L'analyse dialectométrique de données de 'La flexió en els dialectes catalans' d'Antoni Maria Alcover et Francesc de Borja Moll (1929-1932)

Autor(en): **Casassas, Xavier / Goebl, Hans / Smeka, Pavel**

Objekttyp: **Article**

Zeitschrift: **Revue de linguistique romane**

Band (Jahr): **79 (2015)**

Heft 315-316

PDF erstellt am: **25.05.2024**

Persistenter Link: https://doi.org/10.5169/seals-842225

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.

Ein Dienst der *ETH-Bibliothek*
ETH Zürich, Rämistrasse 101, 8092 Zürich, Schweiz, www.library.ethz.ch

http://www.e-periodica.ch

L'analyse dialectométrique des données de 'La flexió verbal en els dialectes catalans' d'Antoni Maria Alcover et Francesc de Borja Moll (1929-1932)

1. Remarques préliminaires

Le présent article constitue un rapport de recherche dont les bases empiriques ont déjà été jetées en 1900. C'est alors que le catalaniste bien connu Mossén Antoni Maria Alcover (1862-1932) a lancé ses enquêtes sur la variabilité géolinguistique dans la morphologie verbale du catalan. Ces recherches – entreprises avec l'encouragement et le soutien du romaniste allemand Bernhard Schädel (1878-1926) – ont été terminées en 1928 et complétées par la publication de leurs résultats, réalisée, entre 1929 et 1932, dans la revue barcelonaise *Anuari de l'Oficina Romànica de Lingüística i Literatura* [AORLL], surtout par les soins de F. de B. Moll (1903-1991)[1].

Cette publication (signalée dorénavant par le sigle LaFV) – étalée sur quatre fascicules de l'AORLL et un total de 368 pages[2] – n'a ni la forme d'un atlas linguistique traditionnel, ni celle de tableaux rangés par colonnes, mais fournit un texte continu, arrangés en alinéas ordonnés: voir les tableaux 1 et 2 qui en montrent deux passages différents (cf. *infra*, p. 364). Il est cependant hors de doute que l'esprit dans lequel le questionnaire et les enquêtes de LaFV ont été conçus et effectués rentre pleinement dans les visées géolinguistiques du temps. Précisons à ce propos qu'A. M. Alcover a deux fois rendu visite à Jules Gilliéron à Paris[3], sans toutefois avoir participé directement à ses fameux cours de géographie linguistique[4].

Par la suite, une petite partie des données verbo-morphologiques de LaFV a été intégrée dans le fameux «Diccionari català-valencià-balear» (DCVB), préparé, élaboré et publié par A. M. Alcover et F. de B. Moll entre 1930 et 1962.

En 1999, toute la documentation de LaFV a été soumise à un réexamen serré de la part de Maria Pilar Perea d'où sont issues une grande monographie

---

1 Voir l'historique que M. P. Perea en a brossé en 2001 (a).

2 Un cinquième fascicule, mis au point par F. de B. Moll et prêt à l'impression, n'a pas pu être publié à cause de l'éclatement de la guerre civile: cf. Sanchis Guarner 1953, 108 et Moll 1962, 306.

3 Ces visites eurent lieu en 1912 et 1913: cf. Perea 2005a, 29-20 et Moll 1962, 89-90.

4 Voir à ce sujet Pop / Pop 1959, 53-63 («Les élèves de Gilliéron de 1883 à 1926»).

en deux volumes (Perea 1999a)[5], une série nourrie d'articles de présentation[6] et surtout la numérisation complète de LaFV. Cette dernière a conduit à l'édition de deux CD-ROM (Perea 2001c et 2005b) et a rendu possible – à partir de 2002 – une coopération directe entre M. P. Perea (de l'Université de Barcelone) et le laboratoire de dialectométrie (DM) de l'Université de Salzbourg dirigé par Hans Goebl.

Cette coopération austro-catalane – engagée à la lumière de la dialectométrie, entendue comme méthode, et aussi du logiciel *Visual DialectoMetry* (VDM), entendu comme puissant outil informatique – a débuté avec l'analyse-DM, très encourageante, de spécimens réduits des données numérisées de LaFV[7] pour être couronnée, en janvier 2007, par une dialectométrisation du total de ces données – faite à Salzbourg même – avec la participation personnelle de Maria Pilar Perea, Xavier Casassas[8] et Slawomir Sobota[9]. En 2013, les calculs-DM faits en 2007 ont été complétés et réajustés par Pavel Smečka.

L'organisation informatique – en sous-corpus de taille variable – alors imposée aux données-LaFV a reçu, par la suite, une valeur 'canonique': elle constitue la base de tous les calculs-DM effectués depuis 2007 et, de ce fait, aussi des résultats présentés dans cet article: voir le tableau 4 (cf. *infra* pp. 371-372).

D'entrée de jeu, la dialectométrisation des données de LaFV comportait un risque d'ordre 'métrologique'. C'est qu'elle devait se servir directement des données de la version numérisée de LaFV qui, elle, retrace exactement l'agencement ordonnateur dans lequel F. de B. Moll avait publié, entre 1929 et 1932, le contenu des carnets d'enquête originaux de Mgr Alcover.

Cet 'agencement ordonnateur' – dont les tableaux 1 et 2 donnent une idée approximative – fut conçu et réalisé par F. de B. Moll uniquement à des fins pratiques, voire économiques[10], et en dehors de toute intention classificatoire majeure telle qu'elle est pratiquée au sein de l''École dialectométrique de

---

5 Il faut y ajouter encore un complément: Perea 1999b.

6 Voir ses contributions respectives entre 2001a et 2008b.

7 Cf. Perea 2008b, où l'on trouve déjà des graphiques créés à l'aide de VDM (ibidem, 71-72). Ceci vaut également pour Perea / Ueda 2010 qui, mise à part une analyse dendrographique-VDM (ibidem, 103), contient une classification quantitative d'une partie des données-LaFV, réalisée moyennant un algorithme particulier, capable de prendre directement en considération les réponses multiples.

8 X. Casassas était collaborateur de recherche de H. Goebl (pour l'atlas ladin ALD et la dialectométrie) entre 2005 et 2008 de même que lecteur de catalan à l'Université de Salzbourg.

9 Slawomir Sobota détenait, à l'Université de Salzbourg, le poste d'informaticien contractuel (pour l'atlas ladin ALD et la dialectométrie) entre 1999 et 2009.

10 Cf. Perea 2005a, 123.

Salzbourg' (DM-S) sous la forme de la 'taxation': voir à ce sujet la figure 3 et l'étape A qui y figure. Toujours est-il qu'il en résulte un certain nombre de 'types', dont le statut empirique et taxométrique est comparable à celui des *taxats* de la DM-S.

Dans la première de ses deux autobiographies [«Els meus primers trenta anys (1903-1934)»], Moll (1970, 242) a mentionné brièvement[11] la genèse de cet arrangement classificateur qui a été réalisé par lui (et le père Josep Calveras) dans la deuxième moitié de l'an 1930:

> «Durant aquelles setmanes passades a Ciutadella [...], no vaig estar ociós. Feia temps, ajudat del P. Calveras[12], havia conjuminat un sistema per a fer publicables els paradigmes de la flexió verbal que mossén Alcover havia recollit de viva veu a 148 pobles[13] de totes les comarques catalanes: ell els tenia escrits amb la seva lletrona menuda en una pila de llibretes, formant centenars de columnes; era impossible consultar-los de manera que es pogués fer-ne una síntesis [*sic*] i reduir-los a tipus bàsics i estructurats. El sistema que vàrem idear per a obtenir aqueixa síntesi, va ser de designar amb un nombre cada de 148 localitats (per no haver d'escriure llurs noms milers de vegades) i fer, per cada persona de temps verbal de cada verb, la relació de desinències diverses, posant al costat de cada una el nombre o nombres indicadors de les localitats on aquella desinència era usada. Això que ara explic amb poques paraules – i que no és fàcil d'entendre per un qui no sigui dialectòleg – representava un treball llarg i minuciosíssim que requeria una gran atenció; però, gràcies a haver-hi dedicat aquelles setmanes d'estiu, vaig deixar molt avançada la redacció en net de l'obra *La flexió verbal en els dialectes catalans*, que s'anava publicant en els Anuaris de l'Oficina Romànica de Barcelona i que va romandre inacabada a causa de la guerra civil[14]. Es un treball de morfologia dialectal únic en el món per la seva amplitud, i una de les moltes proves de la gran constància de mossèn Alcover en la replega de materials dialectològics a base de llargs viatges i de laborioses enquestes.
>
> Aquella feina d'elaboració de *La flexió verbal* m'era pagada per l'Oficina Romànica, i això em rescabalava de la no percepció del sou de mossèn Alcover durant aquells dos mesos, cosa que a ell li resultava un alleutjament ben necessari per la precària situació econòmica en què aleshores es trobava.»

Les tableaux 1 et 2 illustrent très nettement le fait que Moll y avait inséré les données verbo-morphologiques originales dans deux tiroirs: le premier a reçu des formes 'majeures' (avec majuscules et en caractères gras et grands), alors que le second, réservé aux formes 'mineures', ne contient que les formes verbales 'brutes', présentées en notation phonétique et avec des caractères plus petits.

---

[11] Une autre allusion à ce travail complexe se trouve à la page 208 du même ouvrage (Moll 1970).

[12] Josep Calveras (S. J.) était le secrétaire de l'*Oficina Romànica de Lingüística i Literatura*, fondée à Barcelone en 1927 (cf. Moll 1970, 204-5).

[13] Pour le nombre réel des enquêtes faites pour LaFV, voir la note 17.

[14] Voir la note 2.

M. P. Perea avait certainement raison en dénommant, sur ses deux CD-ROM de 2001 et 2005, les formes du premier des deux tiroirs «variants morfològiques» et celles du second, toujours plus nombreuses, «variants fonètiques»[15].

En regardant le tableau 1, l'on constate, en effet, qu'il y a trois «variants morfològiques» et six «variants fonètiques», alors que sur le tableau 2 l'on trouve huit «variants morfològiques» et quatorze «variants fonètiques». Les «variants morfològiques» englobent toujours un certain nombre de «variants fonètiques». Entre ces deux sortes de variantes, il règne donc une relation d'inclusion: toute variante phonétique (A) représente un sous-ensemble d'une variante morphologique (B): B ⊇ A.

Sur les cartes 1 et 2, nous n'avons mis en carte que les «variants fonètiques» définies par Moll, tout en leur conférant ainsi le statut d'un *taxat* suivant les préceptes de la DM-S.

PRESENT D'INDICATIU

1.ª sg. — **Beoh:** **bẹ́k** 1, 6-8, 10-62, 80, 87, 124, 132, 134, 137-141. **bẹ́k** 63-65, 66-70, 72-79, 81-86, 88-117, 148. **bẹ́ḳ** 118, 128-131. **bẹ́k** 119-123, 125-127, 133-136, 140-147. — **Beui:** **bẹ́ui** 2-5, 9. — **Beyoh:** **bẹ́ik** 65a, 71.

Tableau 1: Fac-similé de «La flexió verbal» (LaFV: 1929, II, 74) qui montre le groupement des équivalents dialectaux du catalan standard *bec* "je bois" fait par F. de B. Moll. Voir aussi la carte 1.

PRESENT D'INDICATIU

1.ª sg.—**Canti:** **kánti** 1-24.—**Càntioh:** **kántik** 17, 19, 21, 25.—**Canto:** **kántu** 18-19, 21-30, 33-35, 38-39, 41-62. **kántọ** 60, 63-99.—**Càntooh:** **kántuk** 31-32, 36-37, 40, 44, 46, 61.—**Càntot:** **kántut** 36, 40, 44.—**Cante:** **kántẹ** 100-110, 112-117. —**Canta:** **kánta** 111.—**Cant:** **ḳánt** 118, 129-130. **kánt** 119-126, 132-141, 148. **kẹ́nt** 127. **ḳẹẹ́nt** 128. **ḳẹ́nt** 131. **kán** 142-147.

Tableau 2: Fac-similé de «La flexió verbal» (LaFV: 1929, II, 9) qui montre le groupement des équivalents dialectaux du catalan standard *canto* "je chante" fait par F. de B. Moll. Voir aussi la carte 2.

---

15 Malheureusement, cette bi-partition en deux tiroirs ne réapparaît plus dans la version en ligne de LaFV qui se trouve sur le site officiel de l'*Institut d'Estudis Catalans* (IEC): ‹http://alcover.iec.cat/›. Pour accéder à cette version il faut cliquer plusieurs onglets dans l'ordre suivant: *Treballs dialectals, La flexió verbal, Cartografia*. La liste avec les variantes *phonétiques* 'de mémoire mollienne' apparaît quand on clique l'icône montrant la silhouette de la Catalogne linguistique et le symbole d'un haut-parleur.

La dialectométrisation des données de LaFV ne connaît donc pas la différence faite normalement entre une 'source originale' (= atlas linguistique imprimé) et la 'matrice de données' qui en est dérivée par le biais de la taxation des données brutes de l'atlas linguistique en question. C'est que la version informatisée de l''agencement ordonnateur', opéré par F. de B. Moll en 1930, fait directement fonction de 'matrice de données': voir la figure 3, position 2.

En effet, l'arrangement, voire le classement des données fait par Fr. de B. Moll[16], obéissait, en dernière analyse, largement à des critères d'ordre phonétique, sans pour autant se plier aux exigences étymologiques et classificatoires de la phonétique historique classique.

Du point de vue strictement métrologique, il est indéniable que ce fait peut créer des effets perturbateurs ('erreurs de mensuration'). Néanmoins et au vu des résultats de notre analyse-DM, ces effets semblent être très limités. Cette constatation émane surtout de la comparaison analytique des résultats-DM de plusieurs sous-corpus, relatifs aux différents temps, modes ou personnes des matériaux verbo-morphologiques de base.

Ceci n'empêche qu'il reste très souhaitable de pouvoir soumettre, un jour, les données de LaFV à une analyse taxatoire classique et d'en confronter les résultats-DM avec ceux de l'analyse-DM déjà existante.

Le logiciel de base des deux CD-ROM publiés par M. P. Perea en 2001 et 2005 permettait déjà de projeter les données de LaFV, enfermées d'abord dans des structures tabulaires, dans l'espace, moyennant la mise en carte soit des transcriptions originales, soit des types phonétiques («variants fonètiques») définis par F. de B. Moll. Cette possibilité représentait déjà une sorte de revitalisation empirique de données géolinguistiques alors vieilles de quelque 70 à 80 ans.

Comme le logiciel VDM attribue une importance toute particulière à la confection de cartes linguistiques, l'insertion des données de LaFV dans VDM a permis d'en pousser plus avant la revitalisation et de perfectionner la mise en carte des types 'molliens': voir à ce propos les cartes 1 et 2 qui visualisent les effets du travail ordonnateur accompli par F. de B. Moll pour les réponses dialectales données aux stimuli catalans *bec* "je bois" et *canto* "je chante", tels qu'ils ressortent des tableaux 1 et 2.

Le type de carte utilisé à cet effet par VDM est appelé – en matière de cartographie – '(carte) choroplèthe'. La polygonisation du réseau de recherche

[16] Dans le titre de la publication originale de LaFV, Moll appelle lui-même ce travail «recopilació i sistematització».

– qui comprenait 149 localités[17] déjà du vivant de Mgr Alcover – répond moins à des finalités d'ordre esthétique que d'ordre heuristique: c'est qu'elle permet une comparaison visuelle directe entre cartes *qualitatives* (voir les cartes 1 et 2) et cartes *quantitatives* (voir les cartes 3-38).

Ajoutons que les 38 cartes en couleurs de cet article – qui toutes ont été fabriquées à l'aide du logiciel VDM – ont été reproduites à partir d'autant de captures d'écran à haute résolution. Précisons en outre que nous avons utilisé, pour ce faire, la plus récente édition de VDM (1.12.2.0) dont il n'existe, jusqu'à présent, qu'une version en langue allemande, de même qu'un écran de 30 pouces (résolution: 1600 × 1200 pixels).

2. Genèse et propriétés des corpus-DM de base (matrices de données) utilisés pour la dialectométrisation des données de LaFV

→ Tableaux 1 et 2 ainsi que cartes 1 et 2

Pour parler convenablement d'un corpus-DM et de ses composantes, il est bon de se familiariser, d'entrée de jeu, avec les réalités empiriques de départ et avec la terminologie respective.

À ce propos, commençons par examiner les cartes 1 et 2. Chacune des deux cartes constitue, d'un côté, la visualisation d'un ou deux alinéas de la version imprimée de LaFV[18] et, de l'autre, l'équivalent d'une ligne particulière (ou d'un seul vecteur d'attributs) de la matrice de données[19] lequel, depuis de longues années, est appelé, au sein de la DM-S, *carte de travail* (CT)[20].

Chaque CT est composée d'un certain nombre d'aires ('taxatoires') dont chacune est régie par un type linguistique spécifique, appelé, à Salzbourg, *taxat*. Les aires taxatoires peuvent revêtir des structurations très diversifiées: grandes ou petites, cohérentes ou éparses. En plus, la bigarrure taxatoire des CT est également très multiforme. Théoriquement, elle peut varier entre 2 et le nombre maximal (= N) des points d'enquête du réseau en question. Dans la terminologie de la DM-S, il est alors question de la *poly*-nymie (toujours variable) d'une CT. Une CT avec deux taxats (ou aires taxatoires) est dénom-

---

17 La numérotation originale, utilisée dans LaFV, allait de 1 à 148, alors que celle adoptée par M. P. Perea va de 1 à 149. La raison en est que, dans la numérotation originale, le numéro 65 figure deux fois: pour Esterri d'Àneu (*sub* 65) et pour Boí (*sub* 65a). Dans la numérotation définitive, l'indicatif réservé pour Boí, est passé de 65a à 66: cf. Perea 2005, 123.

18 Voir les tableaux 1 et 2.

19 Voir la figure 3.

20 La forme catalane pour *(la) carte de travail* est *(el) mapa de treball*.

mée *bi*-nyme (ou 2-nyme), alors qu'une CT avec dix taxats s'appelle *déca*-nyme (ou 10-nyme).

Pour un corpus-DM donné, l'observation de la bigarrure (ou: granulation) interne des CT est du plus haut intérêt statistique. La même remarque vaut pour l'observation de la taille des aires taxatoires. Dans les deux cas, l'on constate la présence d'épatantes régularités qui prouvent que les différents domaines dialectaux ne sont pas régis par le hasard événementiel, mais par des lois (linguistiques) comparables aux lois de fréquence qui valent pour la structuration de textes linéaires, et aussi aux lois phonétiques («Lautgesetze»), chères aux néo-grammairiens de Leipzig des années 1870-80.

Quant au caractère changeant de la taille et de la structure des aires taxatoires, un fait – dont il n'existait que des idées relativement vagues auparavant – s'est avéré depuis l'avènement de la DM: à savoir que la coïncidence tant débattue d'un grand nombre d'aires (géo)linguistiques différentes (et surtout des isoglosses qui les entourent) est pratiquement inexistante[21], et que, par contre, il règne, entre les différentes aires taxatoires de la même matrice de données, des 'enchevêtrements particuliers'[22] qui semblent être la clé – dans des données géolinguistiques à base d'atlas – pour le bon fonctionnement interactif et communicatif de réseaux basilectaux tout court.

Précisons en outre que le phénomène de l'enchevêtrement[23] d'un grand nombre d'aires différentes peut se faire sous différentes formes: 'en profondeur' – comme cela arrive pour les données de l'écrasante majorité des atlas linguistiques – et 'à brève distance' – comme c'est le cas pour les données morphologiques de LaFV. C'est que l'imbrication des aires en question ne s'y opère pas 'librement', mais sous la contrainte paradigmatique de systèmes morphologiques.

---

21 Ce phénomène correspond à l'indépendance ontologique attribuée par Jules Gilliéron – et beaucoup d'autres auteurs du début du 20e siècle – aux différents lexèmes (mots) en recourant au slogan suivant: «Chaque mot a son histoire». À la singularité de l'histoire de chaque mot correspond, en matière de géographie linguistique, l'imprévisibilité de la taille et de la structure géographiques des aires taxatoires. Pour l'historique de ce slogan (fr. *Chaque mot a son histoire*, all. *Jedes Wort hat seine eigene Geschichte*), voir les contributions de Jaberg 1908 (6), Christmann 1971 et Malkiel 1967.

22 Cette propriété, valable pour tous les réseaux dialectaux, nous semble être un universel linguistique. Équivalents terminologiques dans d'autres langues: all. *spezielle Verzahnung*, angl. *special entanglement*, cat. *trenat especial*, ital. *intreccio particolare*, esp. *trenzado / entramado especial*.

23 La langue française offre encore les synonymes suivants: *imbrication*, *tuilage*.

3. Présentation des corpus TOT et RÉD utilisés dans cette étude

→ Tableaux 3-5, ainsi que figures 1 et 2

Évidemment, tout corpus relatif à la morphologie verbale dispose, d'entrée de jeu, d'une structuration 'naturelle' selon les catégories des temps, des modes et des personnes, qui, elle, se croise avec les groupements en formes finies et non finies. La version numérisée de LaFV, telle qu'elle avait été utilisée à Salzbourg en 2007, est basée sur 78 verbes et les variations morphologiques respectives: voir le tableau 3.

Pour les subdivisions relatives aux temps, modes et personnes prises en compte pour nos calculs-DM, voir le tableau 4. Au premier coup d'œil, l'on y discerne deux colonnes: TOT (= corpus total) et RÉD (= corpus réduit). Cette bi-partition s'explique par le souci de minimiser les conséquences néfastes d'une propriété omniprésente dans les données de LaFV: il s'agit de l'existence d'un grand nombre de lacunes (« données manquantes, cases vides, missing data »), dont la genèse s'explique soit par la négligence de la part des enquêteurs, soit par l'absence (ou l'obsolescence) réelle de certaines formes verbales dans la compétence des sujets interviewés.

no.^{TOT}	verbe catalan (infinitif): TOT	signification française	no.^{TOT}	retenu dans RÉD	no.^{RÉD}
1	*acudir*	accourir			
2	*ajupir*	se baisser			
3	*anar*	aller	3	anar	1
4	*batre*	battre	4	batre	2
5	*beure*	boire	5	beure	3
6	*bullir*	bouillir	6	bullir	4
7	*cabre*	tenir, contenir			
8	*caldre*	falloir			
9	*cantar*	chanter	9	cantar	5
10	*caure*	tomber, choir	10	caure	6
11	*cloure*	clore, fermer			
12	*collir*	cueillir, ramasser	12	collir	7
13	*conèixer*	connaître	13	conèixer	8
14	*consentir*	consentir	14	consentir	9
15	*córrer*	courir	15	córrer	10
16	*cosir*	coudre	16	cosir	11

no.TOT	verbe catalan (infinitif): TOT	signification française	no.TOT	retenu dans RÉD	no.RÉD
17	*coure*	cuire	17	coure	12
18	*creure*	croire	18	creure	13
19	*cruixir*	craquer, crisser			
20	*defendre*	défendre			
21	*deure*	devoir	21	deure	14
22	*dir*	dire	22	dir	15
23	*doldre*	regretter, plaindre			
24	*dormir*	dormir	24	dormir	16
25	*dur*	porter, amener			
26	*eixir*	sortir, partir			
27	*entendre*	comprendre, entendre	27	entendre	17
28	*escopir*	cracher			
29	*escriure*	écrire			
30	ésser	être	30	ésser	18
31	*estar*	être, rester, demeurer	31	estar	19
32	*fer*	faire	32	fer	20
33	*fondre*	fondre			
34	*fugir*	fuir	34	fugir	21
35	*grunyir*	grogner, gronder			
36	*haver*	avoir	36	haver	22
37	*jaure*	être couché			
38	*llegir*	lire			
39	*lluir*	luire, reluire	39	lluir	23
40	*mentir*	mentir			
41	*merèixer*	mériter	41	merèixer	24
42	*moldre*	moudre			
43	*morir*	mourir	43	morir	25
44	*moure*	mouvoir	44	moure	26
45	*munyir*	traire	45	munyir	27
46	*nàixer*	naître	46	nàixer	28
47	*obrir*	ouvrir	47	obrir	29
48	*oir*	ouïr, entendre			
49	*omplir*	remplir			

no.TOT	verbe catalan (infinitif): TOT	signification française	no.TOT	retenu dans RÉD	no.RÉD
50	*parèixer*	paraître			
51	*partir*	partir	51	partir	30
52	*pèixer*	faire paître, nourrir			
53	*perdre*	perdre	53	perdre	31
54	*plànyer*	plaindre			
55	*plaure*	plaire			
56	*ploure*	pleuvoir	56	ploure	32
57	*poder*	pouvoir	57	poder	33
58	*pondre*	pondre	58	pondre	34
59	*prendre*	prendre	59	prendre	35
60	*prometre*	promettre			
61	*pudir*	puer, sentir mauvais			
62	*rebre*	recevoir	62	rebre	36
63	*remetre*	remettre, renvoyer			
64	*riure*	rire	64	riure	37
65	*rompre*	rompre			
66	*saber*	savoir	66	saber	38
67	*sentir*	sentir, entendre	67	sentir	39
68	*seure*	s'asseoir			
69	*soler*	avoir l'habitude			
70	*sortir*	sortir			
71	*témer*	craindre			
72	*tenir*	tenir	72	tenir	40
73	*traure*	tirer	73	traure	41
74	*valer*	valoir	74	valer	42
75	*venir*	venir	75	venir	43
76	*veure*	voir			
77	*viure*	vivre			
78	*voler*	vouloir	78	voler	44

Tableau 3: Synopse quantitative des infinitifs figurant dans les corpora dialectométriques TOT et RÉD extraits, en 2007, des données brutes de «La flexió verbal» d'A. M. Alcover et F. de Moll (1929-1932) par Maria Pilar Perea et Xavier Casassas.

Comme, en matière de DM, la présence d'un surplus de lacunes dans la matrice de données entraîne inévitablement des distorsions gênantes des calculs taxométriques, il fallait créer à tout prix un corpus aussi complet que possible en opérant une sélection judicieuse dans les données intégrales.

Comme critère de sélection nous avons choisi le principe suivant: n'étaient admises, pour la formation de la matrice de données réduite (RÉD), que des CT pourvues d'un degré de complétude d'au moins 90%. Au vu des 149 points d'enquête du réseau-AM, ceci signifie que toutes les CT disposant de plus de 15 lacunes (cases vides) ont été systématiquement écartées.

		TOT	RÉD
no.	Groupes grammaticaux	∑ cartes de travail (CT)	∑ cartes de travail (CT)
1	*Condicional*	453	241
2	*Futur*	447	229
3	*Gerundi*	78	38
4	*Imperatiu*[24]	199	66
5	*Imperfet Indicatiu*[25]	459	248
6	*Indicatiu 1a persona*	375	169
7	*Indicatiu 2a persona*	375	165
8	*Indicatiu 3a persona*	384	169
9	*Indicatiu 4a persona*	376	165
10	*Indicatiu 5a persona*	375	164
11	*Indicatiu 6a persona*	377	165
12	*Indicatiu total*[26]	2 262[27]	997[28]

[24] Relatif aux personnes 2, 4 et 5.

[25] Relatif aux personnes 1 à 6.

[26] Total des effectifs des classes 6 à 11.

[27] Relatif aux cinq temps suivants: présent, imparfait, futur, conditionnel présent, passé simple (*perfet simple*).

[28] Relatif aux quatre temps suivants: présent, imparfait, futur, conditionnel présent.

		TOT	RÉD
no.	Groupes grammaticaux	∑ cartes de travail (CT)	∑ cartes de travail (CT)
13	*Infinitiu*	78	44
14	*Participi*	77	43
15	*Perfet Simple*	445	0
16	*Present Indicatiu*[29]	459	279
17	*Subjuntiu 1a persona*[30]	152	77
18	*Subjuntiu 2a persona*	150	76
19	*Subjuntiu 3a persona*	154	77
20	*Subjuntiu 4a persona*	150	73
21	*Subjuntiu 5a persona*	150	73
22	*Subjuntiu 6a persona*	150	75
23	*Subjuntiu total*[31]	906[32]	451[33]
	Total des groupes 3, 4, 12-14 et 23	3 600 CT	1 639 CF

Tableau 4: Synopse quantitative des corpora dialectométriques TOT et RÉD extraits, en 2007, des données brutes de «La flexió verbal» d'A. M. Alcover et F. de Moll (1929-1932) par Maria Pilar Perea et Xavier Casassas.

En regardant les tableaux 4 et 5, l'on constate que cette mesure a causé l'abaissement du nombre maximal des CT disponibles de 3 600 (corpus TOT) à 1 639 (corpus RÉD) unités. Heureusement, nous savons depuis plus de 30 ans qu'en matière de DM de telles réductions des dimensions des données

29 Relatif aux personnes 1 à 6.

30 Comprenant toujours le subjonctif présent et imparfait.

31 Total des effectifs des classes 17 à 22.

32 Relatif aux cinq temps suivants: présent, imparfait, futur, conditionnel présent, passé simple (*perfet simple*).

33 Relatif aux quatre temps suivants: présent, imparfait, futur, conditionnel présent.

de base n'entraînent pas automatiquement une détérioration analogue de la qualité des résultats de synthèse. Pour vérifier ces écarts, la DM-S se sert, depuis 2004, de la DM corrélative. En l'occurrence, l'application de ce module statistique confirme l'impression visuelle qui se dégage de la comparaison des cartes 3, 11, 19 (RÉD) et 27 (TOT) ou 6, 14, 22 (RÉD) et 31 (TOT): à savoir que les corpus TOT et RÉD contiennent des structures de profondeur très similaires.

Dans cet article, nous utilisons trois corpus de base pour les démonstrations-DM:

- le corpus RÉD, avec des applications-DM faites à l'aide des trois sous-corpus suivants: *indicatif*, *subjonctif* et *infinitif*,
- le corpus TOT[34], avec les mêmes applications-DM,
- et, pour la comparaison de deux corpus catalans totalement indépendants l'un de l'autre du point de vue linguistique, un corpus tiré de l'analyse-DM des quatre premiers volumes de *l'Atles lingüístic del domini català* (ALDC) de Joan Veny et Lídia Pons i Griera[35].

[34] Signalons encore une particularité technique des opérations-DM réalisées en 2007 par M. P. Perea, X. Casassas et S. Sobota: comme les données originales de LaFV contenaient beaucoup de réponses multiples et qu'il semblait indiqué de ne pas les négliger du point de vue de la DM, il a été décidé de prendre en compte jusqu'à quatre réponses parallèles.
La saisie-DM en a été réalisée de la façon suivante: il a été créé d'abord une matrice de données 'de départ' pour saisir les réponses *uniques* et les *premiers* items des réponses *multiples*. Après son établissement, cette matrice a été triplée pour recevoir, là où c'était nécessaire, les items n° 2, 3 ou 4 des réponses multiples. En distribuant les variantes dans ces quatre matrices, M. P. Perea a essayé de respecter un certain ordre diachronique tout en insérant les variantes les plus anciennes (du point de vue linguistique) dans la première des matrices de données mentionnées (= matrice TOT, ici utilisée) et les variantes plus récentes dans les matrices de données n° 2, 3 et 4.
Ceci n'empêche que les dimensions extérieures de ces quatre matrices de données sont restées identiques: 149 points d'enquête × 3 600 CT.

[35] Voir à ce sujet, Goebl 2013b avec la présentation de quelques résultats de la dialectométrisation des données des volumes I-IV de l'ALDC.

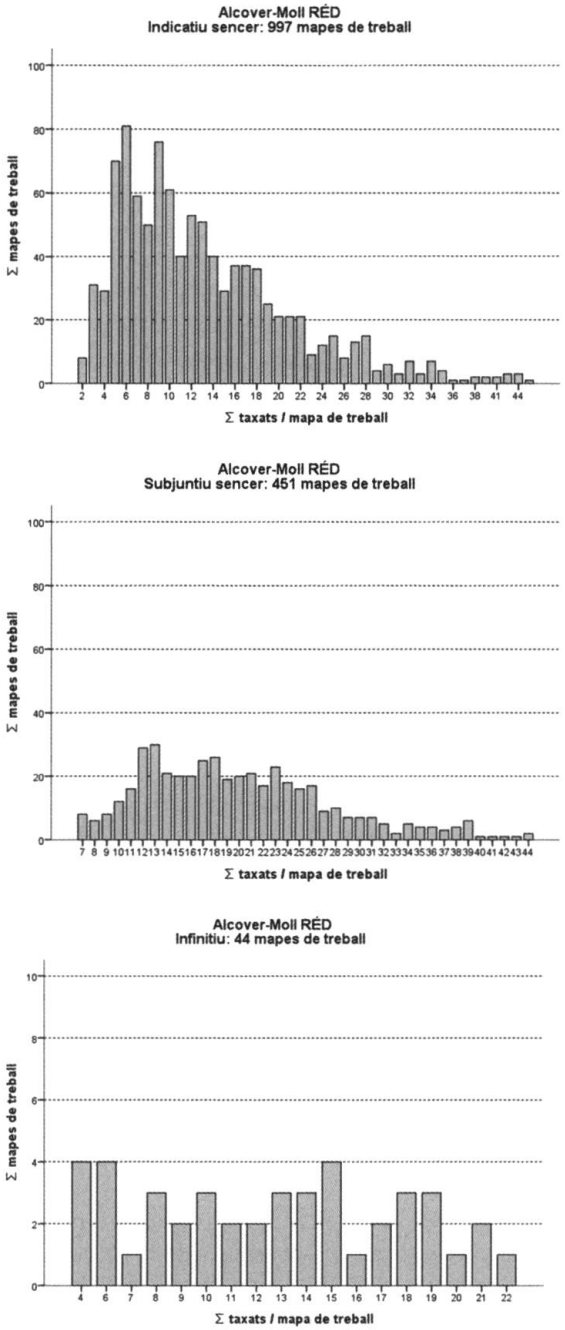


Figure 1: Trois histogrammes montrant la relation entre le nombre de taxats par carte de travail [CT; *mapa / mapes de treball*] (horizontalement, en abscisse) et le nombre absolu des cartes de travail (verticalement, en ordonnée) pour trois groupes grammaticaux tirés du sous-corpus RÉD.
De haut en bas: 997 CT de l'indicatif total [*indicatiu sencer*] (pour quatre temps et six personnes), 451 CT du subjonctif total [*subjuntiu sencer*] (du présent et de l'imparfait et six personnes), 44 CT de l'infinitif [*infinitiu*].
Exemple de lecture pour l'histogramme supérieur: la plus haute des barres verticales signale qu'il y a 81 CT hexa-nymes (pourvus de six taxats).

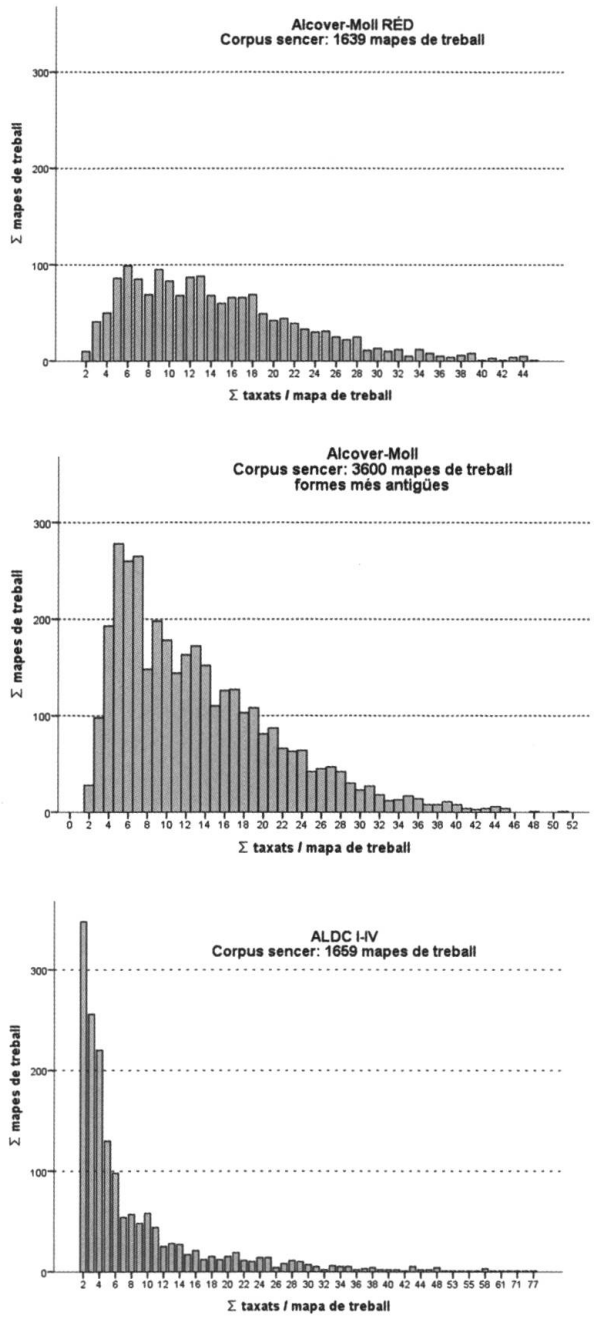


Figure 2: Trois histogrammes montrant la relation entre le nombre de taxats par carte de travail [CT; *mapa / mapes de treball*] (horizontalement, en abscisse) et le nombre absolu des cartes de travail (verticalement, en ordonnée) pour trois corpus analysés. De haut en bas: corpus intégral [corpus sencer] RÉD avec 1 639 CT, corpus intégral TOT avec 3 600 CT [*formes més antigües* – f. les plus anciennes], corpus-ALDC intégral (vols. I-IV) avec 1 659 CT.
Exemple de lecture pour l'histogramme inférieur: la plus haute des barres verticales signale qu'il y a 348 CT bi-nymes (pourvus de deux taxats).

Les figures 1 et 2 permettent de mieux connaître la composition interne de ces trois corpus. Sur chacune de ces deux figures, l'on perçoit trois graphiques à barres verticales ('histogrammes') qui visualisent la relation quantitative qui existe entre la granulation interne (polynymie) des différentes CT et le nombre absolu des CT respectives.

	Corpus-AM total (TOT)			Corpus-AM réduit (RÉD)		
	∑ cartes de travail (CT)	∑ taxats (aires taxatoires)	Variation polynymique (taxats / CT)	∑ cartes de travail (CT)	∑ taxats (aires taxatoires)	Variation polynymique (taxats / CT)
Ind.	2 262	25 943	de 2 à 51	997	13 125	de 2 à 47
Subj.	906	17 848	de 3 à 45	451	8 980	de 7 à 44
Inf.	78	351	de 3 à 23	44	551	de 4 à 22
Total	3 600	48 121	de 2 à 51	1 639	23 926	de 2 à 47

Tableau 5: Synopse quantitative des corpora dialectométriques TOT et RÉD selon les catégories de l'*indicatif*, du *subjonctif* et de l'*infinitif*.

En règle générale, il y a plus de CT *oligo*-nymes ('monotones') que de CT *poly*-nymes ('très bigarrées'). L'on constate, en outre, que le passage entre ces deux pôles s'effectue le long d'une courbe en déclivité continue[36]. Soit dit en passant, les mêmes observations ont pu être faites dans toutes nos analyses-DM antérieures[37].

Pour cinq des six histogrammes présentés sur les figures 1 et 2, le constat de la déclivité continue pour les histogrammes relatifs aux corpus-AM RÉD et TOT est d'autant plus remarquable que les unités de base (= CT) dont se composent les corpus RÉD et TOT sont le résultat non pas d'une analyse classificatoire 'professionnelle' (= taxation), mais – comme ceci a déjà été mentionné plus haut – d'un 'arrangement ordonnateur', fait de manière pragmatique par F. de B. Moll. Il en appert que l'intervention de Moll n'a pas pu éclipser les 'structures de profondeur' originales de LaFV.

36 Cette constatation vaut à l'exception de l'histogramme inférieur de la figure 1, relatif au sous-corpus des infinitifs, caractérisé par une variabilité interne fortement limitée.

37 Voir Goebl 1981 et 1984 I, 206-219.

4. Notices servant à la bonne intelligence des cartes dialectométriques de cet article

Le but de ce paragraphe est de présenter aux lecteurs les outils méthodologiques et conceptuels de base, indispensables pour une bonne compréhension des 36 cartes-DM[38] de cette contribution. Il ressemble d'ailleurs de très près aux explications que nous avons présentées dans Goebl / Smečka 2014, 452-462. C'est pourquoi les lecteurs déjà familiers avec les prérequis de la DM-S, peuvent renoncer à la lecture de ce paragraphe à caractère plutôt introductif et passer directement à celle du paragraphe 5.

4.1. Observations méthodologiques générales

La DM-S vise à déceler les structures de *profondeur* qui, invisibles de prime abord, existent sous la masse des données d'un atlas linguistique. Elle admet en outre que les cartes d'atlas à proprement parler ainsi que leurs subdivisions géolinguistiques – toutes directement accessibles à l'œil de l'observateur – en constituent les structures de *surface*. La démarche logique de la DM-S est inductive: ceci signifie qu'elle se dirige du niveau des faits particuliers vers celui des faits généraux. Dans cette perspective, le niveau des faits *particuliers* comprend les différentes cartes d'un atlas linguistique ainsi que les cartes de travail (CT) qui en sont dérivées (voir, à titre d'exemple, les cartes 1 et 2), alors que le niveau (logique) du *général* embrasse les mises en carte globales de type choroplèthe et isarithmique (voir les cartes 3-38).

Comme la DM-S sert à la découverte (et non pas à l'invention!) de structures globales préexistantes dans les atlas linguistiques analysés, elle représente une discipline exploratoire. En outre, la DM-S suppose que les régularités spatiales mentionnées ci-dessus sont l'émanation directe d'une activité spéciale de l'homme, à savoir de la 'gestion basilectale de l'espace par l'homo loquens'. Ce concept, que nous utilisons explicitement depuis 1993[39], confère non seulement une place centrale à l'activité langagière des dialectophones, mais ouvre également un grand nombre de perspectives interdisciplinaires, ceci surtout à cause du fait que le même acteur, à savoir l'homme, entendu comme *homo faber universalis*, est censé gérer l'espace qu'il habite non seulement de manière linguistique, mais aussi de beaucoup d'autres manières,

---

[38] Les cartes 1 et 2, étant des *cartes de travail* (CT), sont de type *qualitatif*, alors que les 36 cartes restantes, toutes de type *quantitatif*, sont des cartes-DM à proprement parler.

[39] Voir nos contributions de 1993, 277; 2002, 5; 2003, 60 et 2005a où ce concept apparaît même dans le titre de l'article en question.

qu'elles soient de nature matrimoniale, génétique, ecclésiastique, politique, agricole, économique, écologique ou autre.

4.2. Des données d'atlas aux cartes-DM synthétiques, en passant par la taxation, la mensuration de similarités et de distances et les défis d'une visualisation appropriée

→ Figure 3

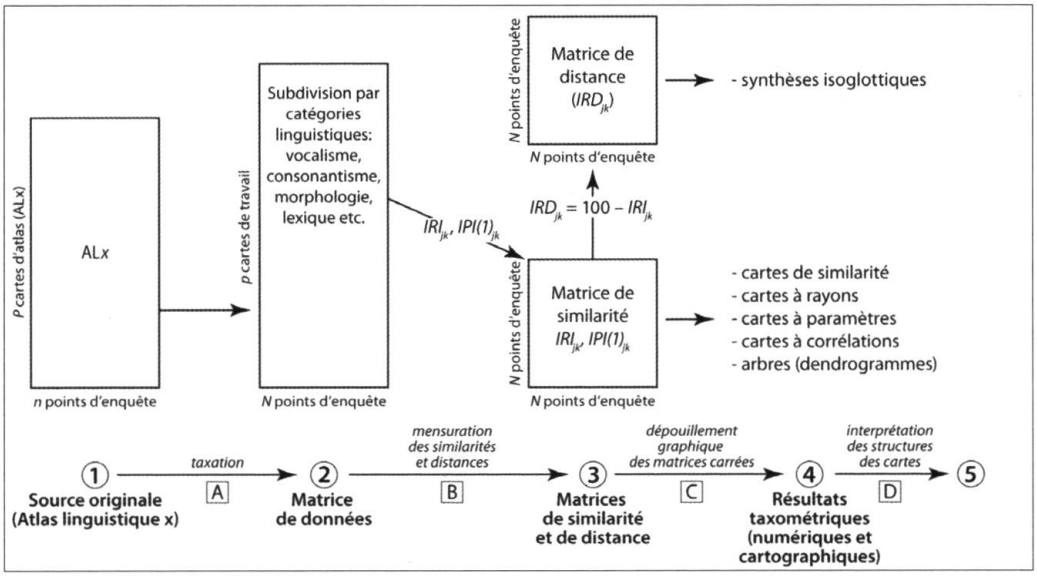


Figure 3: Diagramme de la chaîne méthodique utilisée par l'« Ecole Dialectométrique de Salzbourg »

Les méthodes de la DM-S sont rangées le long d'une chaîne qui s'étend entre les données brutes (figure 3: position 1) et les différentes visualisations synthétiques finales (figure 3: position 5).

L'étape A comprend le traitement typologisant (appelé à Salzbourg *taxation*) des données d'atlas; elle correspond largement à la vieille technique de la typisation des données originales d'un atlas linguistique examiné, appliquée couramment au sein de la romanistique depuis la période de Jules Gilliéron: voir, à titre d'exemple, les cartes typisées (en couleurs) présentées en appendice dans Jaberg 1908. Précisons qu'une opération indispensable pour ce genre de travail était – et continue d'être – le recours à des cartes muettes du réseau d'atlas en question[40].

[40] Le livre de G. Brun-Trigaud, Y. Le Berre et J. Le Dû de 2005, consacré au dépouillement de l'ALF moyennant l'établissement de cartes typisées en couleurs, en fournit quelques 500 spécimens très éloquents.

Le produit final de l'étape A est la matrice de données (N points d'enquête × p CT) dont le contenu se situe, du point de vue métrologique, au niveau de l'échelle nominale (ou: cardinale).

Avec l'étape B commence le volet 'innovateur' de la DM-S où il s'agit de mesurer, à l'aide de différents indices statistiques (ou: métriques), d'abord les similarités et ensuite aussi les distances qui existent entre les N vecteurs d'attributs de la matrice de données. Dans cet article, nous n'utiliserons que l''Indice Relatif d'Identité' (IRI_{jk}) pour la mensuration des similarités, et l''Indice Relatif de Distance' (IRD_{jk}) pour celle des distances entre deux 'vecteurs d'attributs', symbolisés par les lettres *j* et *k*, qui correspondent à autant de points d'enquête de la matrice de données.

Ces deux indices sont des métriques très 'robustes' qui ont prouvé leur grande utilité pour la DM à d'innombrables reprises. Tous les deux sont du type 'adansonien': ceci signifie qu'ils traitent tous les taxats de la matrice de données sur un pied d'égalité (numérique) et ne pratiquent aucune pondération quantitative. Les valeurs calculées à l'aide de ces deux indices sont complémentaires selon la formule suivante: $IRI_{jk} + IRD_{jk} = 100$.

Des présentations plus détaillées se trouvent dans Goebl 1984 I, 74 (en allemand); pour des explications plus fouillées, en français voir Goebl 1981, 2002, 2003 et Goebl / Smečka 2014 (toujours: *passim*); en anglais: 1993 et 2006 (tous les deux: *passim*).

Le but de l'étape C consiste dans la visualisation du contenu numérique des matrices de similarité et de distance. C'est un processus délicat qui doit toujours obéir aux impératifs de la géographie linguistique et ne devrait jamais se faire sous les auspices d'une discipline autre que la (géo)linguistique. Il s'agit donc d'extraire, des deux matrices carrées calculées, certains secteurs censés être utiles ou féconds pour les propos de la géolinguistique, pour les convertir ensuite en cartes ou graphiques appropriés. Ce faisant, la DM-S s'efforce depuis toujours d'opérer un choix équilibré et judicieux et de ne pas confronter – dans un réflexe irréfléchi de modernisme technologique – ses lecteurs avec toute la richesse, souvent déroutante, de la panoplie de la visualistique moderne.

L'outil visualisateur de base de la DM-S est le logiciel VDM ('Visual DialectoMetry'). Il permet – grâce à la possibilité de modifier rapidement la finesse, les teintes et les modalités de l'étalement spatial du symbolisme chromatique utilisé – d'adapter les graphiques-DM produits aux besoins cognitifs et aux capacités physiologiques de l'œil de l'observateur.

L'étape finale D constitue le plus grand défi scientifique parce qu'elle sert à l'interprétation linguistique des cartes préalablement générées, ce qui revient, en dernière analyse, au mariage heureux du rendement heuristique de l'analyse-DM à proprement parler avec les prérequis théoriques et pratiques de la question géolinguistique traitée.

Sur la question, cf. encore Goebl 1984 I, 86*sqq.*; pour des visualisations réalisées en noir-et-blanc: cf. Goebl 1981 et 2006 (*passim*); et en couleurs: 1993, 2002, 2003 (et après).

4.3. La carte de similarité (CS)

→ Paragraphes 5.1.1, 5.2.1, 5.3.1 ainsi que cartes 3-5, 11-13, 19-21 et 27-28

La CS est l'outil heuristique central de la DM-S. C'est un graphique relationnel – centré sur un point de repère à définir au préalable – dont l'utilité est double:

- montrer la position géotypologique (relationnelle), à l'intérieur du réseau d'atlas examiné, du potentiel dialectal ('dialecticité') d'un seul des N points d'enquête, et
- visualiser l'étonnante régularité géographique et quantitative avec laquelle les scores de similarité s'étalent autour du point d'enquête tout en déclinant de façon continue dans l'espace.

C'est que les CS ont permis de découvrir la relation (négative) qui existe entre la distance kilométrique et la similarité linguistique par rapport à un point de repère donné: ceci signifie que la similarité linguistique décroît continuellement avec l'augmentation de la distance kilométrique du point de repère choisi. Évidemment, les modalités de ce déclin varient d'une CS à l'autre, tout en dépendant de la position du point de repère au sein du réseau d'atlas en question.

Pour les deux dimensions de l'espace, cette relation a le statut d'une loi géo-linguistique, comparable aux lois phonétiques («Lautgesetze»), découvertes par les néo-grammairiens de Leipzig à la fin du XIXe siècle, qui, elles, régissent le déploiement des faits linguistiques le long de l'axe (unidimensionnel) du temps.

Quant à leur nature cartographique, les CS appartiennent à la classe des cartes choroplèthes. Un tel type de carte se sert d'un fond de carte polygonisé[41] et d'un certain nombre de couleurs graduées (ordonnées selon la logique du spectre solaire) pour symboliser, par le biais du contraste iconique

[41] Pour tous les problèmes relatifs à la construction et à l'application de réseaux polygonisés, cf. Okabe / Boots / Sugihara 1992.

qui s'instaure entre les différentes plages à couleurs, la variation analogue des données numériques répertoriées dans la matrice de similarité.

Chaque CS a les propriétés suivantes:

- Elle visualise le contenu numérique d'un des N vecteurs de la matrice de similarité. Il existe donc, pour une seule matrice de similarité (aux dimensions N × N), le nombre N de CS.
- Chaque CS dispose d'un point de repère[42] dont le polygone reste toujours en blanc. Ceci signifie que la visualisation à proprement parler d'une CS donnée ne concerne que N-1 polygones.
- Sur toutes les CS, la répartition des couleurs (ou teintes) se fait par voie algorithmique. À cet effet, VDM offre trois solutions différentes, marquées par les sigles MINMWMAX, MEDMW et MED. Ces algorithmes permettent de conférer, au même potentiel numérique, trois sortes de visualisations qui se distinguent par le caractère plus ou moins accidenté de leurs profils choroplèthes.
- Le coloriage des CS peut varier entre 2 et 20 teintes différentes, facilement réglables dans VDM.

Les 36 cartes-DM de cet article utilisent uniquement l'algorithme MEDMW qui crée – des deux côtés de la moyenne arithmétique (MA) tirée des N-1 scores de similarité – des intervalles dont l'ampleur – définie par le nombre de polygones marqués par la même couleur – est sensiblement égale. Cet algorithme génère des profils qui sont, d'un côté, plus équilibrés et, de l'autre, plus accidentés.

Signalons encore que les scores situés au-dessus de la MA (= intervalles 4-6) seront toujours visualisés par des couleurs 'chaudes' (intervalle 4: jaune, 5: orange, 6: rouge), alors que les scores situés au-dessous de la MA (= intervalles 1-3) recevront un coloriage 'froid' (intervalle 1: bleu foncé, 2: bleu moyen, 3: vert).

Les CS comprennent – tout comme les autres graphiques choroplèthes et isarithmiques de cet article – trois parties: le cartogramme à proprement parler, la légende numérique (en bas, à gauche) et l'histogramme (en bas, à droite). Dans la légende, l'on trouve les indications relatives aux seuils numériques des n intervalles et au nombre des points d'enquête répertoriés dans chaque intervalle.

Précisons encore que toutes les cartes-DM synthétiques de cet article disposent de (seulement) *six* paliers chromatiques. En général, cette diversification iconique est suffisamment fine pour permettre une bonne reconnaissance des structures globales des cartes choroplèthes et isarithmiques respectives.

---

[42] Le score-IRI attribué au point de référence (j), est par définition réflexif (IRI_{jj}) et s'élève, de ce fait, toujours à 100 (%).

Au vu de nos cartes choroplèthes, il ne faut pas confondre les pourtours des aires marquées par une des six couleurs étalées sur les polygones, avec des 'limites (ou: frontières) linguistiques'. Ils correspondent plutôt aux notions géo- ou cartographiques des isohypses, isobares ou isohyètes[43] qui représentent autant de variantes du concept général de l'isoligne.

Les histogrammes, comprenant toujours six barres verticales à hauteurs inégales, servent à visualiser le caractère statistique des distributions de fréquence en question. Ils montrent surtout si ces dernières sont symétriques, ou non, et si elles disposent d'un seul ou de plusieurs sommets ou même d'endroits de discontinuité.

L'analyse de la silhouette de l'histogramme est indispensable pour la définition (et la bonne intelligence) des 'cartes à paramètres': voir les cartes 7, 15, 23, 33 et 34, toutes relatives aux 'coefficients d'asymétrie de Fisher' (CAF), et le paragraphe 4.5.

Les lecteurs intéressés trouveront des informations supplémentaires dans Goebl 1984 I, 90*sqq.*; 1981 et 1983 (*passim*). Pour le fonctionnement et l'application des algorithmes de visualisation, cf. 1984 I, 93*sqq.* ainsi que 1981 et 1983 (*passim*).

4.4. L'analyse des distances interponctuelles (cartes isoglottiques)

→ Paragraphes 5.1.2, 5.2.2, 5.3.2, 6.2, ainsi que cartes 6, 14, 22, 31 et 32

L'élément cartographique de base de ce type de cartes n'est plus la superficie d'un polygone, mais l'ensemble des côtés qui l'entourent. Au cours de la visualisation, chacun des côtés du polygone recevra l'empreinte de deux symbolismes graphiques gradués: de l'épaisseur et de la couleur. Le message global de nos cartes isarithmiques[44] dépend donc de l'interaction optique de 343 (AM) ou 453 (ALDC) côtés de polygone, équivalents à autant de scores-IRD interponctuels qui auparavant avaient été extraits de la matrice de distance.

La relation cartographique entre l'épaisseur et la couleur est la suivante: plus les scores-IRD sont grands, plus les côtés de polygone respectifs seront épais et tireront vers le bleu. L'épaisseur des côtés du polygone rappelle donc l'épaisseur des faisceaux d'isoglosses (ou: bourrelets-frontières) de la géogra-

---

[43] Il s'agit, dans les trois cas, de lignes qui, sur une carte de type géographique ou météorologique, relient les points où l'altitude (-hypse), la pression atmosphérique (-bare) et la quantité de précipitations (-hyète) sont égales.

[44] En matière de cartographie, ce genre de cartes, basées sur des éléments linéaires, est appelé (carte) isarithmique. La perception visuelle du profil réticulaire de cartes isarithmiques est plus difficile pour l'œil que celle du profil de cartes choroplèthes qui ont l'avantage de lui présenter une surface continue.

phie linguistique classique. Il en résulte que, de toutes les cartes-DM existantes, la compréhension intuitive des cartes isoglottiques pose le moins de problèmes: voir à ce sujet les superpositions isoglottiques de type traditionnel publiées par le romaniste suédois Arvid Rosenqvist en 1919 (réseau entier de l'ALF) ou par le romaniste autrichien Karl von Ettmayer en 1924 (moitiés nord et sud de l'ALF).

Encore une remarque d'ordre statistique: le rang taxométrique de la synthèse isoglottique est plutôt limité. Regardons à ce propos les cartes 31 et 32: elles reposent respectivement sur 343 et 453 scores de l'IRD. Ces deux effectifs ne correspondent qu'à 3,11 ou 2,52% des scores de distance répertoriés dans les deux matrices de distance à disposition[45]. Cela signifie que la portée classificatoire de cette analyse est plutôt modeste. Évidemment, cette restriction est due au principe de l'interponctualité des mensurations qui, elle, dépend de la considération exclusive de différences linguistiques à très courte distance, tout en écartant toutes les autres possibilités de comparaison.

Sur cette question, voir encore Goebl 1984 I, 183 *sqq.*; 1983 (*passim*).

4.5. Une carte à paramètres: l'analyse des 'coefficients de Fisher' (CAF)

→ Paragraphes 5.1.3, 5.2.3, 5.3.3, 6.3 ainsi que cartes 7, 15, 23, 33 et 34

Dans le paragraphe 4.3., nous avons déjà mentionné le fait que les distributions de fréquence, sur lesquelles reposent les cartes de similarité (CS), servent non seulement à révéler l'étonnante régularité de l'étalement de similarités linguistiques dans l'espace, mais rendent aussi d'autres services exploratoires. C'est que la prise en compte synoptique de leurs 'paramètres caractéristiques' (comme, p. ex., du minimum et du maximum, de la moyenne arithmétique (MA), de l'écart-type, de différents coefficients d'asymétrie, etc.) permet l'analyse et la saisie de certains aspects quantitatifs – et partant aussi linguistiques – du fonctionnement des basilectes (dialectes locaux ou 'locolectes') dans l'espace.

Un des paramètres les plus intéressants du point de vue de la géolinguistique est le 'coefficient d'asymétrie de Fisher' (CAF). Il permet de mesurer le degré de la symétrie (ou de l'asymétrie) d'une distribution de fréquence[46] qui, lui, peut se référer à différentes réalités (géo)linguistiques.

---

[45] En voici les calculs respectifs: dimensions des matrices de distance (N × N): 149^2 et 190^2. Nombre des scores de distance (IRD_{jk}) stockés dans ces deux matrices et prêts à être utilisés (à calculer suivant la formule N × (N – 1)/2): 11 026 et 17 955. Ensuite: 343 (valeurs interponctuelles): 11 026 = 3,11%, ainsi que 453: 17 955 = 2,52%.

[46] Ici, il s'agit de distributions de similarité.

Prenons l'exemple des cartes 3 (CS depuis *Barcelone*, P. 57) et 4 (CS depuis *Lleida*, P. 87). Mis à part le fait que les histogrammes des deux CS offrent des silhouettes relativement semblables, l'on constate – en contrôlant la position des scores de l'IRI des deux côtés de la moyenne arithmétique (MA) respective – que, sur la CS 'orientale' (carte 3), 79 des 148 scores en question (= 53,37%) dépassent la MA, alors que sur la CS de type 'occidental' (carte 4), ce sont seulement 54 scores-IRI sur 148 (= 36,48%) qui, vu leur position au-dessus de la MA, accusent la présence d'une bonne, voire excellente similarité linguistique avec le reste du réseau. En associant le taux de la similarité linguistique à celui de l'interaction communicative en général, l'on est donc en droit de dire que l'insertion communicative du vecteur d'attributs du P. 57 dans le réseau-AM est meilleure que celle du P. 87.

Pour une bonne compréhension de ces différences, il faut se rappeler le fait suivant: chacun des 149 vecteurs d'attributs du réseau-AM est caractérisé par la participation à un grand nombre d'aires taxatoires dont la taille est très variable[47]. Parmi ces nombreuses aires, il peut y en avoir dont l'étendue est petite (= micro-chorique), moyenne (= méso-chorique) ou grande (= méga-chorique).

La spécificité communicative d'un vecteur d'attributs donné dépend donc de la combinaison (ou: co-présence) particulière de ces trois types d'aires. Plus un tel vecteur comprendra des aires méso- et mégachoriques, plus l'impact communicatif en sera grand. Or, la présence de beaucoup d'aires plutôt grandes est le résultat de phénomènes d'irradiation et de propagation plus ou moins intenses, tous survenus dans le passé. En revanche, un surplus d'aires microchoriques est dû à des événements (et comportements) contraires, caractérisés par une certaine immobilité ou un conservatisme accru.

Un indice capable de saisir globalement cette combinaison particulière devrait donc permettre de mesurer le caractère 'dynamique' d'un basilecte donné. Or, nous savons depuis longtemps que cette mensuration est assurée de façon idéale par le 'coefficient d'asymétrie de Fisher' (CAF)[48] qui compte

---

[47] Pour mieux comprendre le fait de la participation (ou appartenance) aux différentes aires taxatoires, il faut se rendre compte de l'évidence suivante: chacun des N vecteurs d'attributs de la matrice de données comprend le nombre p de cellules. Or, chacune de ces p cellules représente un élément d'une aire taxatoire. Comme la taille de ces aires peut varier considérablement, la participation d'un vecteur donné à ces aires est soumise à de fortes fluctuations.

[48] Cf. Goebl 1981, 394-401 et 1984 I, 150-153 (présentation de la formule du CAF et explication de ses fonctions). Ronald A. Fisher (1890-1962) était un biologiste et statisticien anglais.

parmi les meilleurs indices pour la saisie quantitative de l'asymétrie d'une distribution de fréquence.

La particularité (géo)linguistique captée et mesurée par le CAF a été appelée par nous d'abord «Sprachausgleich» en allemand, et baptisée par la suite 'compromis (ou brassage) linguistique' en français.

Quand une distribution de fréquence (ou de similarité) est parfaitement symétrique, le score du CAF est de 0, alors qu'une asymétrie vers la gauche (= cumul des scores au-dessous de la MA) entraîne des scores positifs du CAF, et une asymétrie vers la droite (= cumul des scores au-dessus de la MA) des scores négatifs.

Sur les cartes 7, 15, 23, 33 et 34, les scores minimaux (et négatifs) sont marqués en bleu (foncé et moyen), alors que les scores maximaux (et positifs) sont visualisés en rouge. La signification linguistique de ces deux couleurs (et de leurs teintes) est la suivante:

- Bleu: zone caractérisée par un grand dynamisme linguistique et beaucoup de contacts linguistiques (passifs et actifs), déployés à plus ou moins grande distance. Présence de nombreuses aires taxatoires méga- et mésochoriques.
- Rouge: zone caractérisée par un comportement linguistique conservateur, défensif ou même récessif, et aussi par un nombre (très) réduit de contacts linguistiques établis à grande distance. Présence de beaucoup d'aires taxatoires microchoriques.

La visualisation choroplèthe du CAF résulte de la mise en carte synoptique de N scores-CAF, émanant de N distributions de similarité qui, elles, sont stockées dans les N vecteurs d'une matrice de similarité carrée[49].

Généralement, les profils choroplèthes des cartes-CAF offrent des structurations très claires. Leur importance linguistique est, avant tout, d'ordre diachronique. Précisons que ce type de carte constitue, à côté de la carte à corrélations, l'une des pièces fortes de la DM-S.

La valeur et l'importance du CAF ressortent le mieux de nos études relatives au réseau complet de l'ALF: voir pour cela nos contributions de 2000, 2002 et 2003. Nous renvoyons également à nos travaux de 2004a, 2005b, 2006a, 2007a, 2007b et 2008a ainsi qu'à Goebl / Smečka 2014.

4.6. L'analyse dendrographique (ou: arborescente)

→ Paragraphes 5.1.4, 5.2.4, 5.3.4, 6.4, ainsi que cartes 8, 16, 24, 35 et 36

Ce genre d'analyse utilise la structure ramifiée (et binaire) d'un arbre, pourvu de N 'feuilles' et calculé au moyen d'algorithmes statistiques complexes, pour saisir les relations (généalogiques) de parenté qui existent entre

[49] Voir la figure 3.

les différents dialectes locaux. Comme le recours au concept et à la métaphore de l'arbre dispose d'une longue (et heureuse) tradition en matière de linguistique, rien n'empêche son application aussi en matière de géolinguistique, à condition toutefois que l'outil heuristique de l'arbre soit toujours combiné avec celui de la carte.

La construction algorithmique de schémas dendrographiques (ou arborescences) constitue, au sein de la classification numérique, une vieille pratique, tant sur le plan purement statistique que sur celui de leur application à des fins classificatoires. Il en résulte que la taxométrie moderne dispose d'un nombre relativement grand d'algorithmes dendrographiques, souvent très différents du point de vue mathématique, dont l'utilité classificatoire est également très variable. C'est pourquoi le dialectométricien-classificateur se trouve dans la nécessité d'effectuer, devant cette richesse méthodologique, un choix judicieux, tout en respectant toujours les prérequis et finalités de la géolinguistique.

Dans la foule des algorithmes dendrographiques aujourd'hui disponibles, ce sont ceux qui obéissent aux principes de la 'classification ascendante hiérarchique' (CAH) qui sont les plus utiles pour les propos de la DM et de la géographie linguistique. Ils procèdent, en se basant sur les données d'une matrice de similarité carrée (N × N), par une suite d'agglomérations binaires d'éléments aussi similaires que possible, tout en commençant au niveau des N 'feuilles' de l'arbre pour finir par en rejoindre la racine (ou: le tronc). Métaphoriquement parlant, il s'agit donc d'une 'ascension' à l'intérieur d'un arbre retourné, du bas (à partir de N 'feuilles') vers le haut (vers une seule 'racine'). Du point de vue graphique, les agglomérations se traduisent par la formation de ramifications (ou: bifurcations) toujours binaires qui s'arrangent sous la forme de hiérarchies disjonctives.

Un des algorithmes les plus utiles en matière de DM a été proposé, en 1963, par Joe Ward, Jr.[50] L'avantage en réside dans sa capacité de créer, à l'intérieur des ramifications ('branchages' ou 'ramages') de ses arbres, des groupements à la fois très homogènes vers l'intérieur ('intra-group similarity') et très hétérogènes vers l'extérieur, c'est-à-dire par rapport à des groupements environnants ('inter-group distance').

En matière de DM, l'application courante de la CAH doit parcourir deux étapes: d'abord, il faut identifier, à l'intérieur de l'arbre calculé, certains branchages (appelés à Salzbourg 'dendrèmes'), jugés 'significatifs' du point de vue linguistique, et d'en projeter, par la suite, les éléments dans l'espace. De cette

[50] Joe H. Ward, Jr. (1926-2011) était un statisticien américain.

projection résultent, sur la carte polygonisée du réseau en question, des aires en général très cohérentes (en termes de la DM-S: 'chorèmes') dont la distribution sur la carte offre un grand intérêt linguistique.

En matière d'analyse dendrographique, le logiciel VDM permet non seulement d'utiliser plusieurs algorithmes dendrographiques différents, mais aussi de colorier à discrétion les arborescences ainsi générées, et de projeter, par la suite, les dendrèmes ainsi définis sur la carte. À Salzbourg, ce processus est appelé 'spatialisation' (du contenu dendrématique de l'arbre).

L'interprétation linguistique des arbres-DM peut être faite dans la direction de la racine vers les feuilles (interprétation diachronique) ou dans le sens inverse (interprétation synchronique).

L'interprétation diachronique part de l'idée, chère aux tenants de la lexicostatistique et de la glottochronologie, que tout domaine linguistique avait été, d'entrée de jeu, 'unitaire' et partant homogène, et que les fragmentations spatiales, observables de nos jours, sont le résultat d'une suite de partages postérieurs.

L'interprétation synchronique vise à l'analyse des hiérarchies et imbrications géographiques qui existent entre les différents branchages de l'arbre et leurs pendants chorématiques.

Sur ces questions, cf. encore Goebl 1984 I, 172*sqq.* et 2003, 84-89. Pour une description détaillée du fonctionnement des algorithmes dendrographiques, cf. nos contributions de 1991 et 1992 (en allemand). Pour l'algorithme de Ward, nous renvoyons aux explications qui se trouvent dans trois grands manuels de classification numérique: Sneath / Sokal 1973, 204*sqq.* (en anglais); Bock 1974, 407*sqq.* (en allemand) et Chandon / Pinson 1981 (en français), 123*sq.*

4.7. Les cartes à corrélations

→ Paragraphes 5.1.5, 5.2.5, 5.3.5, 6.5 ainsi que cartes 9-10, 17-18, 25-26, 37 et 38

La DM corrélative sert à analyser le degré de con- et divergence qui existe entre deux stratégies différentes de gérer le même secteur de l'espace, et d'en tirer des leçons linguistiques, en l'occurrence surtout d'ordre diachronique. Évidemment, cette analyse admet l'existence de 'forces (ou énergies) ordonnatrices', qui opèrent dans l'espace tout en résultant de l'action d'agents, anthropiques ou non, censés gérer ce dernier de forme non chaotique. Dans cette perspective, les accouplements en question peuvent comprendre deux catégories linguistiques (p. ex. la phonétique et le lexique), deux

mensurations-DM diverses (p. ex. la mensuration-IRI et la mensuration-IPI[51]) ou la tension qui s'instaure entre la gestion basilectale de l'espace par l'homme, d'un côté, et celle due à des relations non-anthropiques (p. ex. par la géométrie euclidienne), de l'autre.

Du point de vue mathématique, la DM corrélative requiert la mise en place, pour chacune des deux énergies à comparer par voie de corrélation, de matrices (de similarité, etc.) carrées, pourvues, toutes les deux, des mêmes dimensions extérieures (N × N). Moyennant ces deux matrices carrées, il est désormais possible de comparer quantitativement – par le biais de l'application d'indices de corrélation appropriés – N paires de vecteurs (disposant chacun de N scores de similarité, etc.), d'en extraire par la suite N scores de corrélation, pour les visualiser, finalement, dans la forme habituelle d'une synopse choroplèthe.

L'indice de corrélation le plus approprié à cet effet a été proposé successivement par Auguste Bravais (1811-1863, physicien et minéralogiste français) et Karl Pearson (1857-1936, polymathe d'origine anglaise): de nos jours, il s'appelle 'coefficient de corrélation de Bravais-Pearson' [r(BP)]. Ce coefficient varie entre -1 et +1 et saisit la liaison linéaire qui existe entre deux variables quantitatives.

Pour la réalisation du processus de corrélation, le logiciel VDM dispose d'un module approprié à partir de 2004. Depuis cette date, il s'est avéré maintes fois que l'application de ce module mène à la détection de structures spatiales étonnamment bien ordonnées, lesquelles, en plus, se prêtent facilement à des analyses et interprétations linguistiques.

L'intérêt linguistique des cartes à corrélations réside, à l'instar de la synopse choroplèthe des CAF, surtout dans la diachronie. C'est que les structurations spatiales des cartes à corrélation constituent, en dernière analyse, un miroir très fin des dynamismes qui se sont déroulés dans l'espace, avec ou sans la responsabilité directe des humains, dans un passé plus ou moins lointain.

La contribution fondatrice pour la DM corrélative est de 2005(a). Après, l'on trouve des applications et des discussions de la DM corrélative dans presque toutes nos publications postérieures à 2005; voir, en outre, nos travaux de 2007c et 2008b (en allemand), centrés exclusivement sur la DM corrélative.

---

[51] Pour la formule de l'IPI ('Indice Pondéré d'Identité') et son utilité dialectométrique, cf. Goebl 1984 I, 83-86 et 1987 (*passim*).

5. Présentation et interprétation des résultats-DM par catégories linguistiques (*indicatif*, *subjonctif*, *infinitif*)

Le but central de ce paragraphe est de fournir aux lecteurs une idée visuelle approfondie de l'aménagement spatial du domaine catalan tel qu'il est opéré par différentes catégories verbo-morphologiques. Nous avons choisi à cet effet les données-LaFV relatives aux catégories de l'*indicatif*, du *subjonctif* et de l'*infinitif*. Comme cette sélection ne peut constituer qu'un choix plus ou moins arbitraire, elle aurait pu frapper aussi d'autres groupements verbo-morphologiques, répertoriés dans le tableau 4. Il faut toutefois souligner que, dans ce cas, les résultats-DM seraient sensiblement les mêmes.

Tous les échantillons graphiques présentés dans ce chapitre reposent sur le corpus RÉD dont l'atout central est la minimisation de lacunes (cases vides, données manquantes), dont on sait que leur présence massive dans la matrice de données respective peut entraîner des distorsions très gênantes des calculs-DM. Il en résulte que le corpus RÉD, riche de 1 639 CT (à la place des 3 600 CT du corpus TOT), comprend des CT aussi complètes que possible.

Ce but a été atteint par la mesure suivante: le taux des polygones lacunaires ne doit jamais dépasser les 10% (de 149). C'est pourquoi au moins 134 (= 149 – 15) des 149 'cases polygonales' de chaque vecteur-CT doivent contenir des informations linguistiquement valables. Voir, en outre, le tableau 5 et la figure 1 qui nous renseignent sur la polynymie interne des composantes du corpus RÉD: elle va de 2 à 47 taxats par CT.

Nous avons choisi, pour la caractérisation-DM des trois sous-corpus de base (*indicatif*, *subjonctif* et *infinitif*), cinq méthodes-DM très 'expressives' dont le fonctionnement et l'utilité exploratoire ont été expliqués dans le paragraphe précédent: il s'agit de l'analyse de la similarité et des distances interponctuelles, du calcul des coefficients d'asymétrie de Fisher (CAF), de l'analyse dendrographique et de l'analyse corrélative. Les mêmes méthodes-DM seront utilisées aussi dans la section suivante (6) où il s'agit de comparer, par voie-DM, le total des données de LaFV avec celles des volumes I-IV de l'ALDC.

*5.1. Les formes verbales de l'*indicatif

Pour les dimensions du corpus de base (997 CT), voir les tableaux 4 et 5 ainsi que la figure 1. Soulignons encore que ces 997 CT contiennent 13125 aires taxatoires (et autant de taxats) – dûment enchevêtrées entre elles – et aussi que leur polynymie interne varie entre 2 et 47 taxats (ou aires taxatoires) par CT.

Encore deux remarques techniques: a) l'indice de similarité utilisé (IRI_{jk}) est l'indice standard de la DM-S; b) la visualisation de toutes les cartes de cet article s'opère moyennant l'algorithme de visualisation MEDMW dont la vertu principale consiste dans la production de profils choroplèthes équitablement structurés.

Vu l'enchevêtrement 'à brève distance' – et non pas 'en profondeur' – des aires taxatoires du corpus-LaFV, l'utilisation de l'algorithme standard de la DM-S – MINMWMAX – fournit automatiquement des clichés où une seule couleur (bleu foncé: intervalle 1) occupe régulièrement plus d'un tiers des 148 (= N-1) polygones disponibles. Ce fait engendre des profils choroplèthes très déséquilibrés dont la structuration interne réelle est largement occultée. Il est donc préférable de se servir d'un algorithme dont la vertu d'intervallisation est 'plus équitable'.

Le propre de l'algorithme MEDMW est de créer, des deux côtés de la moyenne arithmétique, des classes (groupes, intervalles) sensiblement égales quant au nombre de leurs éléments. Voir, à titre d'exemple, la légende numérique de la carte 3: l'on y constate que chacun des intervalles 1-3, situés au-dessous de la MA de 44,06 (%), dispose de 23 polygones, alors que les effectifs des intervalles 4-6, tous situés au-dessus de la même MA, sont respectivement de 25, 27 et 27 polygones.

Soulignons encore que les 997 CT du sous-corpus de l'*indicatif* se réfèrent aux *six* personnes des quatre temps suivants: *présent*, *imparfait*, *futur* et *conditionnel présent*.

5.1.1. L'analyse de la similarité (cartes de similarité – CS)

→ Paragraphe 4.3 et cartes 3-5

Nous ne présentons que trois des 149 CS à disposition dans la matrice de similarité calculée sur la base de 997 CT. Leurs points de repère se situent respectivement à *Barcelone* (carte 3), *Lleida* (carte 4) et *Palma de Mallorca / Ciutat* (carte 5). C'est ainsi qu'il est possible d'étudier les différences qui s'instaurent, du point de vue de la DM, entre les trois grands domaines linguistiques du catalan: parlers orientaux, parlers (nord-)occidentaux et parlers des îles.

L'on peut qualifier le profil choroplèthe de la carte 3 comme étant typiquement 'oriental' et celui de la carte 4 comme étant typiquement 'occidental', alors que le profil de la carte 5 s'aligne, grosso modo, sur celui de la carte 3, tout en illustrant la grande cohérence verbo-morphologique interne des îles par rapport aux normes verbo-morphologiques de Palma / Ciutat.

Les CS relatives aux sous-corpus du *subjonctif* et de l'*infinitif* (voir les cartes 11-13, ainsi que 19-21) fournissent des résultats très similaires.

5.1.2. L'analyse des distances interponctuelles (carte isoglottique)

→ Paragraphe 4.4 et carte 6

La syntaxe iconique de la carte 6 repose sur la visualisation de 343 scores de distances, calculés à l'aide de l'indice de distance standard – IRD_{jk} – de la DM-S. Cette visualisation s'opère le long d'autant de côtés de polygone tout en les altérant le long de deux variables cartographiques: de l'épaisseur et de la couleur. Ceci signifie que l'épaisseur des côtés de polygone augmente au même rythme que la taille des scores-IRD_{jk} respectifs, et aussi que cette augmentation conditionne le changement gradué du coloriage du rouge (associé à l'épaisseur minimale) au bleu (correspondant à l'épaisseur maximale).

Le profil isarithmique de la carte 6 montre, par le biais des faisceaux isoglottiques épais et bleus, les subdivisions intra-catalanes bien connues (d'ouest en est): parlers de la frange occidentale, parlers occidentaux, parlers orientaux et un morcellement isoglottique poussé, surtout à Majorque[52].

5.1.3. L'analyse des coefficients d'asymétrie de Fisher (CAF)

→ Paragraphe 4.5 et carte 7

L'analyse du CAF sert, avant tout, à saisir les dynamiques – entendues comme enchevêtrement particulier d'un grand nombre d'aires taxatoires soit expansives (ou grandes) soit récessives (ou petites) – qui sous-tendent un réseau géolinguistique donné. En règle générale, la couleur bleu renvoie à une dynamique majeure, alors que la couleur rouge symbolise un certain conservatisme des parlers en question.

Au vu de la carte 7, l'on peut donc dire que les parlers orientaux déploient, face au reste du domaine catalan, un dynamisme verbo-morphologique très marqué, alors que les parlers du catalan nord-occidental (proches des Pyrénées) et ceux de l'extrême sud – tous marqués en rouge sur la carte 7 – adoptent, par contre, une position conservatrice. Sur le plan de la morphologie verbale, le pivot de l'irradiation linguistique se trouve donc à l'intérieur de la Vieille Catalogne. Nous rencontrerons la même situation sur les cartes relatives au *subjonctif* et à l'*infinitif*.

---

[52] Dans ce contexte, il ne faut pas oublier les synthèses isoglottiques entreprises sur le modèle de nos méthodes interponctuelles déjà en 1983 (et reprises en 1992) par Lluís B. Polanco i Roig: voir les cartes respectives: 1983, 29-30; 1992, 25-26, publiées à nouveau dans Adam Aulinas 2006, 35-36.

5.1.4. L'analyse dendrographique

→ Paragraphe 4.6 et carte 8

La grande utilité-DM de l'analyse dendrographique réside dans la possibilité de projeter le bilan classificatoire de l'arbre (binaire et hiérarchisé) dans l'espace ('spatialisation'). C'est ainsi qu'aux 'dendrèmes' de l'arbre font écho, sur la carte, autant de 'chorèmes', pourvus – tous les deux – des mêmes couleurs. Le logiciel VDM permet de réaliser ce transfert pour tous les niveaux (étages) de la structure binaire de l'arbre, en commençant – près de la 'racine' (ou du 'tronc') de l'arbre – avec deux dendrèmes-chorèmes, pour finir – au niveau des 'feuilles' de l'arbre – avec N dendrèmes-chorèmes.

Le même logiciel permet également le calcul et l'application de plusieurs algorithmes dendrographiques parmi lesquels l'algorithme 'WARD' (proposé en 1963 par Joe Ward, Jr.), a prouvé, plus d'une fois, sa grande utilité pour les propos de la géolinguistique. C'est pourquoi toutes les analyses arborescentes de cet article reposent sur lui. Sa grande vertu taxométrique réside dans sa capacité de former des groupements très cohérents à l'intérieur.

Pour permettre une bonne comparaison réciproque de toutes nos analyses dendrographiques, nous avons limité la finesse de l'analyse dendrématique à six classes tout en conférant aux dendrèmes-chorèmes situés dans les mêmes parages un coloriage identique. Ceci n'empêche, bien sûr, que les plages ainsi coloriées varient, à un degré variable, d'une carte à l'autre. En général, nous avons utilisé les couleurs chaudes (rouge, orange, jaune) pour marquer les zones du catalan oriental, et les couleurs froides (marron, vert et bleu) pour caractériser les parlers du catalan occidental.

La carte 8 montre une partition très claire (et cohérente) du domaine catalan en général: les zones orientales, occidentales et insulaires ressortent avec netteté, de même que leurs subdivisions internes.

La carte 16 (relative au *subjonctif*) ne s'en écarte que très peu, alors que la carte 24 (relative à l'*infinitif*) accuse déjà une structure dendrographique quelque peu différente.

5.1.5. L'analyse corrélative

→ Paragraphe 4.7 et cartes 9 et 10

La DM corrélative permet de corréler deux matrices de similarité pourvues des mêmes dimensions extérieures (N × N), et de spatialiser du même coup les N scores de corrélation ainsi calculés. L'indice de corrélation utilisé à cet effet est le 'coefficient de corrélation (linéaire) de Bravais-Pearson' [r(BP)]. Les valeurs en oscillent entre -1 et +1. Son utilité heuristique en

matière de géolinguistique dépend du résultat choroplèthe de la spatialisation des N valeurs-r(BP) issues des calculs de corrélation. Par son coloriage, la carte en question nous renseigne si les deux 'dimensions' (ou: 'énergies, forces') linguistiques (ou non) corrélées aménagent l'espace en question de façon con- ou di-vergente. Dans ce contexte, le concept de convergence renvoie à une certaine harmonie (spatiale) entre les 'forces linguistiques' en lice, alors que le concept de divergence évoque l'idée de la dissonance (spatiale) entre ces deux facteurs.

La carte 9 se réfère à la corrélation-r(BP) qui peut être relevée entre les similarités linguistiques (valables pour les 997 CT du corpus de l'indicatif) et les proximités géographiques, qui, elles, peuvent être calculées – à l'aide du théorème bien connu de Pythagore – entre les 149 points du réseau-AM. Chacune des deux 'énergies' corrélées obéit à certaines modalités de propagation dans l'espace, susceptibles de converger (voir les zones en rouge) ou de diverger (voir les zones en bleu).

La carte 9 montre très nettement que les zones de convergence occupent les zones méditerranéennes de la Vieille Catalogne ainsi que celles du Pays Valencien et de quelques coins des îles, alors que les plages de divergence (toutes coloriées en bleu foncé et bleu moyen) embrassent le domaine du catalan nord-occidental. En règle générale, une telle dissonance entre la 'langue' et la 'géographie' renvoie à des frictions linguistiques majeures, survenues dans un passé plus ou moins lointain.

Évidemment, il est possible de calculer aussi le degré de corrélation qui existe entre deux catégories *linguistiques*, en l'occurrence entre les 997 CT de l'*indicatif* et les 451 CT du *subjonctif*: voir la carte 10. L'on y perçoit avec netteté que les gestions de l'espace des deux catégories convergent largement dans les Pyrénées et la Vieille Catalogne alors qu'elles divergent considérablement le long d'un grand tronçon de la frontière traditionnelle entre le catalan oriental et occidental (et aussi dans les îles).

Ce constat renvoie à des frictions structurelles à l'intérieur de la morphologie verbale dont l'origine et la nature seraient encore à déterminer. Toujours est-il que la structuration régulière de la carte 10 nous dit qu'il s'agit d'un phénomène non-fortuit et bien ordonné.

5.2. Les formes verbales du subjonctif

Le sous-corpus du *subjonctif* comprend 451 CT: voir les tableaux 4 et 5, ainsi que la figure 1. La polynymie interne des 451 CT oscille entre 7 et 44 (taxats / CT). La variabilité interne de ce corpus est assurée par l'enchevêtrement particulier de 8980 aires taxatoires différentes. Ceci signifie que – en

ce qui concerne les effectifs numériques – le sous-corpus du *subjonctif* est légèrement en retrait par rapport à celui de l'*indicatif*.

Ajoutons encore que les 451 CT du sous-corpus du *subjonctif* se réfèrent à toutes les six personnes des quatre temps suivants: *présent*, *imparfait*, *futur* et *conditionnel présent*.

5.2.1. L'analyse de la similarité (cartes de similarité – CS)

→ Paragraphe 4.3 et cartes 11-13

Les profils choroplèthes des trois cartes de similarité disposent de structures très nettes qui ressemblent beaucoup à celles des cartes 3-5, relatives à l'*indicatif*. Ce constat vaut moins pour la carte 13 où l'intensité des relations linguistiques entre la Vieille Catalogne et les îles a été considérablement enrayée par rapport à ce qui ressort de la carte 5, relative aux CT de l'indicatif.

Remarque technique: sur les cartes 11 et 13, plusieurs polygones en bleu foncé sont recouverts de hachures blanches, ce qui signale la présence (parallèle) du score minimal des valeurs-IRI_{jk}. La pluralité de scores minimaux s'explique par l'abaissement des valeurs-IRI_{jk} jusqu'à zéro: voir les légendes numériques des deux cartes en question.

5.2.2. L'analyse des distances interponctuelles (carte isoglottique)

→ Paragraphe 4.4 et carte 14

À nouveau, la structure isarithmique de la carte 14 est très proche de celle de la carte 6, relative à l'*indicatif*. À remarquer la largeur de la gamme des valeurs-IRD_{jk}, qui vont de 1 à 98%. La même constatation vaut d'ailleurs pour la carte 6.

5.2.3. L'analyse des coefficients d'asymétrie de Fisher (CAF)

→ Paragraphe 4.5. et la carte 15

Ici encore, la grande similitude face à la carte 7 saute aux yeux. Ceci signifie que la fonction irradiatrice de la Vieille Catalogne vaut également pour les formes verbales du *subjonctif*.

5.2.4. L'analyse dendrographique

→ Paragraphe 4.6 et la carte 16

Malgré quelques affiliations divergentes du côté des parlers occidentaux (voir les dendrèmes-chorèmes coloriés en marron), l'analyse dendrographique aboutit à des résultats très similaires à ceux de la carte 8.

5.2.5. L'analyse corrélative

→ Paragraphe 4.7 et cartes 17 et 18

La carte 17 montre les corrélations qui existent entre la 'langue' (= formes du subjonctif) et l''espace' (= proximités euclidiennes), alors que la carte 18 se réfère aux corrélations entre les 451 CT du *subjonctif* et les 44 CT de l'*infinitif*. Dans les deux cas, la répartition spatiale des zones de convergence (en rouge) et de divergence (en bleu) obéit à des principes spatiaux très apparentés.

À remarquer que, sur la carte 18, les scores de corrélation varient entre -0,01 et 0,93 et n'occupent, grosso modo, que le secteur positif. À remarquer également que la MA des scores-r(BP) est très élevée: elle se situe à +0,83; voir la légende numérique, à gauche, en bas. Ceci signifie que les stratifications spatiales de la variabilité globale du *subjonctif* et de l'*infinitif* ne s'écartent que très peu. La légende numérique de la carte 10 (pour la corrélation entre l'*indicatif* et le *subjonctif*) signale une situation très similaire.

*5.3. Les formes verbales de l'*infinitif

→ Tableaux 4 et 5 ainsi que figure 1

Au sein du corpus RÉD, le nombre des CT relatives aux formes de l'*infinitif* est très modeste: 44. Ceci vaut également pour le total des aires taxatoires respectives (551) et pour les écarts de la polynymie: de 4 à 22 taxats par CT. À ce sujet, il ne faut pas oublier que la structure des infinitifs n'offre que peu de possibilités pour le déploiement de variations phonétiques: en règle générale, celles-ci ne se manifestent qu'aux désinences ou lors d'alternances du radical (du type *valer* et *valdre* "valoir", *venir* et *vindre* "venir", etc.). Néanmoins, il est très intéressant de voir que la majorité des analyses-DM faites à l'aide des formes de l'*infinitif* s'aligne parfaitement sur celles réalisées à l'aide des sous-corpus de l'*indicatif* et du *subjonctif*, pourtant de loin supérieurs à celui de l'*infinitif* quant aux effectifs de leurs CT.

5.3.1. L'analyse de la similarité (cartes de similarité – CS)

→ Paragraphe 4.3 et cartes 19-21

Ces trois mensurations de la similarité linguistique confirment pleinement le contenu et le message des cartes 3-5 (*indicatif*) et 11-13 (*subjonctif*). Les cartes 19-21 recèlent, tout comme les cartes 11 et 13, l'abaissement des scores-IRI_{jk} à zéro et, partant, la prolifération de hachures blanches sur plusieurs polygones marqués en bleu foncé (= intervalle 1).

5.3.2. L'analyse des distances interponctuelles (carte isoglottique)

→ Paragraphe 4.4 et carte 22

La grande ressemblance de la structure isarithmique de cette carte avec celles des cartes 14 et 6 est remarquable.

5.3.3. L'analyse des coefficients d'asymétrie de Fisher (CAF)

→ Paragraphe 4.5 et carte 23

Ici encore, l'étalement spatial des couleurs chaudes et froides correspond largement à ce qui peut être observé sur les cartes 15 et 7. Le même parallélisme vaut pour la signification linguistique de cette carte : le domaine catalan est sous le coup d'une irradiation tous azimuts à partir de la Vieille Catalogne.

5.3.4. L'analyse dendrographique

→ Paragraphe 4.6 et carte 24

Les effets de cette analyse ne s'alignent que partiellement sur ceux que l'on peut observer sur les cartes 16 et 8. Ce fait est dû surtout à l'extension des chorèmes en rose sur le continent et l'effilochement concomitant des plages de la Vieille Catalogne, marquées en rouge. Cet écart n'empêche pas que l'agencement des bifurcations binaires supérieures de l'arbre correspond largement à ce qui a déjà pu être remarqué sur les cartes 16 et 8.

5.3.5. L'analyse corrélative

→ Paragraphe 4.7 et cartes 25 et 26

En dépit de petites divergences dans les détails, les structures choroplèthes – et aussi les significations linguistiques y ayant trait – des cartes 25 et 26 correspondent largement à celles des cartes 17-18 ainsi que 9-10.

6. Présentation comparative de résultats-DM : *morphologie verbale* (selon LaFV) et *lexique-phonétique* (selon ALDC I-IV)

Évidemment, il est très intéressant d'utiliser les données dialectométrisées de LaFV pour une comparaison-DM plus large, mobilisant à cet effet une quantité aussi grande que possible de CT, de même que des catégories linguistiques très divergentes. Pour ce faire, nous nous sommes servis, d'un côté, du corpus TOT (que nous avons tiré des données-LaFV) et, de l'autre, du corpus-DM (que nous avons constitué lors de la dialectométrisation des cartes des volumes I-IV de l'ALDC, réalisée à Salzbourg entre 2007 et 2009) : pour tous les détails de cette minutieuse analyse, voir Goebl 2013b, *passim*.

Bien que cette comparaison ait l'inconvénient de mettre en rapport deux réseaux à maillages différents[53], elle offre néanmoins le grand avantage de confronter deux macro-catégories linguistiques complètement différentes: à gauche la morphologie verbale, marquée par une variation géolinguistique 'sous contrainte', due à l''emprisonnement paradigmatique' des formes verbales (qu'elles soient finies ou non finies) et, à droite, une combinaison bien équilibrée entre la phonétique (vocalique, consonantique et combinée) et le lexique dans toute sa richesse onomasiologique. Il s'agit donc d'une introspection très avancée dans les profondeurs catégorielles du diasystème des parlers catalans.

Nos expériences-DM antérieures, faites avec des données d'atlas très diversifiées, nous disent que les conséquences du décalage temporel entre les données de LaFV (récoltées entre 1900 et 1928) et de l'ALDC (collectées entre 1964 et 1978) sont de peu d'importance pour ce genre d'études et peuvent donc être tranquillement écartées.

Pour le corpus TOT, voir les tableaux 4 et 5, ainsi que la figure 2 (graphique du milieu); pour le corpus tiré de l'ALDC I-IV, voir également la figure 2 (graphique inférieur) et les indications présentées dans Goebl 2013b, 91-92.

Signalons en outre que le corpus TOT, riche de 3600 CT, a l'avantage d'offrir un nombre très grand de CT, mais aussi l'inconvénient de contenir beaucoup de lacunes (cases vides, données manquantes) dans la matrice de données respective.

Du point de vue linguistique, le même corpus offre, en outre, la particularité suivante: dans les cas où les sujets de Mgr Alcover lui ont fourni plus d'une seule réponse (réponses multiples), notre corpus TOT contient les réponses réputées comme étant 'les plus anciennes' du point de vue linguistique.

Comme le nombre des réponses multiples dans certaines localités est très élevé, il a été décidé, au cours de la dialectométrisation intégrale des données de LaFV, faite à Salzbourg en 2007, de prendre en considération la présence parallèle d'un nombre maximal de quatre réponses par question et d'insérer séparément les items respectifs dans autant de matrices de données différentes, échelonnées selon l'ancienneté linguistique supposée des formes recueillies sur le terrain[54].

Soit dit en passant, les différences entre les structures de profondeur de ces quatre matrices de données particulières sont minimes du point de vue

---

53 Le réseau-AM comprend 149 points d'enquête alors que le réseau de l'ALDC en contient 190 (+ quatre points artificiels).

54 Voir aussi les informations données dans la note 18.

synthétique de la DM: ce fait ressort avec netteté de calculs de corrélation croisés, appliqués à ces quatre corpus.

Signalons encore une différence fondamentale entre le corpus-LaFV et celui de l'ALDC qui concerne la nature de l'enchevêtrement particulier des aires taxatoires respectives. Pour les données de LaFV, cet enchevêtrement ne s'opère avec intensité que dans certaines parties du réseau d'enquête ('enchevêtrement à brève distance'), alors que le même effet touche toute l'étendue du réseau d'enquête pour les données de l'ALDC ('enchevêtrement en profondeur'). Il en résulte des structures choroplèthes plus accidentées pour les analyses effectuées sur les matériaux-AM, et des structures choroplèthes moins escarpées et plus lisses pour les analyses-DM faites à l'aide de l'ALDC.

6.1. L'analyse de la similarité (cartes de similarité – CS)

→ Paragraphe 4.3 et cartes 27-30

Évidemment, les profils choroplèthes des deux paires de cartes – relatives aux points de références *Barcelone* et *Lleida* – évoluent toujours dans la même direction, bien que sous le coup, mentionné ci-dessus, de deux espèces différentes de l'enchevêtrement des aires taxatoires. C'est ainsi que l'allure des profils choroplèthes des cartes 28 et 30 est beaucoup plus lisse et moins accidentée que celle des cartes 27 et 29. Ce qui frappe, sur la carte 29, c'est le grand décalage qui se creuse entre les parlers occidentaux, tous marqués par des couleurs chaudes, et les parlers de l'est et des îles, caractérisés par des couleurs froides. Une rupture similaire se manifeste sur la carte 27.

6.2. L'analyse des distances interponctuelles (carte isoglottique)

→ Paragraphe 4.4 et cartes 31 et 32

À une seule exception près, le parallélisme entre les deux cartes isarithmiques est très considérable. Cette exception concerne l'île de Majorque qui est sillonnée, du côté-AM (carte 31), de gros bourrelets isoglottiques, alors qu'elle en est pratiquement exempte du côté-ALDC (carte 32).

Se pourrait-il que l'attention toute particulière que Mgr Alcover, originaire de Manacor (Majorque orientale), a prêtée aux parlers de sa (petite) patrie, ait engendré un surplus de variation dans les réponses emmagasinées ?

6.3. L'analyse des coefficients d'asymétrie de Fisher (CAF)

→ Paragraphe 4.5 et cartes 33 et 34

Cette paire de cartes offre la plus grande surprise rencontrée au cours de la préparation de cet article. En effet, les deux structures choroplèthes sont

absolument opposées et incompatibles : alors que, sur la carte de gauche (carte 33), le pôle dynamique – représenté par les polygones en bleu foncé – se trouve dans la Vieille Catalogne (voir les polygones en bleu foncé), le même pôle dynamique occupe, sur la carte de droite (carte 34), la frange de transition entre le catalan oriental et le catalan occidental tout en embrassant une partie de ce qu'il est convenu d'appeler le catalan nord-occidental[55]. Il en résulte que le domaine catalan est sillonné, dans la double perspective de cette méthode-DM, de deux gestions différentes de l'espace : la gestion verbo-morphologique atteint le sommet de son dynamisme dans la Vieille Catalogne, alors la gestion phonétique et lexicale du même espace rejoint son maximum à la frontière entre le catalan oriental et occidental et dans les parages du catalan nord-occidental.

Quant aux pôles conservateurs des deux cartes – voir les polygones en rouge –, ceux-ci ne coïncident que dans le Pays Valencien, alors que, le long des Pyrénées, leur implantation géographique diverge : sur la carte 33, le pôle géographique du conservatisme se trouve en Andorre, Pallars et Urgellet, tandis que, sur la carte 34, il se déplace en Roussillon.

Il semble que la raison principale de cette divergence capitale réside dans la nature différente de l'‘enchevêtrement particulier' qui s'instaure entre les nombreuses aires taxatoires des deux stocks de données de départ (*morphologie verbale* versus *phonétique* et *lexique*).

6.4. L'analyse dendrographique

→ Paragraphe 4.6 et cartes 35 et 36

Les deux analyses divergent surtout dans la partie occidentale du domaine catalan : voir, à ce propos, surtout les chorèmes coloriés en bleu et vert dont l'étendue et la position se distinguent considérablement sur les deux cartes. À remarquer toutefois que les deux premiers embranchements des deux arbres embrassent les mêmes régions du domaine catalan : embranchement supérieur : Roussillon, Vieille Catalogne, Îles ; embranchement inférieur : Pays Valencien, Frange occidentale, catalan nord-occidental.

Soulignons encore que, sur toutes les cartes dendrographiques de cet article, l'affiliation du parler de l'Alguer (Alghero) est très variable.

---

[55] Comparer la carte 35 avec les cartes 11 et 12, publiées dans Goebl 2013b (114) : alors que la carte 35 est basée sur 1 659 CT du corpus-ALDC et *toutes les catégories linguistiques* prises en considération, la carte 11 de 2013b repose sur 931 CT *phonétiques* et la carte 12 sur 728 CT *lexicales*.

6.5. L'analyse corrélative

→ Paragraphe 4.7 et cartes 37 et 38

Malheureusement, le maillage différent des deux réseaux d'enquête ne permet pas de corréler entre elles les données du corpus-AM avec celles du corpus-ALDC bien que, théoriquement, une telle corrélation serait réalisable. À cet effet, on devrait réduire le réseau de l'ALDC à la taille du réseau-AM tout en établissant 149 paires de localités (AM-ALDC) prétendument 'identiques'. De cette manière, il serait possible de réajuster la matrice de données du corpus ALDC sur les dimensions de la matrice de données-AM.

En attendant, nous devons nous contenter de présenter une synopsis de deux analyses corrélatives similaires: en l'occurrence, il s'agit de la synopsis de deux corrélations entre la dyade 'langue' et 'espace'. Le profil choroplèthe de la carte 37 rappelle de très près celui des cartes 9 et 16, où les zones de convergence (en rouge) se trouvent aux alentours de Valence (dans le sud) et de Barcelone (dans le nord), alors que les zones de divergence (en bleu foncé) occupent la frange de transition entre le catalan oriental et occidental.

Une structure choroplèthe très similaire peut être repérée sur la carte 38, évidemment – vu le maillage plus étroit de son fond cartographique – d'une manière beaucoup plus fine et nuancée.

L'interprétation linguistique de ces structures choroplèthes renvoie à l'histoire de la langue catalane tout court. Les zones en bleu sont caractérisées par beaucoup de frictions entre l'étalement 'naturel' (= euclidien) et 'anthropique' (c.-à-d. linguistique) des proximités ou similarités inter-dialectales dans l'espace, alors que les zones coloriées en rouge en accusent un déploiement parallèle 'en harmonie'.

Il est très intéressant de constater que, dans cette optique, les trois catégories 'en lice' (*morphologie verbale*, *phonétique* et *lexique*) vont de pair. Et aussi que c'est précisément le nord-ouest du domaine catalan qui sert de théâtre à ces affrontements entre 'langue' et 'géographie'[56]. Notre impression, en tant que dialectométriciens, est que l'espace en question a connu, au cours de son histoire plus ou moins éloignée, un surplus de contacts et conflits linguistiques à plus ou moins longue distance.

---

[56] Voir à ce sujet Veny 1989, 120*sqq.*, Veny / Massanell i Messalles 2002 (*passim*) et aussi l'historique de la classification de ces parlers dans Adam Aulinas 2006, 22-29.

7. Épilogue

À l'issue de ce travail, assez complexe à plus d'un égard, il convient d'en retracer les grandes lignes et d'en souligner les résultats les plus importants.

— Du point de vue empirique:

Rappelons, tout d'abord, que la collecte des données de départ et leur publication moyennant quelque 400 pages de texte linéaire (LaFV, 1929-1932) remontent au premier tiers du XXe siècle et que le recueil de LaFV a déjà connu, peu avant le tournant de ce siècle, une revitalisation décisive sous la forme de sa numérisation intégrale par les soins de M. P. Perea. Il serait cependant injuste et scientifiquement malvenu de critiquer son utilisation pour des analyses-DM comme 'archéologie linguistique' surannée, voire superflue. Et ceci pour trois raisons:

(1) puisque pratiquement toutes les analyses-DM entreprises jusqu'alors se servent des données d'atlas linguistiques relativement anciens, et

(2) parce que les données de tous les atlas linguistiques contiennent, comme conséquence directe des modalités de leur collecte, des structures géolinguistiques à deux volets: le volet visible relatif aux structures de surface, directement accessibles à l'œil du chercheur, et le volet caché, relatif aux structures de profondeur, dont la visibilité doit être assurée par l'application de méthodes quantitatives dans le cadre d'une synthèse inductive.

(3) Les données des atlas linguistiques contiennent des informations qui constituent l'émanation et le reflet direct d'un comportement sémiotique spécifique de l'homme face à l'espace qu'il habite, à savoir de la 'gestion basilectale de l'espace par *l'homo sapiens*'. Ce concept, de nature anthropologique, est bien loin de servir uniquement les propos théoriques de la DM. Il est à la base de beaucoup de sciences spatiales et ouvre donc de nombreuses perspectives interdisciplinaires.

— Du point de vue grammatical:

Dès le début, Mgr Alcover a conçu LaFV comme un trésor de la *morphologie verbale* de tous les dialectes catalans tout en s'inspirant des théories et méthodes de la géolinguistique gilliéronienne. Il est donc absolument légitime d'en comparer les structures (de surface et de profondeur) avec celles d'autres atlas linguistiques catalans, en l'occurrence avec celles de l'ALDC.

Rappelons en outre que des recueils 'atlantographiques' qui portent sur la morphologie verbale sont, au sein des études romanes, extrêmement rares. À notre connaissance, la seule collection qui puisse être comparée directement avec LaVF se trouve à la fin du dernier volume de l'atlas italien AIS de Karl Jaberg et Jakob Jud (vol. VIII, cartes 1682-1701). Comme les données qui y sont répertoriées ont déjà été dialectométrisées avec beaucoup de succès (cf.

Goebl 2008, 64-66, 112*sq.*), une entreprise similaire relative aux données de LaFV semblait être prometteuse d'entrée de jeu. Il est d'ailleurs très rassurant de voir que, dans les deux cas, les résultats vont dans la même direction: à savoir que les structures de profondeur d'un corpus *verbo-morphologique* ressemblent de très près à celles de corpus *phonétiques* ou *lexicaux*. Ceci prouve, une fois de plus, que les structures linguistiques de profondeur imposées à l'espace par les locuteurs disposent d'une grande stabilité (et redondance) interne.

Fait curieux: dans l'histoire de la linguistique catalane, les données de LaFV ont joué un rôle très limité. Et ceci malgré leur insertion – partielle, certes – dans les prestigieux volumes du DCVB, dont le premier a déjà paru en 1930.

En feuilletant quelques manuels faisant autorité en la matière – comme p. ex. les grammaires (ou recueils) historiques de Badia i Margarit (1951), Moll (1952), Gulsoy (1993) ou Pérez Saldanya (1998) – l'on trouve que – quoique LaFV y figure dans les bibliographies – il n'y est fait aucun usage systématique de leur contenu sous la forme de références géolinguistiques précises, pour ne pas parler de la présentation de cartes typisées. La seule exception en la matière semble être le catalaniste anglais Max W. Wheeler qui, dans ses contributions sur la morphologie verbale du catalan, recourt systématiquement aux données de LaFV: voir les neuf contributions rassemblées dans un recueil de ses écrits, publié en 2007.

Wheeler a d'ailleurs le mérite d'avoir utilisé couramment les données de LaFV déjà avant 1999, c'est-à-dire avant la publication des ouvrages de M. P. Perea relatifs à LaFV, et la mise en circulation des deux CD-ROM avec les données numérisées de cette dernière.

Un travail qui reste à faire est la construction d'un pont (théorique et conceptuel) entre les recherches géo-quantitatives ici présentées, et les travaux de linguistes et grammairiens utilisant des méthodes qualitatives.

— Du point de vue dialectométrique:

L'on pourrait croire que la clarté des structures choroplèthes et isarithmiques des 36 cartes-DM de cette contribution rend superflue toute mise en relief de l'utilité et des avantages de la DM en tant que telle. Néanmoins, il nous importe d'évoquer, encore une fois, un fait dont il n'y a aucun reflet sur la figure 3 qui pourtant représente l'agencement linéaire de toutes les méthodes-DM utilisées à Salzbourg.

En l'occurrence, il s'agit de la jonction particulière – appelée par nous *enchevêtrement* (*particulier*) – qui existe entre les nombreuses aires taxatoires sur lesquelles repose n'importe quelle synthèse-DM de type salzbourgeois. Comme les aires taxatoires constituent, en quelque sorte, les 'briques' dont le corps d'une maison est construit, l'analyse de leur taille et du 'mortier' qui les relie est du plus haut intérêt pour comprendre la 'solidité statique' de la maison entière.

Toujours est-il que nos recherches-DM, réalisées moyennant une gamme très vaste de données géolinguistiques d'origine romane, anglaise et allemande, ont montré d'une façon irréfutable que l'ensemble des aires taxatoires tirées d'un atlas linguistique se comporte comme les mots dans un texte linéaire : c'est-à-dire qu'elles sont sujettes à certaines régularités majeures tout en gardant leur apparente 'indépendance', selon le vieil adage que « chaque mot [= aire taxatoire] a sa propre histoire ». Quant à la 'paternité' de ces régularités, il faut se rendre à l'évidence : bien qu'étant des produits directs de la gestion basilectale de l'espace par l'homme, elles émanent, indirectement, du caractère systémique de tous les réseaux géolinguistiques. Nous sommes donc en présence d'un universel linguistique dont on peut présumer qu'il soit effectif (et présent) dans tous les domaines linguistiques de cette planète.

— Du point de vue pratique :

Comme notre analyse-DM de LaFV ne repose pas directement sur une *taxation* particulière de ses données, exécutée selon les règles de la DM-S, il serait du plus haut intérêt de refaire ce travail dans le cadre d'un projet supplémentaire et d'en comparer les résultats-DM avec ceux de la présente analyse. Pour ce faire, il serait indiqué d'extraire de la documentation numérisée de LaFV disponible en ligne[57], pour chacune des 'cartes-LaFV', des listes rangées dans l'ordre ascendant des numéros des 149 points d'enquête de l'entreprise-AM[58], et de les soumettre ensuite à des analyses taxatoires suivant des méthodes et concepts *linguistiques* précis. De ce travail taxatoire naîtrait une nouvelle matrice de données, susceptible de recevoir le même traitement-DM que la matrice de données utilisée dans ce travail, qui avait été tirée directement – c'est-à-dire sans le 'détour' d'une analyse taxatoire à proprement parler – de la version numérisée de LaFV.

---

[57] Voir sous : ‹ http://alcover.iec.cat ›.

[58] Le recours à des listes dûment agencées a d'ailleurs été déjà pratiqué avec beaucoup de succès lors de la dialectométrisation des volumes I-IV de l'ALDC (cf. Goebl 2013b).

La liste de tels «points de vue» pourrait être allongée à discrétion, ou presque.

Deux remarques en guise de conclusion:

(1) Il est surprenant et rassurant à la fois de voir la grande fécondité d'un dialogue avec des recherches scientifiques, vieilles de plusieurs générations.

(2) Depuis longtemps, une question nous inquiète, voire nous hante: que diraient J. Gilliéron, K. Jaberg, J. Jud, A. M. Alcover ou F. de B. Moll, s'ils pouvaient voir les structures de profondeur enfouies, à leur insu, dans les données de leurs chefs-d'œuvre?

Université de Salzbourg — Xavier CASASSAS, Hans GOEBL, Pavel SMEČKA

Université de Barcelone — Maria Pilar PEREA[59]

Remerciements

- dialectométrisation des données-AM: Xavier Casassas, Maria Pilar Perea, Slawomir Sobota, Pavel Smečka
- confection des histogrammes des figures 1 et 2: Pavel Smečka
- confection de la figure 3 ainsi que des cartes 1 et 2: Werner Goebl, Vienne
- toilettage stylistique de notre texte français: Stéphanie Lescure, Marbourg sur Lahn (Allemagne)

À toutes les personnes mentionnées s'adresse l'expression de notre parfaite reconnaissance pour leur précieuse et patiente collaboration.

[59] La participation de Maria Pilar Perea s'est déroulée dans le cadre du projet de recherche FFI2013-41077-P, financé par le ministère espagnol de l'économie et de la compétitivité («Ministerio de Economía y Competitivitad»).

8. Abréviations et références bibliographiques

Abréviations techniques souvent utilisées:

CAF	coefficient d'asymétrie de Fisher: cf. 4.5.
CAH	classification ascendante hiérarchique: cf. 4.6.
CT	carte de travail: cf. 2.
CS	carte de similarité: cf. 4.3.
DM	dialectométrie, dialectométrique, relatif à la dialectométrie
DM-S	dialectométrie / École dialectométrique de Salzbourg
IRD_{jk}	Indice Relatif de Distance (entre les vecteurs d'attributs *j* et *k*): cf. 4.2.
IRI_{jk}	Indice Relatif d'Identité (entre les vecteurs d'attributs *j* et *k*): cf. 4.2.
MA	moyenne arithmétique: cf. 4.3.
MEDMW	algorithme d'intervallisation: cf. 4.3.
r(BP)	coefficient de corrélation de Bravais et Pearson: cf. 4.7.
VDM	Visual DialectoMetry (logiciel dialectométrique créé par Edgar Haimerl): cf. 4.2.

Adam Aulinas, Montserrat, 2006. *El català septentrional de transició: nova visió des de la morfologia*, Barcelona, Institut d'Estudis Catalans.

AIS = Jaberg, Karl / Jud, Jakob, 1928-1940. *Sprach- und Sachatlas Italiens und der Südschweiz*, 8 vols., Zofingen, Ringier (réimpression: Nendeln, Kraus 1971).

ALDC = Veny, Joan / Pons i Griera, Lídia (2001-2014), *Atles lingüístic del Domini Català*, 7 vols., Barcelona, Institut d'Estudis Catalans [en 2015].

ALF = Gilliéron, Jules / Edmont, Edmond, 1902-1910. *Atlas linguistique de la France*, 10 vols., Paris, Champion (réimpression: Bologne, Forni 1968).

ALPI = *Atlas lingüístico de la Península Ibérica, vol. I, Fonética*, Madrid, Consejo Superior de Investigaciones Científicas, 1962.

AM = relatif à «La flexió verbal [ici: LaFV]» d'A. M. Alcover et Fr. de B. Moll.

Badia i Margarit, Antoni, 1951. *Gramática histórica catalana*, Barcelona, Editorial Noguer.

Berschin, Helmut / Felixberger, Josef / Goebl, Hans, 2008[2]. *Französische Sprachgeschichte. Lateinische Basis, interne und externe Geschichte, sprachliche Gliederung Frankreichs. Mit einer Einführung in die historische Sprachwissenschaft*, Hildesheim, Olms.

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

Brun-Trigaud, Guylaine / Le Berre, Yves / Le Dû, Jean, 2005. *Lectures de l'Atlas linguistique de la France de Gilliéron et Edmont. Du temps dans l'espace. Essai d'interprétation des cartes de l'*Atlas linguistique de la France *de Jules Gilliéron et Edmond Edmont augmenté de quelques cartes de l'*Atlas linguistique de la Basse-Bretagne *de Pierre Le Roux*, Paris, CTHS.

Casassas, Xavier, 2008. «Die Dialektometrisierung des ALDC (Atles Lingüístic del Domini Català). Ein Ergebnis der Beziehung von Hans Goebl zum katalanischen Sprachraum», in: Gabriele Blaikner-Hohenwart / Evelyne Bortolotti / Rita Franceschini / Emese Lörincz / Leander Moroder / Gerda Videsott / Paul Videsott (ed.), *Ladinometria. Festschrift für Hans Goebl zum 65. Geburtstag / Miscellanea per Hans Goebl per il 65° compleanno / Publicazion en onour de Hans Goebl en gaujion de si 65 agn*, Salzburg, Universität Salzburg-Fachbereich Romanistik, Bozen/Bolzano/Bulsan, Freie Universität Bozen/Libera Università/Université Ledia, Vich/Vigo di Fassa, Istitut cultural ladin 'majon di fascegn', S. Martin de Tor/S. Martino di Badia, Istitut ladin 'Micurà de Rü', vol. 2, 19-25.

Chandon, Jean-Louis / Pinson, Suzanne, 1981. *Analyse typologique. Théories et applications*, Paris/New York/Barcelone/Milan, Masson.

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

DCVB = *Diccionari català-valencià-balear: inventari lexicogràfic i etimològic de la llengua catalana en totes les seves formes literàries i dialectals*, obra iniciada per Antoni Maria Alcover, redactat per Francesc de B. Moll, amb la col·laboració de Manuel Sanchis Guarner i d'Anna Moll Marquès, 10 vols. [1930-1962], Palma de Mallorca, Editorial Moll (plusieurs réimpressions).

Ettmayer, Karl von, 1924. «Über das Wesen der Dialektbildung, erläutert an den Dialekten Frankreichs», *Denkschriften der Akademie der Wissenschaften in Wien, philosophisch-historische Klasse* 66, 1-56 (avec 7 tableaux).

Goebl, Hans, 1981. «Éléments de l'analyse dialectométrique (avec application à l'AIS)», *RLiR* 45, 349-420.

Goebl, Hans, 1983. «Parquet polygonal et treillis triangulaire: les deux versants de la dialectométrie interponctuelle», *RLiR* 47, 353-412.

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

Goebl, Hans, 1987. «Points chauds de l'analyse dialectométrique: pondération et visualisation», *RLiR* 51, 63-118.

Goebl, Hans, 1991. «Una classificazione gerarchica di dati geolinguistici tratti dall'AIS. Saggio di dialettometria dendrografica», *Linguistica* 31 [Paulo Tekavčić sexagenario in honorem oblata], 341-352.

Goebl, Hans, 1992. «Dendrogramme im Dienst der Dialektometrie. Zwei hierarchisch-agglomerative Klassifikationen von Daten des Sprachatlasses AIS», in: Ursula Klenk (ed.), *Computatio linguae. Aufsätze zur algorithmischen und quantitativen*

Analyse von Sprache, Stuttgart, Steiner, 54-73 (Zeitschrift für Dialektologie und Linguistik, Beiheft 73).

Goebl, Hans, 1993. «Dialectometry. A Short Overview of the Principles and Practice of Quantitative Classification of Linguistic Atlas Data», in: Reinhard Köhler / Burghard B. Rieger (ed.), *Contributions to Quantitative Linguistics*, Dordrecht, Kluwer, 277-315.

Goebl, Hans, 2002. «Analyse dialectométrique des structures de profondeur de l'ALF», *RLiR* 66, 5-63 (avec 24 cartes en couleurs).

Goebl, Hans, 2003. «Regards dialectométriques sur les données de l'Atlas linguistique de la France (ALF): Relations quantitatives et structures de profondeur», *Estudis Romànics* 25, 59-120 (avec 24 cartes en couleurs).

Goebl, Hans, 2005a. «La dialectométrie corrélative. Un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme», *RLiR* 69, 321-367 (avec 24 cartes en couleurs).

Goebl, Hans, 2006a. «Recent Advances in Salzburg Dialectometry», *Literary and Linguistic Computing* 21/4, 411-435.

Goebl, Hans, 2008. «La dialettometrizzazione integrale dell'AIS. Presentazione dei primi risultati, *RLiR* 72, 25-113 (avec 48 cartes en coleurs).

Goebl, Hans, 2010. «Dialectometry and quantitative mapping», in: Alfred Lameli / Roland Kehrein / Stefan Rabanaus (ed.), *Language and Space. An International Handbook of Linguistic Variation*, vol. 2: Language Mapping (Handbücher der Sprach- und Kommunikationswissenschaft [HSK] 30.2.), Berlin, de Gruyter, 1re partie (texte) 433-457, 2e partie (cartes), 2201-2212 (avec 12 cartes en couleurs).

Goebl, Hans, 2012. «Introduction aux problèmes et méthodes de l''École dialectométrique de Salzbourg' (avec des exemples gallo-, italo- et ibéroromans)», in: Afonso Álvarez Pérez / Ernestina Carrilho / Catarina, Magro (ed.), *Proceedings of the International Symposium on Limits and Areas in Dialectology (LimiAr)*, Lisbonne 2011, Lisbonne, Centro de Linguística da Universidade de Lisboa, 117-166 (avec 28 cartes en couleurs relatives à l'ALF) (Actes du congrès en ligne: ‹http://limiar.clul.ul.pt/proceedings_en.html›).

Goebl, Hans, 2013a. «Le Baiser de la Belle au bois dormant ou: des péripéties encourues par la géographie linguistique depuis Jules Gilliéron», in: Rita Caprini (ed.), *Dialectologie. Corpus, atlas, analyses*, Alessandria, Edizioni dell'Orso [= Corpus 12, 2013], 61-84.

Goebl, Hans, 2013b. «La dialectometrització dels quatre primers volums de l'*ALDC*: una breu presentació», *Estudis Romànics* 35, 87-116 (avec 16 cartes en couleurs).

Goebl, Hans, 2013c. «La dialectometrización del ALPI: rápida presentación de los resultados», in: Emili Casanova Herrero / Cesáreo Calvo Rigual (ed.), *Actas del XXVI Congreso Internacional de Lingüística y de Filología Románicas* (Valencia 2010), Berlin, Boston, de Gruyter, vol. VI. 143-154.

Goebl, Hans, 2013d. «Introduzione ai problemi e metodi della dialettometria, applicati ai dati dell'Atlante italo-svizzero AIS (con particolar riguardo ai dialetti del Veneto e dell'Istria)», in: José Enrique Gargallo Gil / Nikola Vuletić (ed.), *Mare Loquens. Études d'étymologie et de géolinguistique romanes à la mémoire de Vojmir*

Vinja (1921-2007), (Posebna izdanja Centra za jadranska onomastička istraživanja – Knjiga 3), Zadar, Sveučilište u Zadru, 171-225 (avec 24 cartes en couleurs).

Goebl, Hans, 2014. «L'impact de la polynymie des cartes d'atlas sur le résultat de calculs dialectométriques», in: Leszek Bednarczuk *et al.* (ed.), *Linguistique romane et linguistique indo-européenne. Mélanges offerts à Witold Mańczak à l'occasion de son 90e anniversaire*, Cracovie, PAU-UJ, 243-271 (avec 10 cartes en couleurs).

Goebl, Hans / Smečka, Pavel, 2014. «L'analyse dialectométrique des cartes de la série B de l'ALF», *RLiR* 78, 439-497 (avec 32 cartes en couleurs).

Gulsoy, Joseph, 1993. *Estudis de gramàtica històrica*, València, Institut Universitari de Filologia Valenciana, Barcelona, Publicacions de l'Abadia de Montserrat.

Jaberg, Karl, 1908. *Sprachgeographie. Beitrag zum Verständnis des Atlas linguistique de la France*, Aarau, Sauerländer (traduction espagnole: *Geografía lingüística. Ensayo de interpretación del 'Atlas lingüístico de Francia'*. Traducción de A. Llorente y M. Alvar, Granada, Universidad de Granada, Secretariado de Publicaciones, 1959).

LaFV = Alcover, Antoni Maria / Moll, Francesc de Borja, 1929-1932. «La flexió verbal en els dialectes catalans. [Materials recollits per Mn. Antoni Maria Alcover; recopilació i sistematització, pròleg i notes per Francesc de B. Moll]», *Anuari de l'Oficina Romànica de Lingüística i Literatura* [Barcelona] II [1929], 73-184; III [1930], 73-168; IV [1931], 9-104; V [1932], 9-72.

Malkiel, Yakov, 1967. «Each Word has a History of its Own», *Glossa* 1, 137-149.

Moll, Francesc de B., 1952. *Gramática histórica catalana*, Madrid, Gredos (nouvelle édition, corrigée: València, Universitat de València, 2006).

Moll, Francesc de B., 1962. *Un home de combat (Mossèn Alcover)*, Palma de Mallorca, Editorial Moll.

Moll, Francesc de B., 1970. *Els meus primers trenta anys (1903-1934)*, Mallorca, Editorial Moll.

Moll, Francesc de B., 1975. *Els altres quaranta anys (1935-1974)*, Mallorca, Editorial Moll.

Okabe, Atsuyuki / Boots, Barry / Sugihara, Kokichi, 1992. *Spatial Tesselations. Concepts and Applications of Voronoi Diagrams*, Chichester/New York/Brisbane/Toronto/Singapour, Wiley & Sons.

Perea, Maria Pilar, 1997. «El balear a 'La flexió verbal en els dialectes catalans'», in: Massot i Muntaner, Josep (ed.), *Estudis de llengua i literatura en honor de Joan Veny*, Barcelona, Publicacions de l'Abadia de Montserrat, vol. I, 329-355.

Perea, Maria Pilar, 1999a. *Compleció i ordenació de 'La flexió verbal en els dialectes catalans' d'A. M. Alcover i F. de B. Moll*, Barcelona, Institut d'Estudis Catalans, 2 vols. (+ CD-ROM).

Perea, Maria Pilar, 1999b. *Complements a 'La flexió verbal en els dialectes catalans'*, Barcelona, Publicacions de l'Abadia de Montserrat.

Perea, Maria Pilar, 2001a. «La metodologia de l'enquesta dialectal d'A. M. Alcover i F. de B. Moll aplicada a 'La flexió verbal en els dialectes catalans'», *Zeitschrift für Katalanistik* 14, 98-127.

Perea, Maria Pilar, 2001b. «Les dades de 'La flexió verbal en els dialectes catalans' d'A. M. Alcover i F. de B. Moll. Una aproximació sociolingüística», *Catalan Review. International Journal of Catalan Culture* 15.1, 53-77.

Perea, Maria Pilar, 2001c. *La flexió verbal en els dialectes catalans d'Antoni M. Alcover i F. de B. Moll. Les dades i els mapes*, Palma de Mallorca, Conselleria d'Educació i Cultura (CD-ROM).

Perea, Maria Pilar, 2004. «New Techniques and Old Corpus: 'La flexió verbal en els dialectes catalans' (Alcover-Moll, 1929-132). Systematisation and Mapping of a Morphological Corpus», *Dialectologia et Geolinguistica* 12, 25-45.

Perea, Maria Pilar, 2005a. *Antoni M. Alcover. Dialectòleg, gramàtic, polemista*, Castelló de la Plana, Fundació Germà Colón Domenech/Barcelona, Publicacions de l'Abadia de Montserrat.

Perea, Maria Pilar, 2005b. *Antoni M. Alcover. Dades dialectals*, Palma de Mallorca, Conselleria d'Educació i Cultura (CD-ROM).

Perea, Maria Pilar, 2006. «From the notebook to the computer: the systematisation of a dialectal corpus», in: A. Timuška (ed.), *Proceedings of the 4th International Congress of Dialectologists and Geolinguists* (Riga, 28.7.-2.8.2003), Riga, Latvian Language Institute, University of Latvia, 400-412.

Perea, Maria Pilar, 2008a. «Mapping morphological and phonetic features of Catalan: a general template for contemporary atlases and corpus», *Dialectologia* [Barcelona] 1, 121-135 (revue en ligne: ‹http://www.publicacions.ub.es/revistes/dialectologia1/›).

Perea, Maria Pilar, 2008b. «Catalan verb morphology and dialectometric analysis», in: Gabriele Blaikner-Hohenwart / Evelyne Bortolotti / Rita Franceschini / Emese Lörincz / Leander Moroder / Gerda Videsott / Paul Videsott (ed.), *Ladinometria. Festschrift für Hans Goebl zum 65. Geburtstag / Miscellanea per Hans Goebl per il 65° compleanno / Publicazion en onour de Hans Goebl en gaujion de si 65 agn*, Salzburg, Universität Salzburg-Fachbereich Romanistik, Bozen/Bolzano/Bulsan, Freie Universität Bozen/Libera Università/Université Ledia, Vich/Vigo di Fassa, Istitut cultural ladin 'majon di fascegn', S. Martin de Tor/S. Martino di Badia, Istitut ladin 'Micurà de Rü', vol. 2, 61-78.

Perea, Maria Pilar / Ueda, Hiroto, 2010. «Applying Quantitative Analysis Techniques to 'La flexió verbal en els dialectes catalans'», *Dialectologia et Geolinguistica* 18, 99-114.

Pérez Saldanya, Manuel, 1988. *Del llatí al català. Morfosintaxi verbal històrica*, València, Universitat de València.

Polanco Roig, Lluís B., 1984. «Llengua o dialecte: solucions teòriques i aplicació al cas català», in: *Sociolinguistique des langues romanes. Actes du XVIIe Congrès International de Linguistique et Philologie Romanes (Aix-en-Provence 1983)*, Aix-en-Provence, Université de Provence/Marseille, Laffitte, vol. 5, 13-31.

Polanco Roig, Lluís B., 1992. «Llengua o dialecte: una aplicació dialectomètrica a la llengua catalana», in: Antoni Ferrando (ed.), *Miscel·lània Sanchis Guarner*, València, Universitat de València (Departament de Filologia Catalana)/Barcelona, Publicacions de l'Abadia de Montserrat, vol. 3, 5-28.

Pop, Sever / Pop, Rodica Doina, 1959. *Jules Gilliéron. Vie, enseignement, élèves, œuvres, souvenirs*, Louvain, Centre international de dialectologie générale.

Rosenqvist, Arvid, 1919. «Limites administratives et division dialectale de la France», *Neuphilologische Mitteilungen* 20, 87-119 (carte publiée à nouveau in: Berschin / Felixberger / Goebl 2008[2], 261).

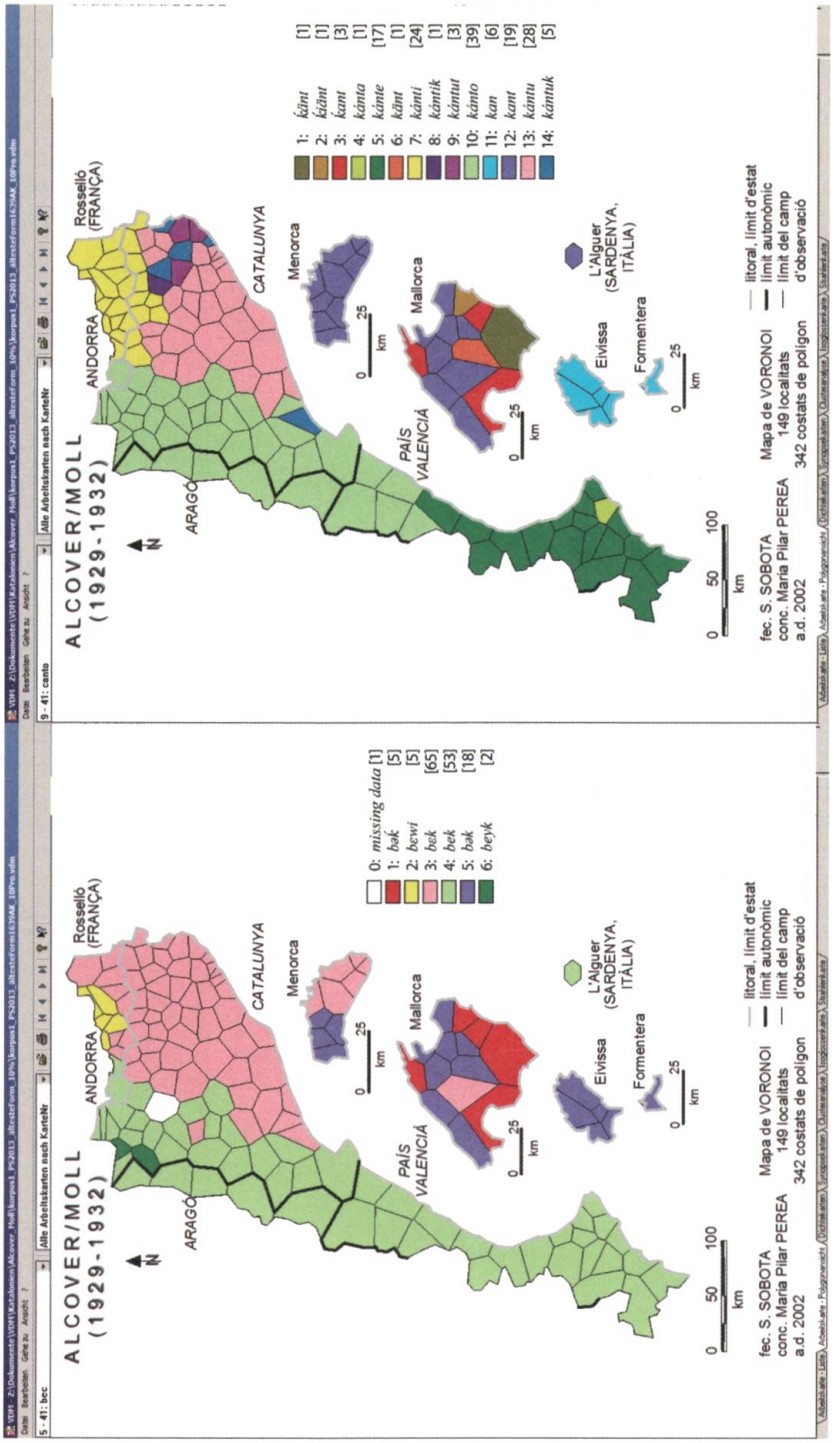
Sanchis Guarner, Manuel, 1953. «Le dictionnaire historique et dialectal du catalan 'Alcover-Moll'. Travaux, problèmes et méthodes», *Orbis* 2, 104-112.

Sneath, Peter A. / Sokal, Robert R., 1973. *Numerical Taxonomy. The Principles and Practice of Numerical Classification*, San Francisco, Freeman.

Veny, Joan, 1989[8]. *Els parlars catalans*, Palma de Mallorca, Editorial Moll.

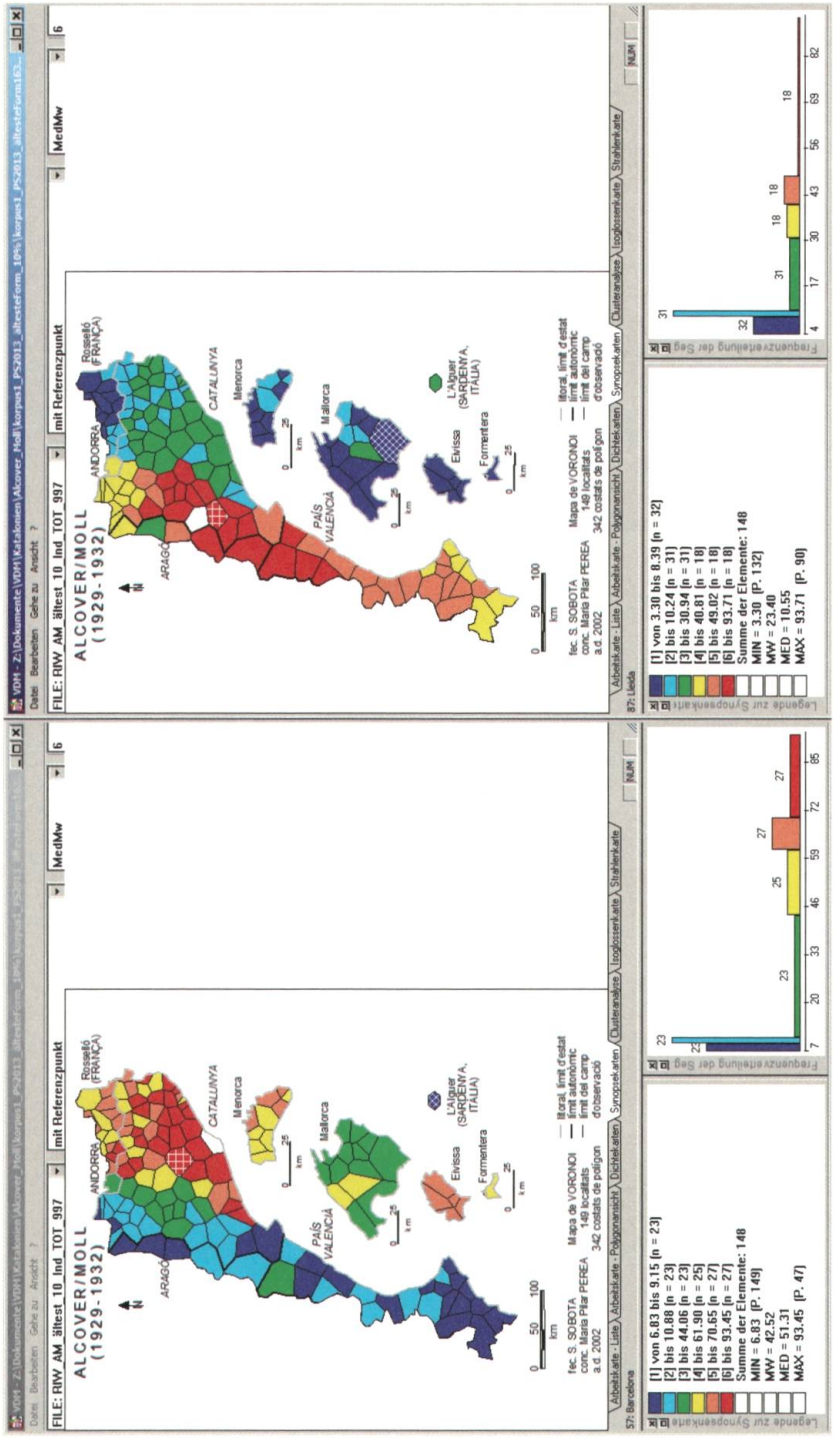
Veny, Joan / Massanell i Messalles, Mar, 2002. *Dialectes occidentals. Català nord-occidental i valencià*, Barcelona, Universitat Oberta de Catalunya.

Wheeler, Max W., 2007. *Morfologia i fonologia catalana i romànica: estudis diacrònics*, Alacant, Institut Interuniversitari de Filologia valenciana/Barcelona, Publicacions de l'Abadia de Montserrat.



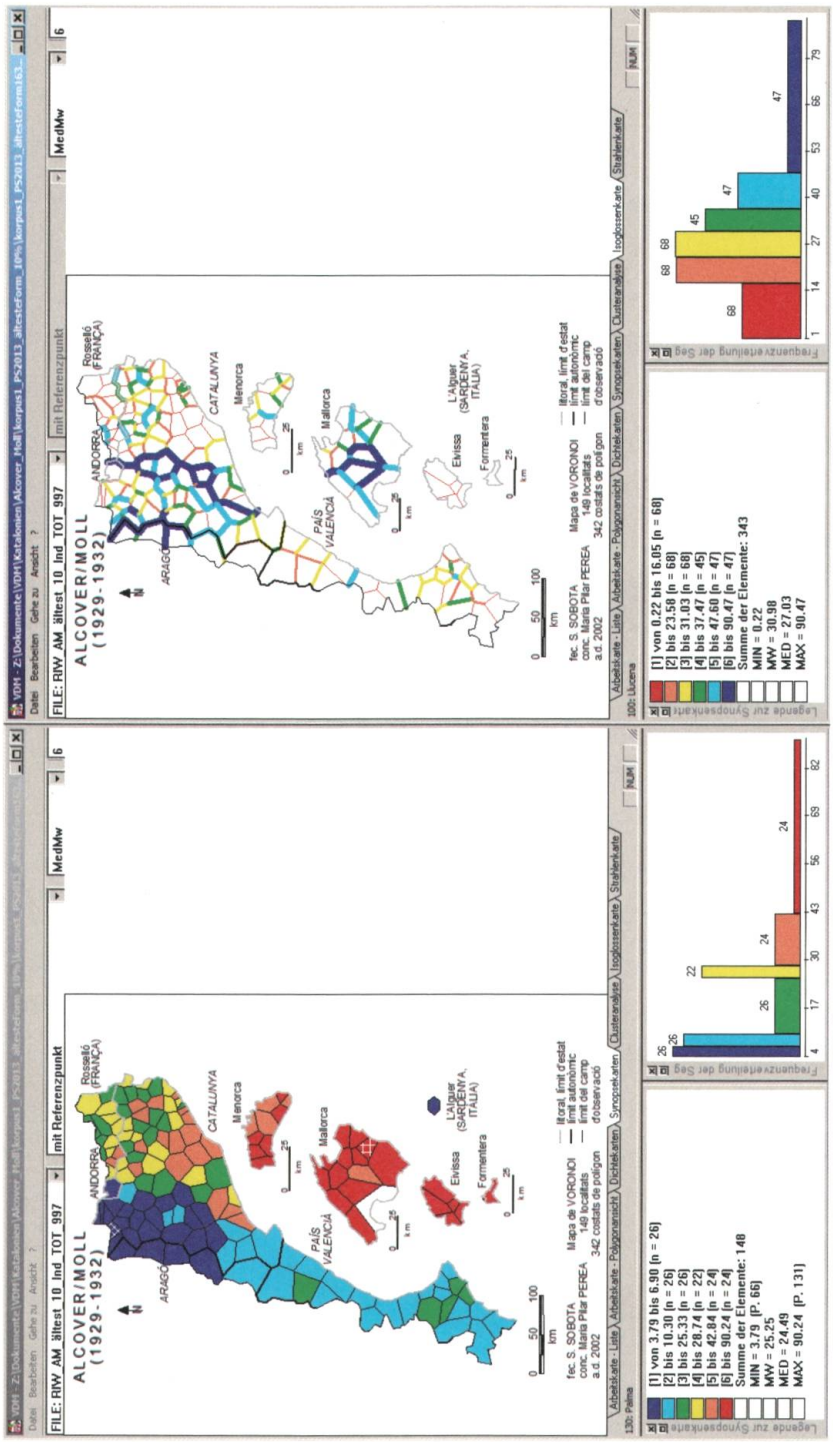
Carte 1: Exemple d'une carte de travail 6-nyme (d'après le groupement-LaFV): les équivalents dialectaux du catalan standard *bec* "je bois". Voir aussi le tableau 1.

Carte 2: Exemple d'une carte de travail 14-nyme (d'après le groupement-LaFV): les équivalents dialectaux du catalan standard *canto* "je chante". Voir aussi le tableau 2.



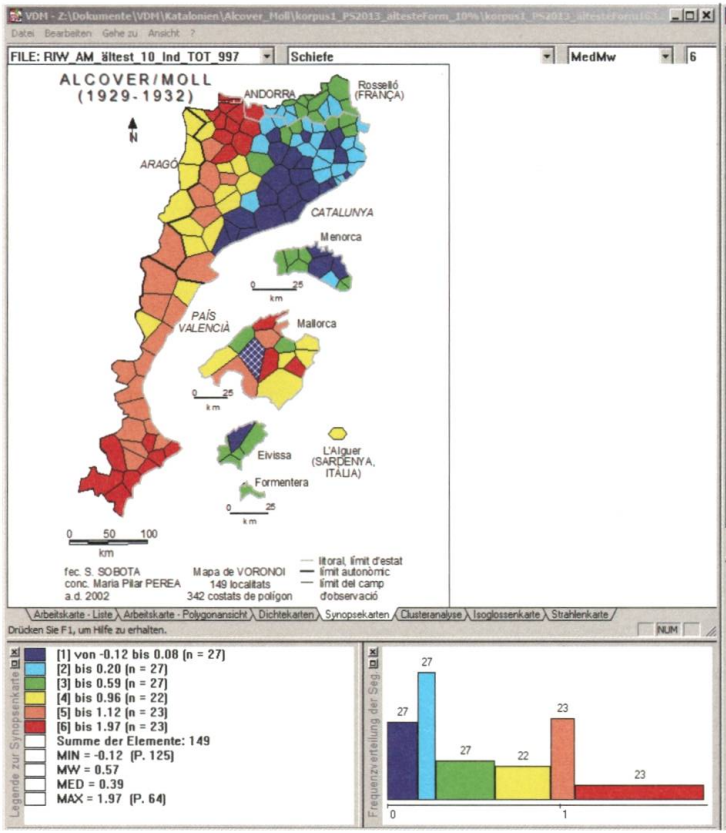
Carte 3: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-AM 57 (Barcelone)
Corpus: 997 CT du corpus RÉD relatives aux formes dialectales de l'*indicatif* (de toutes les personnes du *présent*, de l'*imparfait*, du *futur* et du *conditionnel présent*)
Algorithme d'intervallisation: MEDMW 6-tuple

Carte 4: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-AM 87 (Lleida)
Corpus: 997 CT du corpus RÉD relatives aux formes dialectales de l'*indicatif* (de toutes les personnes du *présent*, de l'*imparfait*, du *futur* et du *conditionnel présent*)
Algorithme d'intervallisation: MEDMW 6-tuple

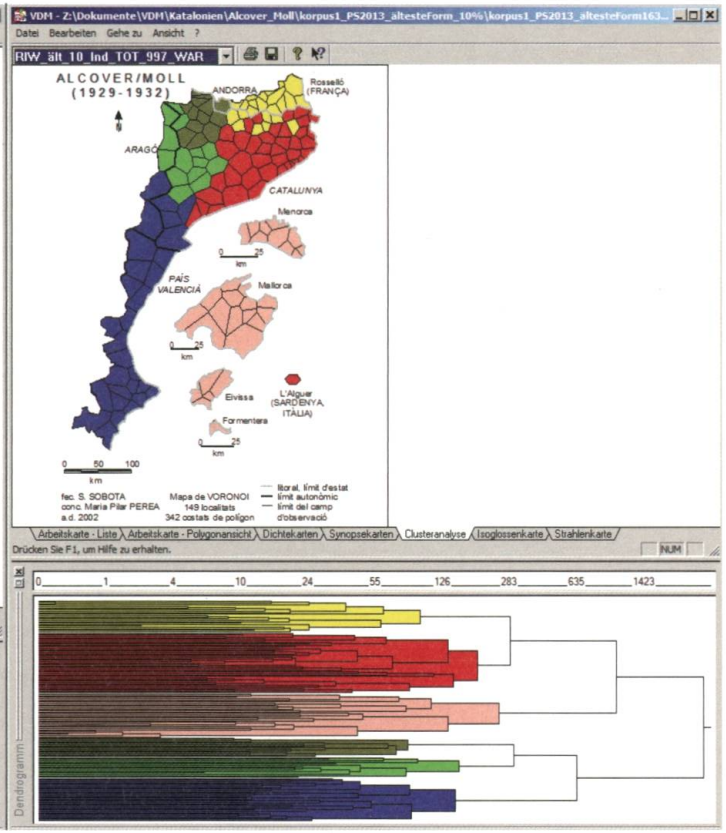


Carte 5: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-AM 130 (Palma de Mallorca / Ciutat)
Corpus : 997 CT du corpus RÉD relatives aux formes dialectales de l'*indicatif* (de toutes les personnes du *présent*, de l'*imparfait*, du *futur* et du *conditionnel présent*)
Algorithme d'intervallisation : MEDMW 6-tuple

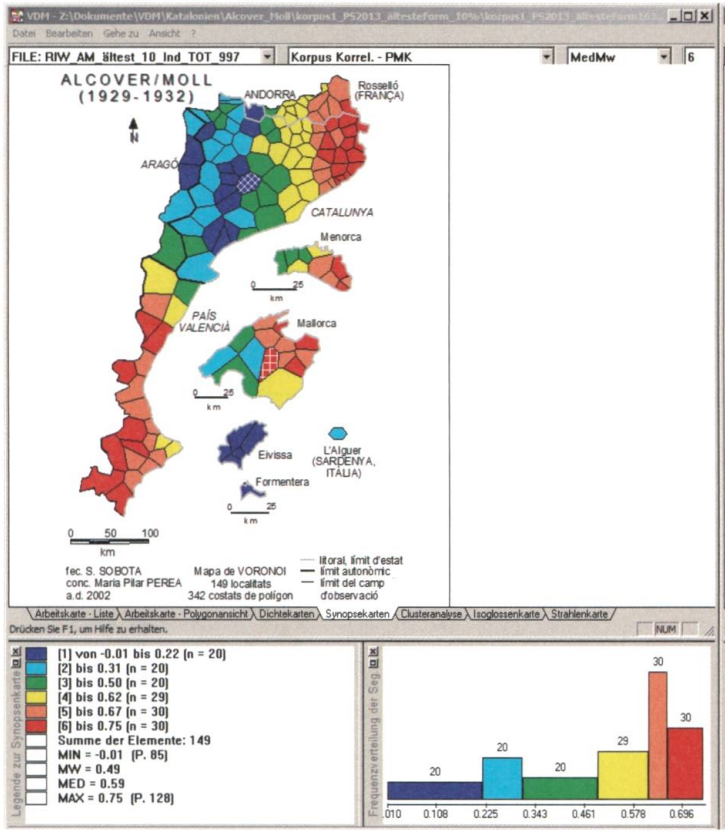
Carte 6: Synthèse isoglottique ('carte à cloisons')
Corpus : 997 CT du corpus RÉD relatives aux formes dialectales de l'*indicatif* (de toutes les personnes du *présent*, de l'*imparfait*, du *futur* et du *conditionnel présent*)
Indice de distance : IRD_{jk}
Algorithme d'intervallisation : MEDMW 6-tuple



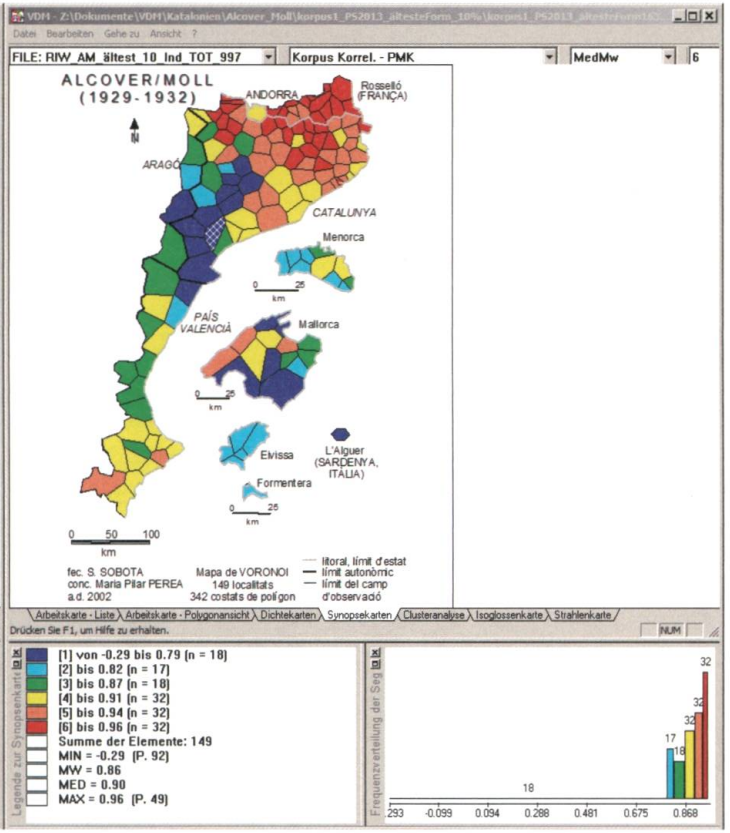
Carte 7: Profil choroplèthe de la synopse de 149 scores-CAF ('coefficient d'asymétrie de Fisher')
Corpus: 997 CT du corpus RÉD relatives aux formes dialectales de l'*indicatif* (de toutes les personnes du *présent*, de l'*imparfait*, du *futur* et du *conditionnel présent*)
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MEDMW 6-tuple



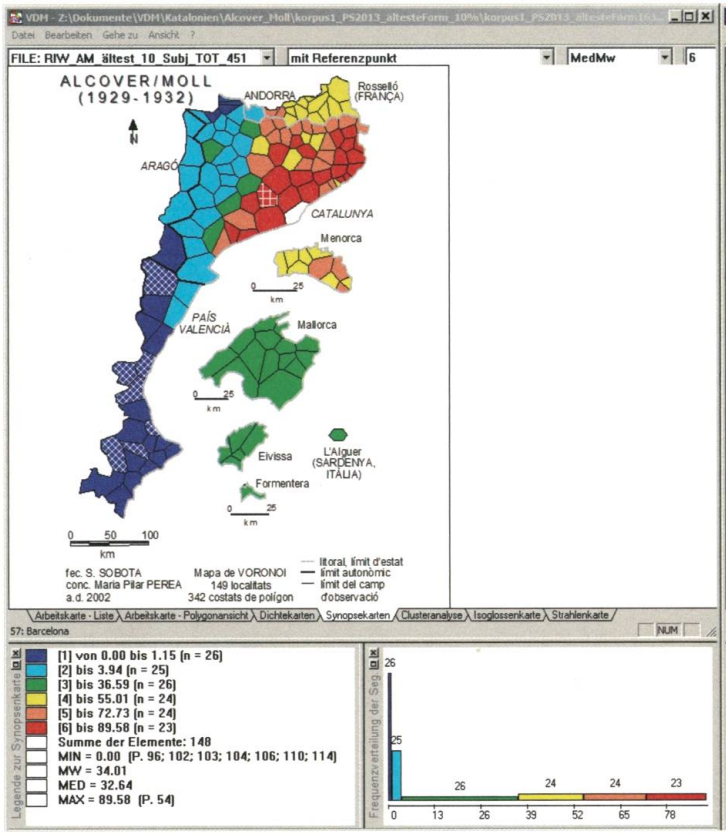
Carte 8: Classification ascendante hiérarchique (méthode de Joe Ward, Jr.)
Corpus: 997 CT du corpus RÉD relatives aux formes dialectales de l'indicatif (de toutes les personnes du *présent*, de l'*imparfait*, du *futur* et du *conditionnel présent*)
Indice de similarité: IRI_{jk}
Nombre des chorèmes (en haut) et dendrèmes (en bas) coloriés: 6



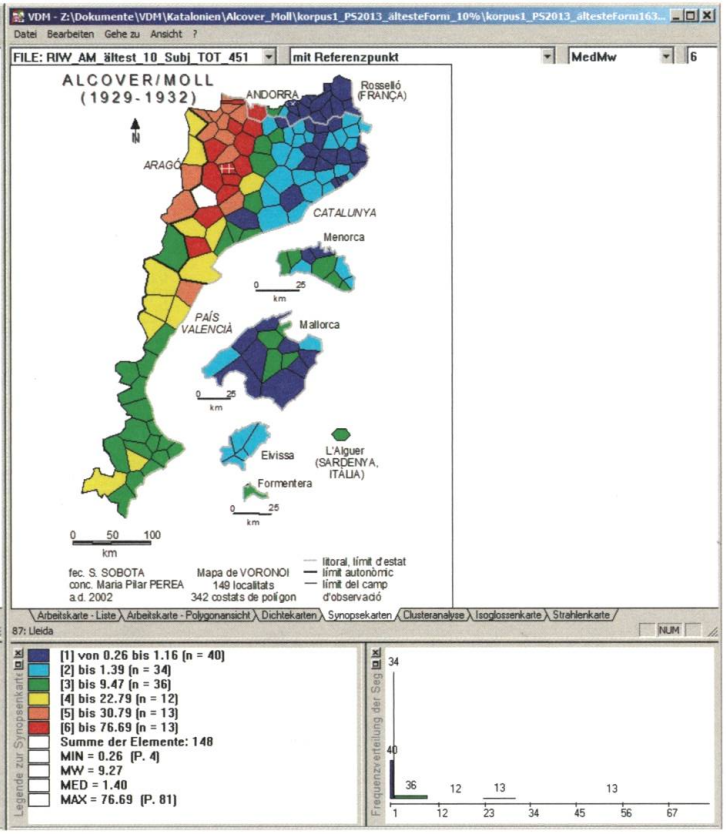
Carte 9: Profil choroplèthe d'une carte à corrélations
Matrices carrées corrélées:
matrice de similarité (selon IRI_{jk})
matrice de proximité géographique (selon le théorème de Pythagore)
Corpus linguistique: 997 CT du corpus RÉD relatives aux formes dialectales de l'*indicatif* (de toutes les personnes du *présent*, de l'*imparfait*, du *futur* et du *conditionnel présent*)
Algorithme d'intervallisation: MEDMW 6-tuple



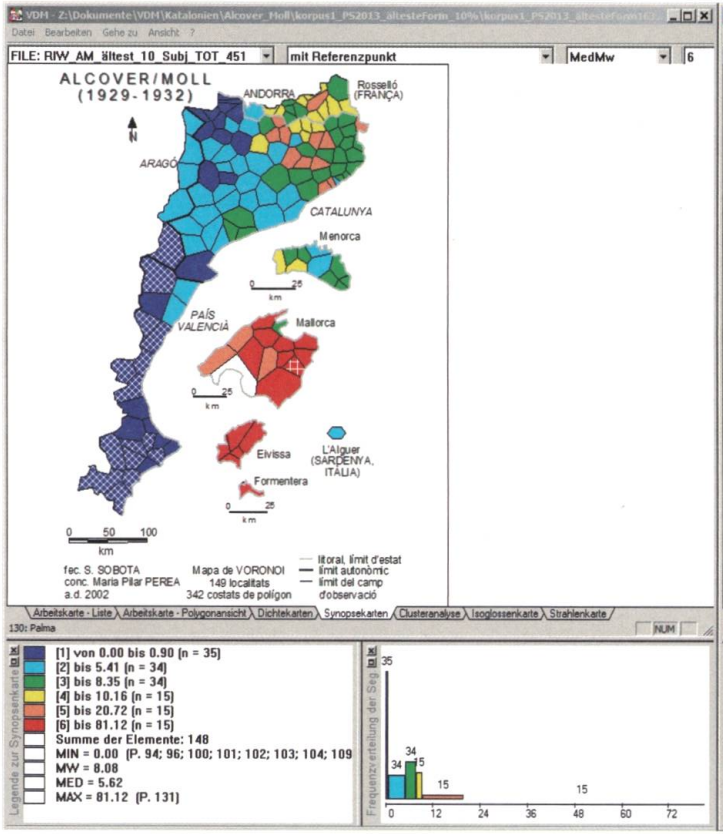
Carte 10: Profil choroplèthe d'une carte à corrélations
Matrices carrées de similarité corrélées (selon IRI_{jk}):
matrice A: 997 CT du corpus RÉD (*indicatif*)
matrice B: 451 CT du corpus RÉD (*subjonctif* du *présent* et de l'*imparfait*)
Algorithme d'intervallisation: MEDMW 6-tuple



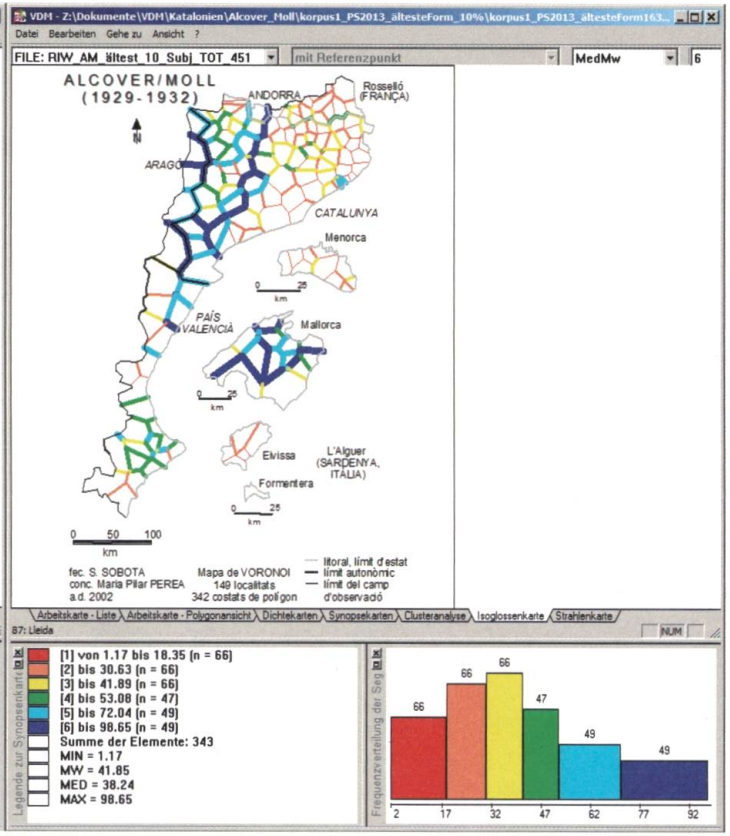
Carte 11: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-AM 57 (Barcelone)
Corpus: 451 CT du corpus RÉD relatives aux formes dialectales du *subjonctif* (de toutes les personnes, tant du *présent* que de *l'imparfait*)
Algorithme d'intervallisation: MEDMW 6-tuple



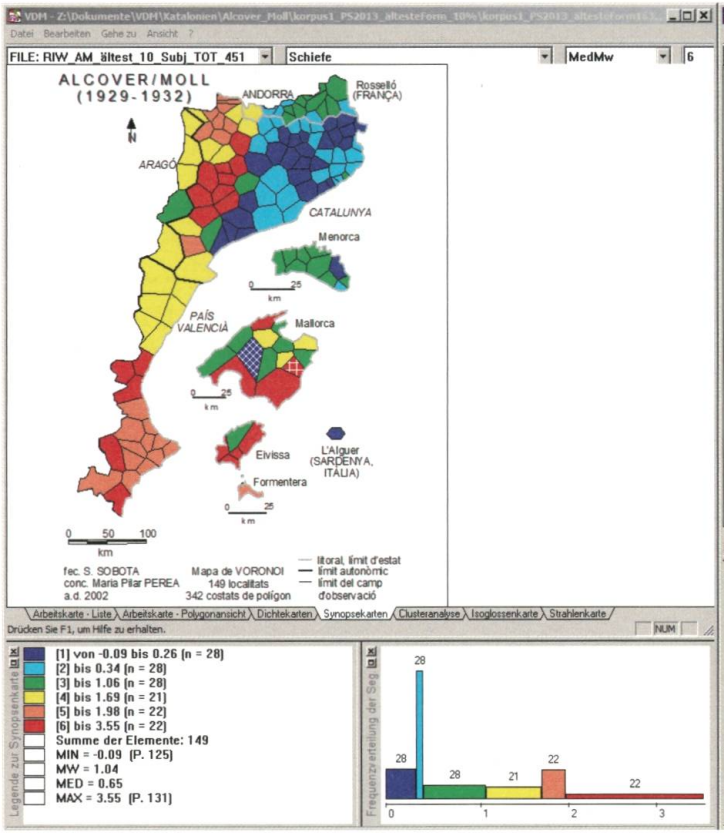
Carte 12: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-AM 87 (Lleida)
Corpus: 451 CT du corpus RÉD relatives aux formes dialectales du *subjonctif* (de toutes les personnes, tant du *présent* que de l'*imparfait*)
Algorithme d'intervallisation: MEDMW 6-tuple



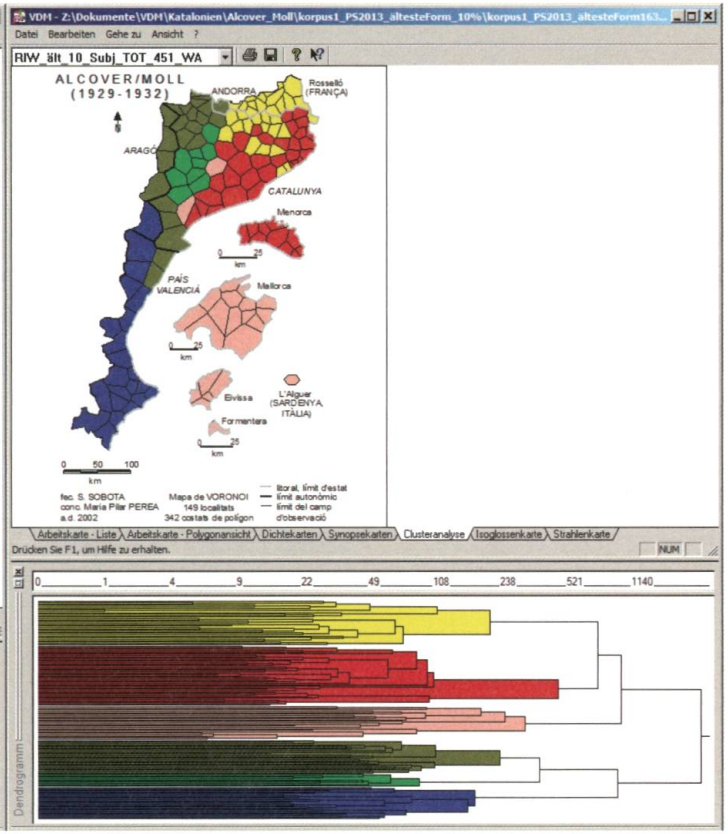
Carte 13: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-AM 130 (Palma de Mallorca / Ciutat)
Corpus: 451 CT du corpus RÉD relatives aux formes dialectales du *subjonctif* (de toutes les personnes, tant du *présent* que de l'*imparfait*)
Algorithme d'intervallisation: MEDMW 6-tuple



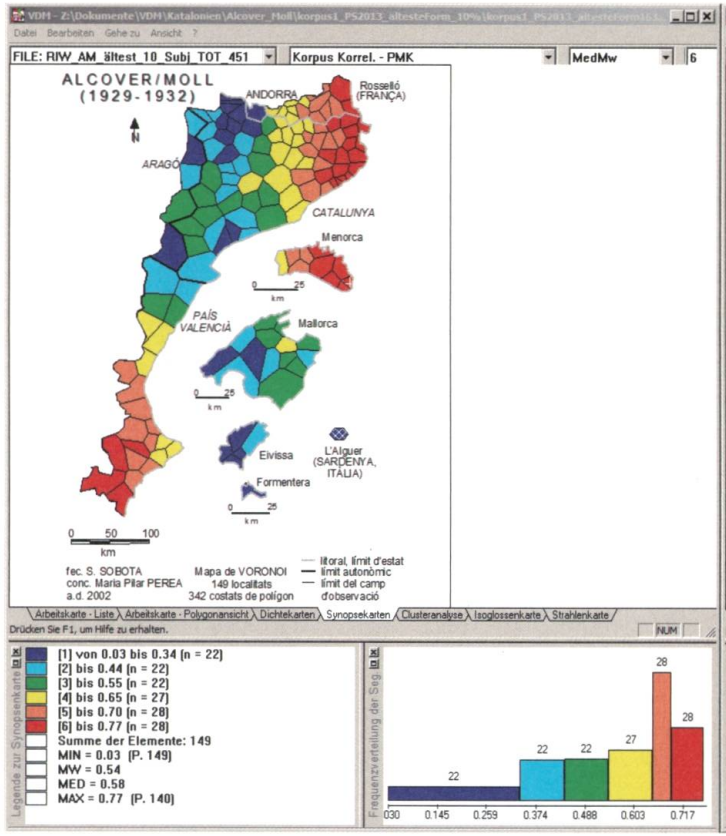
Carte 14: Synthèse isoglottique ('carte à cloisons')
Corpus: 451 CT du corpus RÉD relatives aux formes dialectales du *subjonctif* (de toutes les personnes, tant du *présent* que de l'*imparfait*)
Indice de distance: IRD_{jk}
Algorithme d'intervallisation: MEDMW 6-tuple



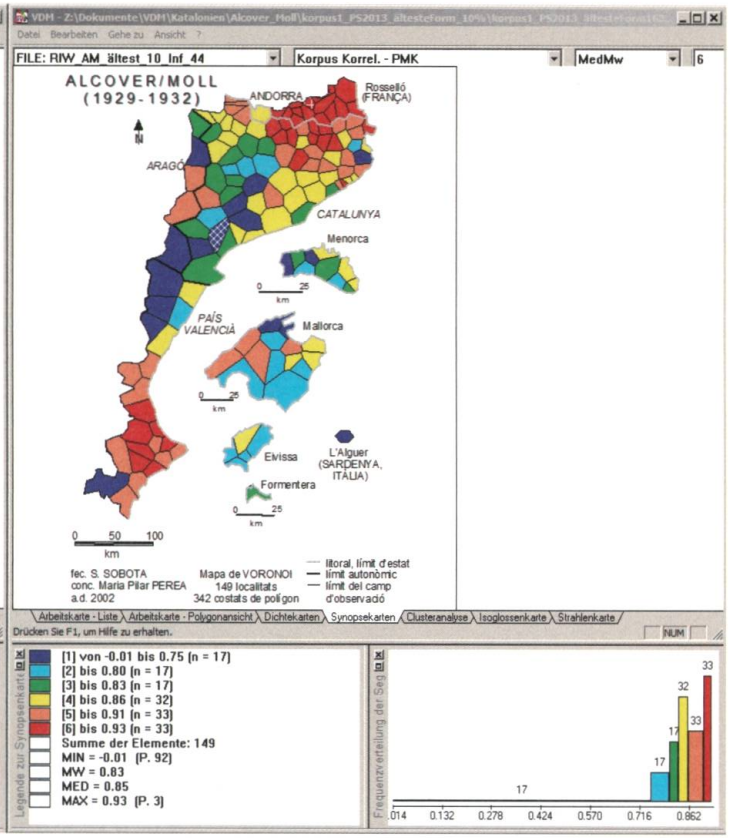
Carte 15: Profil choroplèthe de la synopse de 149 scores-CAF ('coefficient d'asymétrie de Fisher')
Corpus: 451 CT du corpus RÉD relatives aux formes dialectales du *subjonctif* (de toutes les personnes, tant du *présent* que de l'*imparfait*)
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MEDMW 6-tuple



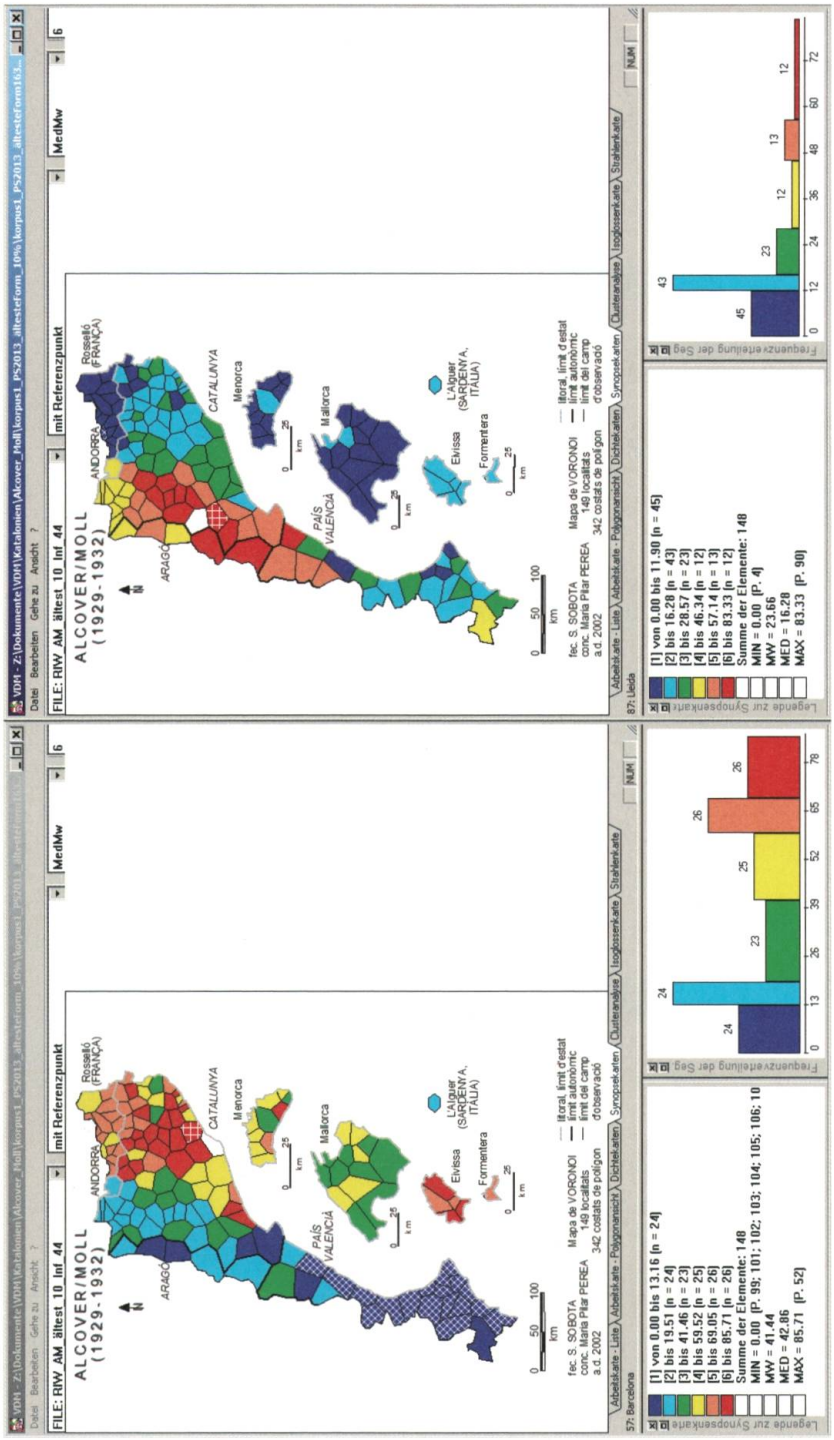
Carte 16: Classification ascendante hiérarchique (méthode de Joe Ward, Jr.)
Corpus: 451 CT du corpus RÉD relatives aux formes dialectales du *subjonctif* (de toutes les personnes, tant du *présent* que de l'*imparfait*)
Indice de similarité: IRI_{jk}
Nombre des chorèmes (en haut) et dendrèmes (en bas) coloriés: 6



Carte 17: Profil choroplèthe d'une carte à corrélations
Matrices carrées corrélées:
matrice de similarité (selon IRI_{jk})
matrice de proximité géographique (selon le théorème de Pythagore)
Corpus: 451 CT du corpus RÉD relatives aux formes dialectales du *subjonctif* (de toutes les personnes, tant du *présent* que de l'*imparfait*)
Algorithme d'intervallisation: MEDMW 6-tuple

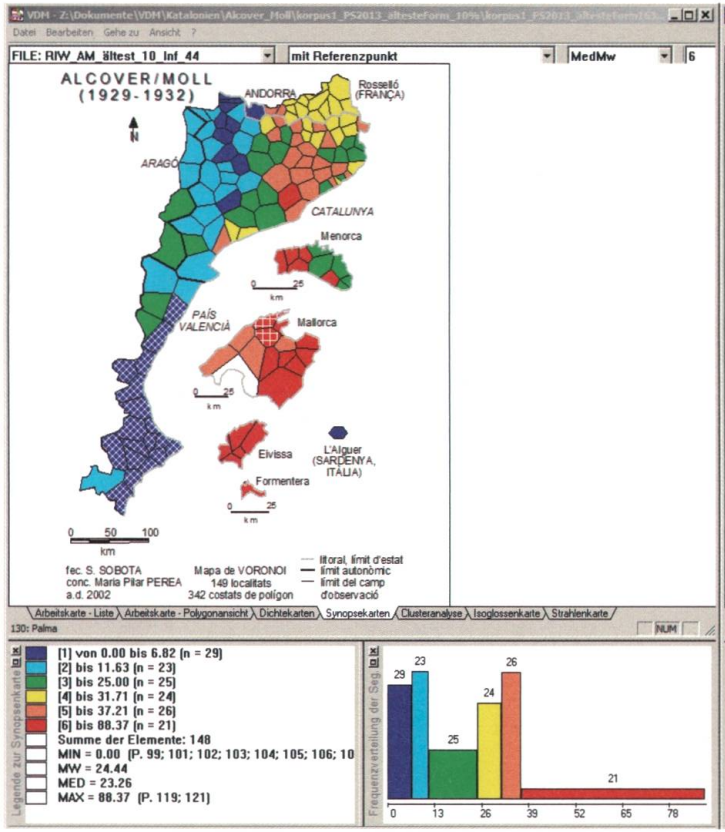


Carte 18: Profil choroplèthe d'une carte à corrélations
Matrices carrées de similarité corrélées (selon IRI_{jk}):
matrice A: 451 CT du corpus RÉD (*subjonctif*)
matrice B: 44 CT du corpus RÉD (*infinitif*)
Algorithme d'intervallisation: MEDMW 6-tuple

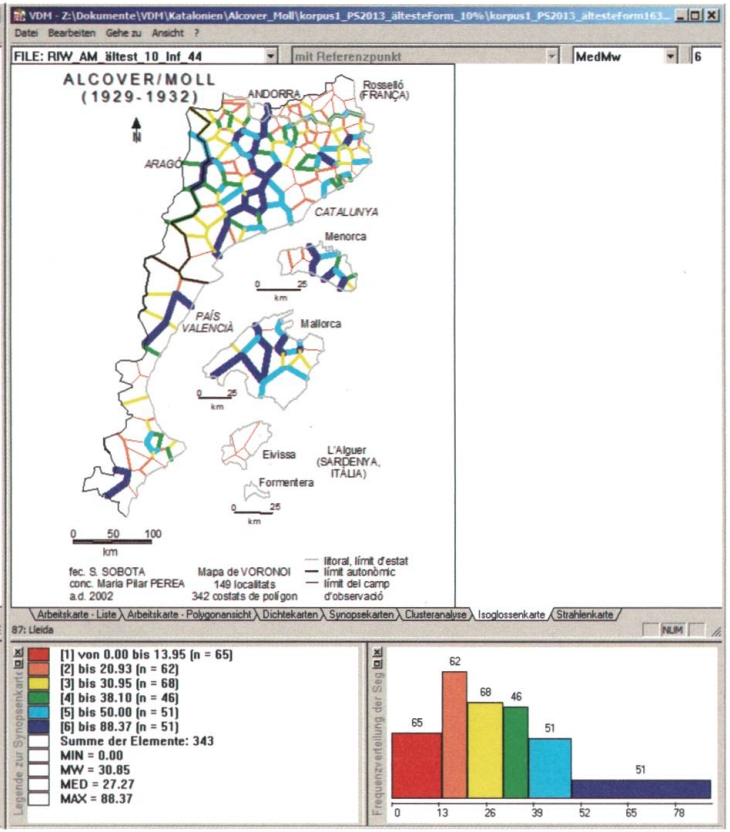


Carte 19: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-AM 57 (Barcelone)
Corpus: 44 CT du corpus RÉD relatives aux formes dialectales de l'*infinitif*
Algorithme d'intervallisation: MEDMW 6-tuple

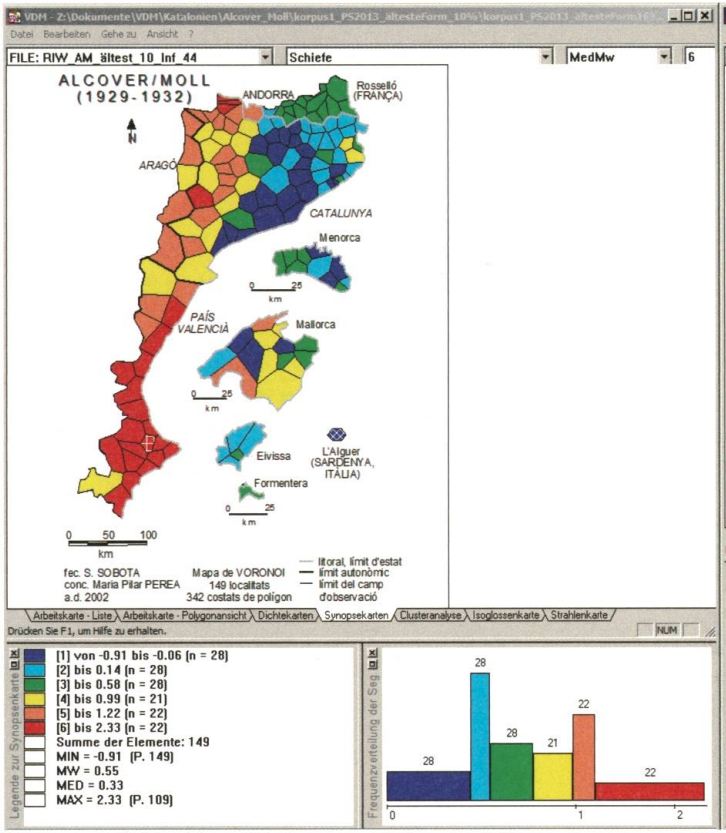
Carte 20: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-AM 87 (Lleida)
Corpus: 44 CT du corpus RÉD relatives aux formes dialectales de l'*infinitif*
Algorithme d'intervallisation: MEDMW 6-tuple



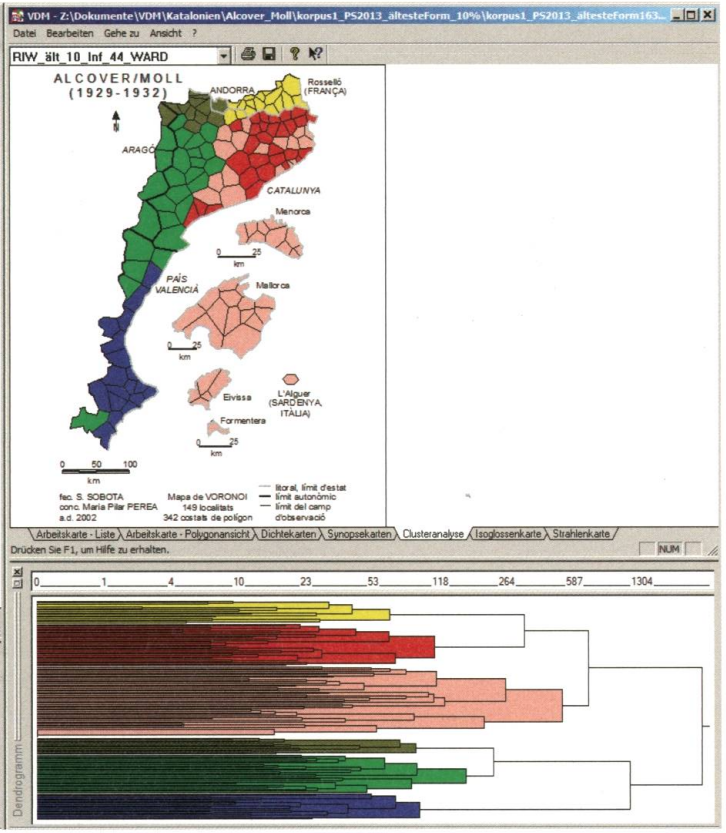
Carte 21: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-AM 130 (Ciutat/Palma de Mallorca)
Corpus: 44 CT du corpus RÉD relatives aux formes dialectales de l'*infinitif*
Algorithme d'intervallisation: MEDMW 6-tuple



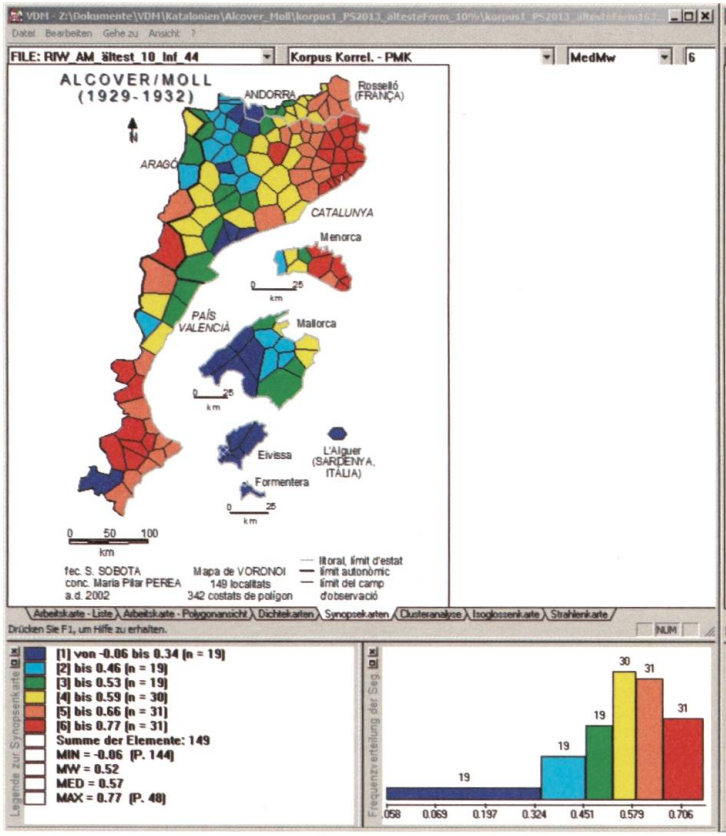
Carte 22: Synthèse isoglottique (« carte à cloisons »)
Corpus: 44 CT du corpus RÉD relatives aux formes dialectales de l'*infinitif*
Indice de distance: IRD_{jk}
Algorithme d'intervallisation: MEDMW 6-tuple



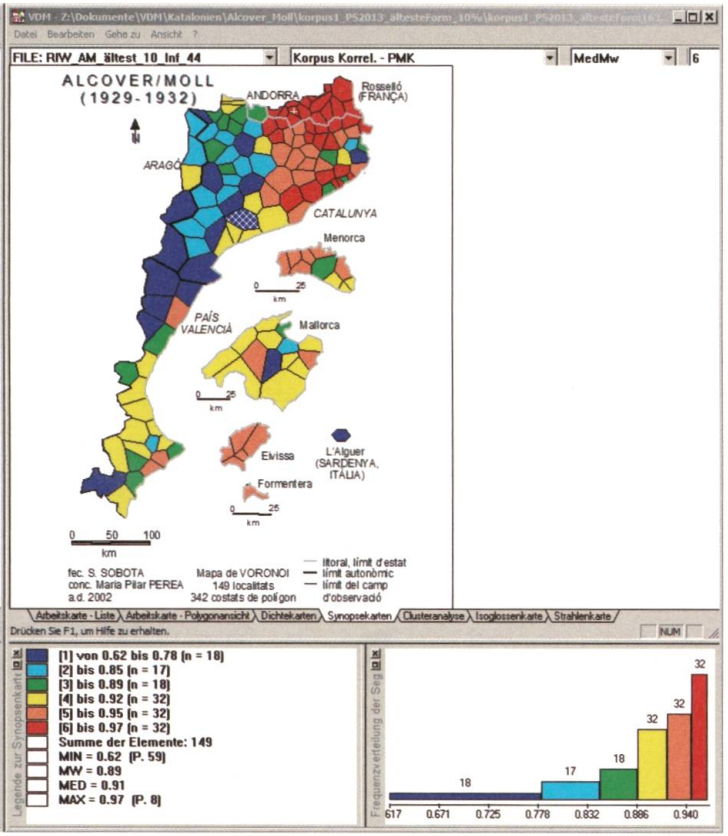
Carte 23: Profil choroplèthe de la synopse de 149 scores-CAF ('coefficient d'asymétrie de Fisher')
Corpus: 44 CT du corpus RÉD relatives aux formes dialectales de l'*infinitif*
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MEDMW 6-tuple



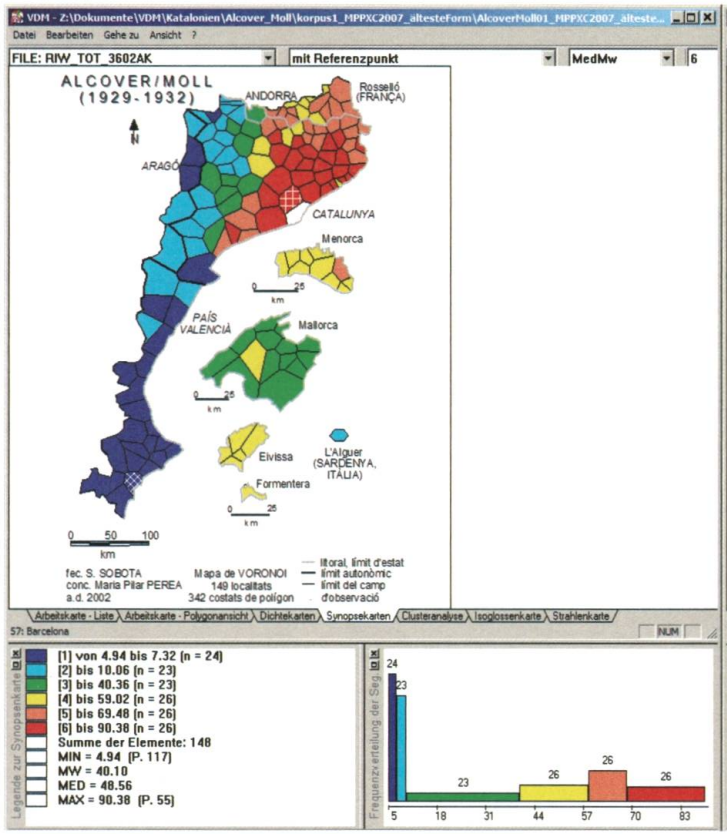
Carte 24: Classification ascendante hiérarchique (méthode de Joe Ward, Jr.)
Corpus: 44 CT du corpus RÉD relatives aux formes dialectales de l'*infinitif*
Indice de similarité: IRI_{jk}
Nombre des chorèmes (en haut) et dendrèmes (en bas) coloriés: 6



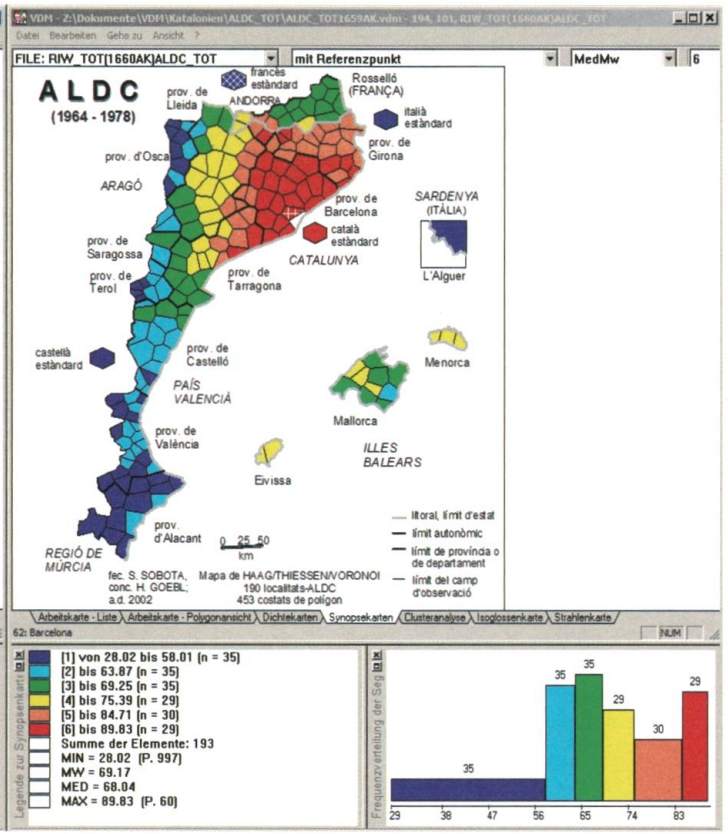
Carte 25 : Profil choroplèthe d'une carte à corrélations
Matrices carrées corrélées :
matrice de similarité (selon IRI_{jk})
matrice de proximité géographique (selon le théorème de Pythagore)
Corpus : 44 CT du corpus RÉD relatives aux formes dialectales de l'*infinitif*
Algorithme d'intervallisation : MEDMW 6-tuple



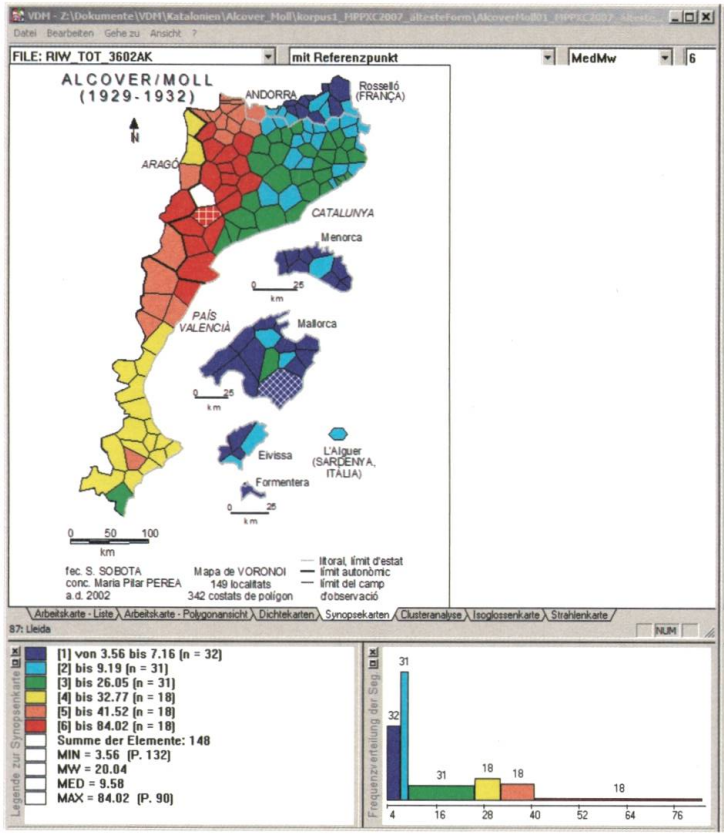
Carte 26 : Profil choroplèthe d'une carte à corrélations
Matrices carrées de similarité corrélées (selon IRI_{jk}) :
matrice A : 44 CT du corpus RÉD (*infinitif*)
matrice B : 997 CT du corpus RÉD (*indicatif*)
Algorithme d'intervallisation : MEDMW 6-tuple



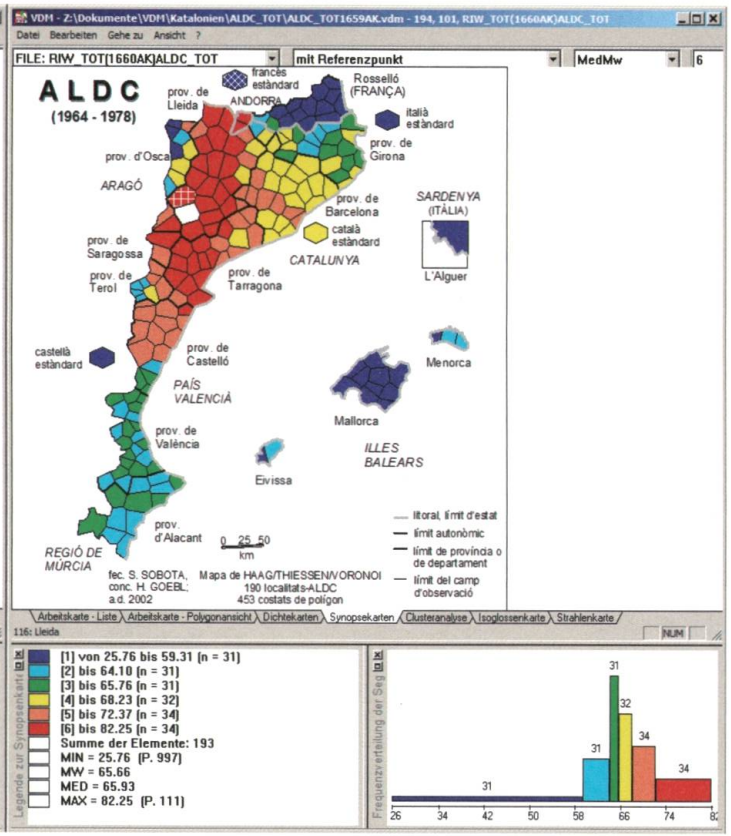
Carte 27: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-AM 57 (Barcelone)
Corpus: 3 602 CT du corpus TOT relatives aux formes dialectales de toutes les catégories de la *morphologie verbale* selon LaFV
Algorithme d'intervallisation: MEDMW 6-tuple



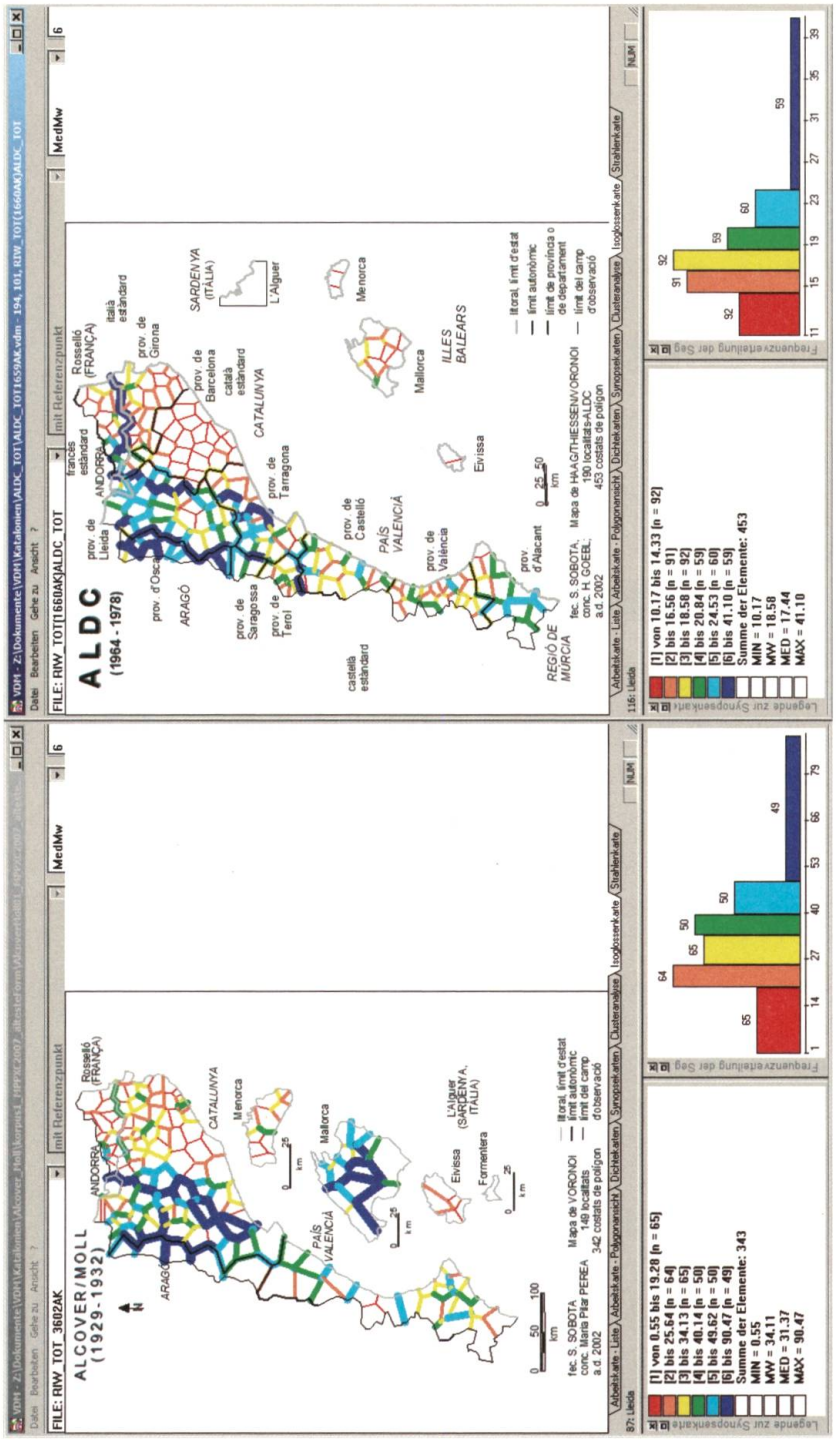
Carte 28: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-ALDC 62 (Barcelone)
Corpus: 1 659 CT de nature *phonétique* et *lexicale* (selon ALDC I-IV)
Algorithme d'intervallisation: MEDMW 6-tuple



Carte 29: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-AM 87 (Lleida)
Corpus: 3 602 CT du corpus TOT relatives aux formes dialectales de toutes les catégories de la *morphologie verbale* selon LaFV
Algorithme d'intervallisation: MEDMW 6-tuple

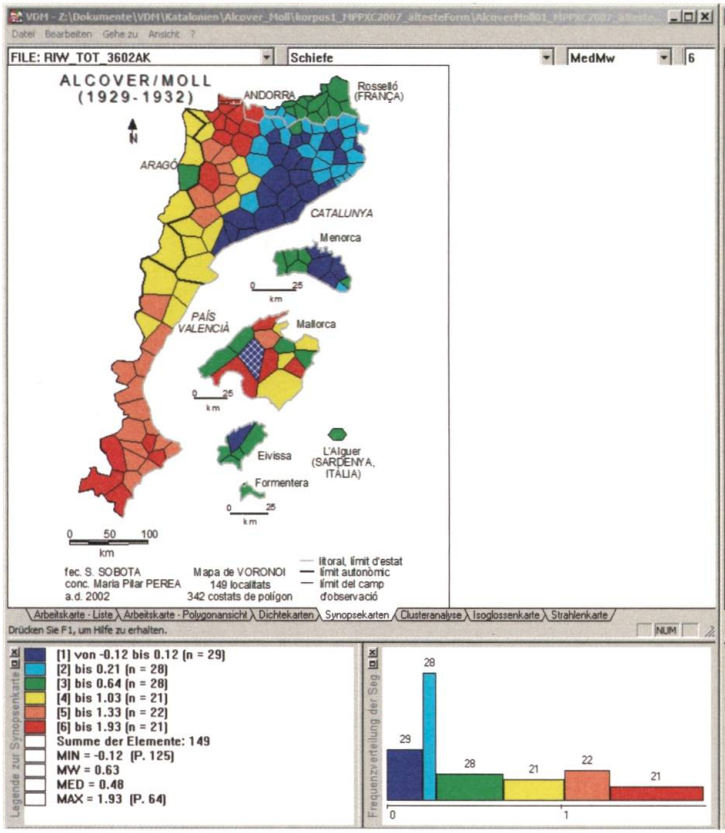


Carte 30: Profil choroplèthe d'une carte de similarité (selon IRI_{jk}) relative au point-ALDC 116 (Lleida)
Corpus: 1 659 CT de nature *phonétique* et *lexicale* (selon ALDC I-IV)
Algorithme d'intervallisation: MEDMW 6-tuple

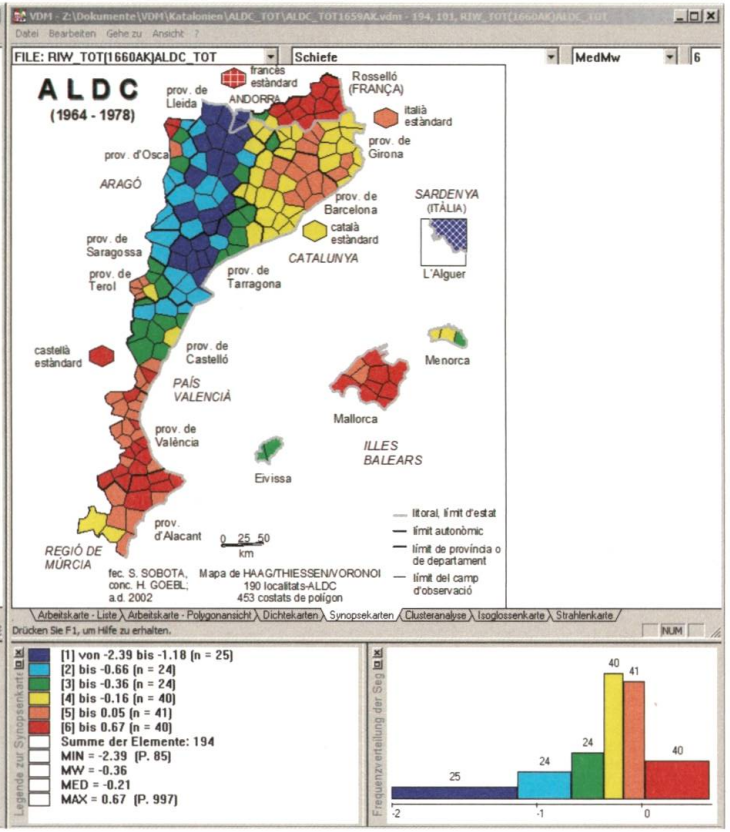


Carte 31: Synthèse isoglottique ('carte à cloisons')
Corpus: 3 602 CT du corpus TOT relatives aux formes dialectales de toutes les catégories de la *morphologie verbale* selon LaFV
Indice de distance: IRD_{jk}
Algorithme d'intervallisation: MEDMW 6-tuple

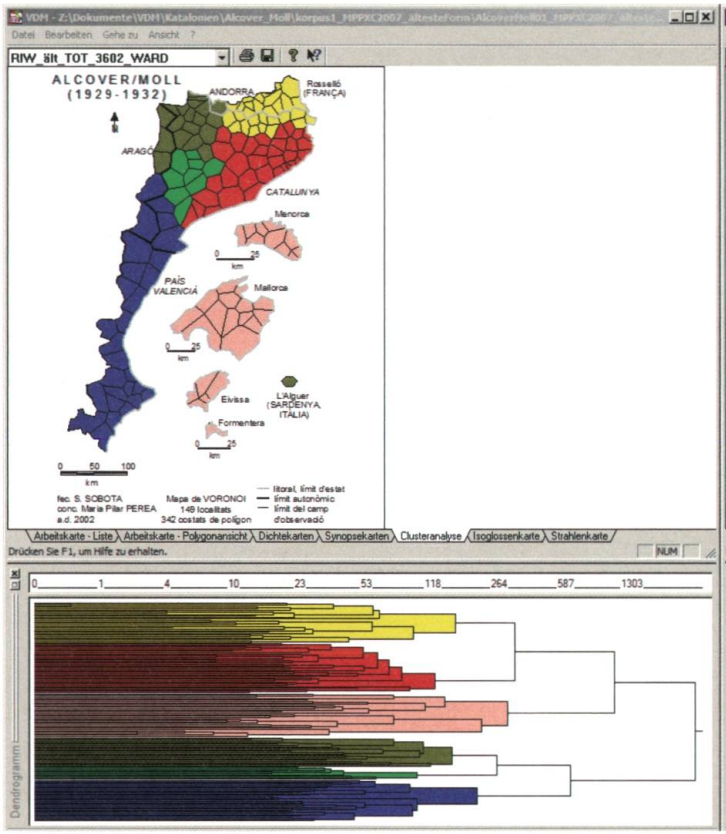
Carte 32: Synthèse isoglottique ('carte à cloisons')
Corpus: 1 659 CT de nature *phonétique* et *lexicale* (selon ALDC I-IV)
Indice de distance: IRD_{jk}
Algorithme d'intervallisation: MEDMW 6-tuple

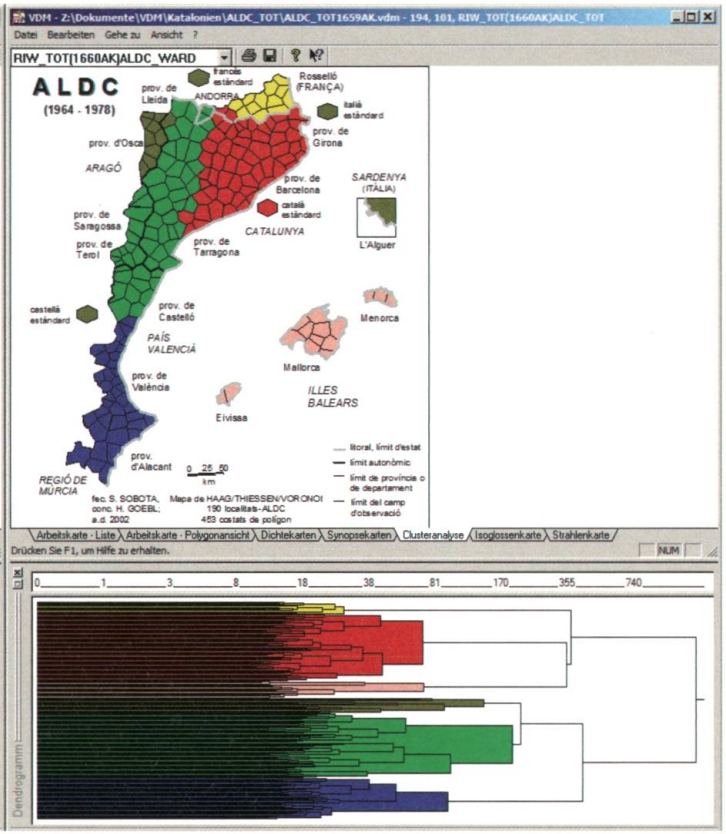


Carte 33: Profil choroplèthe de la synopse de 149 scores-CAF ('coefficient d'asymétrie de Fisher')
Corpus: 3 602 CT du corpus TOT relatives aux formes dialectales de toutes les catégories de la *morphologie verbale* selon LaFV
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MEDMW 6-tuple



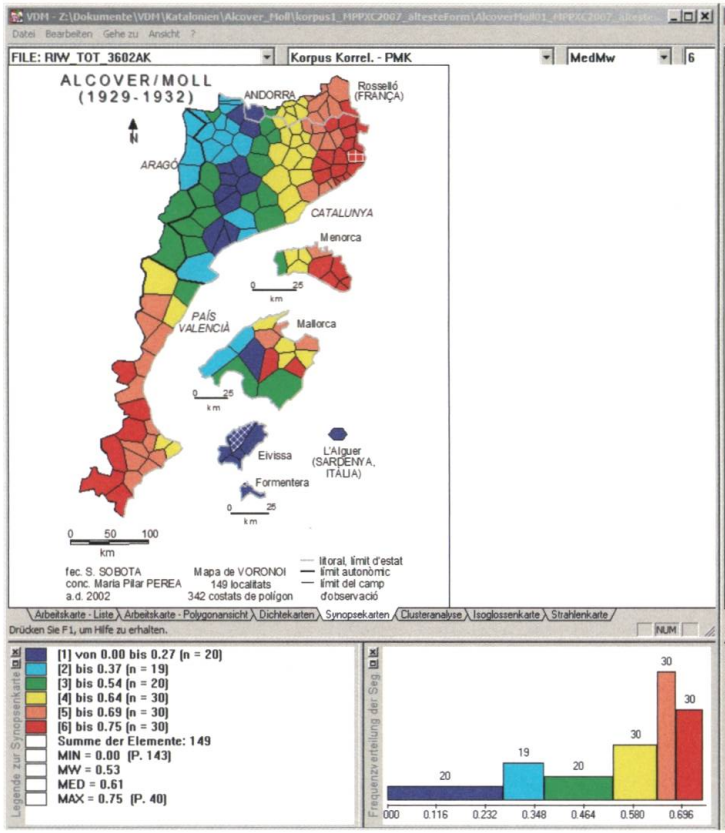
Carte 34: Profil choroplèthe de la synopse de 149 scores-CAF ('coefficient d'asymétrie de Fisher')
Corpus: 1 659 CT de nature *phonétique* et *lexicale* (selon ALDC I-IV)
Indice de similarité: IRI_{jk}
Algorithme d'intervallisation: MEDMW 6-tuple



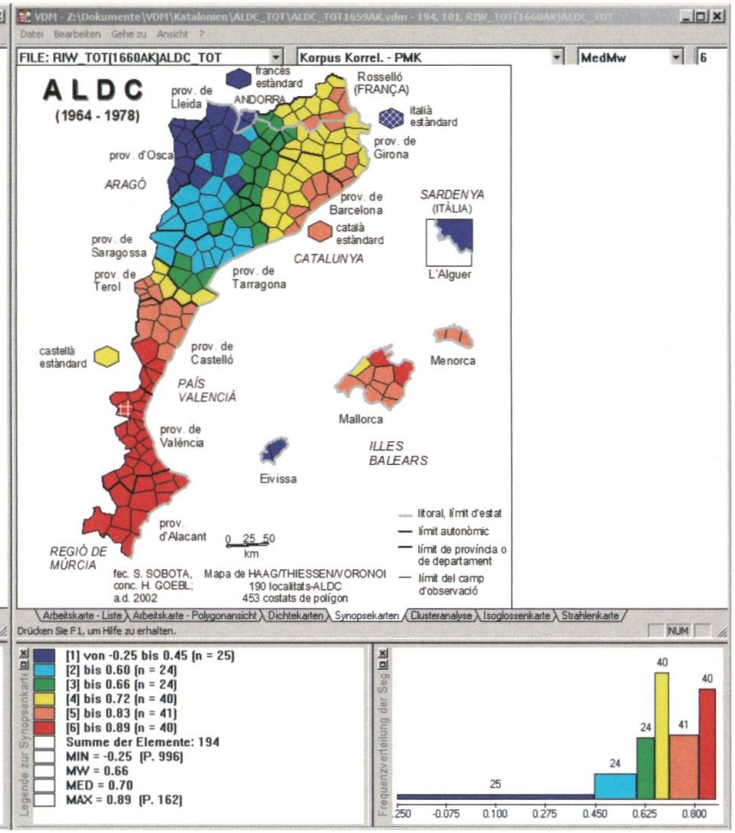


Carte 35: Classification ascendante hiérarchique (méthode de Joe Ward, Jr.)
Corpus: 3 602 CT du corpus TOT relatives aux formes dialectales de toutes les catégories de la *morphologie verbale* selon LaFV
Indice de similarité: IRI_{jk}
Nombre des chorèmes (en haut) et dendrèmes (en bas) coloriés: 6

Carte 36: Classification ascendante hiérarchique (méthode de Joe Ward, Jr.)
Corpus: 1 659 CT de nature *phonétique* et *lexicale* (selon ALDC I-IV)
Indice de similarité: IRI_{jk}
Nombre des chorèmes (en haut) et dendrèmes (en bas) coloriés: 6



Carte 37: Profil choroplèthe d'une carte à corrélations
Matrices carrées corrélées:
matrice de similarité (selon IRI_{jk})
matrice de proximité géographique (selon le théorème de Pythagore)
Corpus: 3 602 CT du corpus TOT relatives aux formes dialectales de toutes les catégories de la *morphologie verbale* selon LaFV
Algorithme d'intervallisation: MEDMW 6-tuple



Carte 38: Profil choroplèthe d'une carte à corrélations
Matrices carrées corrélées:
matrice de similarité (selon IRI_{jk})
matrice de proximité géographique (selon le théorème de Pythagore)
Corpus: 1 659 CT de nature *phonétique* et *lexicale* (selon ALDC I-IV)
Algorithme d'intervallisation: MEDMW 6-tuple