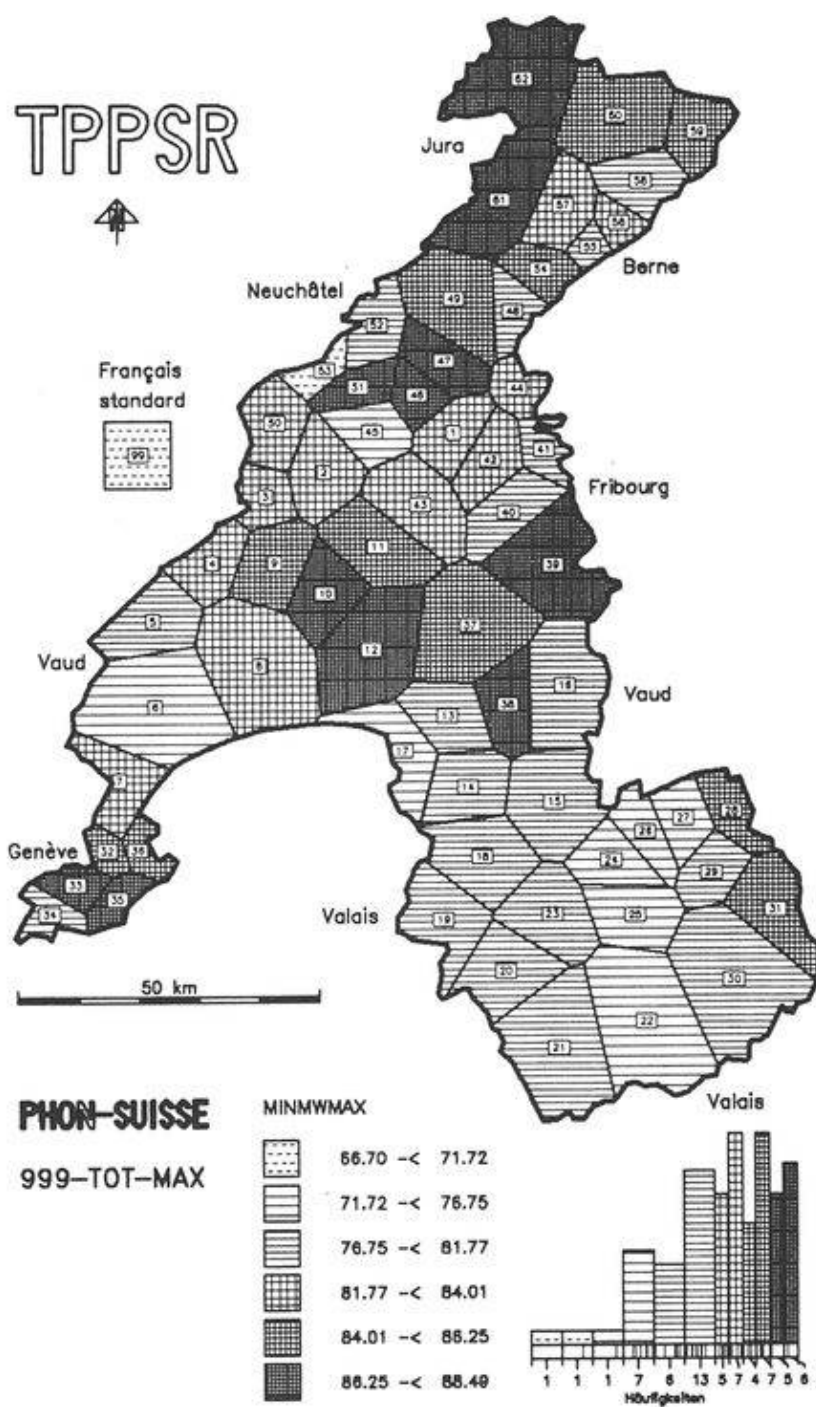


Fig. 17: Carte choroplèthe de la synthèse des maxima des 63 distributions de similarité (selon IRI_{jk}).
 Algorithme d'intervallisation: MINMWMAX à 6 intervalles.
 Voir aussi la Fig. 27.

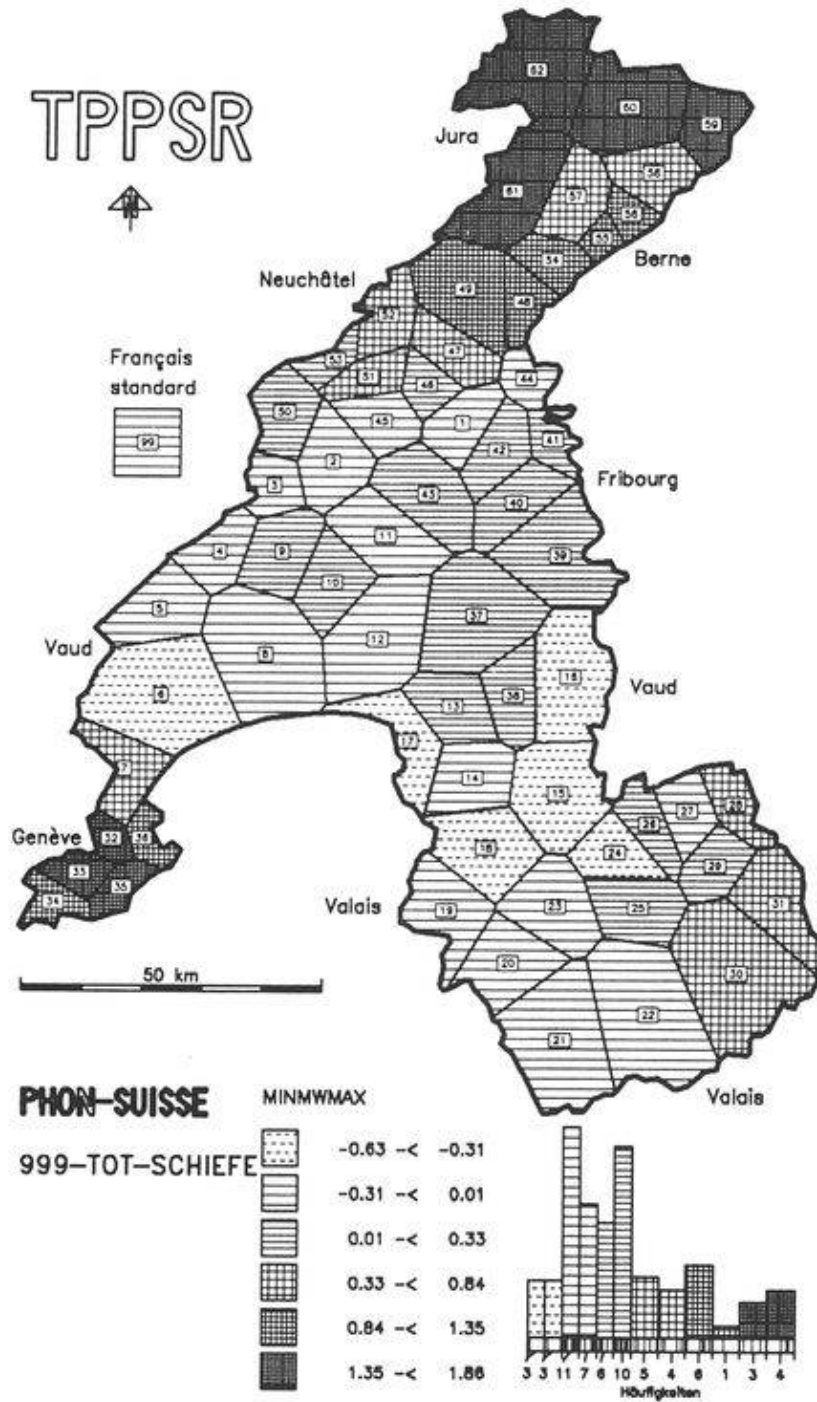


Fig. 18: Carte choroplèthe de la synopsis des coefficients d'asymétrie (de Fisher) des 63 distributions de similarité (selon IRI_{jk}).
 Algorithme d'intervallisation: MINMWMAX à 6 intervalles.
 Voir aussi la Fig. 28.

