Anàlisi de la variació dialectal del català amb el programa VDM

Esteve Clua (esteve.clua@upf.edu)
Maria-Rosa Lloret (mrosa.lloret@ub.edu)

COL·LOQUI NACS 2013, TORONTO

INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA

FFI2010-22181-C03-02
FFI2010-22181-C03-03

UNIVERSITAT DE BARCELONA

Objectius de la comunicació

- Presentar alguns dels resultats més rellevants de l'anàlisi dialectomètrica de les dades del COD (Corpus oral dialectal del català contemporani) realitzada amb el VDM (Visual Dialectometry)

- Contrastar els resultats d'aquesta anàlisi amb els obtinguts anteriorment amb l'MCOD

Esquema de la comunicació

1. Anàlisi dialectomètrica de les dades del COD.
2. Comparació de resultats: MCOD / VDM
3. Un exemple: les varietats del nord de Castelló.

1. ANÀLISI DIALECTOMÈTRICA DE LES DADES DEL COD

El corpus oral dialectal (COD) del català contemporani

(http://www.ub.edu/lincat)

Característiques:

- qüestionari de 600 ítems
- textos espontanis (10’)
- 82 capitals de comarca de l’àmbit lingüístic català
- 3 informants de 30 a 45 anys
- base de dades amb 135.480 registres fonètics i 532.508 morfològics.

Territoris de parla catalana

82 capitals de comarca o equivalents

COD en versió CD-ROM

Bases de dades	Registres
Verbs	68.262
Fonètica	55.270
Clítics pronominals	24.766
Articles	5.049
Possessius	5.223
Pronoms personals	1.608
Demostratius	1.512
Locatius	803
TOTAL	162.493

Anàlisi dialectomètrica i representació de la distància lingüística

Aplicació de tècniques d'anàlisi multivariant (cluster analysis) a les dades fonològiques i morfològiques del COD per definir la distància lingüística entre varietats i obtenir classificacions dialectals.

Tècniques de descripció i d'anàlisi d'un conjunt d'elements a partir de diverses variables observades, sense que, a priori, s'atorgui un estatus qualitatiu especial a cap d'elles.

Concepte de distància lingüística (DL)

Quantificació de les diferències lingüístiques observades entre varietats.

La DL permet establir clasificacions de varietats a partir del tractament sintètic i global del conjunt de característiques lingüístiques analitzades.

El mètode COD (MCOD)

És una proposta d'anàlisi quantitativa de la variació lingüística que determina la distància lingüística entre varietats a partir d'una anàlisi lingüística prèvia.

Índex de similitud / distància lingüística

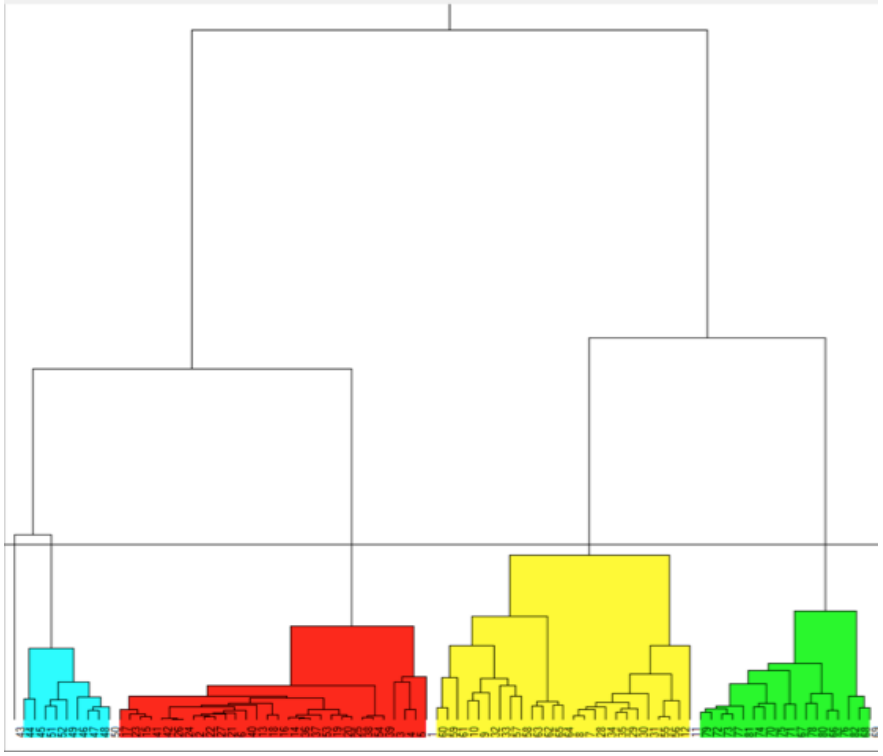
$$dist(i, j) = \frac{\sum_{k=1}^{long} dif_k(i, j)}{long} \times 100$$

És a dir, la distància lingüística entre dues varietats (*i, j*) és igual al sumatori (Σ) de les diferències quant a una variable k entre les varietats (*i, j*), dividit per *long*, que és la longitud (nombre de sons) de cada segment morfològic comparat.

Representació gràfica de la DL

Mètodes de representació:

- Dendrogràfica (Cluster Analysis)
- Cartográfica



Visual Dialectometry (VDM)

- El VDM és un paquet estadístic dissenyat per a l'anàlisi i la representació de la variació lingüística geogràfica, i implementat a l'Escola Dialectomètrica de l'Institut de Romanística de la Universitat de Salzburg, un dels centres de recerca europeus capdavanters en aquest tipus anàlisi, dirigit pel professor Hans Goebl.

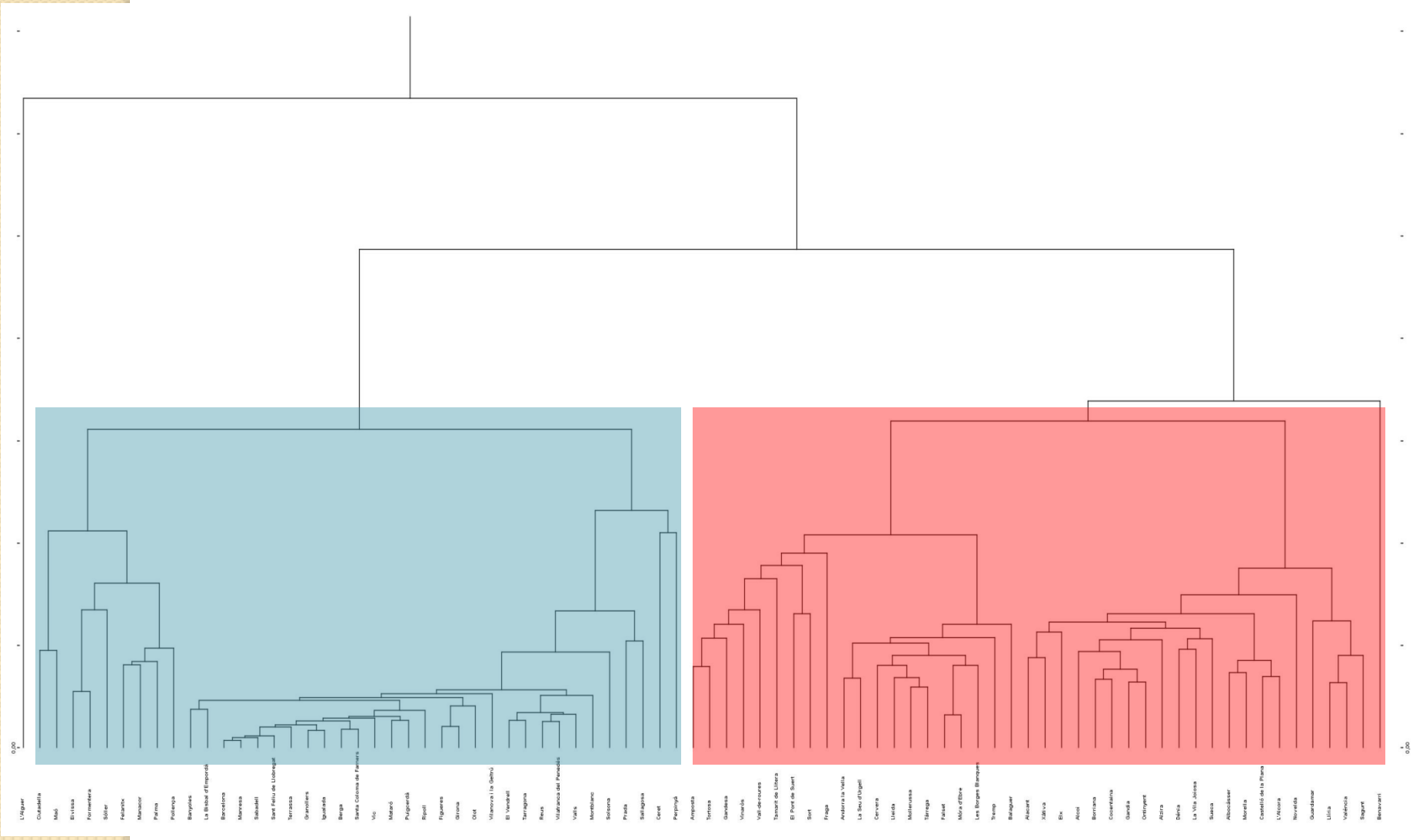
Aplicació a les dades del COD del VDM: representacions dendrogràfiques i representacions espacials mitjançant mapes diversos

- Mapes de similitud
- Mapes de paràmetres
- Mapes interpuntuals en funció discriminatòria (síntesi isoglòtica)
- Mapes interpuntuals en funció comunicativa (mapes de raigs)
- Arbres
- Mapes de correlació

2. COMPARACIÓ DE RESULTATS

MCOD / VDM

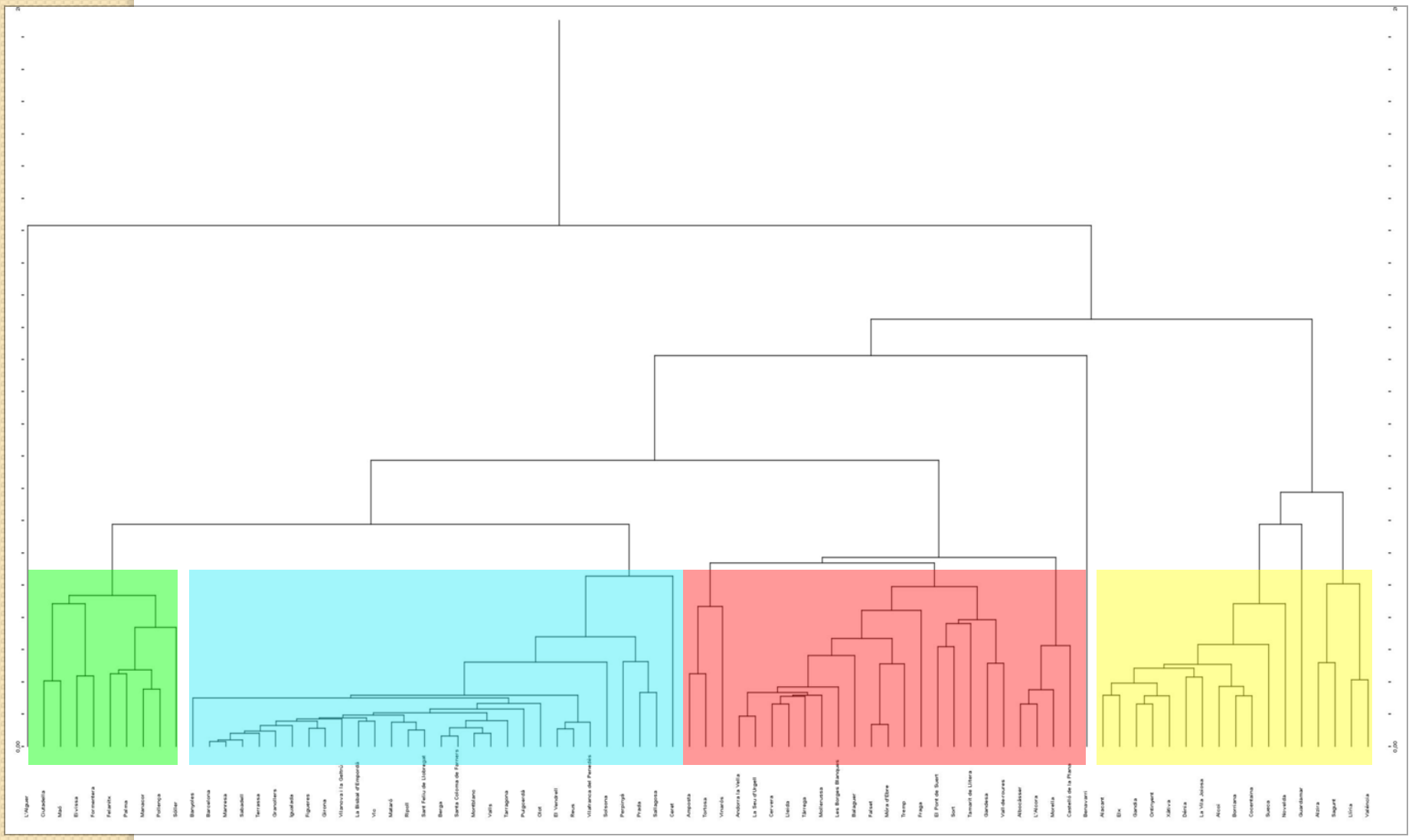
COD / dades fonètiquess



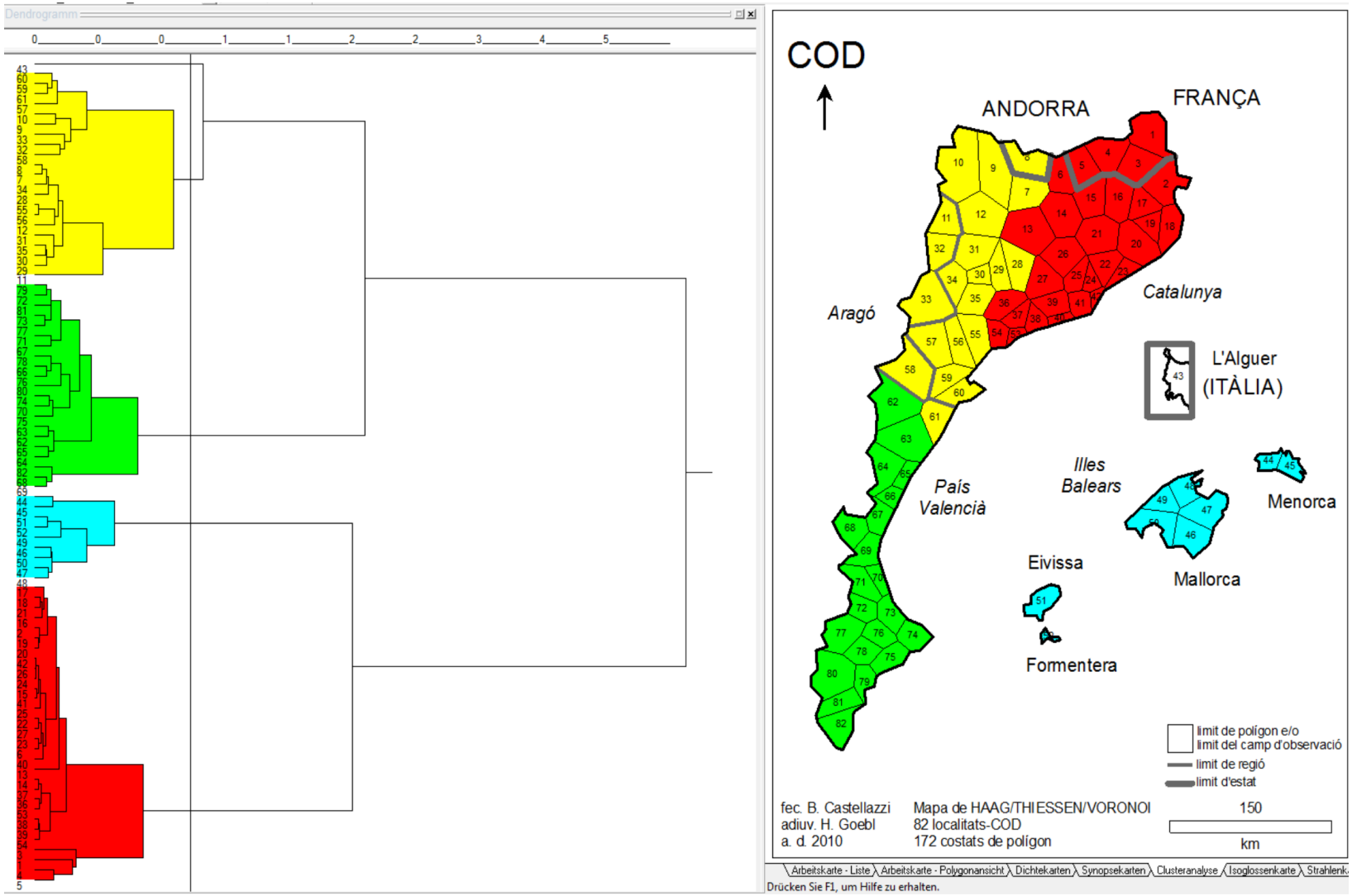
Català Oriental

Català Occidental

COD / dades analitzades fonològicament

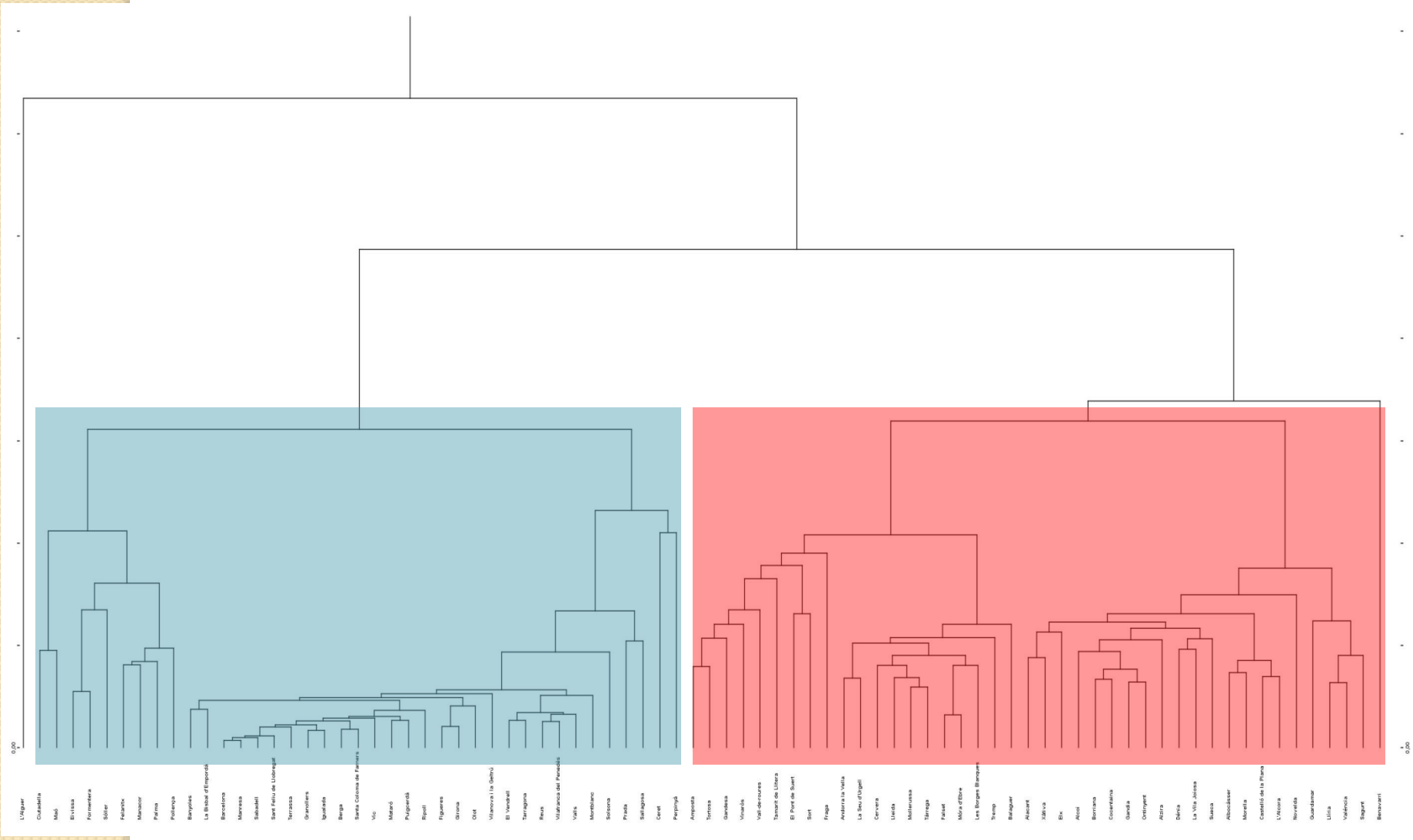


V. Balears V. Centrals V. Nord-occidentals V. Valencianes



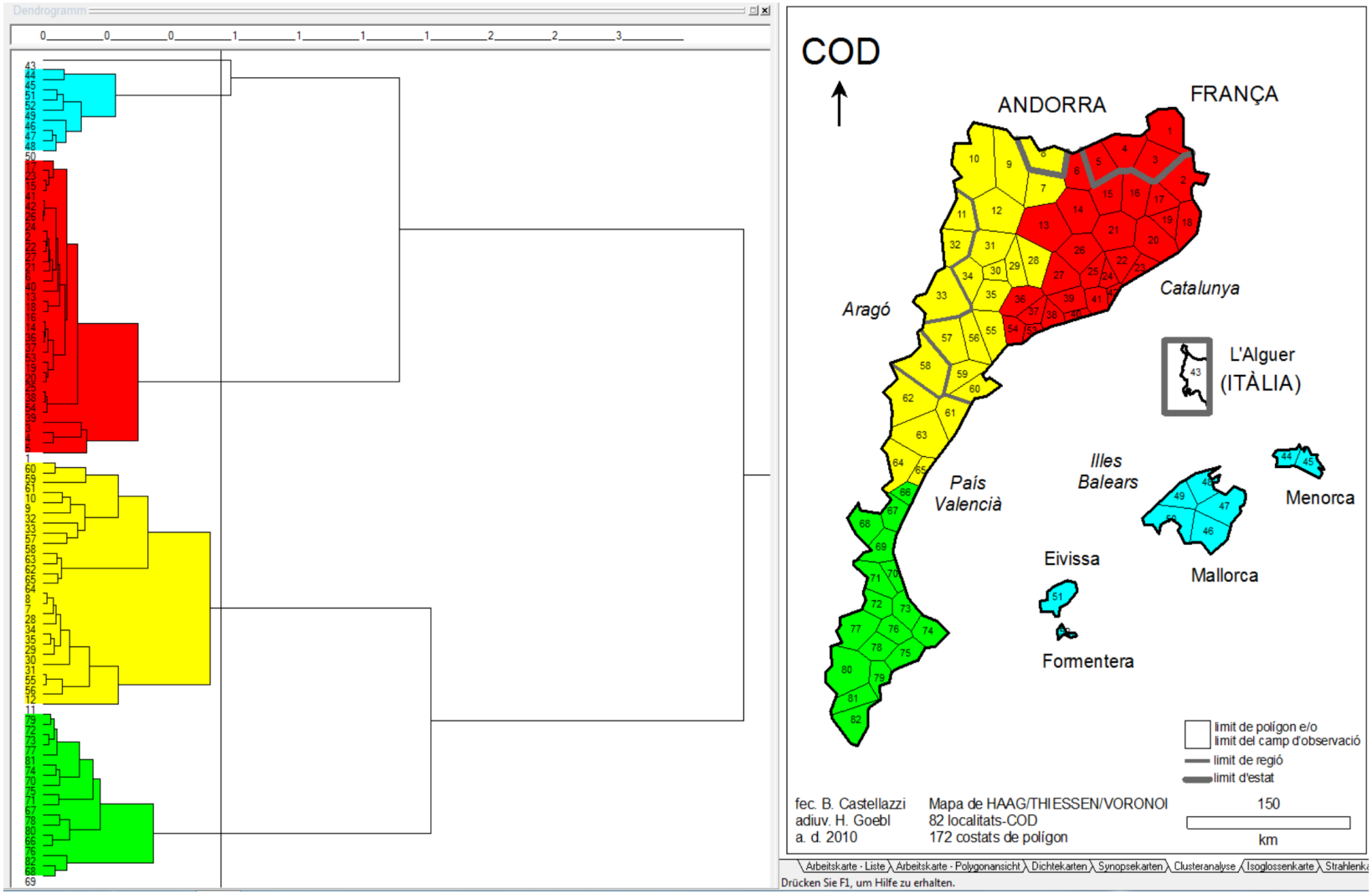
CJA amb el mètode Ward / Índex de distància: MCOD / Corpus: COD dades fonètiques / Número de clústers: 5

COD / dades fonètiques



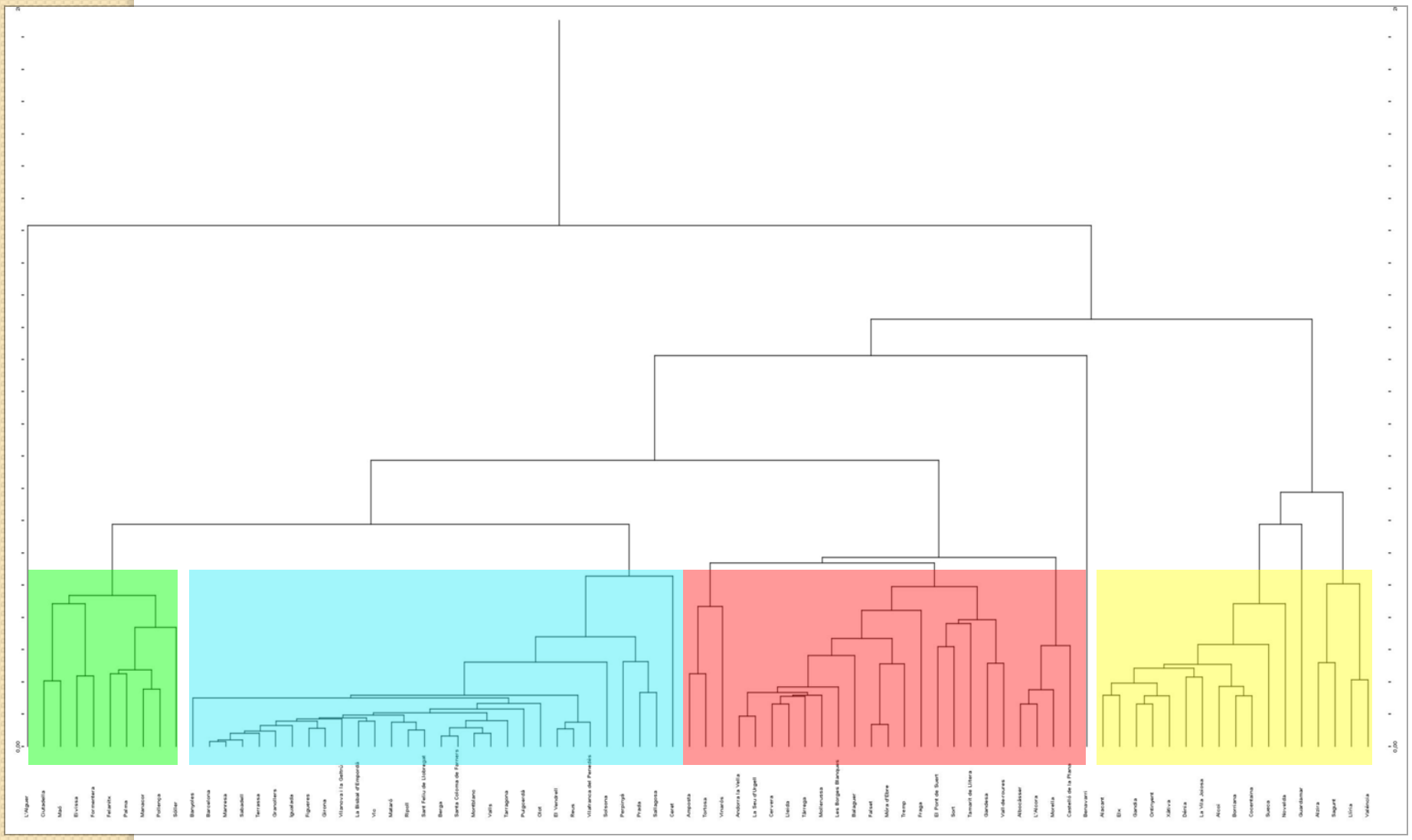
Català Oriental

Català Occidental



Classificació jeràrquica aglomerativa (CJA) amb el mètode Ward / Índex de distància: MCOD / Corpus: COD amb anàlisi fonològica / Número de clústers: 5

COD / dades analitzades fonològicament



V. Balears V. Centrals V. Nord-occidentals V. Valencianes

UN EXEMPLE :
LES VARIETATS DEL NORD DE CASTELLÓ

Classificació dialectal tradicional

Alguer

Les varietats del nord de Castelló

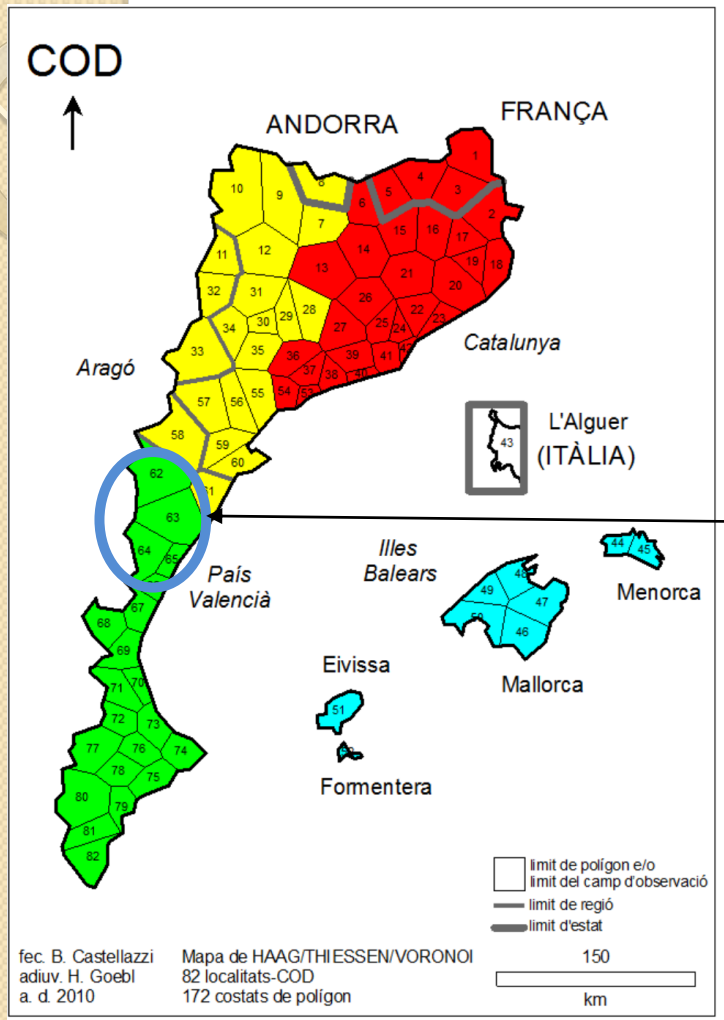
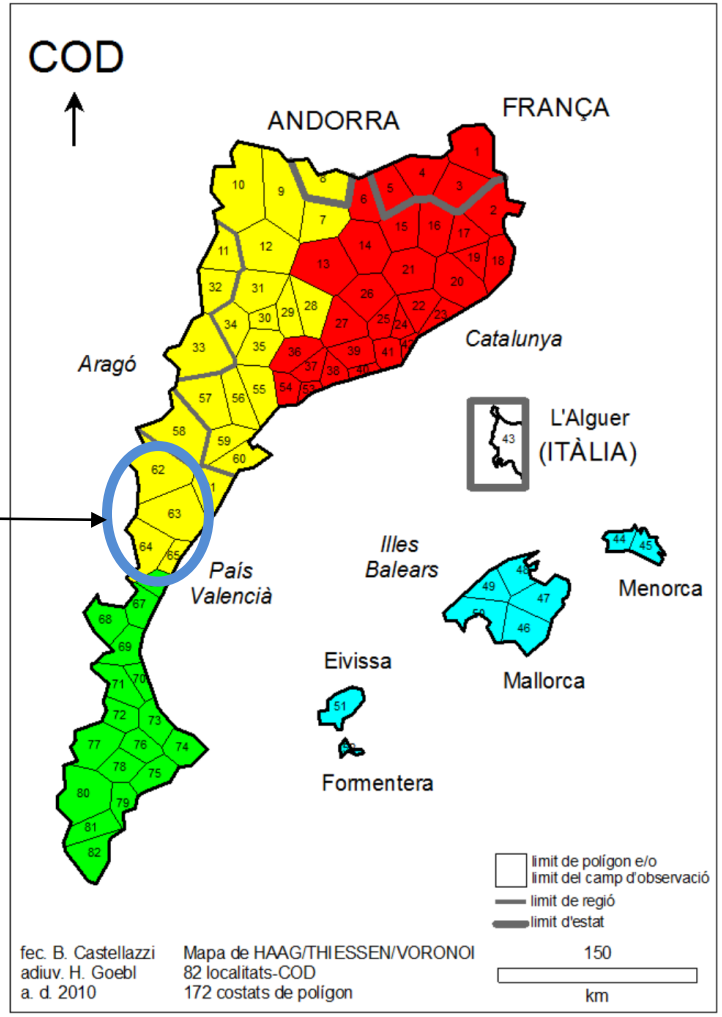


Figura 6. CJA del COD, dades fonètiques



CJA del COD, dades fonològiques

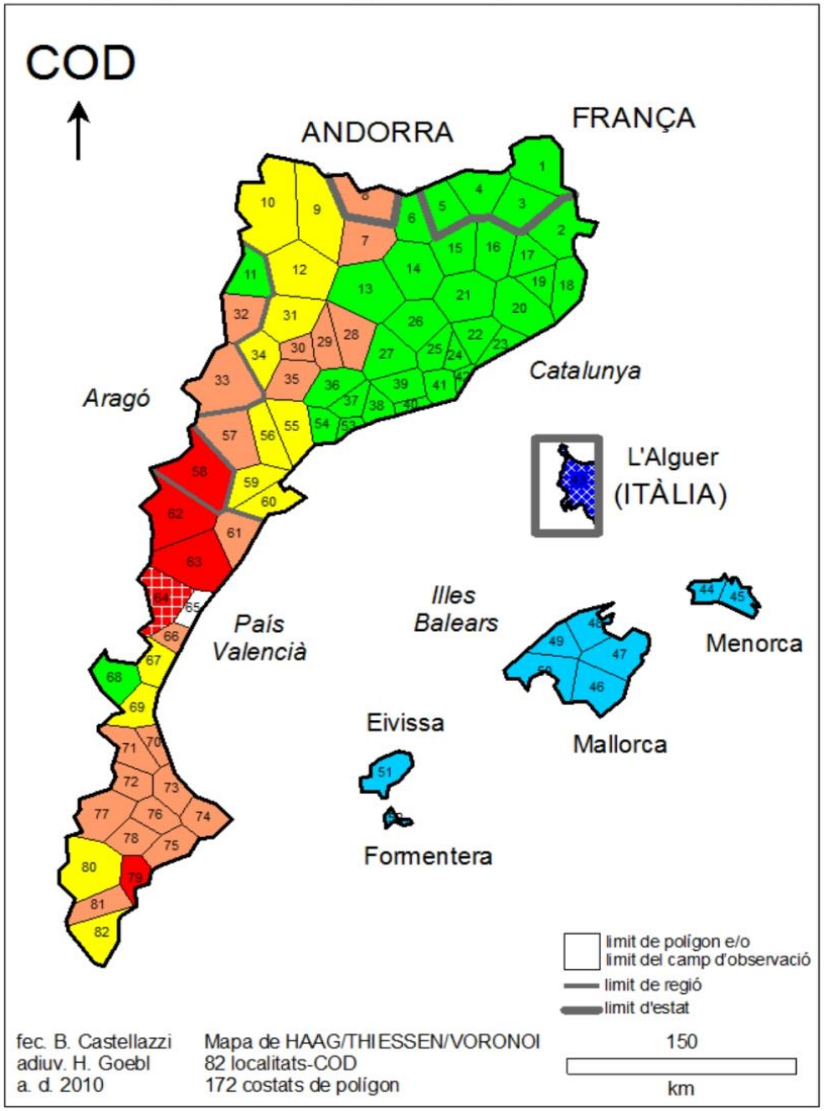
Mapes de similitud cloropetes

Representen la distància lingüística de forma cromàtica:

colors més càlids → varietats més similars

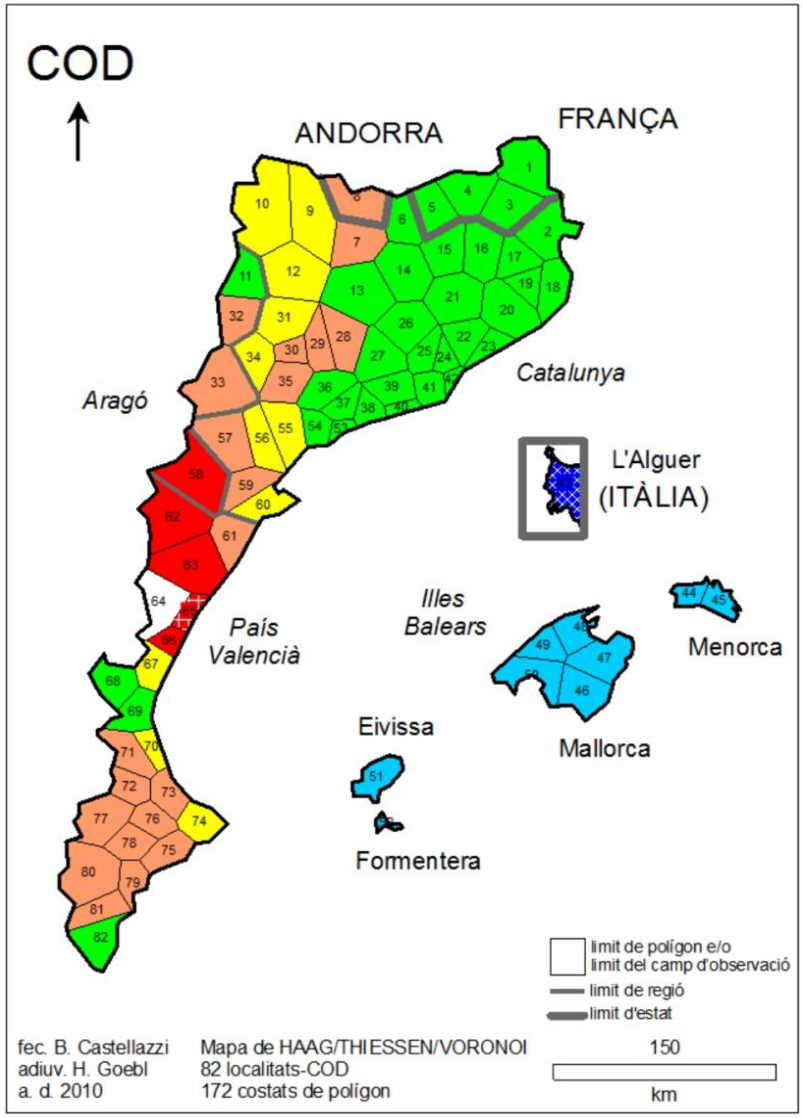
colors més freds → varietats més distants

+ distància ⟷ + similitud



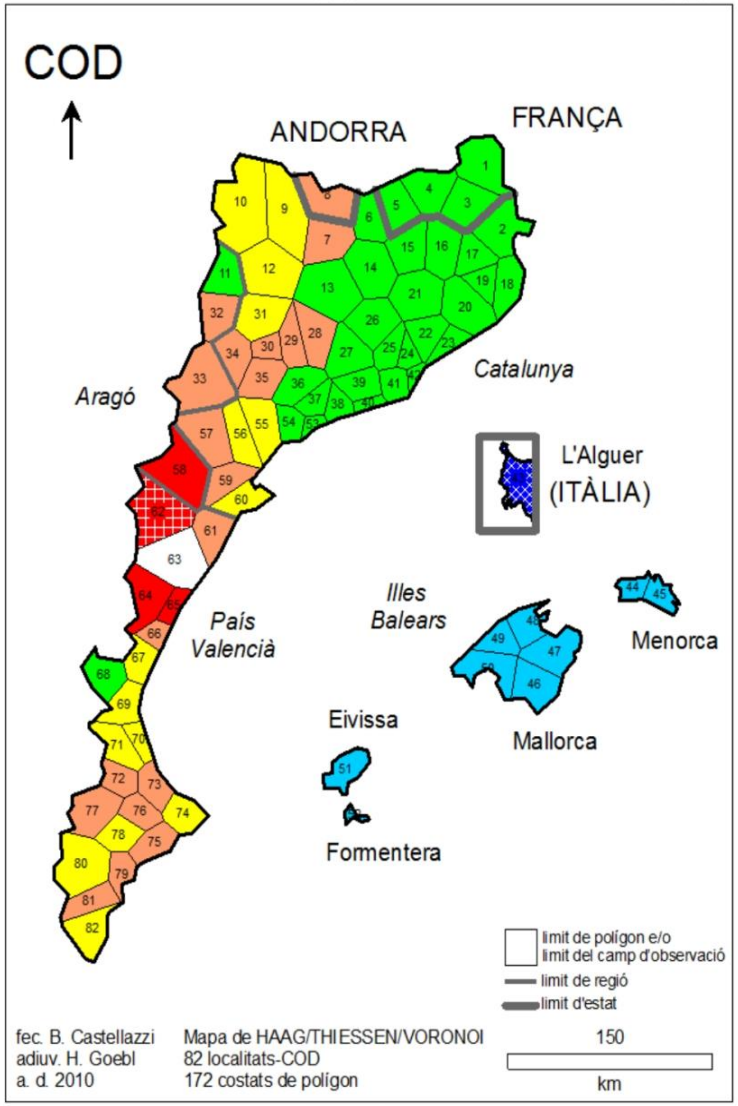
Mapa de similituds de Castelló

Índex de distància: MCOD / Corpus: COD amb anàlisi fonològica / Algorisme d'intervalització MINMWMAX



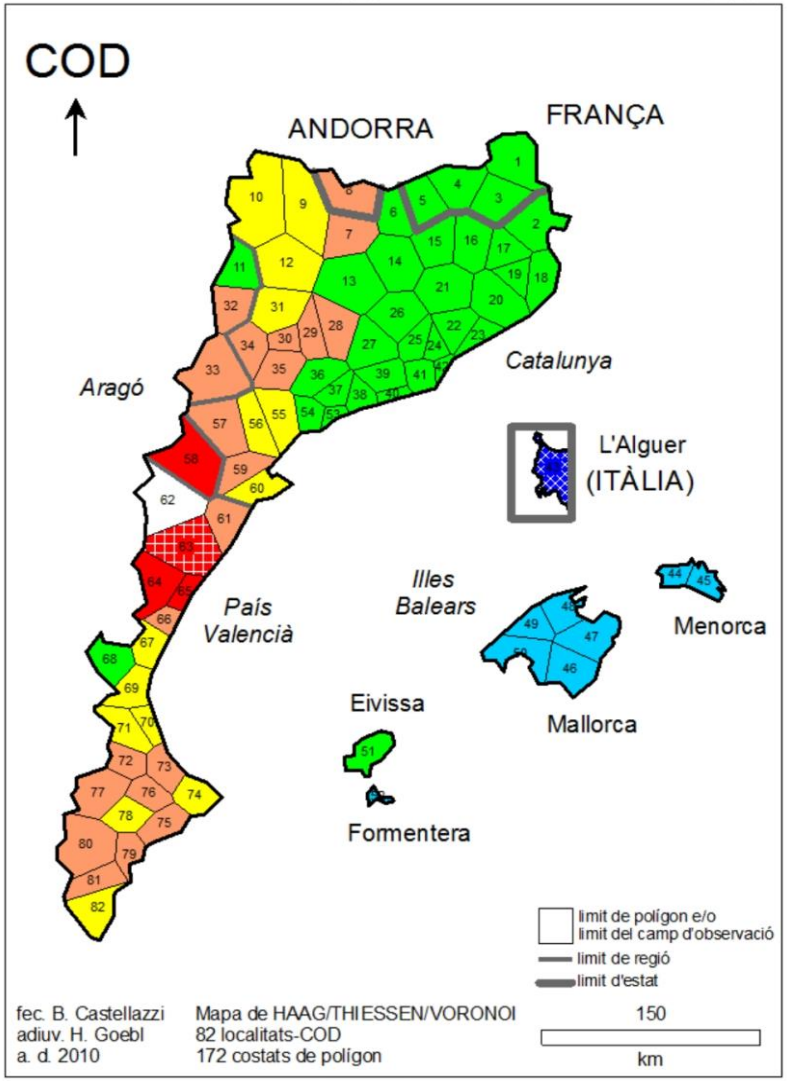
Mapa de similituds de l'Alcora

Índex de distància: MCOD / Corpus: COD amb anàlisi fonològica / Algorisme d'intervalització MINMWMAX



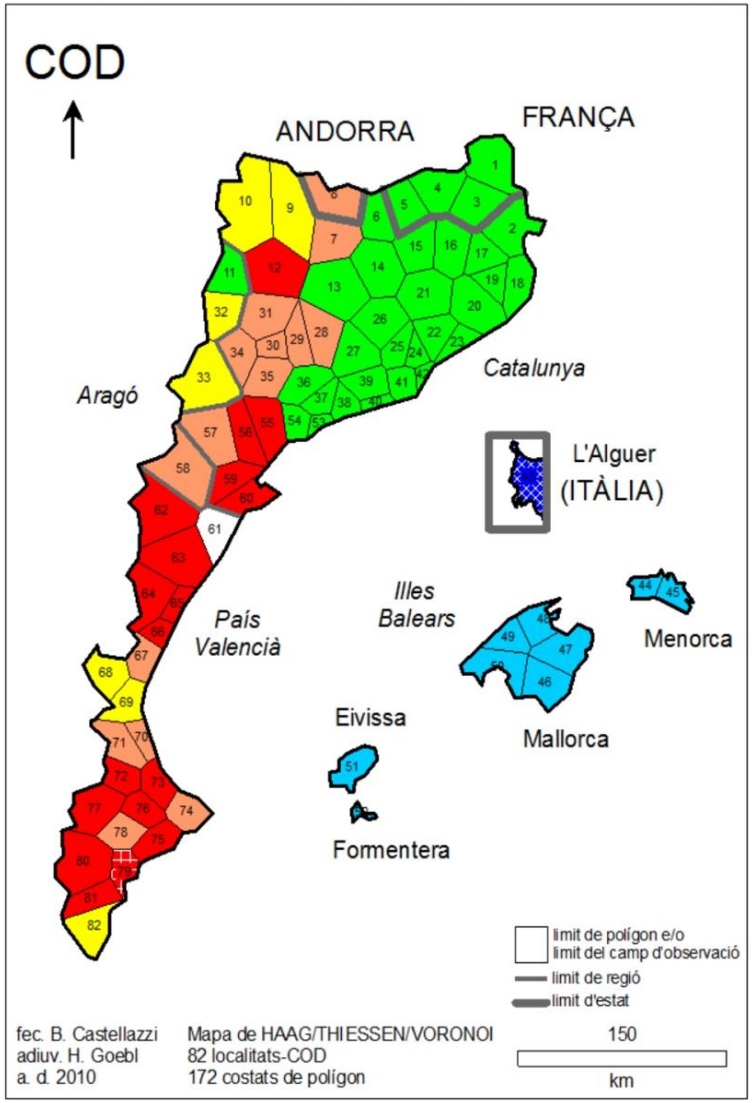
Mapa de similituds d'Albocàsser

Índex de distància: MCOD / Corpus: COD amb anàlisi fonològica / Algorisme d'intervalització MINMWMAX



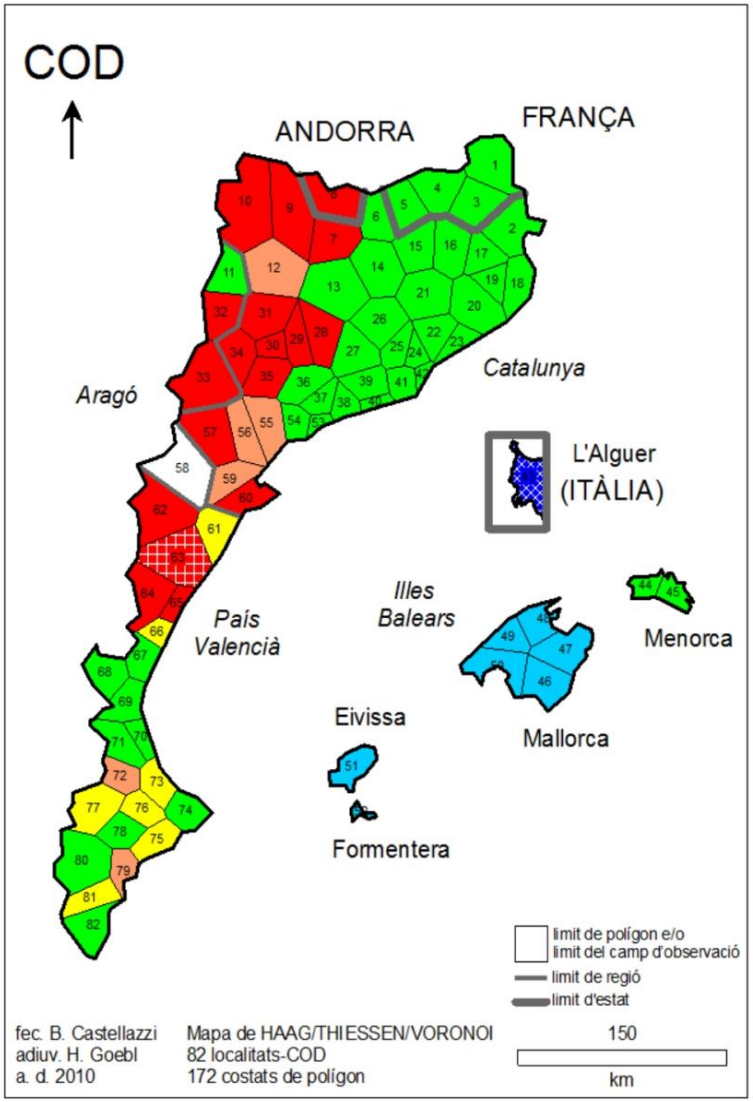
Mapa de similituds de Morella

Índex de distància: MCOD / Corpus: COD amb anàlisi fonològica / Algorisme d'intervalització MINMWMAX



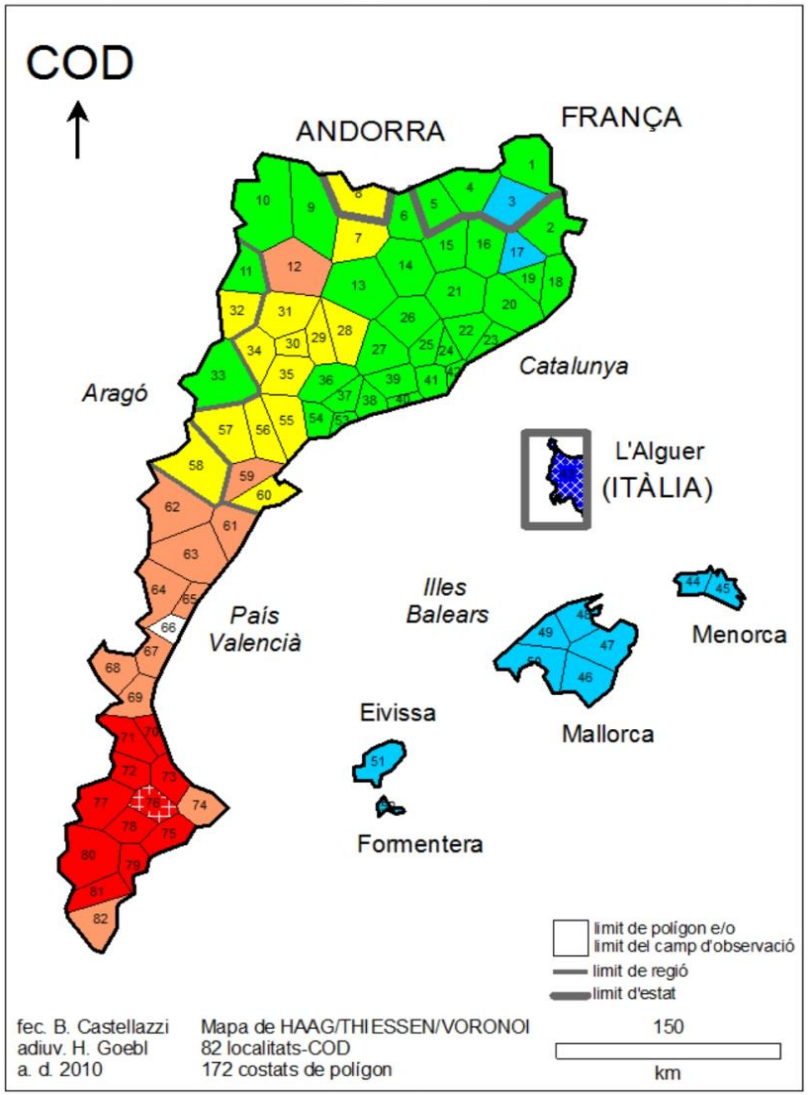
Mapa de similituds de Vinaròs

Índex de distància: MCOD / Corpus: COD amb anàlisi fonològica / Algorisme d'intervalització MINMWMAX



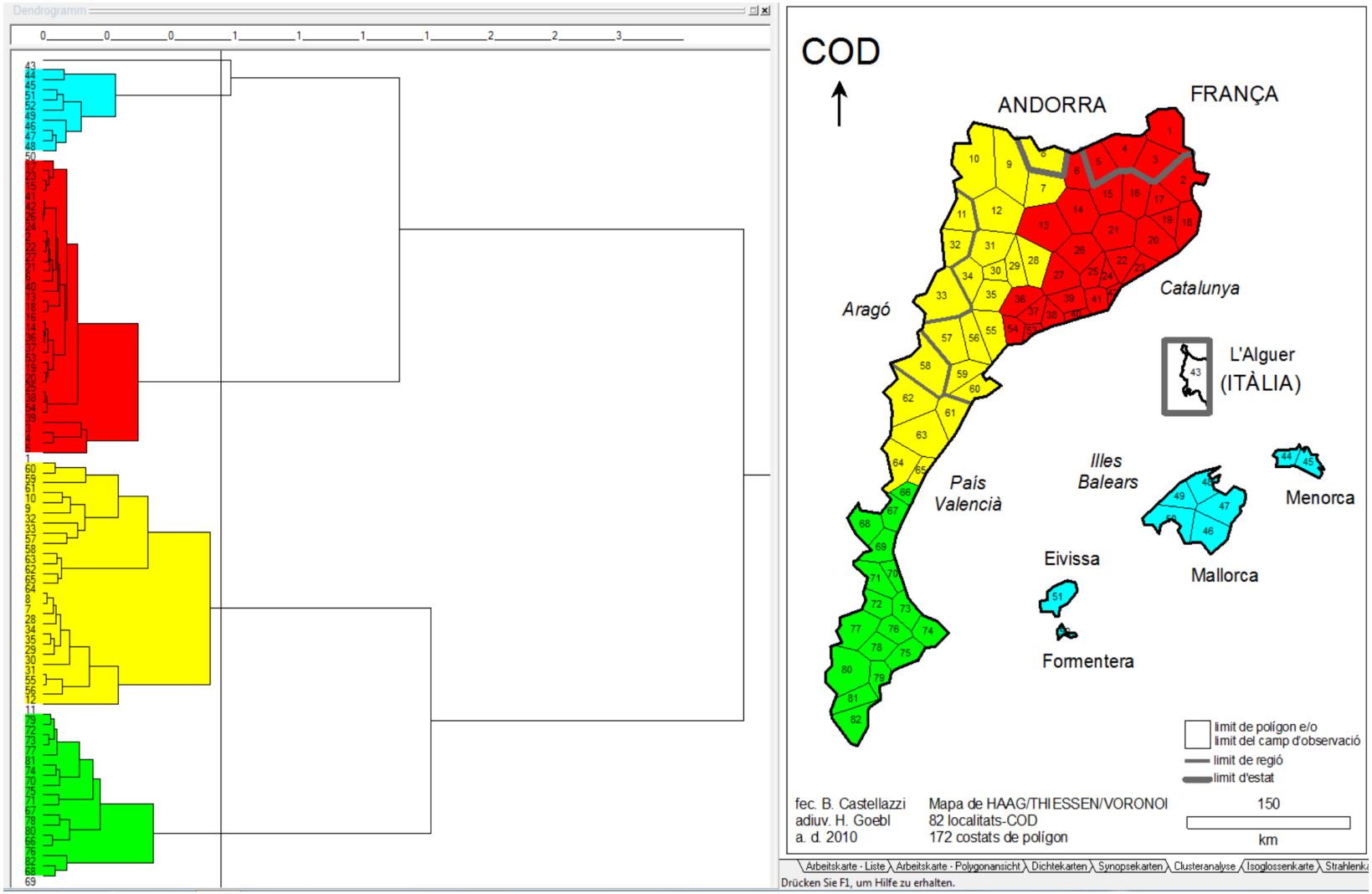
Mapa de similituds de Vall-de-roures

Índex de distància: MCOD / Corpus: COD amb anàlisi fonològica / Algorisme d'intervalització MINMWMAX



Mapa de similituds de Borriana

Índex de distància: MCOD / Corpus: COD amb anàlisi fonològica / Algorisme d'intervalització MINMWMAX



Classificació jeràrquica aglomerativa (CJA) amb el mètode Ward / Índex de distància: MCOD / Corpus: COD amb anàlisi fonològica / Número de clústers: 5

Moltes gràcies per la vostra atenció

mrosa.lloret@ub.edu

esteve.clua@upf.edu

Referències bibliogràfiques

- Chambers, J. K. i P. Trudgill (1980): *Dialectology*. Cambridge University Press, Cambridge.
- Clua, E. (1999): "Distància lingüística i classificació de varietats dialectals", *Caplletra*. *Revista Internacional de Filologia*, núm. 26. Publicacions de l'Abadia de Montserrat / Institut de Filologia Valenciana, Barcelona / València, pp.11-26.
- Clua, E. (2010): "Relevancia del análisis lingüístico en el tratamiento cuantitativo de la variación dialectal", dins Aurrekoetxea, G. i J. L. Ormaetxea (eds.), *Tools for Linguistic Variation*. Universidad del País Vasco, Bilbao, pp.151-166.
- Clua, E. i M.-R. Lloret (2006): "New tendencies in geographical dialectology: The Catalan Corpus Oral Dialectal (COD)", dins Montreuil, J.P. (ed.), *New Perspectives on Romance Linguistics*. Vol. 2: *Phonetics, phonology and dialectology*. John Benjamins, Amsterdam / Philadelphia, pp. 31-47.

- Clua, E.; M-R. Lloret i E. Valls (2009): “Análisis lingüístico y dialectométrico del corpus oral dialectal (COD)”, dins Cantos Gómez, P. i A. Sánchez Pérez (eds.), *A survey on corpus-based research*. Asociación Española de Lingüística del Corpus, Murcia, pp. 1033-1045.
- Colomina i Castanyer, J. (1999*): Dialectologia catalana. Introducció i guia bibliogràfic*a. Servei de Publicacions de la Universitat d’Alacant, Alacant.
- Goebl, H. (2010): “Introducción a los problemas y métodos según los principios de la escuela dialectométrica de Salzburgo”, dins Aurrekoetxea, G. i J. L. Ormaetxea (eds.), *Tools for Linguistic Variation*, Universidad del País Vasco, Bilbao, pp. 3-39.
- Goebl, H. (2010): “Dialectometry and quantitative mapping”, dins: Lameli, A., R. Kehrein i E. Rabanus (eds.): *Language and Space. An International Handbook of Linguistic Variation*, vol. 2: Language Mapping (Handbücher der Sprach- und Kommunikationswissenschaft [HSK] 30.2.), Berlin: Mouton de Gruyter; 1st part: 433-457 (text), 2nd part (maps), pp. 2201-2212.