

ATLAS, MATRICES ET SIMILARITÉS: PETIT APERÇU DIALECTOMÉTRIQUE¹

HANS GOEBL

collaboration technique: S. Selberherr, W.-D. Rase et H. Pudlatz

The subject of this article is a brief survey of dialectometry, which is defined according to the following equation: linguistic geography + numerical taxonomy = dialectometry. First dealt with are problems concerning the processing of raw data, the coding, scaling, measurement and compilation of the data matrix (atlas points \times atlas maps). After the description of the similarity coefficient used and the compilation of the similarity matrix it will be shown how the individual vectors of the similarity matrix can be visualized in the form of choropleth maps and three-dimensional surfaces. These visualized vectors of the similarity matrix are described as similarity maps. After consideration of the usefulness of such similarity maps for genuine linguistic geography there is a discussion of other possible fields which will be opened up as a result of intensified dialectometrical research (e.g. graph theory, network theory, game theory). This article is supplemented by 9 figures, 2 tables and an extensive bibliography.

0. Géographie linguistique + taxonomie numérique = dialectométrie

La *dialectométrie* telle qu'elle sera présentée dans les lignes qui suivent, constitue un ensemble bidisciplinaire basé sur la *géographie linguistique* d'une part et la *taxonomie* (ou *taxinomie*) *numérique* de l'autre. Alors que la méthodologie de la géographie linguistique semble être définie depuis longtemps et jouit, de ce fait, d'une certaine standardisation sur le plan international, voire mondial (cf. Dauzat, 1922; Jaberg et Jud, 1928;

Jaberg, 1936; Pop, 1950), la taxonomie numérique (autres dénominations, en français: analyse typologique, classification automatique, analyse des données, taxométrie; en anglais: numerical taxonomy, mathematical taxonomy; en allemand: numerische Taxonomie, numerische Klassifikation, etc.) est une discipline en plein développement et, paraît-il, vouée à un avenir prometteur. Elle

apporte une réponse au problème suivant: comment décomposer une population donnée d'individus ou d'objets, décrits par un ensemble de caractéristiques, en un certain nombre de sous-groupes homogènes.

Cette préoccupation est commune à de nombreuses disciplines: biologie, zoologie, archéologie, botanique, géographie, géologie, agronomie, médecine, psychiatrie, psychologie, sociologie, anthropologie, linguistique, documentation automatique, intelligence artificielle, sciences de la gestion des entreprises. (Chandon et Pinson, 1981, p. IX).

La première synthèse taxométrique à proprement parler date de 1963 (Sokal et Sneath). Elle a été remplacée entretemps par le manuel capital de Sneath et Sokal (1973). On consultera aussi avec profit les ouvrages suivants:

- en français: Bertier et Bourroche (1975), Jambu (1978), Benzécri et al. (1980), Bourroche et Saporta (1980), Chandon et Pinson (1981);
- en anglais: Jardine et Sibson (1971) (ainsi que Sokal et Sneath, 1963, et Sneath et Sokal, 1973); Anderberg (1973), Hartigan (1975), Tukey (1977);
- en allemand: Bock (1974), Sodeur (1974), Vogel (1975), Opitz (1980).

Vu les rapports apparents entre la géolinguistique et la géographie et, par conséquent, entre la dialectométrie (en tant que géolinguistique quantitative) et la 'Nouvelle Géographie' (ou géographie quantitative), tout dialectométricien aura intérêt à en suivre les progrès et à s'inspirer

Hans Goebel est Privatdozent à l'Institut für Romanistik de l'Université de Ratisbonne (Regensburg).

des méthodes qui y sont utilisées ; cf. à ce sujet les orientations générales de Ablér et al. (1977), French et Racine (1971) et Claval (1977), de même que les études (voire introductions) méthodiques de Bahrenberg et Giese (1975), Béguin (1979), Fischer (1980), Hammond et McCullagh (1975), Racine et Reymond (1973), et Smith (1975).

Alors que beaucoup de sciences tant exactes que humaines se sont prévaluées jusqu'alors de l'appui de la taxométrie, la linguistique n'en a fait qu'un usage fort restreint. En ce qui concerne l'utilisation de méthodes taxométriques pour l'analyse de données géolinguistiques, je renvoie aux contributions de Reed et Spicer (1952), Cowan (1964), Houck (1969) et Rensch (1970), parmi lesquels Reed et Spicer, et Cowan méritent une mention particulière vu la précision méthodologique avec laquelle ils envisagent les problèmes de la classification de données dialectales.

La dialectométrie à proprement parler a pour pères fondateurs les linguistes français H. Guiter (1973, 1979) et J. Séguy (1971, 1973a et b) ; c'est d'ailleurs J. Séguy (1973a, p. 1) qui a forgé le néologisme de 'dialectométrie'.

Parmi les contributions récentes situées à cheval entre la linguistique et la classification numérique, je cite encore Altmann et Lehfeldt (1973) (première application explicite de la taxométrie à la linguistique en général), Shaw (1974), Fronzaroli (1975) et Viereck (1980).

Quant à mes propres travaux dialectométriques je ne renvoie qu'aux articles parus entre 1976 et 1978 et à une présentation sommaire de date récente (1981).

1. De l'atlas linguistique à la matrice de données

En classification numérique, le processus taxométrique commence par le choix des *objets* (individus ou éléments) à classer, et par le choix des *attributs* (caractéristiques ou variables) les caractérisant. Pour le dialectométricien, ce double choix est facile : les objets correspondent aux points d'enquête d'un atlas linguistique, et les attributs aux données dialectales répertoriées dans l'atlas linguistique à côté des points d'enquête. L'étape suivante consiste dans l'application d'une *mesure adéquate* pour tirer de la bigarrure prime-

Matrice de données

		Objets (points d'atlas)					
		1	2	3	4	5	6
Attributs (cartes d'atlas)	4	j	j	j	j	j	j
	3	i	h	h	×	i	i
	2	e	f	f	e	g	g
	1	a	b	b	c	a	d

× case vide

Fig. 1. Matrice de données constituée de 6 objets (points d'atlas) et 4 attributs qualitatifs (cartes d'atlas). En gris : l'équivalent matriciel d'une carte d'atlas.

sautière des données d'atlas, une description codée appelée *matrice* (ou *tableau*) de données. C'est un tableau à double entrée comprenant n objets (ou points d'atlas) et p attributs (ou cartes d'atlas). Voir la Fig. 1.

Le codage effectué par le dialectométricien revient au découpage des données brutes de l'atlas linguistique dépouillé en un certain nombre d'unités taxatoires (appelées 'taxats'). Ce découpage se fait sur l'échelle de mesure la plus simple (échelle nominale) et en fonction d'un critère linguistique discriminatoire préalablement choisi (différences lexicales et morpho-syntaxiques ; cf. 1.1).

1.1. Structures de la matrice de données utilisée dans ce travail

- Données dialectales brutes : Vol. I, II et IV de l'atlas italo-suisse AIS.
- Objets (points d'atlas) : 247 points-AIS + 4 points artificiels :
 - p. 154 - seconde enquête faite à Turin (cf. Jaberg et Jud, 1928, pp. 56-57) ;
 - p. 262 - seconde enquête faite à Milan (cf. Jaberg et Jud, 1928, p. 69) ;
 - p. 524 - seconde enquête faite à Florence (cf. Jaberg et Jud, 1928, pp. 100-101) ;
 - p. 999 - équivalent de l'italien standard tel qu'il figure dans les en-têtes des planches de l'AIS.

- Critère discriminatoire utilisé pour l'isolement des taxats: différences lexicales et morpho-syntaxiques².
- Attributs (cartes codées): 696³.
- Nombre des taxats consignés dans la matrice de données: 4836.
- Nombre de taxats/carte codée: 6,9483.

1.2. Pondération des attributs

Tous les attributs ont été soumis au même traitement taxométrique. En ceci, mon analyse obéit au principe taxométrique de l'adansonisme (ou de l'isocratie); cf. Sneath et Sokal (1973, p. 5).

1.3. Cases vides

Les lacunes que contiennent tous les atlas linguistiques (et ceci pour des raisons que les dialectologues connaissent bien) causent, lors du codage des données d'atlas, l'apparition de cases vides dans les colonnes (ou lignes) de la matrice de données. L'homogénéité de l'exploitation taxométrique de la matrice de données s'en trouve plus ou moins affectée, si bien que l'on doit en tenir compte lors de l'interprétation des indices de similarité consignés ou bien dans la matrice de similarité ou bien sur les cartes de similarité.

Pour l'utilisation d'un questionnaire réduit (Qr, ca. 800 questions) lors de la collecte des données de l'AIS, cf. Jaberg et Jud, 1928, p. 38 et pp. 175-176. Soit dit en passant que le Qr a été utilisé exclusivement dans des villes importantes où, pour des raisons évidentes (ignorance du vocabulaire rural), le recours au questionnaire normal (Qn, ca. 2000 questions) n'aurait pas eu de sens.

En ce qui concerne le traitement taxométrique des cases vides nous avons renoncé à toute pondération artificielle des données incomplètes étant donné le fait que la présentation des résultats de notre analyse se fait exclusivement dans le contexte spatial des données originales. Or, l'agencement spatial des résultats dialectométriques permet de dépister rapidement, parmi les valeurs numériques présentées en synopse spatiale, celles qui s'insèrent mal dans le décalage des autres. Voir la remarque 'punti con corpo ridotto' sur la carte choroplèthe (Fig. 7) et 5.1 (infra).

Vu la quantité remarquable des vecteurs-

attributs considérés (696) il est possible d'en exclure ceux qui contiennent des données lacunaires. L'on obtient ainsi une matrice de données 'réduite' (251 points d'atlas fois ca. 250 cartes de travail) dont la fiabilité taxométrique n'est d'ailleurs que très légèrement en retrait sur celle de la matrice complète. Cf. Goebel (1977b, pp. 350-359). Pour le traitement taxométrique de données manquantes (cases vides), cf. Sneath et Sokal, 1973, pp. 178-181.

2. Mesure de similarité et matrice de similarité

Comme l'intérêt majeur de la géolinguistique traditionnelle est (et était de tout temps) de préciser les rapports qui existent entre les différents points d'enquête d'un atlas linguistique, la procédure taxométrique se poursuit par la mesure de la similarité (similitude, proximité, ressemblance) réciproque de tous les points d'atlas (comparaison des points d'atlas par paires).

Comme le nombre des indices de similarité susceptibles d'être utilisés est considérable et que, de surcroît, chacun d'eux répond à des besoins classificatoires très précis, il est de toute première urgence que le dialectométricien règle son choix sur les acquis conceptuels de la géolinguistique traditionnelle en matière de similarité entre points d'atlas. En voici une proposition déjà fort ancienne qui reflète fort bien l'opinion commune des dialectologues encore de nos jours: "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 de Gros, 1889, p. 63).

L'indice qui y correspond le mieux se définit comme suit:

$$s = 100 \left(\frac{\text{nombre des co-identités d'attributs par paire d'objets}}{\text{nombre des co-présences d'attributs par paire d'objets}} \right)$$

Cet indice peut prendre toutes les valeurs (scores) entre 0 et 100%. Du point de vue algébrique il ressemble au 'Simple Matching Coefficient' de la

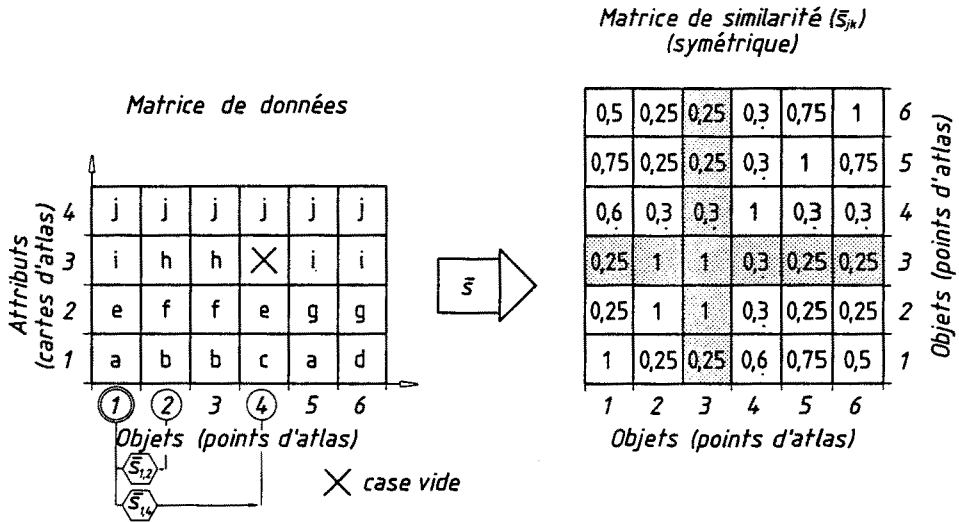


Fig. 2. Génération de la matrice de similarité (\bar{s}_{jk}) à partir de la matrice des données par l'application d'un indice de similarité \bar{s} . En gris: L'équivalent matriciel d'une carte de similarité.

taxonomie numérique (cf. Sneath et Sokal, 1973, p. 132; Chandon et Pinson, 1981, p. 74; Bock, 1974, p. 51). Je l'appelle 'Indice Relatif d'Identité' (IRI) dans mes contributions dialectométriques en français (Goebel, 1976–1978, 1981) et 'Relativer Identitätswert' (RIW) dans celles en langue allemande (Goebel, 1977a et b).

L'indice utilisé dans la Fig. 2 (\bar{s}) se distingue de l'IRI comme suit:

$$100 \cdot \bar{s} = \text{IRI}.$$

Vérification de $\bar{s}_{1,2}$ (voir la Fig. 2):

- nombre des co-identités d'attributs: 1 (par le taxat *j*)
- nombre des co-présences d'attributs: 4 (par les paires de taxats *a/b*, *e/f*, *i/h*, *j/j*)
- $\bar{s}_{1,2} = 1/4 = 0,25$.

Vérification de $\bar{s}_{1,4}$ (voir la Fig. 2):

- nombre des co-identités d'attributs: 2 (par les taxats *e* et *j*)
- nombre des co-présences d'attributs: 3 (par les paires de taxats *a/c*, *e/e*, *j/j*)
- Remarquer que la case vide dans le vecteur du point 4 cause la disparition d'un co-présence d'attributs.
- $\bar{s}_{1,4} = 2/3 = 0,6$.

Comme, sur n points comparés par paires, le nombre des indices calculables est de n^2 , les

dimensions de la matrice de similarité (s_{jk}) sont de n^2 et, partant, symétriques. En plus, il faut se rendre compte que chacun des n^2 scores est calculé deux fois, vu la symétrie algébrique de l'indice en question: $s_{jk} = s_{kj}$, et que, finalement, les scores répertoriés le long de la diagonale sont tous de 1 (ou de 100%). Il en résulte que le nombre des valeurs réellement utilisables à des fins taxométriques est de

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

En principe, tous les procédés qu'offre l'analyse multivariée⁴ peuvent être utilisés pour le traitement consécutif de la matrice de similarité. Il est bon cependant de régler l'exploitation taxométrique de la matrice de similarité sur les contraintes conceptuelles que crée la contiguïté spatiale des points d'atlas⁵, et de commencer par faire une analyse partitionnelle de l'information numérique consignée dans les différents vecteurs (colonnes et/ou lignes) de la matrice de similarité.

Chaque vecteur de la matrice de similarité comprend n valeurs, dont une égale à 1 et $n - 1$ valeurs sont réparties entre 0 et 1. L'ensemble des $n - 1$ valeurs sera appelé dorénavant 'distribution de similarité'. Voir la Fig. 2 où la distribution de similarité du vecteur no. 3 (en gris) contient les valeurs $\{0, 25/1/0, 3/0, 25/0, 25\}$. Cette distribution

reflète la similitude respective (\bar{s}) entre le point de référence no. 3 et les points comparés 1, 2, 4, 5 et 6. De la Fig. 2 il apparaît en outre que la matrice de similarité (\bar{s}_{jk}) contient en tout six distributions de similarité différentes, donc une pour chacun des six points d'atlas.

Comme le réseau d'atlas examiné se compose de 251 ($=n$) points-AIS, il va de soi d'un côté que la distribution de similarité respective comprendra 250 ($=n-1$) scores situés au-dessous de 100%, et de l'autre que la matrice de similarité contiendra 251 ($=n$) distributions de similarité différentes. Par la suite il ne sera présenté qu'une seule de ces 251 distributions de similarité tant sous forme numérique que sous forme visualisée.

3. Notice servant à l'intelligence de la carte choroplèthe (Fig. 7) et de la surface statistique lissée (Fig. 8)

3.1. Algorithme d'intervallisation MINMWMAX (voir la Fig. 7)

Parmi les nombreuses méthodes utilisées en cartographie thématique pour la visualisation de différences numériques au sein d'un réseau de données ponctuelles, une des meilleures est, sans aucun doute, la méthode de la carte choroplèthe. Le principe consiste à traduire une variation (quasiment continue) d'un grand nombre de valeurs numériques en une variation (discrète) d'un nombre fort limité de paliers visuels ordonnés (teintes, couleurs, hachures, etc.). La mise en carte des valeurs numériques présuppose le choix (et l'application consécutive) d'un algorithme de transposition qui, en outre qu'il assure la répartition des valeurs originales en un nombre limité d'intervalles, devra aussi opérer un agencement spatial 'adéquat' des données ponctuelles; pour plus de détails, cf. Dickinson, 1973, p. 83 sqq., ainsi que les contributions de Jenks (1963), Jenks et Coulson (1963), Kishimoto (1972) et de Wonka (1980).

Au cours de mes travaux dialectométriques j'ai expérimenté toute une série d'algorithmes d'intervallisation, parmi lesquels l'algorithme présenté par la suite (sigle: MINMWMAX) répond le

mieux à mes intentions classificatoires⁶. En voici les propriétés statistiques :

- nombre des intervalles: 6; 3 au-dessus et 3 au-dessous de la moyenne arithmétique de la distribution de similarité à visualiser;
- limites d'intervalle: division des écarts entre le minimum et la moyenne, et entre le maximum et la moyenne respectivement par trois.

Exemple (voir le Tableau 1)

Moyenne arithmétique	65,173
Minimum	-45,072
	$\frac{20,101}{3} = 6,7003$
Taille des intervalles 1, 2 et 3	: 6,7003.
Maximum	89,286
Moyenne arithmétique	-65,173
	$\frac{24,113}{3} = 8,0376$
Taille des intervalles 4, 5 et 6	: 8,0376.

L'algorithme MINMWMAX a l'avantage de créer (par l'insertion de certaines valeurs dans les intervalles 5 et 6) des agrégats spatiaux fort homogènes, qui épousent de très près les contours de ce que la dialectologie traditionnelle appelle communément les 'aires dialectales'. Dialectométrie et aréologie traditionnelle peuvent donc être fort bien réconciliées.

En termes de taxométrie, l'application d'un algorithme d'intervallisation équivaut à une analyse partitionnelle en classes disjonctives.

3.2. Histogramme (voir la Fig. 7)

L'histogramme figurant en bas de la Fig. 7 permet de contrôler rapidement quelques-uns des paramètres caractéristiques (position, dispersion, symétrie, etc.) de la distribution de similarité visualisée. Pour la construction géométrique de l'histogramme on peut se servir des données numériques fournies par l'algorithme MINMWMAX tout en divisant les écarts entre le minimum et la moyenne, voire entre le maximum et la moyenne non pas par trois, mais par six. Au lieu de six intervalles on en obtient ainsi douze équivalents à autant de colonnes de l'histogramme (voir le Tableau 2). Pour la construction de chacune des 12 (j) colonnes de l'histogramme on porte:

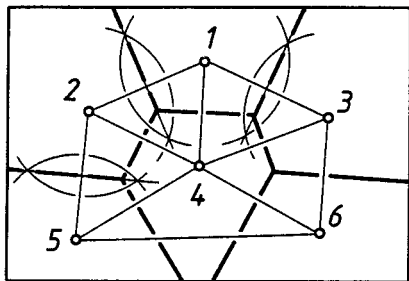


Fig. 3. Construction de polygones de Thiessen dans un réseau de 6 points d'atlas.

- en abscisse: la taille (d_j) de l'intervalle considéré,
- en ordonnée: le quotient (p_j/d_j) entre la fréquence relative des points d'atlas affectés à l'intervalle considéré (p_j)⁷ et la taille de cet intervalle (d_j).

Pour les données numériques utilisées voir le Tableau 2.

3.3. Construction des polygones de Thiessen

Pour assurer la transposition des données ponctuelles du réseau examiné (représentation discrète) en un pavage disjonctif de surfaces géométriques (représentation quasi-continue), il est nécessaire de soumettre le fond de carte à un traitement géométrique particulier. Parmi les nombreuses possibilités qu'offre la cartographie thématique en la matière, la plus adéquate à nos propos est le découpage du réseau examiné en polygones-Thiessen (Fig. 3).

Les étapes constructives sont les suivantes:

- (a) triangulation des points du réseau examiné,
- (b) construction des médiatrices pour chacun des segments de droite reliant deux points du réseau,
- (c) prolongement des médiatrices jusqu'à ce qu'elles se rencontrent;
- (d) Les points de rencontre des médiatrices constituent les angles des polygones de Thiessen.

Références bibliographiques: Thiessen (1911), Rhynsburger (1973), Pudlatz (1977), et Brassel et Reif (1979).

3.4. Surface statistique lissée (Fig. 8)

A côté de la carte choroplèthe il y a - en cartographie thématique - bon nombre d'autres pro-

cédés de visualisation dont la valeur heuristique en dialectométrie doit être jugée avec circonspection par le dialectométricien. En passant de l'expression bidimensionnelle qu'offre la carte choroplèthe, à la perspective tridimensionnelle (2 dimensions de l'agencement spatial des points d'atlas + 1 dimension relative aux scores de l'IRI traduits en surélévations du relief), l'on obtient un relief stéréographique dont les discontinuités ont été lissées par voie électronique et selon des procédés mathématiques plus ou moins compliqués. Il convient cependant de ne pas surestimer l'importance ou la valeur cartographique de telles surfaces statistiques lissées. Quant à l'expression des détails et à la représentation exacte des relations numériques, la carte choroplèthe l'emporte de loin sur le stéréogramme. Ce n'est que dans des cas particuliers (distribution des masses en bordure d'image; cf. Goebel, 1981, Figs. 10, 18, 25 et 28) que le recours à l'effet stéréographique est pleinement justifié. Pour une discussion approfondie de ces problèmes, cf. Monmonier, 1978, et Peucker, 1977.

4. Notice servant à l'intelligence de la carte numérique (Fig. 9)

La carte numérique contient une en-tête (reproduite en haut de page, à gauche et à droite) et des indications numériques relatives aux 251 points d'atlas considérés (Fig. 4).

Les indications contenues dans l'en-tête doivent être interprétées de la manière suivante:

PUNKT:172 point de référence 172 (AIS)
 MW moyenne arithmétique (all. Mittelwert, angl. mean)
 écart type (all. Standardabweichung, angl. standard deviation)
 N-KART nombre des attributs disponibles (taxats) dans le vecteur du point 172 de la matrice de données. Comme tous les vecteurs de la matrice de données disposent de 696 cases (cf. Section 1.1) et que de ces 696 cases 693 seulement détiennent des attributs analysables (taxats), il en résulte que ce vecteur dispose de 3 cases vides.

*KARTE 999-TOT PUNKT: 172
 MW = 65.173 S = 7.141 N-KART = 693

MINMWMAX	45.072	51.773	58.473	65.173	73.210	81.248	89.286
MEDMW	45.072	61.939	63.821	65.262	67.128	71.925	89.286
MED	45.072	61.739	63.411	64.689	66.281	70.954	89.286

	11-3	7-1	21-101	56-25	7-12	4-2
	24-24	24-24	24-24	17-18	18-18	18-17
	20-21	21-21	21-21	21-21	21-21	21-20

Fig. 4. En-tête de la carte numérique de la distribution de similarité relative au point de référence 172. Voir aussi les Tableaux 1 et 2, et la Fig. 9.

Tableau 1

Schéma de lecture relatif à l'algorithme d'intervallisation MINMWMAX (à six intervalles).
 Voir aussi les Figs. 4 et 9, et le Tableau 2

Intervalle	de	à	Taille de l'intervalle	Points d'enquête/intervalle
1	45,072 ^a	51,773	6,7003	11 + 3 = 14
2	> 51,773	58,473	6,7003	7 + 1 = 8
3	> 58,473	65,173 ^b	6,7003	21 + 101 = 122
4	> 65,173	73,210	8,0376	56 + 25 = 81
5	> 73,210	81,248	8,0376	7 + 12 = 19
6	> 81,248	89,286 ^c	8,0376	4 + 2 = 6

^aMinimum.

^bMoyenne arithmétique (valeur continue).

^cMaximum.

Tableau 2

Schéma de lecture relatif à l'algorithme d'intervallisation MINMWMAX (à douze intervalles).
 Voir aussi les Figs. 4 et 9, et le Tableau 1

Intervalle	de	à	Taille de l'intervalle	Points d'enquête/intervalle
1	45,072 ^a	48,422	3,35016	11
2	> 48,422	51,773	3,35016	3
⋮	⋮	⋮	⋮	⋮
6	> 61,823	65,173 ^b	3,35016	101
7	> 65,173	69,192	4,01883	56
⋮	⋮	⋮	⋮	⋮
11	> 81,248	85,267	4,01883	4
12	> 85,267	89,286 ^c	4,01883	2

^aMinimum.

^bMoyenne arithmétique (valeur continue).

^cMaximum.

Les six lignes qui suivent fournissent des indications relatives à trois algorithmes d'intervallisation (MINMWMAX, MEDMW, MED). Le Tableau 1 permet d'en extraire les valeurs appartenant à MINMWMAX (à 6 intervalles).

Pour l'algorithme MINMWMAX à 12 inter-

valles, utilisé pour la construction de l'histogramme (cf. Section 3.2 et voir la Fig. 7), voir le Tableau 2.

Quant aux données numériques reproduites à l'emplacement des différents points d'atlas, elles doivent être interprétées comme suit (Fig. 5):

(1) indicatif du point d'atlas (176);

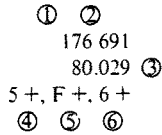


Fig. 5. Données numériques relatives à un point d'atlas.

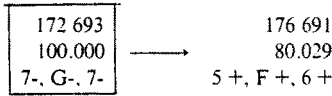


Fig. 6. Mesure de similarité (selon IRI) entre les points-AIS 172 (point de référence) et 176 (point comparé). Voir aussi la Fig. 5.

(2) nombre des attributs disponibles (taxats) dans le vecteur du point 176 de la matrice de données;

(3) Indice Relatif d'Identité (IRI) entre le point de référence (172) et le point comparé (176);

(4) classement par intervalles selon l'algorithme d'intervallisation MINMWMAX:

- classement dans l'intervalle 5 selon MINMWMAX à 6 intervalles;
- classement dans l'intervalle 5+ (=intervalle 10) selon MINMWMAX à 12 intervalles;

(5) classement par intervalles selon l'algorithme d'intervallisation MEDMW (pas traité dans cet article)⁶;

(6) classement par intervalles selon l'algorithme d'intervallisation MED (pas traité dans cet article)⁶.

Pour pouvoir apprécier correctement chacun des 250 scores-IRI contenus dans la Fig. 9, l'on doit toujours consulter simultanément les indications numériques du point de référence (en l'occurrence celles du point-AIS 172) et du point comparé; voir la Fig. 6.

Il s'agit d'une comparaison numérique (selon IRI) entre les vecteurs des points 172 et 176 de la matrice de données. Comme le nombre des attributs (taxats) calculables dans le vecteur du point 172 est de 693, et de 691 dans celui du point 176, il y a, entre les vecteurs 172 et 176, 691 co-présences d'attributs. Parmi ces 691 co-présences d'attributs il y a 80,029% – soit 553 (=691 · 0,80029) – de co-identités d'attributs. La similarité taxométrique entre les points 172 et 176 ($IRI_{172,176}$) est donc de 80,029%.

5. Carte de similarité relative au point de référence 172 (Villafalletto, province de Coni/Cuneo; Piémont/Piemonte)

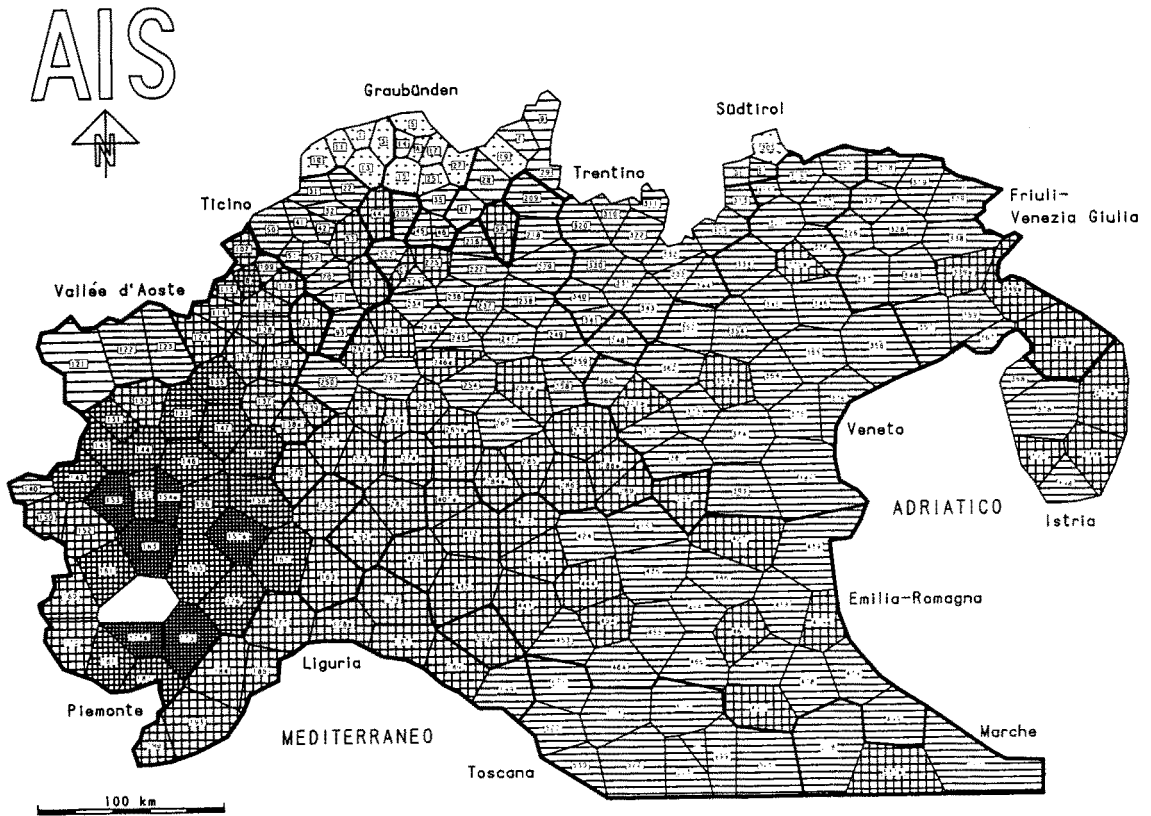
Voir les Figs. 7, 8 et 9. Quant au point-AIS 172, cf. Jaberg et Jud, 1928, p. 59.

5.1. Interprétation géolinguistique

L'agencement des plages choroplèthes de la Fig. 7 reflète la distribution spatiale des similarités typologiques des 250 points-AIS par rapport au point de référence 172. Au premier coup d'oeil l'on constate que les polygones-Thiessen classés dans les intervalles 5 et 6 forment un espace cohérent dont les contours coïncident largement avec ceux du Piémont linguistique⁸ tel qu'il nous est décrit par la dialectologie classique (cf. p. ex. Simon, 1967). De la Fig. 7 il ressort en outre que le bloc piémontais (intervalles 5 et 6) repose – au sein du réseau examiné – sur des fondements typologiques géographiquement bien définis (polygones dans l'intervalle 4), et qui englobent le Nord du Piémont politique (province de Novare/Novara), la Lombardie méridionale (province de Pavie/Pavia), l'Emilie occidentale (province de Plaisance/Piacenza) et les vallées occitano-provençales du versant oriental de l'arc alpin. Les antipodes typologiques (intervalles 1 et 2) du bloc piémontais se retrouvent en Vallée d'Aoste (où l'on parle des dialectes franco-provençaux) et dans des régions rhétoromanes (Grisons et Tyrol du Sud).

On remarquera en outre qu'un certain nombre de polygones classés dans l'intervalle 4 et marqués d'un astérisque ('punti con corpo ridotto'), ne s'insèrent que très mal dans le modèle du décalage spatial des indices de similarité (IRI). Ces polygones correspondent à des points d'atlas dont les vecteurs matriciels contiennent beaucoup de cases vides (cf. 1.3). Il en résulte des distorsions considérables des scores numériques respectifs (IRI) qu'il est donc bien d'exclure de l'interprétation globale de la carte choroplèthe.

Voir aussi la Fig. 8 (surface statistique lissée) qui, bien que d'une façon plus estompée et, partant, moins précise, visualise l'agencement spatial des indices de similarité (IRI) d'une manière tridimensionnelle. Remarquer que l'élévation



LEX - ITALIEN

999-T0T-172

Punkte mit Buchstaben

- A = 10
- B = 11
- C = 12
- D = 13
- E = 14
- F = 15



Kartographie: W.-D. Rose

Choroplethenkarte
MINNMAX 6-fach

1		45.073 - 51.773
2		58.473
3		65.173
4		73.211
5		81.248
6		89.286

Häufigkeitsverteilung
MINNMAX 12-fach

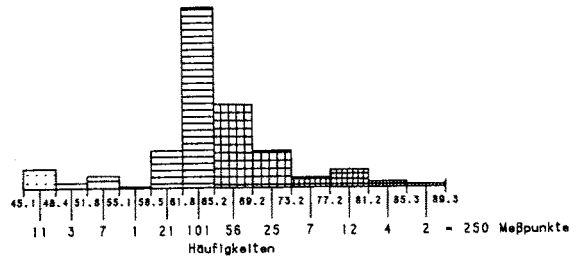


Fig. 7. Carte choroplèthe de la distribution de similarité relative au point de référence 172 (Villafalletto, province de Coni/Cuneo; Piémont/Piemonte). Voir aussi les Figs. 8 et 9.

Choroplethenkarte – carte choroplèthe; Häufigkeitsverteilung – distribution de fréquence; Häufigkeiten – fréquences (absolues); Meßpunkte – points d'atlas; 6-fach, 12-fach – à 6 intervalles, à 12 intervalles.

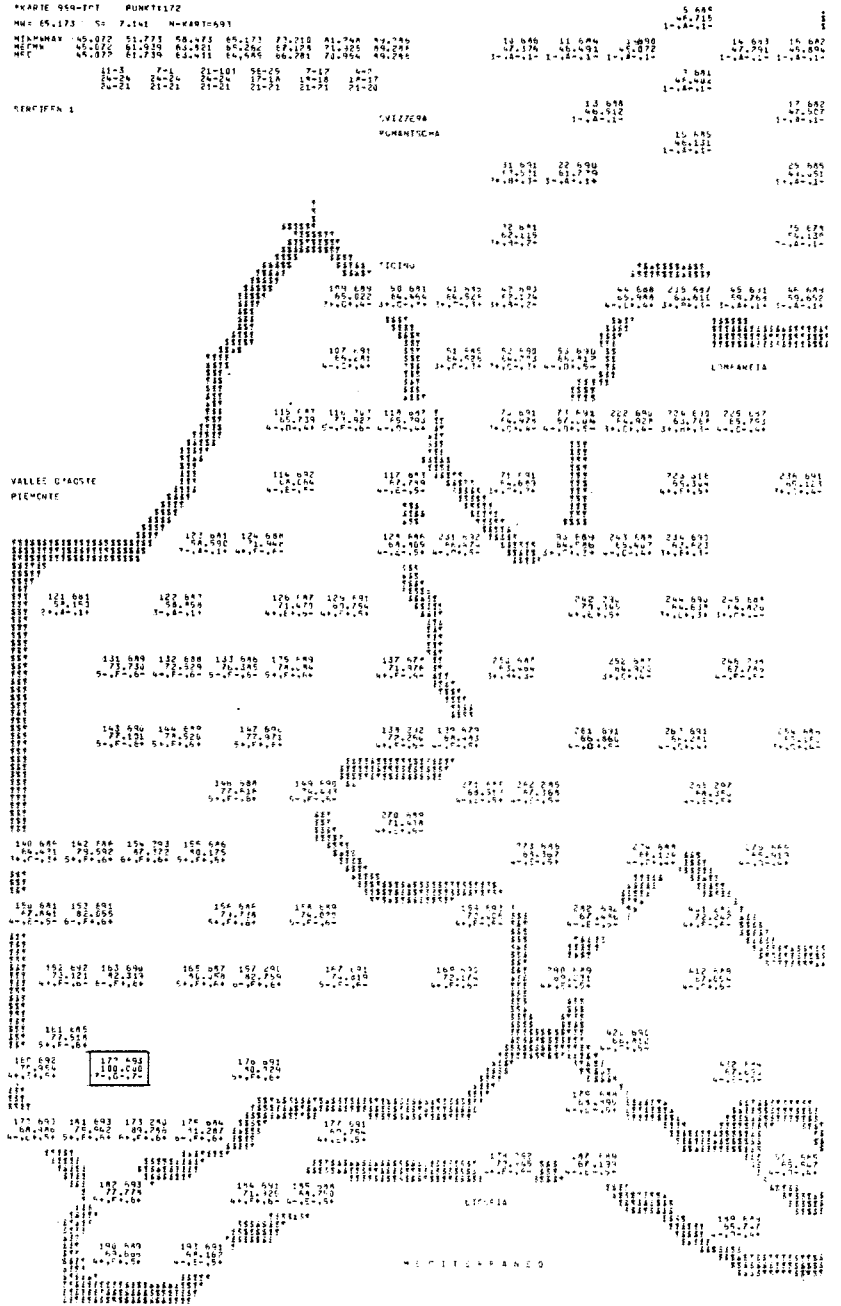


Fig. 9. Carte numérique de la distribution de similarité relative au point de référence 172 (Villafalletto, province de Coni/Cuneo; Piémont/Piemonte). Voir aussi les Figs. 7 et 8 ainsi que les Tableaux 1 et 2.

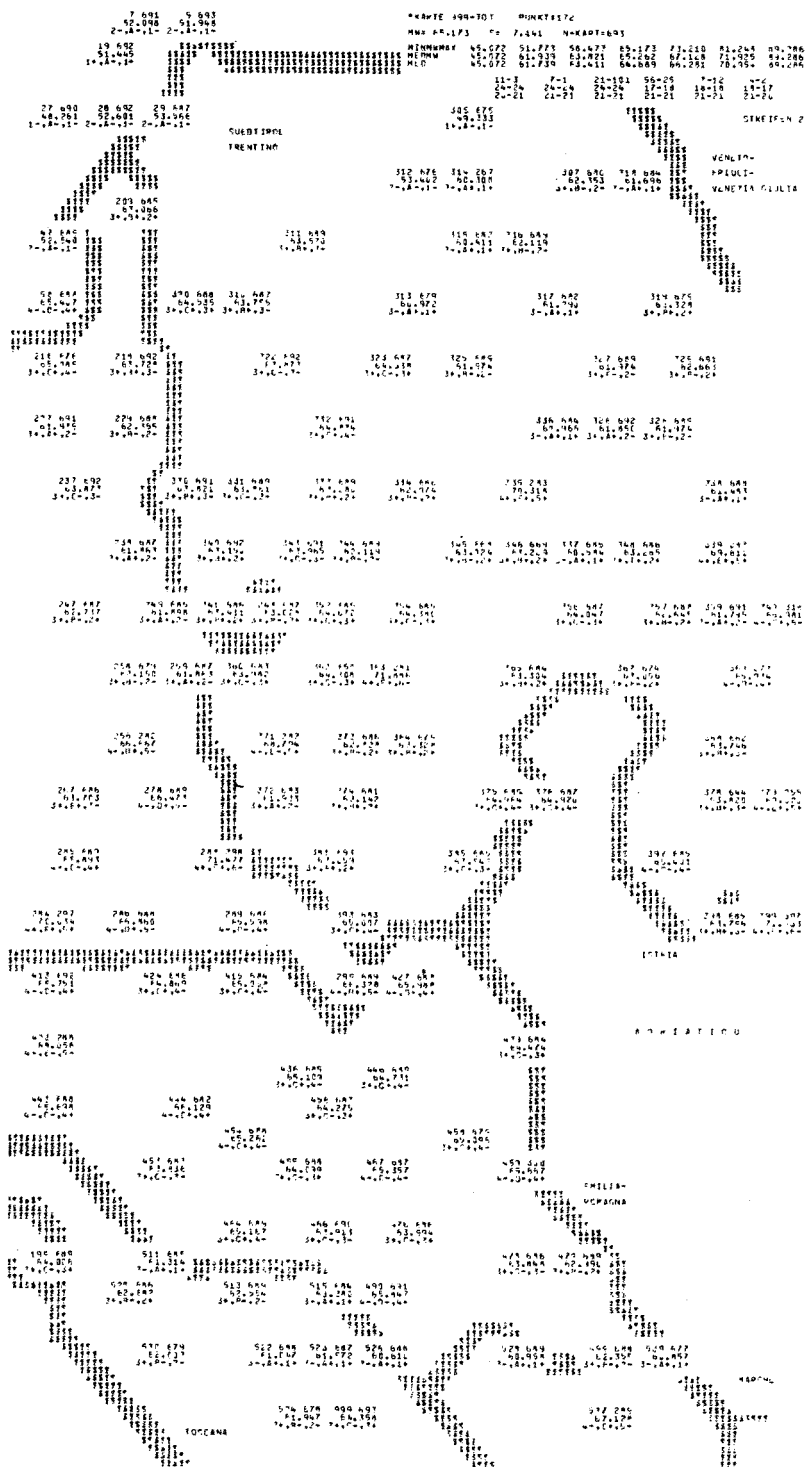


Fig. 9 (continued).

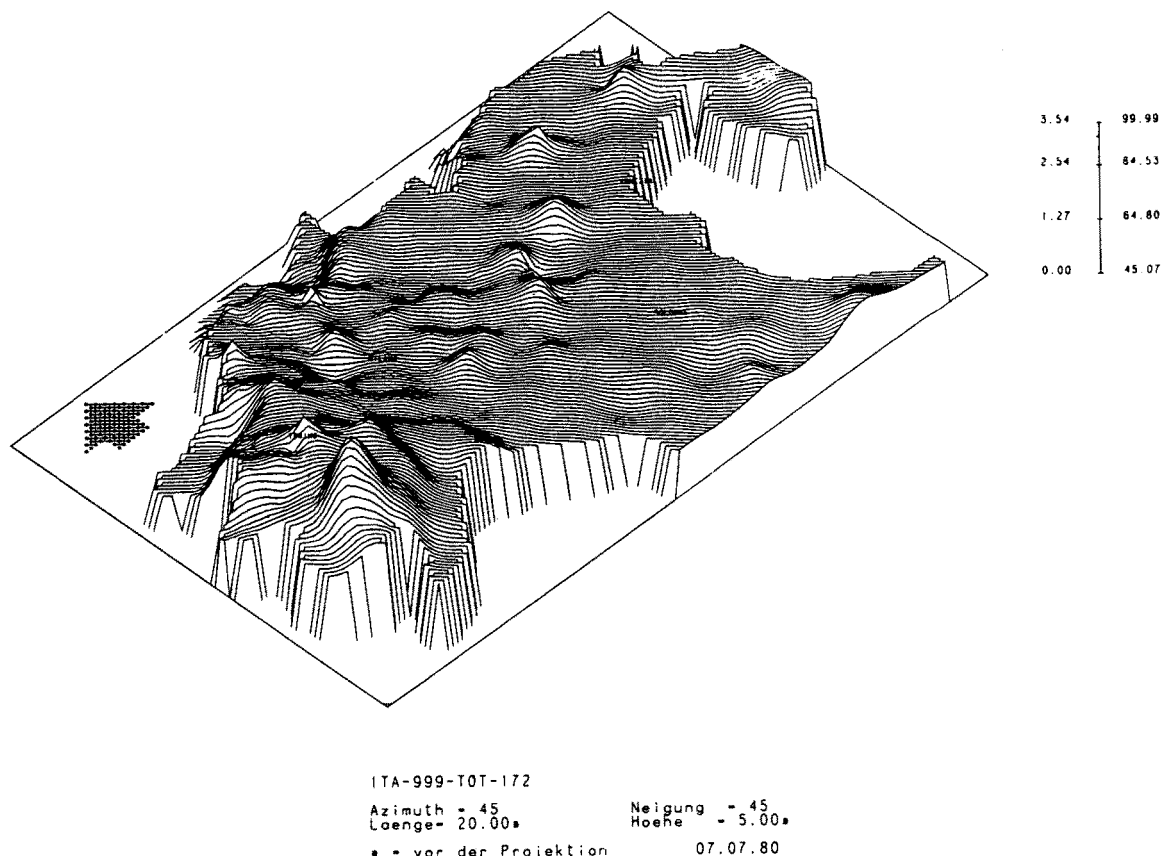


Fig. 8. Surface statistique lissée de la distribution de similarité relative au point de référence 172 (Villafalletto, province de Coni/Cuneo; Piémont/Piemonte). Voir aussi les Figs. 7 et 9.

Länge - longueur; Neigung - inclinaison, élévation; Höhe - altitude; vor der Projektion - avant la projection.

majeure du relief correspond à la valeur répertoriée au point de référence 172 ($IRI_{172,172} = 100\%$) et que la perspective cavalière se fait à partir du sud-ouest.

5.2. Dialectométrie et cartographie dynamique

Quand on déplace progressivement le point de référence à l'intérieur d'un domaine dialectal typologiquement fort cohérent (p. ex. au coeur du Piémont linguistique), l'on obtient des profils de proximité fort similaires (et qui ressemblent donc beaucoup à celui de la Fig. 7). Des changements majeurs du type iconique de la carte de similarité n'interviennent qu'après le déplacement du point de référence par-delà une barrière linguistique (= faisceau d'isoglosses) d'une certaine impor-

tance. Plus la barrière linguistique franchie est importante, plus l'agencement des plages choroplèthes change. Alors que, lors d'un déplacement du point de référence de 172 à 163 par exemple, les changements intervenus n'affectent que quelques isochores qui passent de l'intervalle 6 à 5 (ou vice versa), le type iconique est totalement inversé lors d'un déplacement du point de référence de 172 à, mettons, 10 (Grisons).

On peut se servir de ces alternances iconiques pour produire un effet cinématographique très éloquent. Cet effet repose sur la projection consécutive de plusieurs profils de similarité alignés le long d'un parcours d'épreuve préalablement choisi, et partant, plus ou moins différents l'un de l'autre. L'effet cinématographique en question réussit particulièrement bien quand la projection

consécutives des profils de similarité se fait en fondu enchaîné à l'aide de deux projecteurs de diaporama. L'impression de dynamisme cartographique que perçoit le spectateur est alors parfaite. J'ai présenté à plusieurs reprises de telles projections diaporamiques (cf. Goebel, 1977b, p. 335 et 1978b). Pour plus de détails voir les contributions exemplaires de Gould et Bassett (1967) et de Rase (1974).

5.3. *Interprétation extralinguistique et perspectives de recherche*

Tout réseau géolinguistique peut être comparé à n'importe quel autre réseau interactif. Si l'on remplace la notion linguistique de *parenté* (ou *affinité*) étymologique (toute parenté linguistique revient, en dernière analyse, à un rapport extralinguistique) par la notion géographique d'*interaction* (ou par la notion sociologique de *communication*), on peut dire, en regardant la Fig. 7, que les flux (ou flots) d'interaction (de transport, de trafic, d'échange ou de communication) qui émanent du point de référence 172 (ou qui convergent vers lui), accusent des structures très précises, susceptibles d'être interprétées aussi sous l'angle de la géographie, de la sociologie, de l'économie, etc. Il y a là un carrefour interdisciplinaire du plus haut intérêt.

Pour pousser plus avant nos connaissances de ce réseau interactif, il est nécessaire non seulement de comparer tous les 251 profils de similarité entre eux, mais aussi d'élargir l'exploitation taxométrique de la matrice de similarité. Soit dit en passant, des possibilités taxométriques très prometteuses s'offrent à partir d'une considération synoptique des paramètres caractéristiques (maximum, moyenne, écart type, coefficient de symétrie de Fisher, etc.) des distributions de similarité (cf. Goebel, 1981 passim).

L'utilisation des procédés taxométriques non seulement à des fins purement classificatoires mais aussi à des fins exploratoires (et visant, par là, à instaurer un circuit rétroactif entre la genèse et la vérification ou falsification d'hypothèses ou de théories nouvelles) amène le dialectométricien à envisager les réseaux dialectaux sous l'angle de la théorie des graphes (cf. Berge, 1967; Flament,

1968; Kaufmann, 1968, Sneath et Sokal, 1973, pp. 253-256), de la théorie des réseaux de transport (cf. Haggett et Chorley, 1969; Boissevain et Mitchell, 1973) et de la théorie des jeux (cf. Rapoport, 1970a et b). S'éloignant de l'orientation carrément empirique qu'elle avait au début, la dialectométrie gagne, par l'intermédiaire de ces disciplines mathématiques, rapidement en profondeur théorique et en intérêt pluridisciplinaire. La tâche, à peine commencée, devrait être poursuivie dans un esprit d'interdisciplinarité qui, lui, permettrait à la dialectologie de se défaire définitivement d'une tare désormais séculaire et caractérisée fort bien par le slogan anglais suivant : *data rich and theory poor*.

Notes

- [1] Nos travaux dialectométriques ont bénéficié de l'aide et du soutien des fondations de recherche suivantes :

Hochschuljubiläumsstiftung der Stadt Wien (Vienne);
Jubiläumsfonds der Österreichischen Nationalbank zur Förderung der Forschungs- und Lehraufgaben der Wissenschaft (Vienne);
Fonds zur Förderung der wissenschaftlichen Forschung in Österreich (Vienne);
Deutsche Forschungsgemeinschaft (Bonn).

Que les organismes mentionnés ci-dessus reçoivent ici l'expression de ma profonde reconnaissance. Mes remerciements vont également aux MM. Selberherr (calcul électronique), Rase (cartographie électronique) et Pudlatz (polygones de Thiessen) pour leur collaboration efficace et compétente, ainsi qu'à Mme F. Schmittlein (Summary) et à M.L. Karlson (toilette stylistique de mon texte français).

- [2] Le codage descriptif sur échelle nominale correspond à la 'typisation' des données dialectales brutes, telle qu'elle a été pratiquée de tout temps par les géolinguistes, les lexicologues et onomasiologues; cf. à ce sujet le passage suivant de K. Jaberg (1947, p. 6):

"Le géographe-linguiste, pour ne pas se perdre dans le détail, se voit obligé de 'typiser' les nombreuses variantes d'un mot qui forment le tapis bariolé d'une carte linguistique. Il crée ainsi des mots-types, dont il symbolise les aires par des couleurs ou par des hachures. Cela lui permet de voir en étendue ce que l'étude étymologique voit en profondeur. Le mot-type peut être une base étymologique, un mot littéraire, un mot régional, voire un mot inexistant qu'on a forgé arbitrairement - ce qui importe, c'est qu'il résume les caractères essentiels d'un faisceau de formes similaires. Si l'on veut symboliser un ensemble de mots qui diffèrent par leurs éléments morphologiques, on peut se contenter d'en indiquer le radical. On procédera de

- même quand il s'agira de réunir des mots qui se rapprochent par la similitude de leur phonétisme expressif."
- [3] Il faut distinguer les cartes-AIS originales (plan des données brutes) des cartes codées (ou analysées: plan du codage, voire de la matrice de données). Cf., par exemple, la carte-AIS 703 soffro assai qui a été codée deux fois, le premier codage portant sur l'onomasologie du concept 'soffro' ('soffro, vado male, patisco, sto male, etc.'), le second sur celle du concept 'assai' ('assai, tanto, molto, bene, abbastanza, etc.'). Il en résulte deux cartes codées (équivalent à autant de vecteurs d'attributs de la matrice de données), tirées d'une seule carte-AIS.
- [4] J'ai obtenu de très bons résultats par l'application de l'analyse des clusters (méthode hiérarchique, 'complete linkage'). Cf. Anderberg (1973, pp. 138-139). Les logiciels utilisés portent les sigles suivantes:
SUBROUTINE OCLINK (perform a cluster analysis),
SUBROUTINE USTREE (print a binary tree),
et peuvent être consultés à Houston (Texas): *International mathematical and statistical libraries, Library 3*. Soit dit en passant que l'utilisation d'arborescences hiérarchiques est très répandue en matière de linguistique depuis le siècle passé quand le philologue A. Schleicher illustra le premier la parenté génétique des langues par la construction d'un arbre généalogique des langues indo-européennes. A l'époque, le recours à des schémas arborescents était chose courante sous l'impulsion de la pensée biologique de Ch. Darwin dont Schleicher était d'ailleurs un admirateur fervent.
- [5] C'est ainsi que l'on reprend, en termes taxométriques, le problème classique (et souvent discuté en dialectologie traditionnelle) de la position d'un dialecte *A* dans un ensemble d'autres dialectes {*B, C, D, E, etc.*}. Signalons encore que le même procédé (calcul d'une matrice de similarité et visualisation des données d'un vecteur de cette matrice) a déjà été utilisé par Reed et Spicer (1952, p. 357) et Cowan (1964, p. 6 sq.).
- [6] Pour les algorithmes MEDMW et MED (pas traités dans cet article), cf. Goebel, 1977b, pp. 345-346. Dans la Fig. 9 le lecteur intéressé trouvera d'ailleurs aussi les résultats des intervallisations selon les algorithmes MEDMW et MED. Voir les Fig. 4 et 5 qui permettent un décodage rapide des valeurs répertoriées dans la Fig. 9 aussi pour les algorithmes MEDMW et MED. La valeur centrale, c'est-à-dire celle qui équivaut au seuil supérieur de l'intervalle 3, est – pour l'algorithme MEDMW – l'indice de similarité (*IRI*) le plus proche de la moyenne arithmétique, alors que – pour l'algorithme d'intervallisation MED – c'est la médiane. Pour MEDMW les limites d'intervalles s'obtiennent par une partition des scores-*IRI* situés des deux côtés de la moyenne arithmétique (et qui ont été arrangés préalablement en fonction de leur importance numérique) en 2 × 3 classes égales. Quant à l'algorithme MED, le même procédé se fait à partir de la médiane.
- Il va de soi que la syntaxe iconique de la carte choroplète dépend directement du choix de l'algorithme d'intervallisa-

tion. Comme il semble acquis que les procédés dialectométriques décrits dans cet article sont *au service exclusif* de la *géographie linguistique*, il va de soi que le dialectométricien cherchera à *rapprocher* – pour autant que faire se peut – les *résultats* de son *analyse taxométrique des concepts géolinguistiques* de la *dialectologie traditionnelle*. D'une comparaison synoptique des résultats dûment visualisés des algorithmes MINMWMAX, MEDMW et MED il ressort clairement que l'algorithme MINMWMAX répond le mieux aux exigences de la géographie linguistique traditionnelle, alors que l'algorithme MED s'en écarte le plus. Cette expérience a d'ailleurs été faite plus d'une fois au sein de la *géographie quantitative* (ou: 'Nouvelle Géographie') ou, plus exactement, dans le cadre de la *cartographie statistique*: cf. Béguin (1979, pp. 199-246), Jenks (1963), Jenks et Coulson (1963), Kishimoto (1972) et Wonka (1980).

- [7] La fréquence relative (p_j) se calcule comme suit :

$$p_j = \frac{\text{points d'enquête/intervalle}}{250}$$

- [8] Il faut distinguer le Piémont *linguistique* du Piémont *politique* (= une des 19 régions administratives de l'Italie moderne).

Bibliographie et abréviations

- Abler, R., J.S. Adams et P. Gould (1977). *Spatial Organization, The Geographer's View of the World* (Prentice Hall, Englewood Cliffs, NJ).
- AIS: Sprach- und Sachatlas Italiens und der Südschweiz, ed. K. Jaberg et J. Jud. Vol. I-VIII (Ringier & Co., Zofingen) 1928-1940.
- Altmann, G. et W. Lehfeldt (1973). *Allgemeine Sprachtypologie. Prinzipien und Meßverfahren* (Fink, Munich).
- Anderberg, M.R. (1973). *Cluster Analysis for Applications* (Academic Press, New York, London).
- Bahrenberg, G. et E. Giese (1975). *Statistische Methoden und ihre Anwendung in der Geographie* (Teubner, Stuttgart).
- Béguin, H. (1979). *Méthodes d'analyse géographique quantitative* (Librairies Techniques, Paris).
- Benzécri, J.-P., F. Benzécri et alii (1980). *Pratique de l'analyse des données. 1. Analyse des correspondances. Exposé élémentaire* (Dunod, Paris).
- Berge, Cl. (1967). *Théorie des graphes et ses applications* (Dunod, Paris).
- Bertier, P. et J.-M. Bourroche (1975). *Analyse des données multidimensionnelles* (Presses Universitaires de France, Paris).
- Bock, H.H. (1974). *Automatische Klassifikation. Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten (Cluster-Analyse)* (Vandenhoeck & Rupprecht, Göttingen).
- Boissevain, J. et J.C.I. Mitchell (1973). *Network Analysis. Studies in Human Interaction* (Mouton, La Kaye, Paris).
- Bourroche, J.M. et G. Saporta (1980). *L'analyse des données* (Presses Universitaires de France, Paris).

- Brassel, K.E. et D. Reif (1979). A procedure to generate Thiessen Polygons. *Geographical Analysis* 11, 289-303.
- Chandon, J.L. et S. Pinson (1981). *Analyse typologique. Théories et applications* (Masson, Paris, New York, Barcelone, Milan).
- Claval, P. (1977). *La nouvelle géographie* (Presses Universitaires de France, Paris).
- Cowan, H.K.J. (1964). Old Limburgian and the quantitative classification of Westgermanic dialects, *Leuvense Bijdragen* 53, 1-21.
- Dauzat, A. (1922). *La géographie linguistique* (Flammarion, Paris).
- Dickinson, G.C. (1973). Statistical mapping and the presentation of statistics (Arnold, Londres).
- Durand, J.P. (1889). Notes de philologie rouergate (suite), *Revue des Langues Romanes* 33, 47-84.
- Fischer, M.M. (1980). Regional taxonomy. A comparison of some hierarchic and non-hierarchic strategies, *Regional Science and Urban Economics* 10, 503-537.
- Flament, Cl. (1968). *Théorie des graphes et structures sociales* (Mouton, Paris).
- French, H.M. et J.-B. Racine (1971). Quantitative and qualitative geography. La nécessité d'un dialogue (University of Ottawa Press, Ottawa, Ont.).
- Fronzaroli, P. (1975). Problemi di classificazione delle lingue su base quantitativa, in: *Colloquio sul tema: Le tecniche di classificazione e loro applicazione linguistica* (Firenze 1972), *Accademia nazionale dei Lincei, anno 372-1975, Rome 1975*, 123-141.
- Goebel, H. (1976). La dialectométrie appliquée à l'ALF (Normandie), in: A. Várvaro, ed., *XIV Congresso internazionale di linguistica e filologia romanza* (Macchiaroli et Benjamins, Naples, Amsterdam). Vol. II, 165-195.
- Goebel, H. (1977a). Rätoromanisch versus Hochitalienisch versus Oberitalienisch. *Dialektometrische Beobachtungen innerhalb eines Diasystems, Ladinia* 1, 39-71.
- Goebel, H. (1977b). Zu Methoden und Problemen einiger dialektometrischer Meßverfahren, *Germanistische Linguistik* 3-4/77, 335-365.
- Goebel, H. (1978a). Analyse dialectométrique de quelques points de l'AIS (italien standard, valdotain, provençal alpin, turinois, milanais), in: G.P. Clivio et G. Gasca Queirazza, eds., *Lingue e dialetti nell'arco alpino occidentale. Atti del Convegno internazionale di Torino* (1976) (Centro Studi Piemontesi, Turin) pp. 282-294.
- Goebel, H. (1978b). Dialektometrie - Methoden und Probleme einer empirischen Betrachtung von Sprache im Raum (Tondiaschau), in: W.U. Dressler et W. Meid, eds., *Proceedings of the twelfth International Congress of Linguists* (Vienna 1977) (Innsbrucker Beiträge zur Sprachwissenschaft, Vienne) pp. 611-613.
- Goebel, H. (1981). *Éléments d'analyse dialectométrique (avec application à l'AIS)*, *Revue de Linguistique Romane* 45, 349-420.
- Gould, P.R. et K. Bassett (1967). *The spatial dynamics of regional taxonomy*, Pennsylvania State University, University Park, PA.
- Guiter, H. (1973). Atlas et frontières linguistiques, in: G. Straka et P. Gardette, eds., *Les dialectes romans de France à la lumière des atlas régionaux (colloque de Strasbourg 1971)* (Centre national de la recherche scientifique, Paris) 1973, 61-109.
- Guiter, H. (1979). Critique et limite d'une méthode, *Mélanges à la Mémoire de Louis Michel* (Déhan, Montpellier) 261-272.
- Haggett, P. et R.J. Chorley (1969). *Network Analysis in Geography* (Arnold, Londres).
- Hammond, R. et P. McCullagh (1974). *Quantitative Techniques in Geography: An Introduction* (Clarendon, Oxford).
- Hartigan, J.A. (1975). *Clustering Algorithms* (Wiley, New York, Londres, Sydney, Toronto).
- Houck, Ch.L. (1969). A statistical and computerized methodology for analyzing dialect material, *Diss., Iowa* (University Microfilms, Ann. Arbor).
- IRI: Indice Relatif d'Identité; cf. 2.
- Jaberg, K. (1936). *Aspects géographiques du langage* (Droz, Paris).
- Jaberg, K. (1975). *Géographie linguistique et expressivisme phonétique. Les noms de la balançoire en portugais*, *Revista portuguesa de filologia* 1 (1947) 1-44 (aussi in: S. Heinemann et K. Jaberg, eds., *Sprachwissenschaftliche Forschungen und Erlebnisse*, N.F. (Franche, Berne) 1975, pp. 63-90).
- Jaberg, K. et J. Jud (1928). *Der Sprachatlas als Forschungsinstrument. Kritische Grundlegung und Einführung in den Sprach- und Sachatlas Italiens und der Südschweiz* (Niemeyer, Halle).
- Jambu, M. (1978). *Classification automatique pour l'analyse des données. I. Méthodes et algorithmes* (Dunod, Paris).
- Jardine, N. et R. Sibson (1971). *Mathematical taxonomy* (Wiley, London).
- Jenks, G.F. (1963). Generalization in statistical mapping, *Annals of the Association of American Geographers* 53, 15-26.
- Jenks, G.F. et M.R.C. Coulson (1963). Class intervals for statistical maps, *Internationales Jahrbuch für Kartographie* 3, 119-134.
- Kaufmann, A. (1968). *Introduction à la combinatoire en vue des applications* (Dunod, Paris).
- Kishimoto, H. (1972). Ein Beitrag zur Klassenbildung in statistischer Kartographie unter besonderer Berücksichtigung der maschinellen Herstellung von Choroplethenkarten, *Kartographische Nachrichten* 22, 224-239.
- MED: algorithme d'intervallisation; cf. 3.1.
- MEDMW: algorithme d'intervallisation; cf. 3.1.
- MINMWMAX: algorithme d'intervallisation; cf. 3.1.
- Monmonier, M.S. (1978). Viewing azimuth and map clarity, *Annals of the Association of American Geographers* 68, 180-195.
- Opitz, O. (1980). *Numerische Taxonomie* (Fisher, Stuttgart, New York).
- Peucker, T.K. (1977). Die Strukturen der digitalen Erfassung und Darstellung topographischer Oberflächen, in: I. Kretschmer, ed., *Beiträge zur theoretischen Kartographie* (Deuticke, Vienne) pp. 125-140.
- Pop, S. (1950). *La dialectologie. Aperçu historique et méthodes*

- d'enquêtes linguistiques, Louvain-Gembloux, 2 vol.
- Pudlatz, H. (1977). Automatische Erzeugung von Isoglossen auf dem Plotter mit Hilfe von Thiessen-Polygonen, *Germanistische Linguistik* 3-4/77, 245-258.
- Ricine, J.-B. et H. Reymond (1973). *L'analyse quantitative en géographie* (Presses Universitaires de France).
- Rapoport, A. (1970a). *Fights, Games and Debates* (Univ. of Michigan Press, Ann Arbor, MI).
- Rapoport, A. (1970b). *N-person Game Theory. Concepts and Applications* (Univ. of Michigan Press, Ann Arbor, MI).
- Rase, W.-D. (1979). Kartographische Darstellung dynamischer Vorgänge in computergenerierten Filmen, *Kartographische Nachrichten* 24, 210-215.
- Reed, D.W. et J.L. Spicer (1952). Correlation methods of comparing idiolects in a transition area, *Language* 28, 348-359.
- Rensch, K.H. (1970). Die dia-synchronische Verwandtschaft von Phonemsystemen. Ein Beitrag zur strukturellen Dialektologie, *Orbis* 19, 94-101.
- Rhynsburger, D. (1973). Analytic delineation of Thiessen Polygons, *Geographical Analysis* 5, 133-144.
- Séguy, J. (1971). La relation entre la distance spatiale et la distance lexicale, *Revue de Linguistique Romane* 35, 335-357.
- Séguy, J. (1973a). La dialectométrie dans l'Atlas linguistique de la Gascogne, *Revue de Linguistique Romane* 37, 1-24.
- Séguy, J. (1973b). *Atlas linguistique et ethnographique de la Gascogne*, Vol. V-VI/1+2 (Centre national de la recherche scientifique, Paris).
- Shaw, D. (1974). Statistical analysis of dialectal boundaries, *Computers and the Humanities* 8(3), 173-177.
- Simon, H.J. (1967). *Beobachtungen an Mundarten Piemonts* (Winter, Heidelberg).
- Smith, D.M. (1975). *Patterns in Human Geography* (Penguin, Harmondsworth).
- Sneath, P.H.A. et R.R. Sokal (1973). *Numerical Taxonomy. The Principles and Practice of Numerical Classification* (Freeman, San Francisco, CA).
- Sodeur, W. (1974). *Empirische Verfahren zur Klassifikation* (Teubner, Stuttgart).
- Sokal, R.R. et P.H.A. Sneath (1963). *Principles of Numerical Taxonomy* (Freeman, San Francisco, CA).
- Thiessen, A.H. (1911). Precipitation averages for large areas, *Monthly Weather Reviews* 39, 1082-1084.
- Tukey, J.J.W. (1977). *Exploratory Data Analysis* (Addison-Wesley, Reading, Menlo Park, Londres, Amsterdam, Don Mills, Sydney).
- Viereck, W. (1980). Dialektometrie und englische Dialektologie, *Grazer linguistische Studien* 11/12, 335-356.
- Vogel, F. (1975). Probleme und Verfahren der numerischen Klassifikation unter besonderer Berücksichtigung von Alternativmerkmalen (Vandenhoeck & Rupprecht, Göttingen).
- Wonka, E. (1980). Methoden der Wertstufenbildung und ihre Eignung für die thematische Kartographie, *Mitteilungen der Österreichischen Geographischen Gesellschaft* 122, 70-96.