Explotación de corpus orales: la experiencia del COD

(**C**orpus **O**ral **D**ialectal del catalán)

Esteve Clua
Universitat Pompeu Fabra
esteve.clua@upf.edu

Maria-Rosa Lloret
Universitat de Barcelona
mrosa.lloret@ub.edu

Resultados del proyecto ECOD

Explotación de un **C**orpus **O**ral **D**ialectal: análisis de la variación lingüística y desarrollo de aplicaciones informáticas

HUM2004-1504/FILO (1ª fase), 2004-07

HUM2007-65531/FILO (2ª fase), 2007-10

P. web del proyecto: www.ub.edu/lincat

Territorios de habla catalana

EQUIPO INVESTIGADOR

Universitat de Barcelona (9)

- **Maria-Rosa Lloret** (IP)
- **Joaquim Viaplana**
- **Maria-Pilar Perea**
- Montserrat Adam
- Sebastià Bonet
- Andreu Bosch
- Josefina Carrera
- Josep Moran
- Clàudia Pons

Universitat Pompeu Fabra (1)

- **Esteve Clua**

Universitat de València (1)

- Jesús Jiménez

Universitat Catòlica de València (1)

- Ricard Herrero (HUM2007)

Euskal Herriko Unibertsitatea (1)

- Gotzon Aurrekoetxea (HUM2004)

Universität Salzburg (2)

- Xavier Casassas (HUM2007)
- Hans Goebl

Universität Basel (1)

- Germà Colon

Meertens Instituut (Amsterdam) (1)

- Franz Hinskens

Becarios predoctorales (U Barcelona) (9)

- Elisenda Campmany (2003-06; dir.: Lloret)
- Laia Querol (2003-06; dir.: Lloret)
- Laia Rosàs (2004-07; dir.:Perea)
- Núria Busquet (2005-08; dir.:Perea)
- Joan-Pau Callejón (2005-08; dir.: Bonet)
- Roser Marí (2006-09; dir.: Perea)
- Maria Cabrera (2007-10; dir: Carrera & Pons)
- Esteve Valls (2007-10; dir.: Clua & Viaplana)
- Margalida Adrover (2008-11; dir.: Lloret)

OBJETIVOS

1. Completar e informatizar los **datos** del *Corpus Oral Dialectal* (COD) del catalán contemporáneo, recopilado en proyectos anteriores. (J. Viaplana, junto con M-R. Lloret, M-P. Perea y E. Clua)

2. Análisis de los **fenómenos lingüísticos** en relación con la variación dialectal. (M-R. Lloret)

3. Análisis **dialectométrico** y representación de la **distancia lingüística**. (E. Clua)

4. Elaboración de un **atlas sonoro** fonético-morfológico. (M-P. Perea)

¿De dónde provienen los datos del COD?

- **Proyectos anteriores**, dirigidos por J. Viaplana:
  - PB94-0951 (1994-1997)
  - PB97-0889 (1998-2001)
  - BFF2001-3798 (2002-2004)

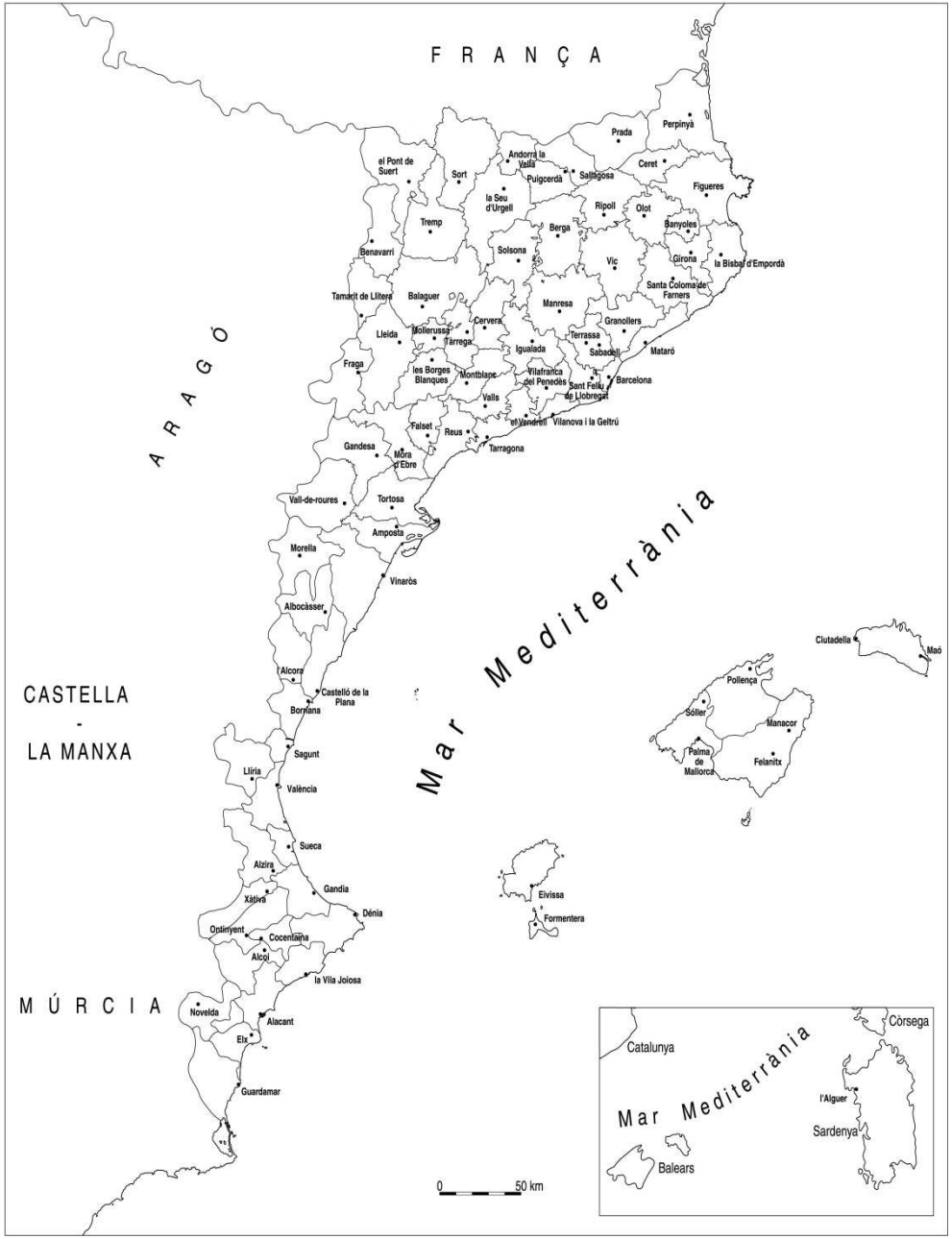
- **Encuestas**, básicamente entre 1994-96:
  82 localidades (capitales de comarca o similares)
  303 entrevistas (mínimo 3 por localidad)

- **Cuestionario** fonético-morfológico (aprox. 600 ítems)
  Texto libre (10')

- **Informantes** 30-45 años, no escolarizados en catalán; nivel cultural medio

- Distinto a trabajos geolingüísticos tradicionales (Veny & Pons)

Territorios de habla catalana



COD

82 localidades

303 entrevistas

Cuestionario

1. COMPLETAR

Ej.: VERBOS REGULARES: I conjugación (*cantar*):

- Cada dia al matí jo …… una cançó. [canto]
- Mentre ens esperàvem, jo ….. una cançó. [cantava]

2. IDENTIFICAR

Ej.: FONÉTICA

- [Imagen] (perla) ……

3. TRADUCIR del español, francés o italiano

Ej.: COMBINACIÓN DE PRONOMBRES PERSONALES

- Si tus hijos quieren ese coche, *se lo* compras. [els el / els hi compres]

Metodología

- Grabación digital.

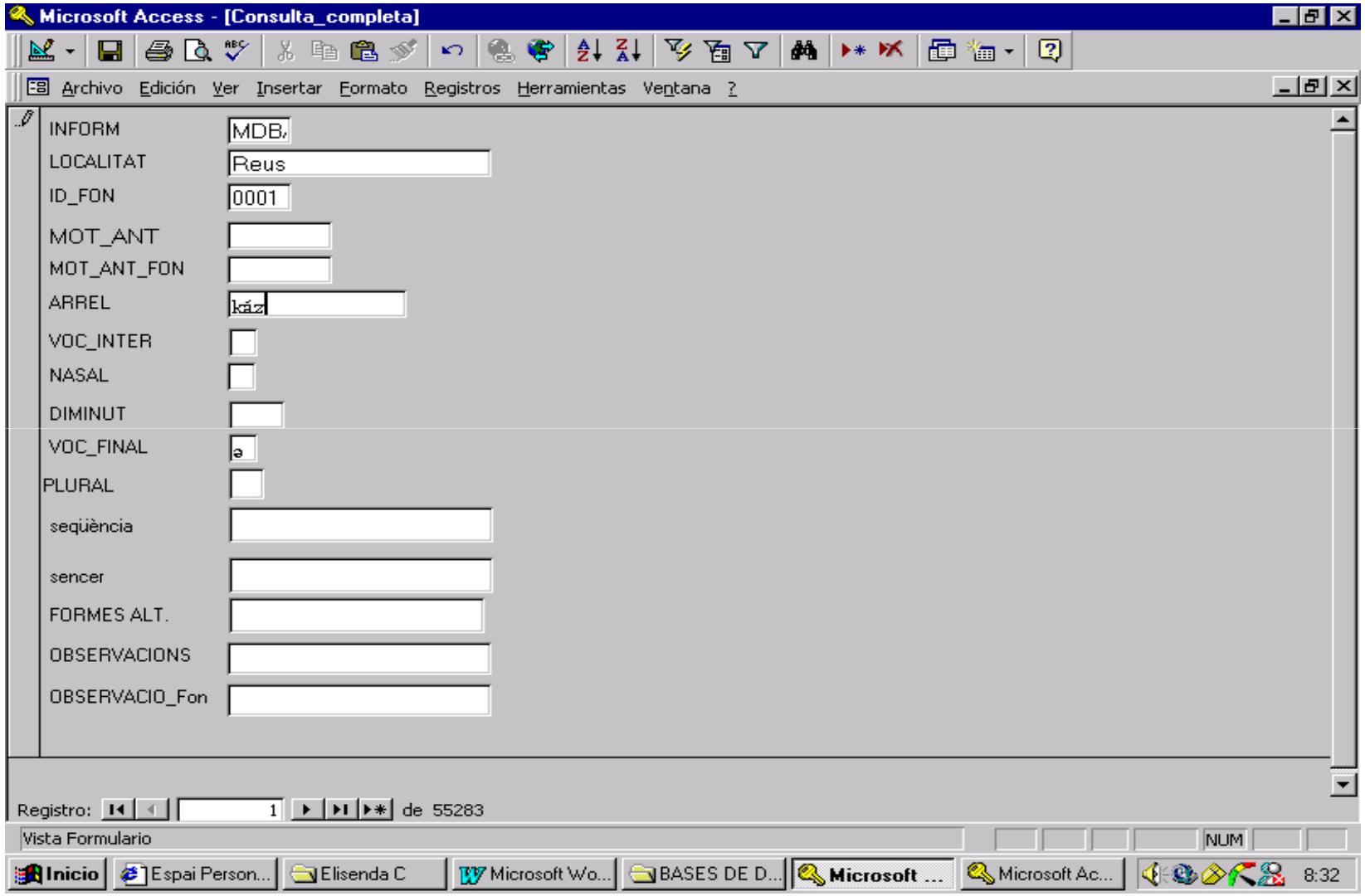
DATOS DEL CUESTIONARIO:

- Transcripción ortográfica y fonética de los datos segmentados morfológicamente. Bases de datos diseñadas con Microsoft ® Access, **específicas** para cada conjunto de datos.
- Supervisión de los datos transcritos.
- Exportación de los datos a base de datos (© ORACLE) para poder **relacionar** cada registro con registro sonoro.
- Selección automática de la forma más ***frecuente*** (3>1).

Ejemplo base de datos creada para los verbos

Fonética	Raíz	Ep	Ext	Tema	Modo	MT	Ep2	NP
kəntəɾíə	kənt			ə		ɾíə		
pərðəɾɛ́m	pərð	ə				ɾɛ́		m
bəɣént	bə		ɣ	é	nt			
báwɾə	báw				ɾ		ə	
sénten	sént						e	n
seɾβíʃko	seɾβ		íʃk			o		

Base de datos (fonética)



Base de datos (clíticos pronominales)

Microsoft Access - [Total]

Archivo Edición Ver Insertar Formato Registros Herramientas Ventana ?

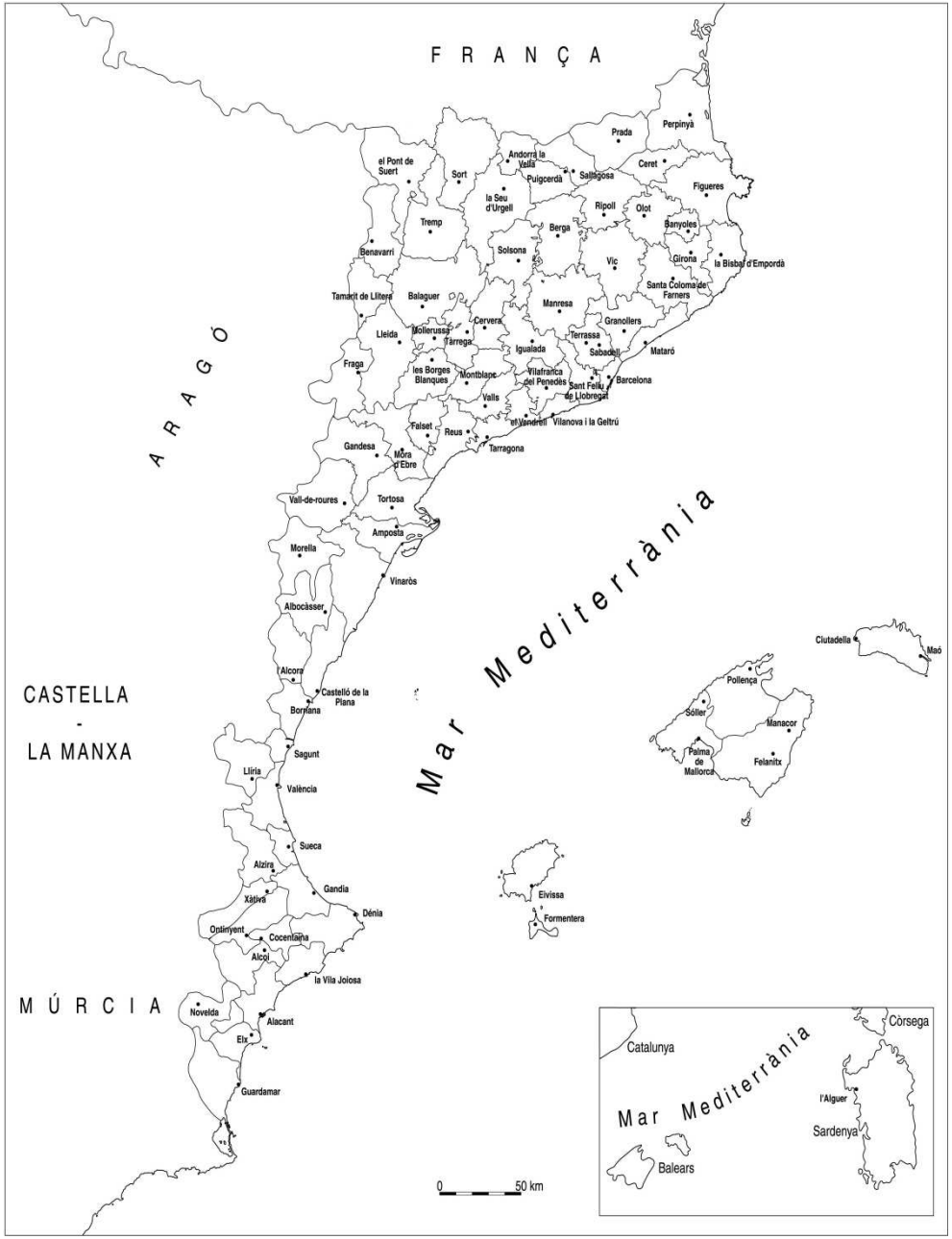
Clítics

INFOR	DFM	PLUR2		VO_FE1	
LOC	La Bisbal	V_FIPOST2		PL1	
ID_PF	0001	VOC_VERB		VO_FST1	
MOTS_ANT	Cada dia	CONS_VERB	réntu	VO_I2	
MOTS_ANT	káðə ðí	VERB_VOC		P_2	
		VERB_CONS		VO_FE2	
V_I1	ə	VERB_W		PL2	
PRO1	m	VERB_R		VO_ST2	
V_FIPRE1		VERB_NT		MOTS_POST	la cara
PLUR1		VO_I1		MOTS_POST	le káre
V_FIPOST1		P_1			
V_I2				SENCER	am
PRO2				Observacions	
V_FIPRE2					

Registro: 1 de 24376

Vista Formulario NUM

Inicio | Espai Person... | Elisenda C | Microsoft Wo... | BASES DE D... | Microsoft Ac... | Microsoft ... | 8:37



COD

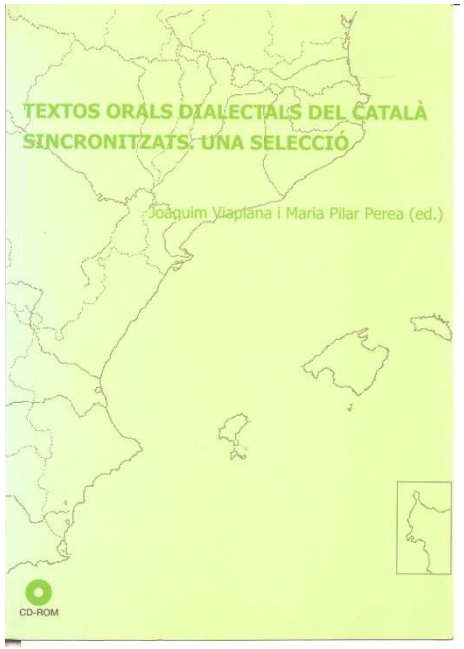
82 localidades
303 entrevistas

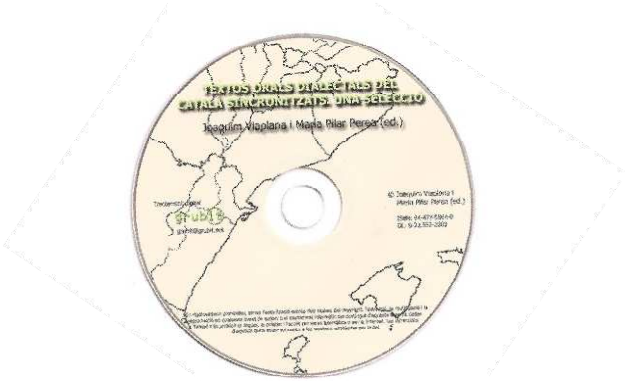
Bases de datos	Núm. registros
Verbos	68262
Fonética	55270
Clíticos pronominaels	24766
Artículos	5049
Posesivos	5223
Pronombres personales	1608
Demostrativos	1512
Locativos	803
TOTAL	**162.493**

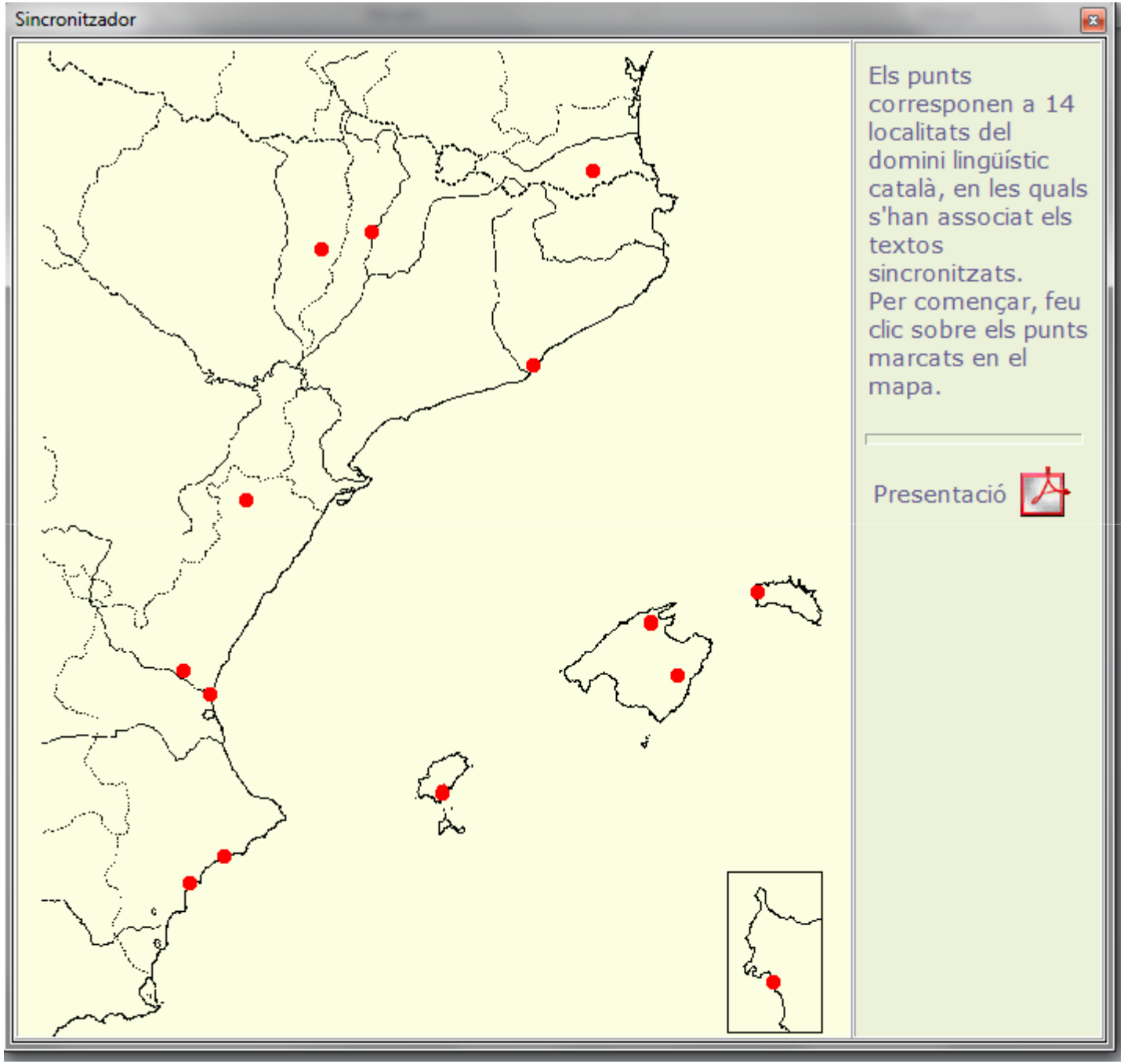
DATOS TEXTOS LIBRES:

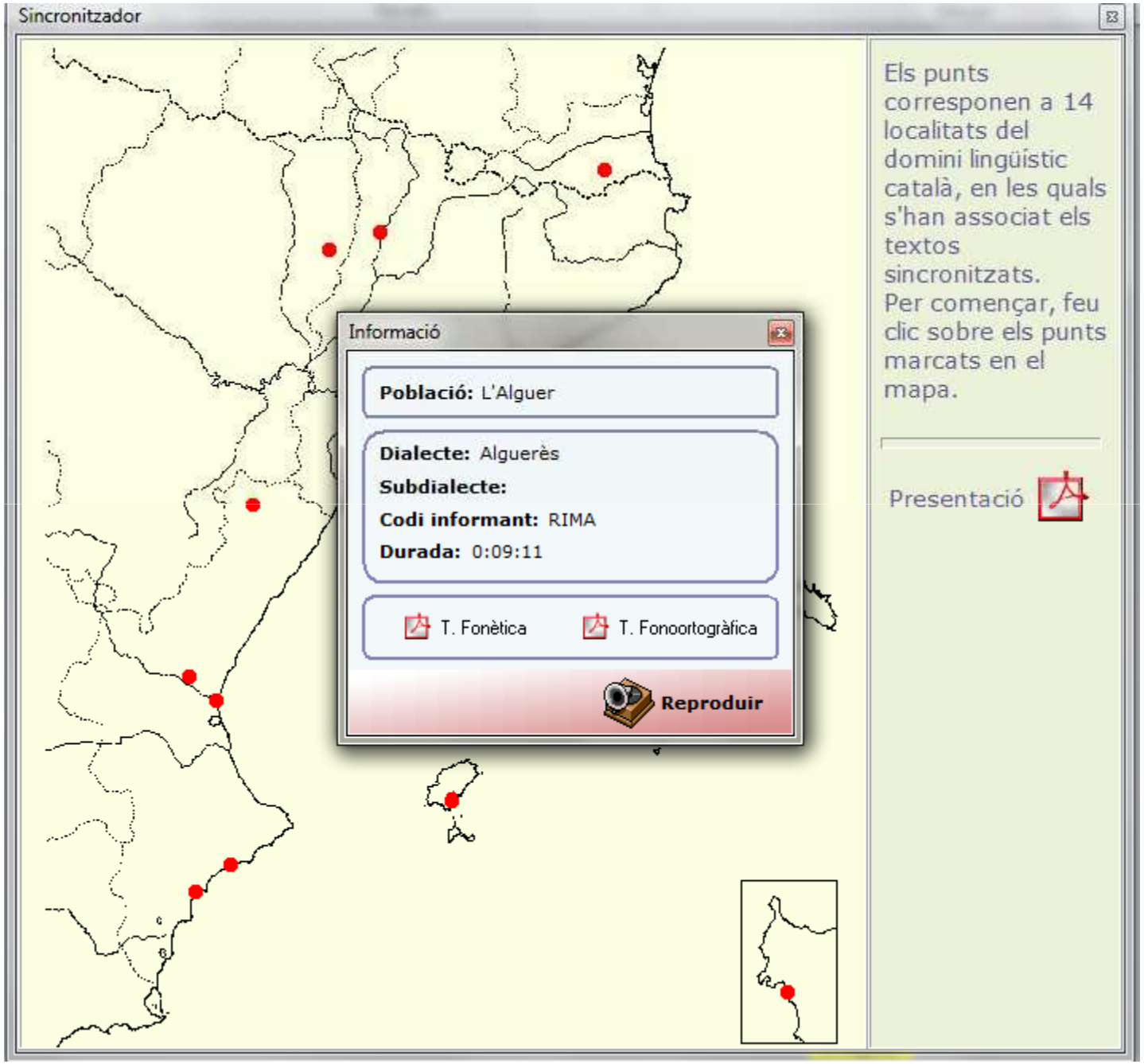
- En principio, para comprobar resultados encuestas.
- Aplicación docente: Transcripción ortográfica y fonético-ortográfica de **13 textos de distintos dialectos** publicados en versión sincronizada, en papel y en CD-ROM:

Viaplana,J.; Perea, M. P. (eds.) (2003): *Textos orals dialectals del català sincronitzats. Una selecció*. Barcelona: PPU.






Sincronitzador
Els punts corresponen a 14 localitats del domini lingüístic català, en les quals s'han associat els textos sincronitzats.
Per començar, feu clic sobre els punts marcats en el mapa.
Presentació


Sincronitzador
Els punts corresponen a 14 localitats del domini lingüístic català, en les quals s'han associat els textos sincronitzats.
Per començar, feu clic sobre els punts marcats en el mapa.
Presentació
Informació
Població: L'Alguer
Dialecte: Alguerès
Subdialecte:
Codi informant: RIMA
Durada: 0:09:11
T. Fonètica
T. Fonoortogràfica
Reproduir

Transcripció fonètica: *L'Alguer*

1 'ez um be'ʎisiʎ 'ʎo̯k (.) e: 'tens de 'na a 'vera 'tota las kus'teraz
2 ðe ɾ al'ɣe (.) de 'na (.) a de 'ka de 'kasa (.) a l aspa'ransa a'po̯lina
3 'sez 'ma e'naɾ a 'po̯lina (.) 'tens i de 'vera: la: lu la lukali'tat i 'k ez
4 a'po̯lina ðe 'ro̯kas (.) 'ez e: (.) uʎ 'ʎo̯k a'saj pejza'd͡ʒistik (.) i 'tu
5 'tens i ðe 'nal 'vewra (.)

6 E: I tant.

7 e 'j a wna jɣre'zjeta 'mẽntre 'tu 'vaz e ke s a'kira l aspa'ransa (.)
8 'aṇti a lu 'vujt i ðe se'tembra sa 'fa la: (.) 'j ez e: la preɣaru'ria
9 ðe 'tu ɾus kampa'ɲoruz (.) 'j ez la 'misa (.) i: a la 'siz de 'taḷda la
10 'diw tam'be wŋ̍ kapa'ʎa (.) la 'diw aŋ̍ kata'ra (.)

11 E: En català.

12 ner 'se (.) e: (.) 'i e'z una 'koza (.) 'ez un ra'kɔlt as'torio de 'tota

Transcripció fonoortogràfica: *L'Alguer*

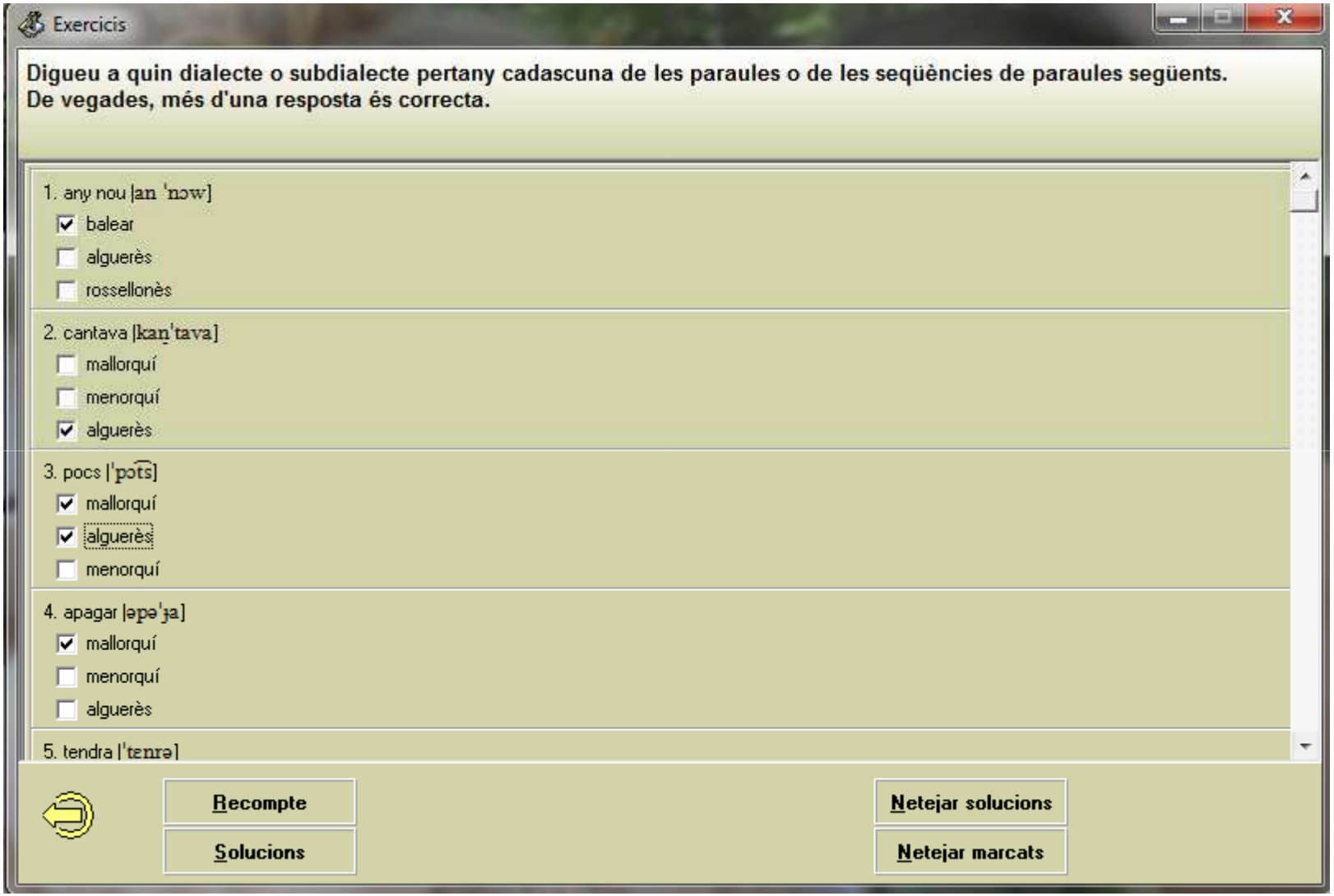
1 És un bellissim lloc. E... tens d'anar a vere totes les costeres
2 de l'Alguer, d'anar a... de ca- de Caça a l'Esperança, Pòllina.
3 Ets mai anada a Pòllina? Tens de vera la lo- la localitat que és
4 a Pòllina de Roques, és e... un lloc assai paisatgistic i tu
5 tens d'anà-l veure.

6 E: I tant.

7 E... hi ha una iglesieta mentre tu vas, e... que s'aquidra l'Esperança.
8 *Anti* a lo vuit de setembre se fa la..., hi és... la pregadoria
9 de tots los campagnolos, hi és la missa, i... a les sis de tarda la
10 diu també un capellà, la diu en català.

11 E: En català.

12 Per se-, e... i és una cosa, és un record històric de totes
13 les, e... los costarells i se fanan a la vigilia de les festes
14 se mengen alli, la primera só io que previnc cada any


Exercicis
Digueu a quin dialecte o subdialecte pertany cadascuna de les paraules o de les seqüències de paraules següents.
De vegades, més d'una resposta és correcta.
1. any nou [an 'nɔw]
balear
alguerès
rossellonès
2. cantava [kan̪'tava]
mallorquí
menorquí
alguerès
3. pocs ['pɔt͡s]
mallorquí
alguerès
menorquí
4. apagar [əpə'ɟa]
mallorquí
menorquí
alguerès
5. tendra ['tɛnrə]
Recompte
Solucions
Netejar solucions
Netejar marcats

Exercicis

Digueu a quin dialecte o subdialecte pertany cadascuna de les paraules o de les seqüències de paraules següents.
De vegades, més d'una resposta és correcta.

1. any nou [an 'nɔw]
   - [x] balear
   - [ ] alguerès
   - [ ] rossellonès

   ✗ **alguerès**

2. cantava [kan̪'tava]
   - [ ] mallorquí
   - [ ] menorquí
   - [x] alguerès

   ✓

3. pocs ['pɔ̃ts]
   - [x] mallorquí
   - [x] alguerès
   - [ ] menorquí

   ✓

4. apagar [əpə'ʒa]
   - [x] mallorquí
   - [ ] menorquí
   - [ ] alguerès

   ✓

5. tendra ['tɛnrə]

Recompte | **Solucions** | **Netejar solucions** | **Netejar marcats**

Encerts: 3
Errades: 1
En blanc: 90

EXPLOTACIÓN DEL COD

(HUM2004, HUM2007)

1. **Difusión** de los datos del COD (acceso on-line)
   a) Datos del cuestionario
   b) Textos orales

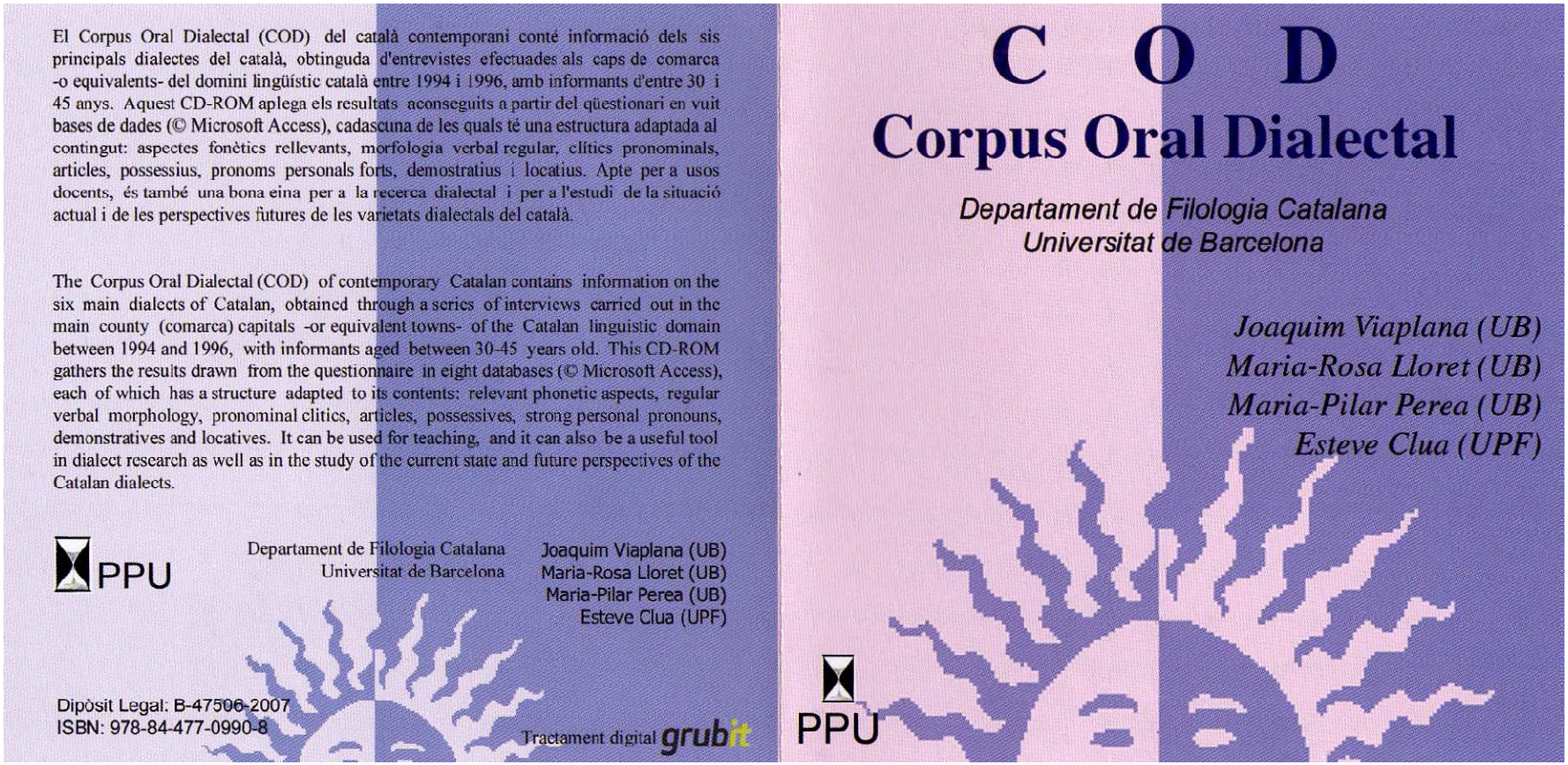
2. **Análisis** de los datos:
   a) Análisis de **fenómenos lingüísticos** en relación con la variación dialectal
   b) Análisis **dialectométrico** y representación de la distancia lingüística

1. Difusión de los datos del COD

Coord.: Viaplana, con Lloret, Perea y Clua

a) Bases de datos (cuestionario)

1ª fase: Publicación de las bases de datos en CD-Rom (2007)



El Corpus Oral Dialectal (COD) del català contemporani conté informació dels sis principals dialectes del català, obtinguda d'entrevistes efectuades als caps de comarca -o equivalents- del domini lingüístic català entre 1994 i 1996, amb informants d'entre 30 i 45 anys. Aquest CD-ROM aplega els resultats aconseguits a partir del qüestionari en vuit bases de dades (© Microsoft Access), cadascuna de les quals té una estructura adaptada al contingut: aspectes fonètics rellevants, morfologia verbal regular, clítics pronominals, articles, possessius, pronoms personals forts, demostratius i locatius. Apte per a usos docents, és també una bona eina per a la recerca dialectal i per a l'estudi de la situació actual i de les perspectives futures de les varietats dialectals del català.

The Corpus Oral Dialectal (COD) of contemporary Catalan contains information on the six main dialects of Catalan, obtained through a series of interviews carried out in the main county (comarca) capitals -or equivalent towns- of the Catalan linguistic domain between 1994 and 1996, with informants aged between 30-45 years old. This CD-ROM gathers the results drawn from the questionnaire in eight databases (© Microsoft Access), each of which has a structure adapted to its contents: relevant phonetic aspects, regular verbal morphology, pronominal clitics, articles, possessives, strong personal pronouns, demonstratives and locatives. It can be used for teaching, and it can also be a useful tool in dialect research as well as in the study of the current state and future perspectives of the Catalan dialects.

PPU

Departament de Filologia Catalana
Universitat de Barcelona

Joaquim Viaplana (UB)
Maria-Rosa Lloret (UB)
Maria-Pilar Perea (UB)
Esteve Clua (UPF)

Dipòsit Legal: B-47506-2007
ISBN: 978-84-477-0990-8

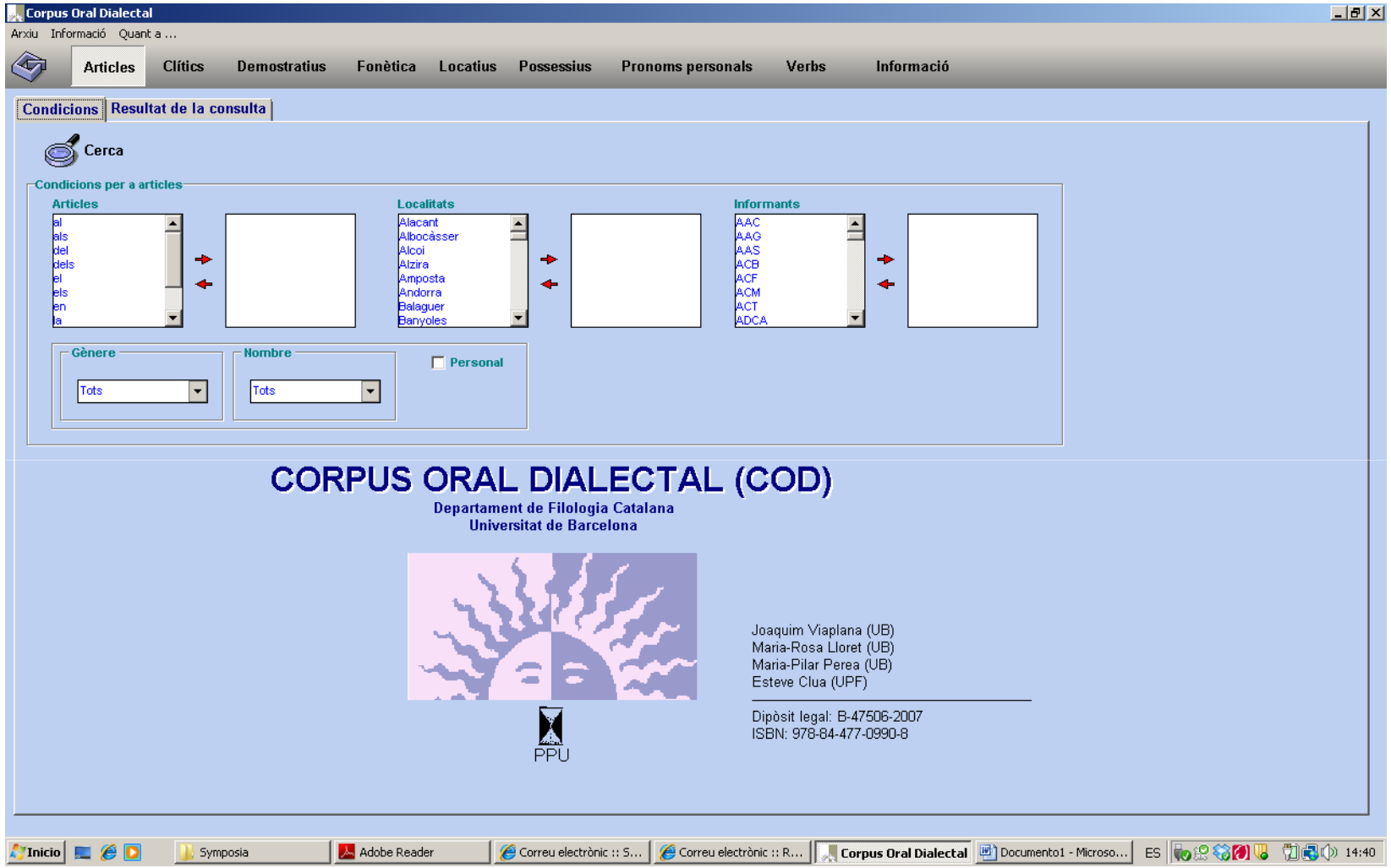
Tractament digital grubit

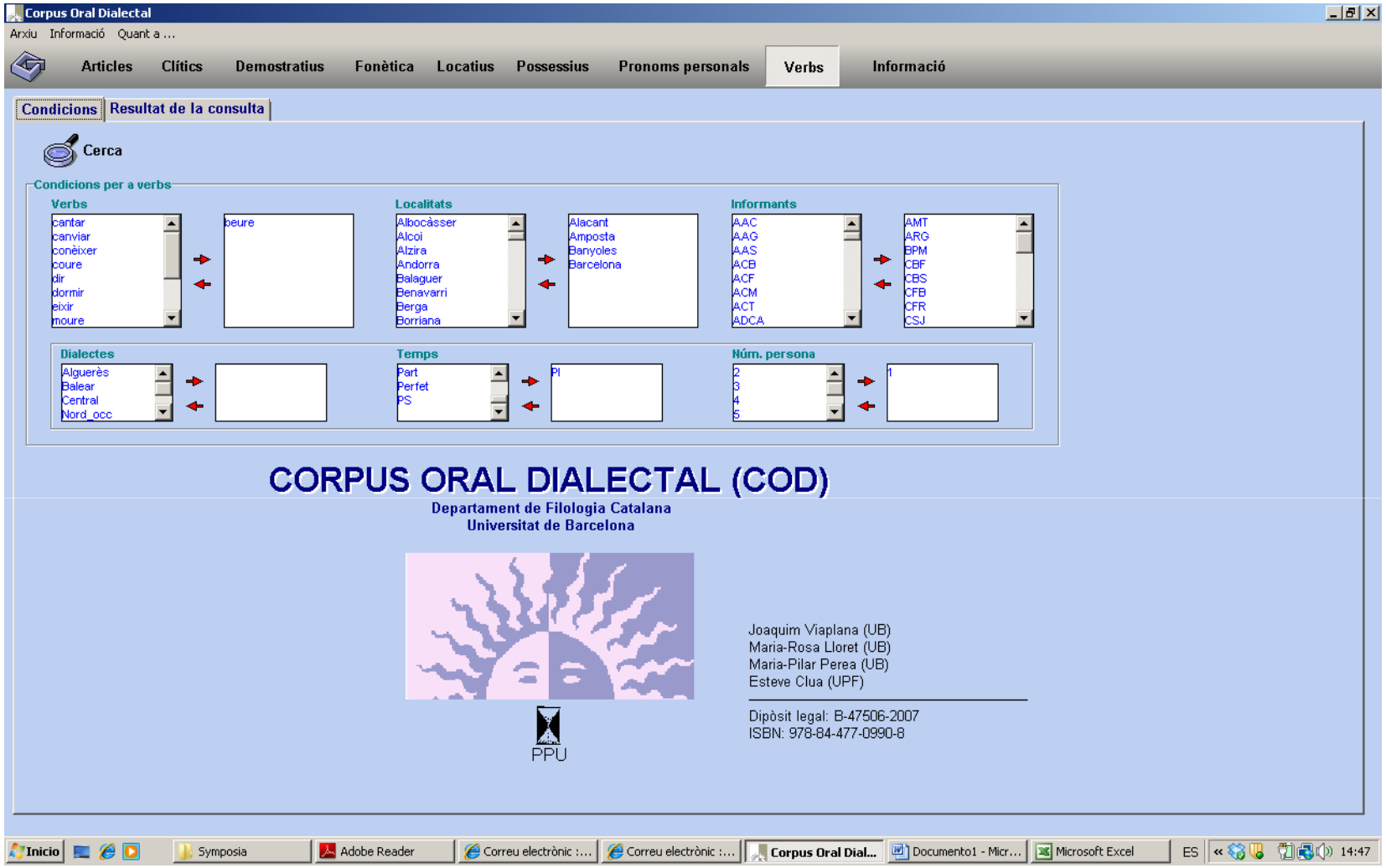
C O D
Corpus Oral Dialectal

Departament de Filologia Catalana
Universitat de Barcelona

Joaquim Viaplana (UB)
Maria-Rosa Lloret (UB)
Maria-Pilar Perea (UB)
Esteve Clua (UPF)

PPU


Corpus Oral Dialectal
Arxiu Informació Quant a ...
Articles
Clítics
Demostratius
Fonètica
Locatius
Possessius
Pronoms personals
Verbs
Informació
Condicions
Resultat de la consulta
Cerca
Condicions per a articles
Articles
al
als
del
dels
el
els
en
la
Localitats
Alacant
Albocàsser
Alcoi
Alzira
Amposta
Andorra
Balaguer
Banyoles
Informants
AAC
AAG
AAS
ACB
ACF
ACM
ACT
ADCA
Gènere
Tots
Nombre
Tots
Personal
CORPUS ORAL DIALECTAL (COD)
Departament de Filologia Catalana
Universitat de Barcelona
PPU
Joaquim Viaplana (UB)
Maria-Rosa Lloret (UB)
Maria-Pilar Perea (UB)
Esteve Clua (UPF)
Dipòsit legal: B-47506-2007
ISBN: 978-84-477-0990-8
Inicio
Symposia
Adobe Reader
Correu electrònic :: S...
Correu electrònic :: R...
Corpus Oral Dialectal
Documento1 - Microso...
ES
14:40


Corpus Oral Dialectal
Arxiu Informació Quant a ...
Articles
Clítics
Demostratius
Fonètica
Locatius
Possessius
Pronoms personals
Verbs
Informació
Condicions
Resultat de la consulta
Cerca
Condicions per a verbs
Verbs
cantar
canviar
conèixer
coure
dir
dormir
eixir
moure
beure
Localitats
Albocàsser
Alcoi
Alzira
Andorra
Balaguer
Benavarri
Berga
Borriana
Alacant
Amposta
Banyoles
Barcelona
Informants
AAC
AAG
AAS
ACB
ACF
ACM
ACT
ADCA
AMT
ARG
BPM
CBF
CBS
CFB
CFR
CSJ
Dialectes
Alguerès
Balear
Central
Nord_occ
Temps
Part
Perfet
PS
PI
Núm. persona
2
3
4
5
1
CORPUS ORAL DIALECTAL (COD)
Departament de Filologia Catalana
Universitat de Barcelona
PPU
Joaquim Viaplana (UB)
Maria-Rosa Lloret (UB)
Maria-Pilar Perea (UB)
Esteve Clua (UPF)
Dipòsit legal: B-47506-2007
ISBN: 978-84-477-0990-8
Inicio
Symposia
Adobe Reader
Correu electrònic :...
Correu electrònic :...
Corpus Oral Dial...
Documento1 - Micr...
Microsoft Excel
ES
14:47

Articles | Clítics | Demostratius | Fonètica | Locatius | Possessius | Pronoms personals | Verbs | Informació

Condicions | Resultat de la consulta

Exporta | Imprimeix

Que la consulta inclogui les paraules següents:
beure
Que la consulta inclogui les localitats següents:
Alacant
Amposta
Banyoles
Barcelona
Que la consulta inclogui els informants següents:

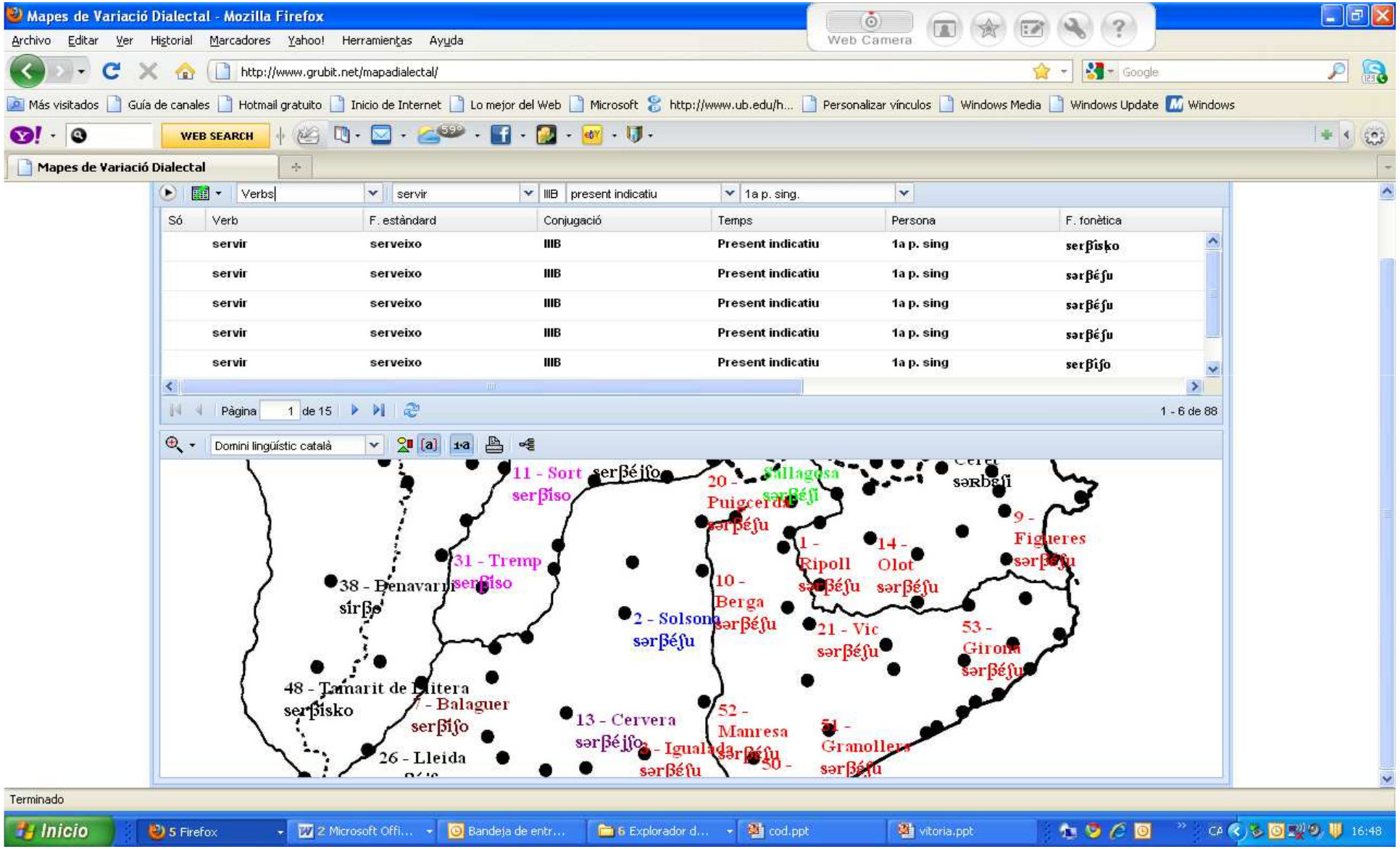
Localitat	Inf.	Conj.	Dialecte	Ort. 1	Ort. 2	Temps	Persona	Fon.	Arrel	Ep.	Ext.
Alacant	BPM	2b	Valencià	bec	beure	PI	1	bék	bé		k
Alacant	CBS	2b	Valencià	bec	beure	PI	1	bék	bé		k
Alacant	FJCP	2b	Valencià	bec	beure	PI	1	bék	bé		k
Amposta	JPM	2b	Nord_occ	bec	beure	PI	1	bék	bé		k
Amposta	CFR	2b	Nord_occ	bec	beure	PI	1	bék	bé		k
Amposta	ARG	2b	Nord_occ	bec	beure	PI	1	bék	bé		k
Banyoles	JST	2b	Central	bec	beure	PI	1	bék	bé		k
Banyoles	PGG	2b	Central	bec	beure	PI	1	bék	bé		k
Banyoles	CBF	2b	Central	bec	beure	PI	1	bék	bé		k
Barcelona	CSJ	2b	Central	bec	beure	PI	1	bék	bé		k
Barcelona	CVB	2b	Central	bec	beure	PI	1	bék	bé		k
Barcelona	DAC	2b	Central	bec	beure	PI	1	bék	bé		k
Barcelona	ECC	2b	Central	bec	beure	PI	1	bék	bé		k
Barcelona	FFS	2b	Central	bec	beure	PI	1	bék	bé		k
Barcelona	MAHC	2b	Central	bec	beure	PI	1	bék	bé		k
Barcelona	MIBF	2b	Central	bec	beure	PI	1	bék	bé		k
Barcelona	MFI	2b	Central	bec	beure	PI	1	bék	bé		k

pàgines del llistat

Verbs

Localitat	Inf.	Dialecte	Conj.	Temps	Verb	Pers.	Fonètica	Arrel	Ep	Ext	Tema	Mode	I
Barcelona	CSJ	Central	2b	PI	beure	1	bék	bé		k			
Barcelona	CVB	Central	2b	PI	beure	1	bék	bé		k			
Banyoles	PGG	Central	2b	PI	beure	1	bék	bé		k			
Barcelona	ECC	Central	2b	PI	beure	1	bék	bé		k			
Barcelona	FFS	Central	2b	PI	beure	1	bék	bé		k			
Barcelona	MAHC	Central	2b	PI	beure	1	bék	bé		k			
Barcelona	MIBF	Central	2b	PI	beure	1	bék	bé		k			
Barcelona	MFI	Central	2b	PI	beure	1	bék	bé		k			
Barcelona	PAC	Central	2b	PI	beure	1	bék	bé		k			
Barcelona	MPT	Central	2b	PI	beure	1	bék	bé		k			
Barcelona	TEG	Central	2b	PI	beure	1	bék	bé		k			
Barcelona	GPC	Central	2b	PI	beure	1	bék	bé		k			
Barcelona	CFB	Central	2b	PI	beure	1	bék	bé		k			
Barcelona	JBT	Central	2b	PI	beure	1	bék	bé		k			
Barcelona	JFM	Central	2b	PI	beure	1	bék	bé		k			
Barcelona	MAA	Central	2b	PI	beure	1	bék	bé		k			
Barcelona	MABA	Central	2b	PI	beure	1	bék	bé		k			
Barcelona	MOBA	Central	2b	PI	beure	1	bék	bé		k			
Barcelona	MBM	Central	2b	PI	beure	1	bék	bé		k			
Banyoles	CBF	Central	2b	PI	beure	1	bék	bé		k			
Amposta	JPM	Nord_occ	2b	PI	beure	1	bék	bé		k			
Amposta	CFR	Nord_occ	2b	PI	beure	1	bék	bé		k			
Alacant	CBS	Valencià	2b	PI	beure	1	bék	bé		k			
Alacant	FJCP	Valencià	2b	PI	beure	1	bék	bé		k			
Barcelona	AMT	Central	2b	PI	beure	1	bék	bé		k			
Barcelona	DAC	Central	2b	PI	beure	1	bék	bé		k			
Banyoles	JST	Central	2b	PI	beure	1	bék	bé		k			
Amposta	ARG	Nord_occ	2b	PI	beure	1	bék	bé		k			
Alacant	BPM	Valencià	2b	PI	beure	1	bék	bé		k			

2ª fase: Mapas interactivos on-line (coord. Perea, 2010)



- Versión on-line de las bases de datos del COD publicadas anteriormente en CD-Rom.
- Mapas automáticos (a la carta) y asociación con la secuencia sonora.

Tareas:

1. Transcripciones:
   a) ortográfica
   b) fonética

2. Corte de la secuencia sonora.

3. Asociación de la secuencia sonora con las transcripciones correspondientes.

www.ub.edu/mapadialectal

o a través p. web del proyecto COD:

www.ub.edu/lincat

C O D

Corpus Oral Dialectal

Mapes de variació dialectal

Maria Pilar Perea (coordinació)

Departament de Filologia Catalana

Universitat de Barcelona

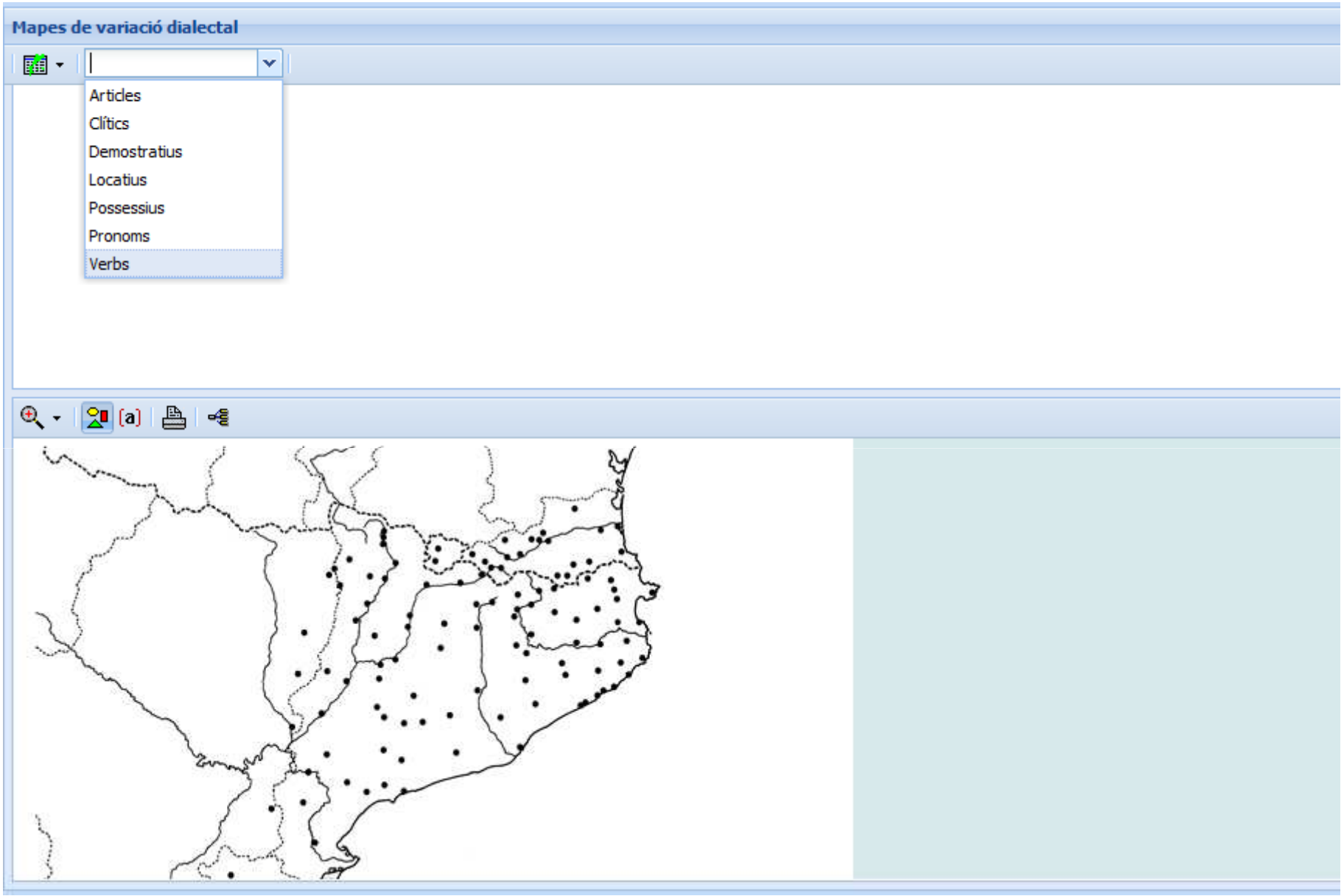
Desenvolupat per grubit

Català Anglès

El Corpus Oral Dialectal (COD) del català contemporani conté informació dels sis principals dialectes del català, obtinguda d'entrevistes efectuades als caps de comarca -o equivalents- del domini lingüístic català entre 1994 i 1996, amb informants d'entre 30 i 45 anys. La web representa els resultats aconseguits a partir del qüestionari en vuit bases de dades, cadascuna de les quals té una estructura adaptada al contingut: aspectes fonètics rellevants, morfologia verbal regular, clítics pronominals, articles, possessius, pronoms personals forts, demostratius i locatius.

Entra a l'aplicació

Finançat pels projectes "Explotació d'un corpus oral dialectal" [HUM2004-01504/FILO] i "Explotació d'un corpus oral dialectal: anàlisi de la variació lingüística i desenvolupament d'aplicacions informàtiques per a la transcripció automatitzada (ECOD)" [HUM2007 65531/FILO].


Mapes de variació dialectal
Articles
Clítics
Demostratius
Locatius
Possessius
Pronoms
Verbs

Mapes de variació dialectal

Verbs | beure | IIB | present indicatiu | 1a p. sing.

So	Verb	F. estàndard	Conjugació	Temps	Persona	F. fonètica
	beure	bec	IIB	Present indicatiu	1a p. sing	béʧ
	beure	bec	IIB	Present indicatiu	1a p. sing	béʧ
	beure	bec	IIB	Present indicatiu	1a p. sing	béʧ
	beure	bec	IIB	Present indicatiu	1a p. sing	béʧ
	beure	bec	IIB	Present indicatiu	1a p. sing	béʧ

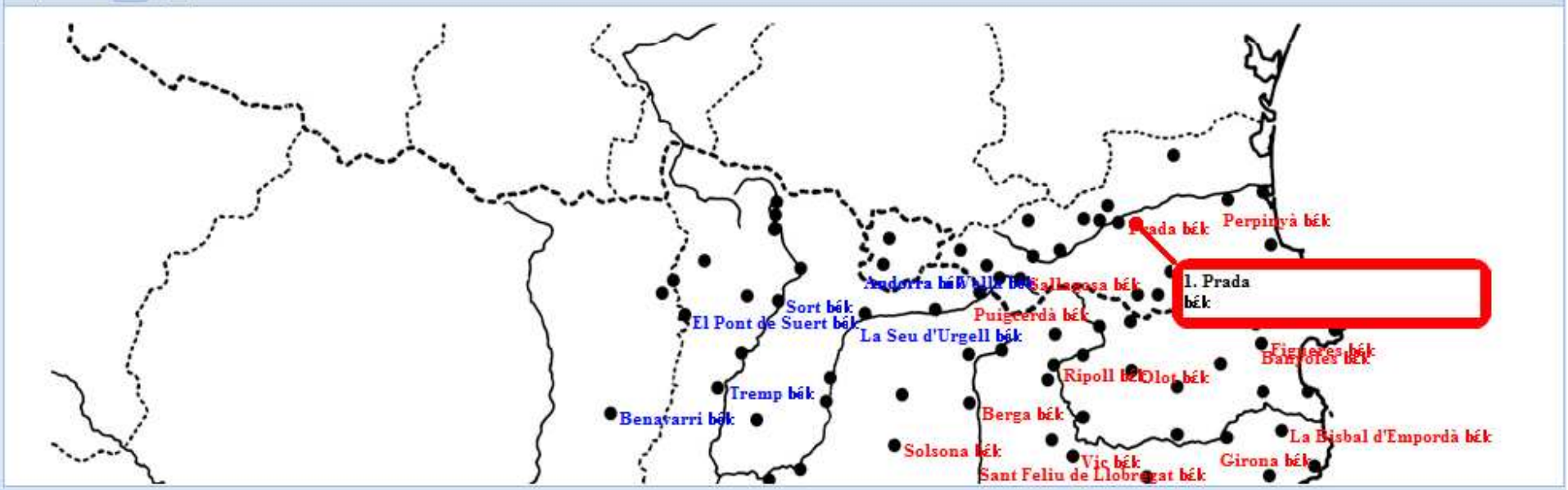
Pàgina 1 de 15 — 1 - 6 de 89

Mapes de variació dialectal

Verbs | beure | IIB | present indicatiu | 1a p. sing.

So	Verb	F. estàndard	Conjugació	Temps	Persona	F. fonètica
	beure	bec	IIB	Present indicatiu	1a p. sing	bék
	beure	bec	IIB	Present indicatiu	1a p. sing	bék
	beure	bec	IIB	Present indicatiu	1a p. sing	bék
	beure	bec	IIB	Present indicatiu	1a p. sing	bék
	beure	bec	IIB	Present indicatiu	1a p. sing	bék

Pàgina 1 de 15 — 1 - 6 de 89



b) Textos orales espontáneos

Accesibilidad on-line de diferentes colecciones de textos: Transcritos en versión **fonética** y **fono-ortográfica**, y **archivo sonoro** correspondiente (MP3):

Dipòsit digital de la Universitat de Barcelona (http://diposit.ub.edu/dspace/handle/2445/10413), y también desde la p. web del proyecto (http://www.ub.edu/lincat) y la del "Corpus de Català Contemporani de la Universitat de Barcelona (CCUB)" (http://www.ub.edu/cccub/):

- **2008-09: 13 textos de distintos dialectos** publicados como Viaplana & Perea (eds.) (2003) en papel y CD-Rom.
- **2009: 12 textos del balear** (Pons & Viaplana, eds.).
- **2010: 14 textos de distintos dialectos** (Perea & Viaplana, eds.).
- **2010: 8 textos del noroccidental** (Carrera & Viaplana, eds.).

diposit.ub.edu/dspace/handle/2445/10413

o a través p. web del “Corpus de Català Contemporani de la UB”:

www.ub.edu/cccub

o a través p. web del proyecto COD:

www.ub.edu/lincat

Corpus de Català Contemporani de la Universitat de Barcelona (CCCUB)

El CCCUB és un arxiu de corpus de llengua catalana oral contemporània que ha estat confegit per a l'estudi de la variació dialectal, social i funcional en la llengua catalana.

Aquesta pàgina posa a disposició dels investigadors la transcripció i una part dels arxius sonors dels textos del CCCUB, que han estat dipositats al Dipòsit Digital de la Universitat de Barcelona.

Alguns documents ja havien estat publicats amb anterioritat en llibre i CD-Rom. En aquests casos, els CD-Rom corresponents contenen materials addicionals que no han estat dipositats al Dipòsit Digital de la UB: bases de dades, arxius sonors i la descripció detallada dels materials.

La publicació a internet dels materials del CCCUB és un procés obert: alguns dels documents publicats amb anterioritat en CD-Rom no han completat encara el procés de publicació a la xarxa; a més, es preveu afegir nous materials inèdits al CCCUB. Consulteu les novetats.

NOVETATS (31.01.2011):

- Textos orals del nord-occidental (J. Carrera i J. Viaplana): Móra d'Ebre, Sort, Tamarit de Llitera, Tàrrega.

- Corpus Oral Dialectal (COD)
- Corpus de Varietats Socials (COS)
- Corpus Oral de Conversa Col·loquial (COC)
- Corpus Oral de Registres (COR)

Corpus Oral Dialectal (COD)

Pàgina principal del CCCUB

Viaplana, Joaquim i Maria Pilar Perea (ed.) (2003): *Textos orals dialectals del català sincronitzats. Una selecció*. Barcelona: Promociones y Publicaciones Universitarias (PPU). [Llibre i CD-ROM]
ISBN: 84-477-0864-0

Viaplana, Joaquim; Maria-Rosa Lloret, Maria-Pilar Perea i Esteve Clua (2007): COD. Corpus Oral Dialectal. Barcelona: Promociones y Publicaciones Universitarias (PPU). [CD-ROM]
ISBN: 978-84-477-0990-8

Pons, Clàudia i Joaquim Viaplana (ed.) (2009): *Corpus oral dialectal (COD). Textos orals del balear .* Dipòsit Digital de la UB.

Perea, Maria-Pilar i Joaquim Viaplana. *Corpus Oral Dialectal (COD). Selecció de textos.* Dipòsit Digital de la UB [Arxiu obert]

Carrera-Sabaté, Josefina i Joaquim Viaplana (ed.): *Corpus Oral Dialectal (COD). Textos orals del nord-occidental.* Dipòsit Digital de la UB.

Copyright: Grup d'Estudi de la Variació, Departament de Filologia Catalana, Universitat de Barcelona, 2008

COD, Textos orals del balear (2009)

Pàgina principal del CCCUB
Pàgina principal del COD

- Ciutadella
- Eivissa 1
- Eivissa 2
- Felanitx
- Formentera 1
- Formentera 2
- Manacor
- Maó
- Palma
- Pollença 1
- Pollença 2
- Sóller

Pons, Clàudia i Joaquim Viaplana (ed.) (2009): *Corpus oral dialectal (COD). Textos orals del balear.* Dipòsit Digital de la UB.

Copyright: Grup d'Estudi de la Variació, Departament de Filologia Catalana, Universitat de Barcelona, 2008

DIPÒSIT DIGITAL DE LA UB

UNIVERSITAT DE BARCELONA

Buscar en el Repositorio digital

Ir

Búsqueda avanzada

- Página de inicio
- Qué es?
- Cómo publicar en el Repositorio
- Política de Acceso Abierto UB
- Estadísticas

Listar

- Comunidades y colecciones
- Fecha Publicación
- Autor
- Título
- Materia

Servicios

- Alertas
- Mi repositorio usuarios autorizados
- Editar perfil

- Ayuda
- Aviso legal

Dipòsit Digital de la UB > Recerca > Grup d'Estudi de la Variació (GEV) - Corpus de Català Contemporani de la Universitat de Barcelona (CCCUB) > Corpus Oral Dialectal (COD) >

Por favor, use este identificador para citar o enlazar este ítem: http://hdl.handle.net/24

Título : Corpus Oral Dialectal (COD). Ciutadella1

Autor : Pons, Clàudia
Viaplana, Joaquim, 1942-

Matèria: Linguistics
Spoken Catalan
Dialectology
Balearic
Menorquí language
Ciudadela (Balearic Islands)
Lingüística
Català parlat
Dialectologia
Balear
Menorquí
Ciutadella (Balears)

Fecha de publicación : 29-sep-2009

Resumen : Aquest document conté la transcripció fonètica, la fonoortogràfica i l'arxiu de so d'un fragment de conversa lliure amb un informant de C del Corpus de Català Contemporani de la Universitat de Barcelona (CCCUB), un arxiu de corpus de llengua catalana oral contemporània finalitat de contribuir a l'estudi de la variació dialectal, social i funcional en la llengua catalana. Una selecció de materials del CCCUB ha www.recercat.net), i també és accessible a través del web del CCCUB: http://www.ub.edu/cccub.

URI : http://hdl.handle.net/2445/11627

Aparece en las colecciones: Corpus de Català Contemporani de la Universitat de Barcelona (CCCUB)
Corpus Oral Dialectal (COD)

Ficheros en este ítem:

Fichero	Descripción	Tamaño	Formato	
COD_Ciutadella_Fon.pdf		51,88 kB	Adobe PDF	Visualizar/Abrir
COD_Ciutadella_For.pdf		39,62 kB	Adobe PDF	Visualizar/Abrir
COD_Ciutadella_MP3.mp3		9,52 MB	Unknown	Visualizar/Abrir

RefWorks

Mostrar el registro Dublin Core completo del ítem

Recomiende este ítem

Estadísticas de este documento

Resumen de estadísticas [Año/mes] [País] [Dominio]

" **Estadísticas del ítem** " Corpus Oral Dialectal (COD). Ciutadella1 "

Período [este mes] [este año] [último año] [todos los años]

tadísticas

	Visitas	Descargas
	120	80

País	Visitas	Descargas
unknown	50	27
Spain	19	22
United States	18	13
United Kingdom	0	6
Germany	5	5
China	0	3
France	1	2
Taiwan	1	1
Lithuania	0	1
Korea, Republic of	10	0
Italy	8	0
Ukraine	5	0
Czech Republic	1	0
Sweden	1	0
Ireland	1	0

Dominio	Visitas	Descargas
unknown	63	30
yandex.ru	19	12
ecs.soton.ac.uk	0	6
Red-83-43-254.dynamicIP.rima-tde.net	1	4
fil.ub.es	1	4
137.99.40.188.clients.your-server.de	0	4
Red-81-39-247.dynamicIP.rima-tde.net	1	3

2. Análisis de los datos del COD:

a) Análisis de fenómenos lingüísticos en relación con la variación dialectal

Coord.: M-R. Lloret

i) Estudio de mecanismos que permitan relativizar los efectos de las generalizaciones lingüísticas y análisis de la variación lingüística.

Ámbito fonético

- VOCALES: Descripción y estudio acústico de:
  a) La **armonía vocálica** en valenciano.
  b) Las **vocales medianas tónicas** en catalán central y noroccidental.

  Resultados más importantes: diferencias microdialectales en el timbre de las vocales; **consecuencias para la armonía vocálica**; aspectos **sociolingüísticos**.

- CONSONANTES: Estudio de la producción y percepción de las realizaciones **fricativas** y **africadas** de las palatales en catalán noroccidental.

  Resultados más importantes: aspectos **sociolingüísticos** que determinan la producción y la percepción.

(V. trabajos de Carrera, Herrero y Jiménez)

Ámbito fonológico

- Estudio comparativo entre distintos dialectos del catalán y fenómenos afines en otras lenguas románicas:

  a) VOCALES: procesos de **armonía vocálica**, **apertura** de vocales átonas y **epéntesis**.

  b) CONSONANTES: **asimilaciones** de sonoridad y entre consonantes.

  Resultados más importantes: efectos de **prominencia** (segmental y estructural) en los fenómenos fonológicos; importancia de la **similitud** articulatoria en las asimilaciones; adecuación de la **Teoría de la Optimidad (TO)** para explicar la variación existente.

- Descripción y análisis de fenómenos consonánticos del alguerés relativos al **rotacismo** y distribución de **líquidas**, y a las **africadas**.

  Resultados más importantes: **nuevos** datos sobre esta variedad; efectos de **prominencia** analizados en el marco de la **TO**.

(V. trabajos de Bosch, Cabrera, Jiménez, Lloret y Pons)

Ámbito morfológico (flexivo)

- Estudios sobre distintos dialectos del catalán:
  a) Naturaleza morfológica de las **vocales finales** de los elementos nominales y de los **segmentos verbales posradicales** velares y palatales.
  b) Predictibilidad en **alomorfía**.
  c) Sobreaplicación y bloqueo de fenómenos fonológicos por **efectos paradigmáticos**.

  Resultados más importantes: **analogías condicionadas** por cuestiones fonológicas y por efectos de uniformidad paradigmática; **migración morfológica** condicionada; adecuación de la **Teoría de la Optimidad** (**TO**) y de la **Morfología Natural** para explicar estos fenómenos de interfaz.

- Descripción de la variación existente en la **flexión verbal** del catalán septentrional.

  Resultados más importantes: **nuevos** datos sobre esta variedad; estudio de **cambios en curso**; adecuación de la **Morfología Natural** para explicar estos cambios.

(V. trabajos de Adam, Adrover, Lloret, Marí, Perea, Pons, Querol y Viaplana)

ii) Estudio de mecanismos que permitan diferenciar las características del estilo formal vs. informal y de las variedades geográficas vs. estándar.

- Diferencias entre el estilo **formal** y el **informal**.

  Resultados más importantes: efectos de **marcadez** (naturalidad) en el estilo informal; adecuación de la **Teoría de la Optimidad** para explicar la variación.

- Constatación de la progresiva **desdialectalización**:

  a) General en algunas áreas noroccidentales: estudio comparativo global de las características obtenidas a partir de los datos del COD y estudio de zonas de frontera específicas.

  b) Pérdida del artículo derivado de IPSU en algunos dialectos del catalán.

  c) Pérdida de características específicas del alguerés.

  Resultados más importantes: estudio de causas **sociolingüísticas** en los procesos de desdialectalización.

(V. trabajos de Bosch, Busquet, Cabrera, Campmany, Clua, Rosàs, Valls y Viaplana)

EJEMPLO 1: Estudio comparativo sobre vocalismo

Correlación entre prominencia estructural y segmental:
posiciones fuertes/débiles y elementos más/menos sonantes
(v. trabajos de Lloret &Jiménez)

Variedad	Característica que determina la prominencia	Efecto de la prominencia	Ejemplos
Gallego	Posición pretónica	Mantenimiento de la apertura	b**ɛ**lleza, **ɔ**sudo, **ɔ**siño, **ɔ**bx**ɛ**ctivo (en vez de [e], [o] átonas)
Catalán (Mallorca)	Sílaba átona inicial	No reducción a [ə] (mantenimiento de [e])	p**e**dreta, p**e**drota (en vez de [ə] átona)
Catalán, gallego, italiano	Sílaba tónica	Apertura de vocales en préstamos y cultismos	est**ɔ**p, **ɛ**tica, t**ɛ**cnica b**o**la vs. b**ɔ**la (‘esfera’) (gall.)
Catalán	Sílaba tónica inicial	Apertura de *o*	**ɔ**lla, h**ɔ**ra, fl**ɔ**r (en vez de [ó])
Español, catalán ...	Sílaba inicial	Preferencia por la vocal [a]	NOVACULA > n**a**vaja (esp.), n**a**valla (cat.)
Español, catalán ...	Sílaba inicial	Mantenimiento de hiatos	b**i.ó**logo vs. ra.d**ió**.logo (esp.)
Valenciano (Canals)	Posición inicial absoluta de palabra (sílaba átona)	Mantenimiento de la apertura	**ɔ**brim ‘abrimos’, **ɔ**mplim ‘llenamos’ vs. p**o**rtem ‘llevamos’
Catalán occidental	Posición inicial absoluta de palabra	Vocal epentética: [a]	[**a**]spina vs. pobr[**e**]
Valenciano (Pedreguer)	Posición proclítica	Vocal epentética: [a]	[**a**t] parle ‘te hablo’ vs. parlant-[t**e**] ‘hablándote’

En teoría de la optimidad, se interpretan como
efectos de **Marcadez Posicional** y **Fidelidad Posicional**.

EJEMPLO 2: Estudio comparativo sobre armonía vocálica

¿Articulatorio o perceptivo?

UNIFORMIDAD GESTUAL (base articulatoria):

- **Fidelidad Posicional:**
- El rasgo que se esparce aparece en una posición fuerte y se esparce hacia posiciones débiles **para minimizar el esfuerzo articulatorio**.
- Ej.: Tónico → Átono ˈɛ … e → ˈɛ … ɛ

LEGITIMACIÓN PERCEPTIVA (base funcional):

- **Marcadez Posicional:**
- El rasgo que se esparce aparece en una posición débil y se esparce hacia una posición más fuerte **para ser más perceptible**.
- Ej.: Átono → Tónico ˈɛ … e → ˈe … e

EJEMPLO 2: Estudio comparativo sobre armonía vocálica
(v. trabajos de Jiménez & Lloret)

	ARTICULATORIO	PERCEPTIVO	ELEMENTOS AFECTADOS
Español (andaluz)	0 general: ˈkɔ̱mikɔ̱ʰ variación: ˈtɾɛ̱βolɛ̱ʰ ~ ˈtɾɛ̱βɔ̱lɛ̱ʰ	1	a) Sílaba acentuada (ˈtɾɛ̱βolɛ̱ʰ) b) Sílaba acentuada y postónica interna (ˈtɾɛ̱βɔ̱lɛ̱ʰ) [+ beneficio articulatorio] c) Palabra (mɔ̱ˈmɛ̱ntɔ̱ʰ) [+ beneficio articulatorio]
Asturiano (Lena)	0 ˈpe̱ʃaɾu̱	1	Sílaba acentuada (ˈpe̱ʃaɾu̱ vs. ˈpaʃaɾa)
Asturiano (montañés Tudanca)	2	1	Sílaba acentuada y postónica interna (centralización) (oˈɾE̱gA̱nU̱)
Asturiano (montañés pasiego)	2	1	Palabra (centralización) (rI̱ˈfɾjA̱tI̱kU̱)
Catalán (valenciano)	1	2	Sílaba postónica adyacente (tovaˈʎɔ̱lɔ̱ vs. ˈmɛ̱ðika)
Catalán (sur de Alicante)	1	2	Palabra (tɔ̱vɔ̱ˈʎɔ̱lɔ̱)
Catalán (Tortosa)	1	2	Sílabas pretónicas adyacentes (ri̱ˈvi̱sta, i̱ni̱ˈmi̱k, di̱si̱ˈkat vs. ˈbi̱men, ˈdi̱ɣe)

1 = factor detonante de la armonía; 2 = beneficio secundario de la armonía; 0 = factor/beneficio no relevante

Algunos trabajos disponibles en:

www.uv.es/foncat/

Coord.: J. Jiménez

2. Análisis de los datos del COD

b) Análisis dialectométrico y representación de la distancia lingüística

Coord.: E. Clua

i) Completar y contrastar la clasificación de las variedades del catalán obtenida mediante la metodología del *Cluster Analysis* con la aplicación de otras técnicas clasificatorias

- Clasificación **dendrométrica** a partir de la información de las bases de datos del COD, con análisis **fonológico** previo

- **Comparación** de estos resultados con los obtenidos a partir de las formas **fonéticas** (en línea con el trabajo de las escuelas dialectométricas de **Salzburgo** y de **Groninga**)

(V. trabajos de Clua y Valls)

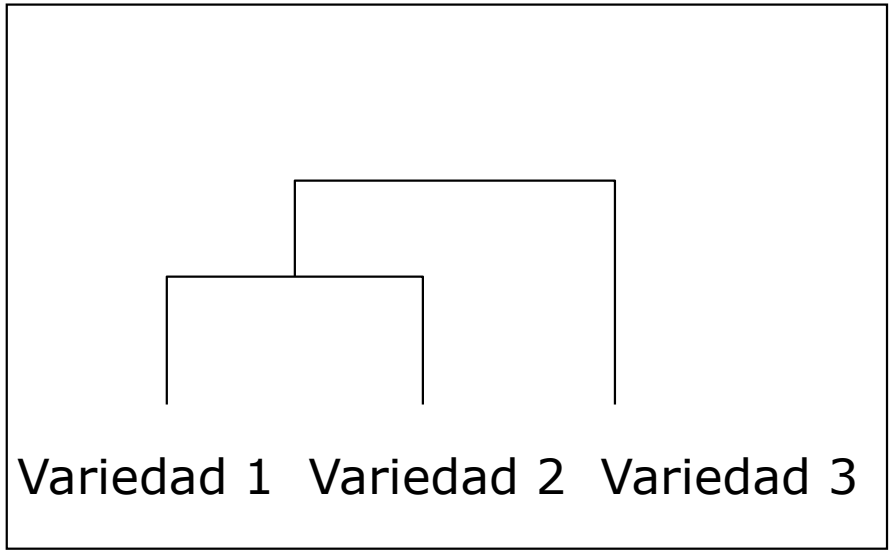
Ejemplo de pronombres clíticos del valenciano

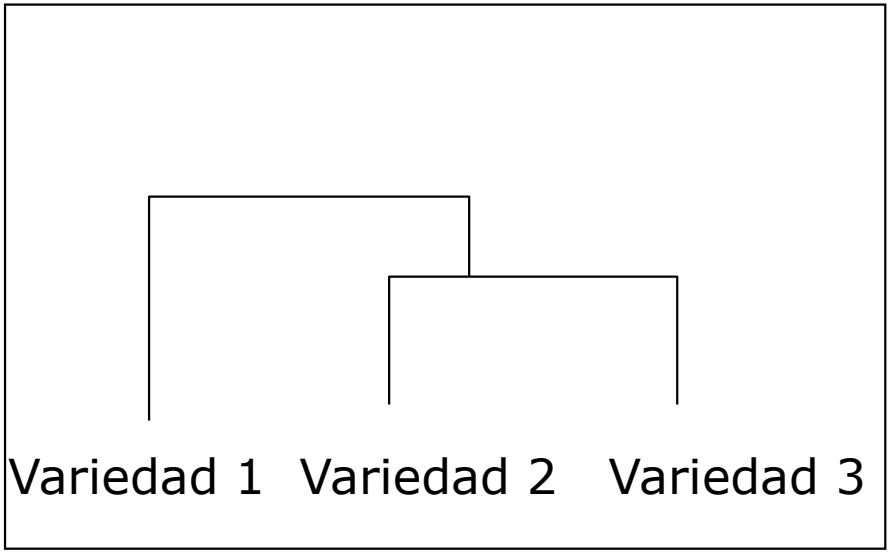
Ortografía		Variedad 1	Variedad 2	Variedad 3
et *parle*	‘te hablo’	**[et]** parle	**[te]** parle	**[te]** parle
t’*anime*	‘te animo’	**[t]** anime	**[t]** anime	**[t]** anime
*animant-**te***	‘animándote’	animant **[te]**	animant **[te]**	animant **[te]**
*anima’**t***	‘anímate’	anima **[t]**	anima **[t]**	anima **[te]**

Distancia lingüística basada en datos fonéticos:

	Variedad 1	Variedad 2	Variedad 3
Variedad 1			
Variedad 2	1		
Variedad 3	2	1	

Representación de la distancia lingüística basada en datos fonéticos



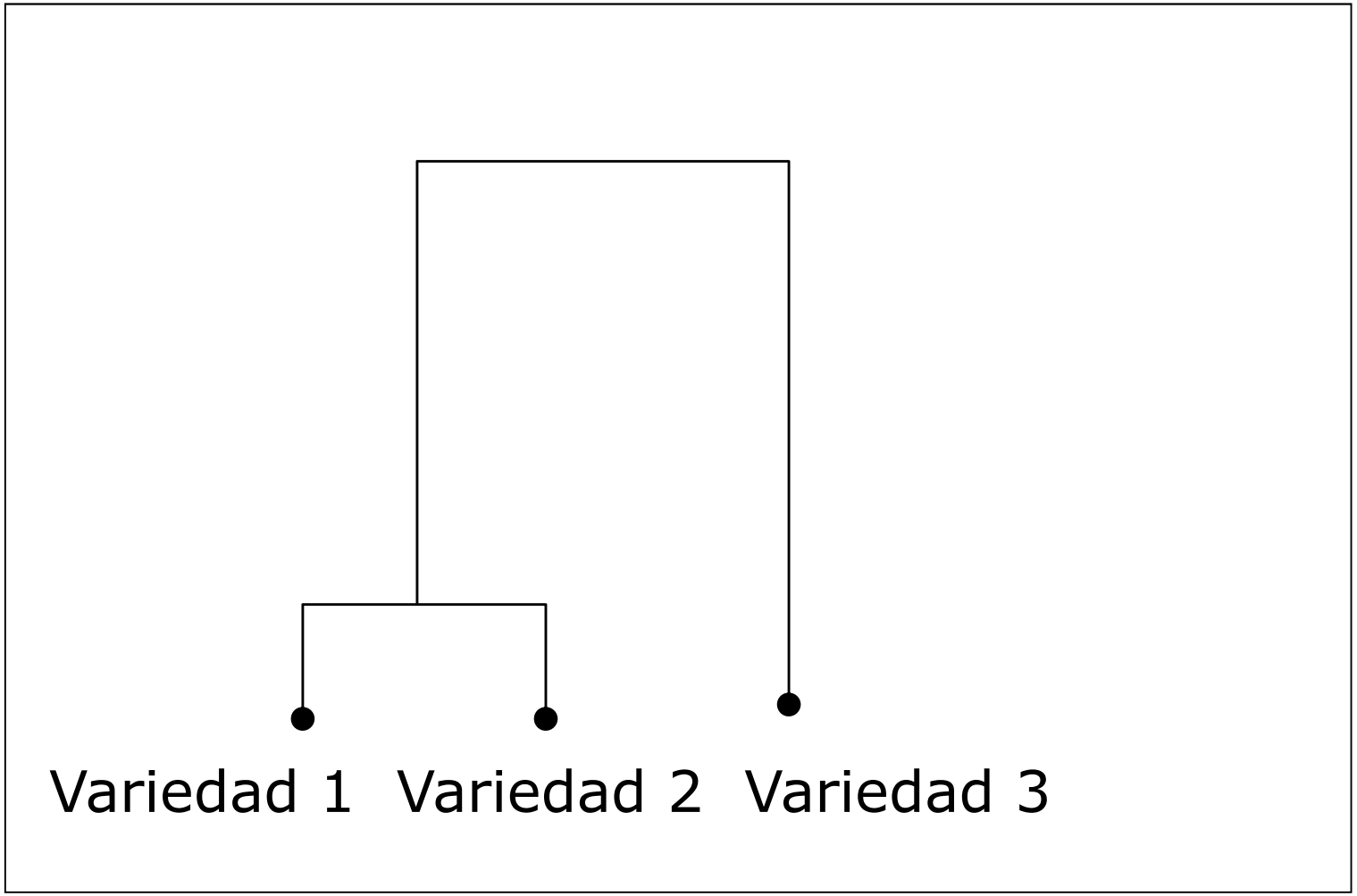


Ortografía		Variedad 1	Variedad 2	Variedad 3
et *parle*	‘te hablo’	**[et]** parle	**[te]** parle	**[te]** parle
t’*anime*	‘te animo’	**[t]** anime	**[t]** anime	**[t]** anime
*animant-**te***	‘animándote’	animant **[te]**	animant **[te]**	animant **[te]**
*anima’**t***	‘anímate’	anima **[t]**	anima **[t]**	anima **[te]**
Interpretación fonológica		**/t/** epéntesis periférica	**/t/** epéntesis a la derecha	**/te/** elisión de vocales en contacto

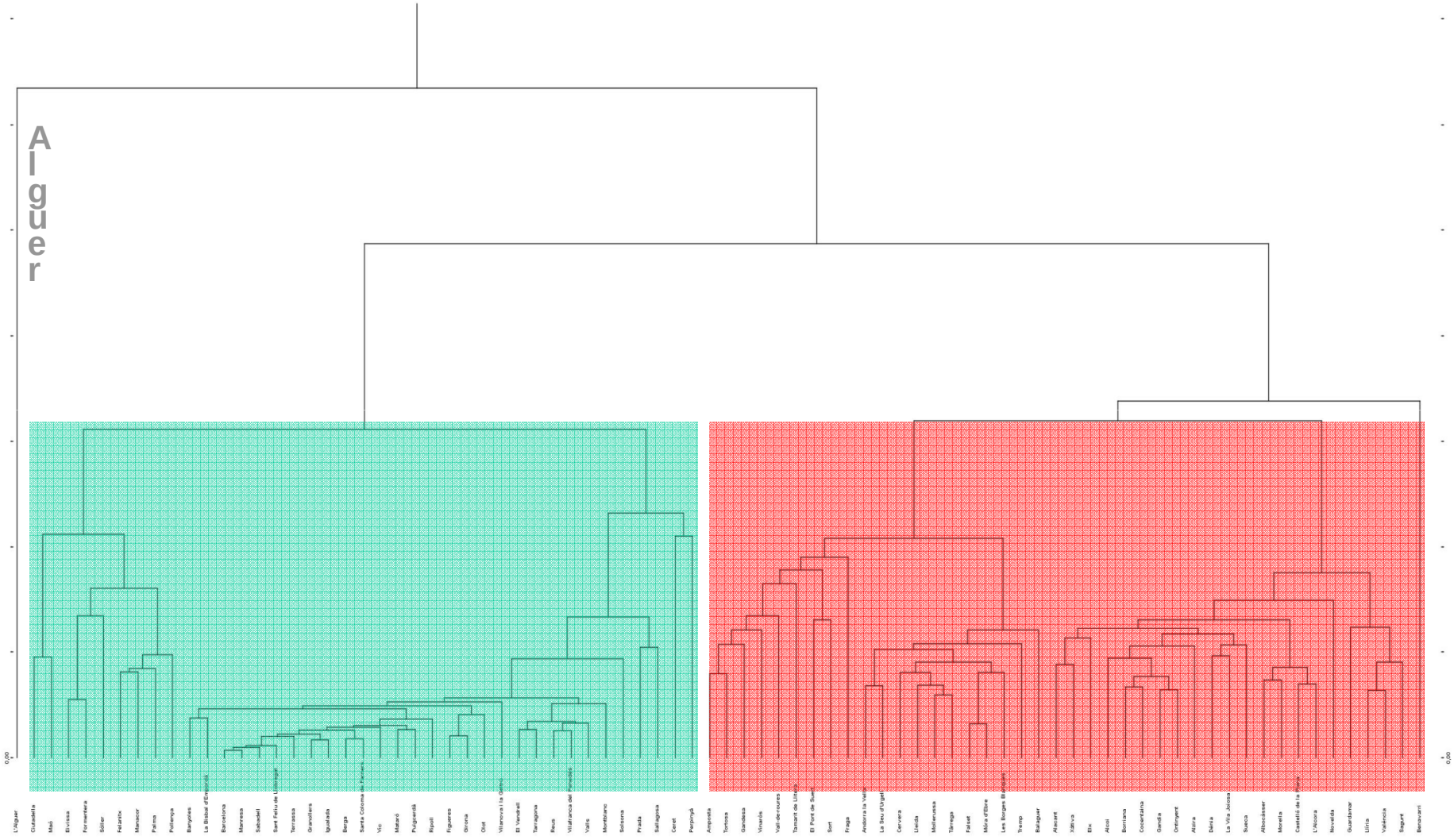
Distancia lingüística basada en formas fonológicas:

	Variedad 1	Variedad 2	Variedad 3
Variedad 1			
Variedad 2	0		
Variedad 3	4	4	

Representación de la distancia lingüística basada en el análisis fonológico



Análisis dialectométrico del COD (datos fonéticos)



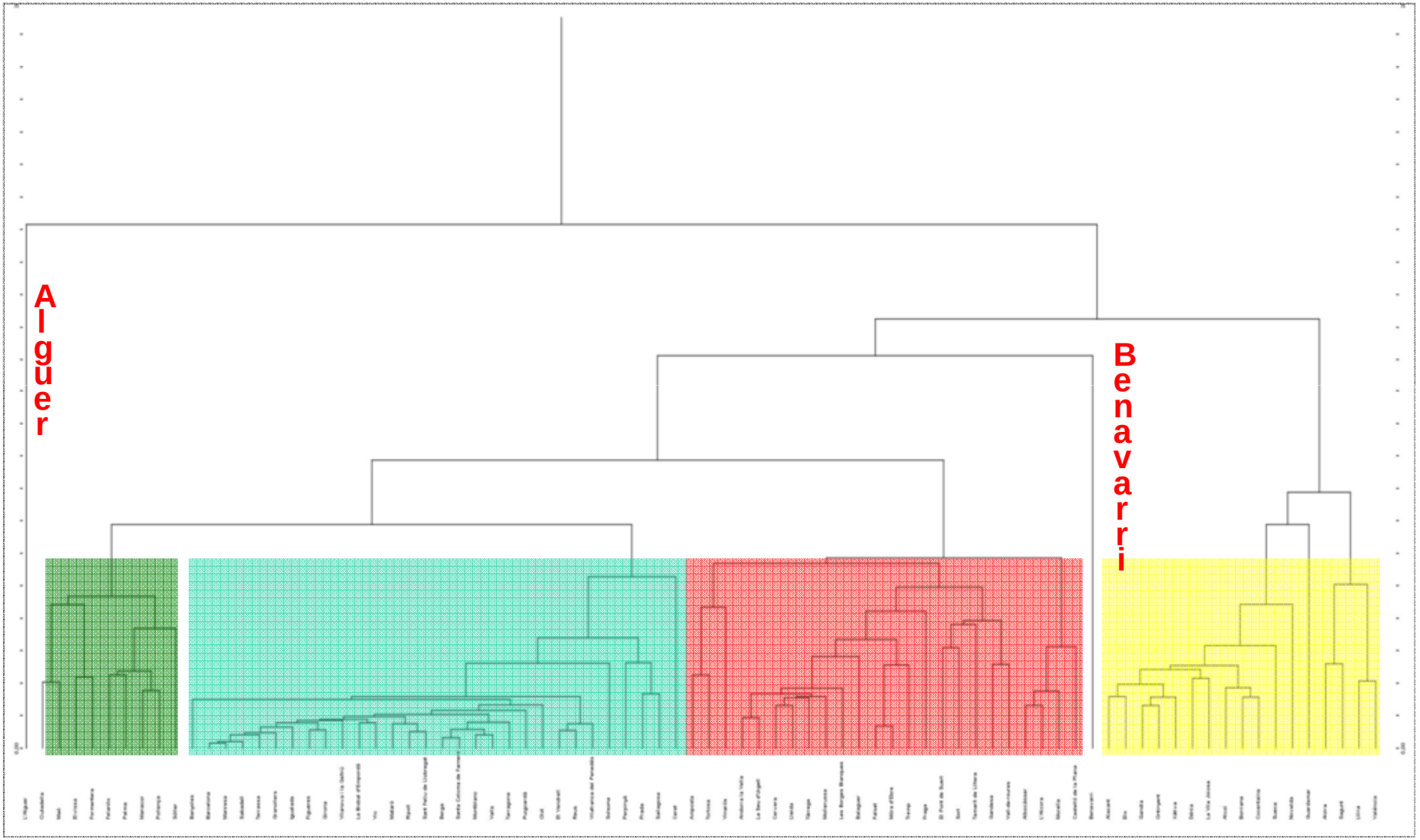
Catalán oriental

Catalán occidental

Clasificación dialectal tradicional

Alguer

Análisis dialectométrico del COD (análisis fonológico previo de los datos). Representación dendrográfica



V. baleares

V. centrales i septentrionales.

V. noroccidentales

V. valencianas

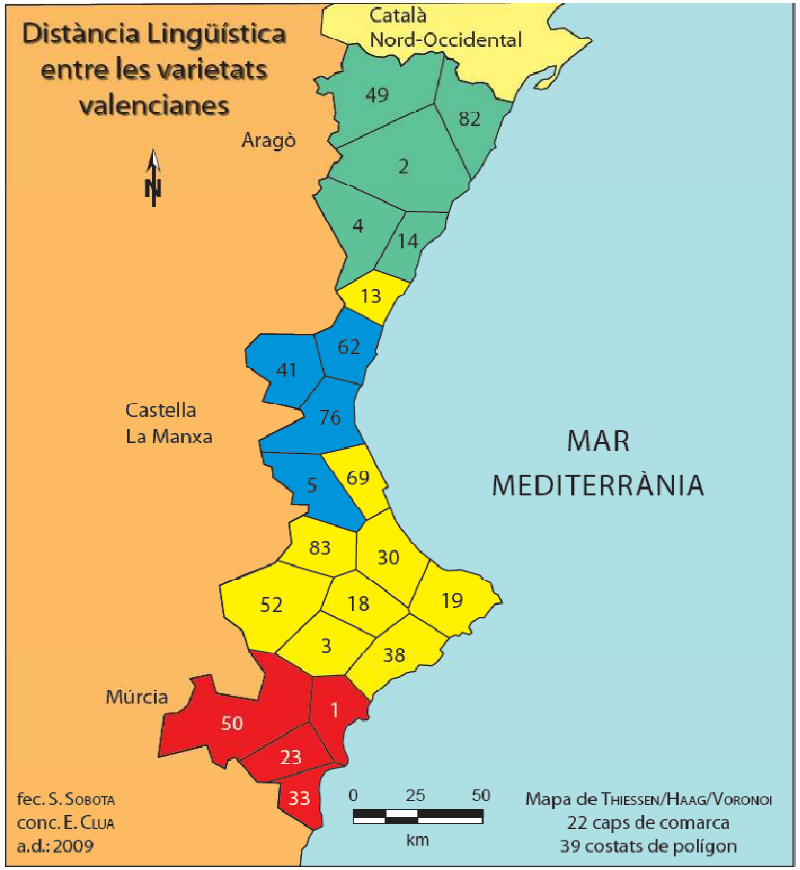
ii) Aplicación a los datos del COD del *Visual Dialectometry Method* (VDM): representaciones dendrográficas y representaciones espaciales mediante mapas cloropetas

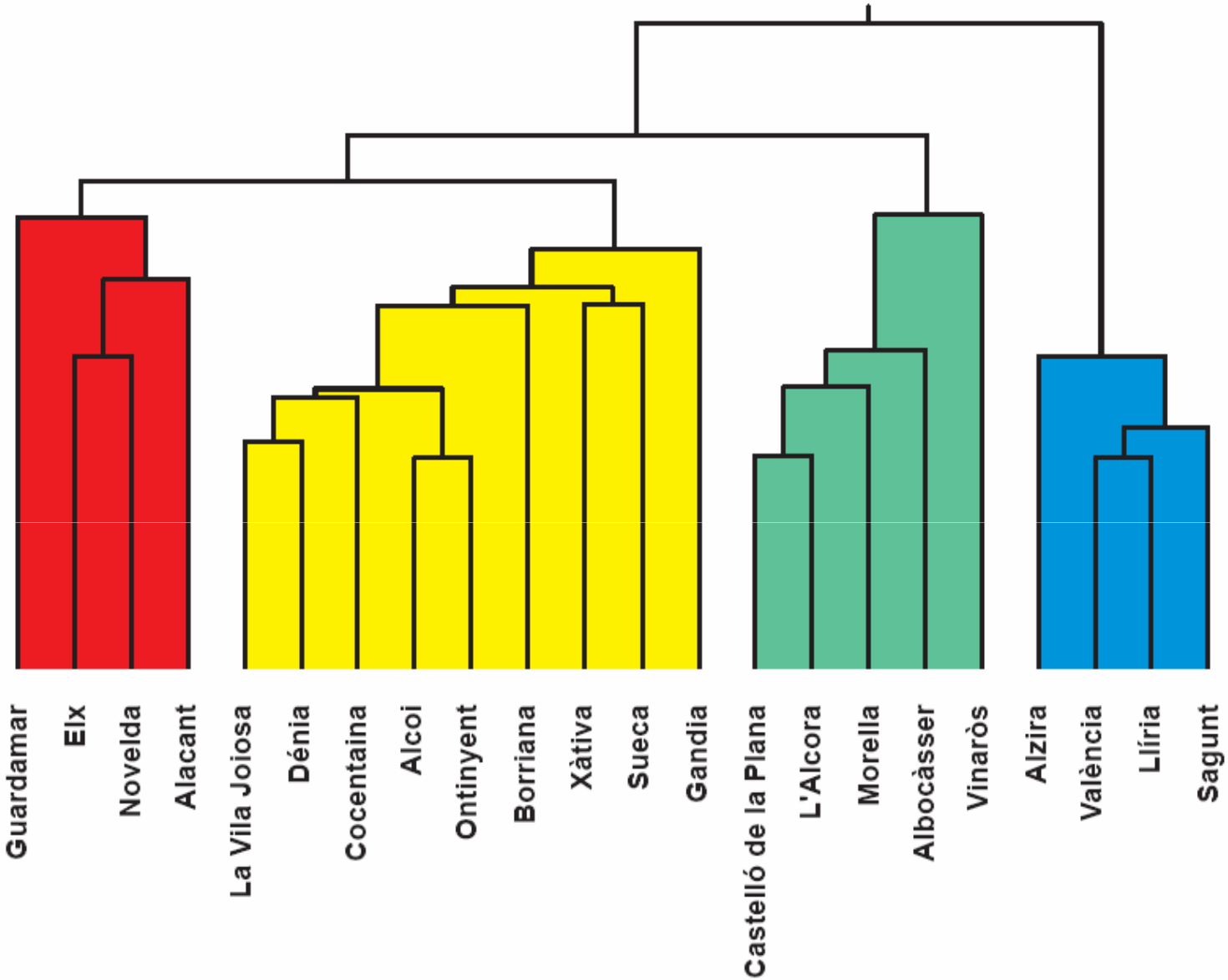
Mapas cloropetas: Reflejan la distancia lingüística de forma cromática:

colores más cálidos → variedades más similares

colores más fríos → variedades má distantes

- En colaboración con la escuela de Salzburgo (Goebl *et al.*), se ha creado una **Applet** que permite visualizar los resultados interactivament on-line de las variedades valencianas y se han analizado todos los datos del COD.


Distància Lingüística
entre les varietats
valencianes
Català
Nord-Occidental
Aragò
N
49
82
2
4
14
13
62
41
76
Castella
La Manxa
5
69
MAR
MEDITERRÀNIA
83
30
52
18
19
3
38
Múrcia
50
1
23
33
0
25
50
km
fec. S. Sobota
conc. E. Clua
a.d.: 2009
Mapa de Thiessen/Haag/Voronoi
22 caps de comarca
39 costats de polígon


Guardamar
Elx
Novelda
Alacant
La Vila Joiosa
Dénia
Cocentaina
Alcoi
Ontinyent
Borriana
Xàtiva
Sueca
Gandia
Castelló de la Plana
L'Alcora
Morella
Albocàsser
Vinaròs
Alzira
València
Llíria
Sagunt

Maria-Rosa Lloret, Joaquim Viaplana, Maria Pilar Perea i Esteve Clua

Applet Dialectomètrica

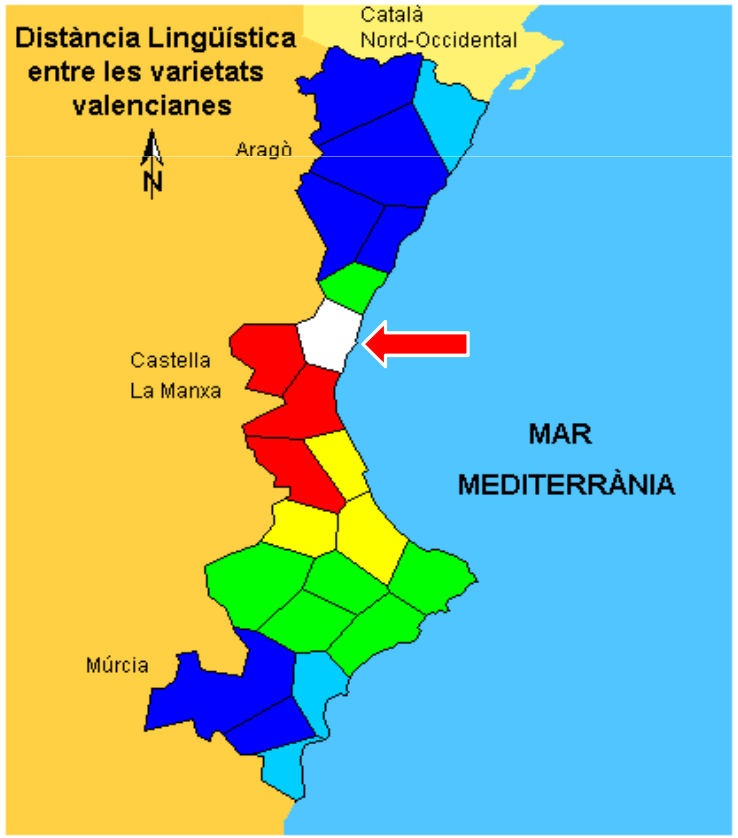
Distància Lingüística entre les varietats valencianes

Coordinació: Esteve Clua

Desenvolupat per: Slawomir Sobota

Global: Mapa de similitud

Algoritme d'intervalització MINMWMAX 6 Grups Per veure el codi de les localitats en el mapa



Localitat de referència escollida:

P. 62 Sagunt

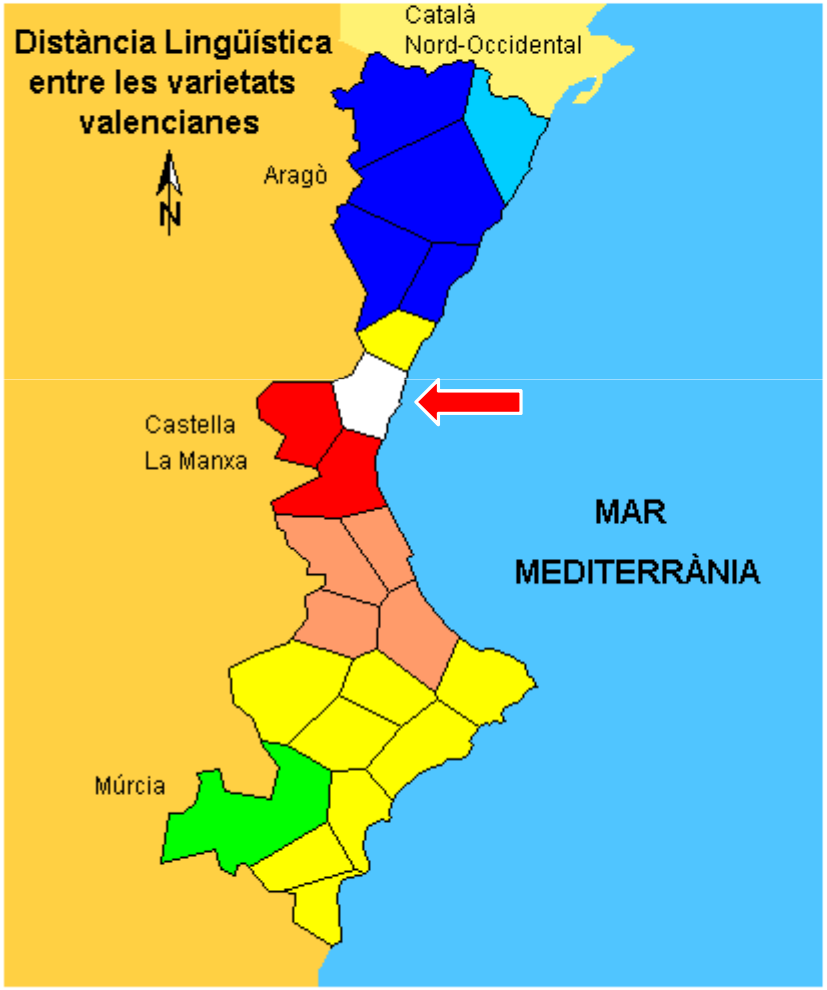
Distribució de similituds:

1		63,90 - 67,35 %	(Σ = 6)
2		67,36 - 70,79 %	(Σ = 3)
3		70,80 - 74,24 %	(Σ = 6)
4		74,25 - 82,10 %	(Σ = 3)
5		82,11 - 89,97 %	(Σ = 0)
6		89,98 - 97,84 %	(Σ = 3)

Distància Lingüística entre les varietats valencianes

Flexió verbal: Mapa de similitud

Algoritme d'intervalització MINMWMAX 6 Grups Per veure el codi de les localitats en el mapa



Localitat de referència escollida:

P. 62 Sagunt

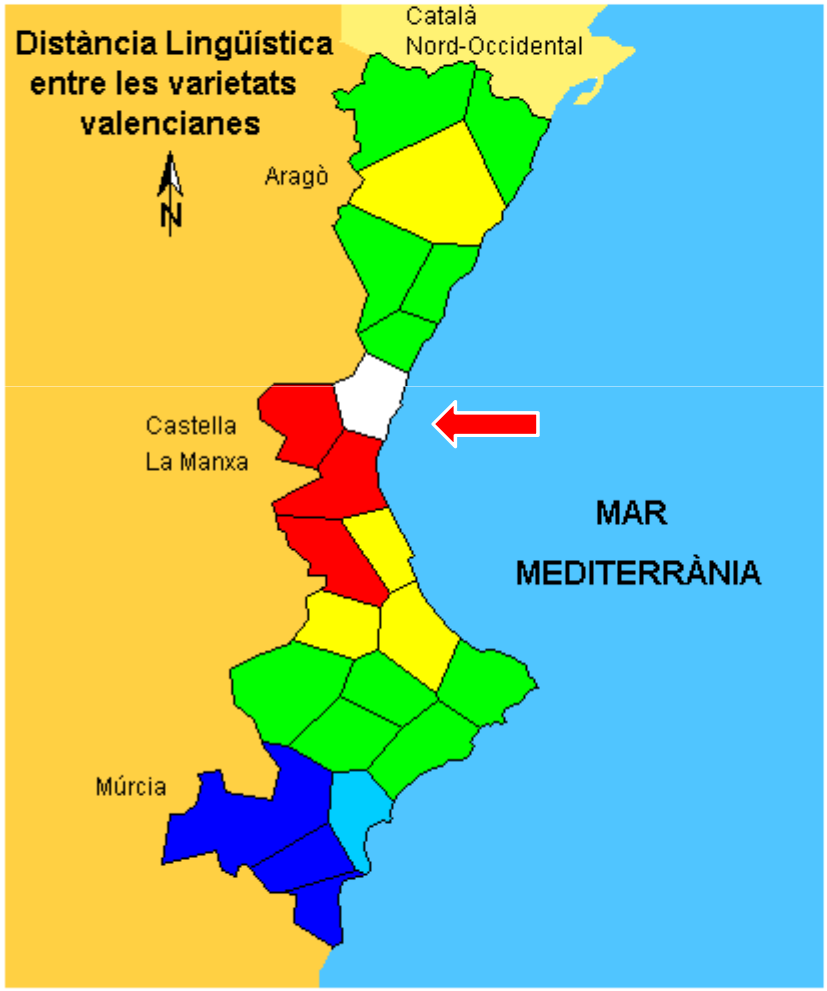
Distribució de similituds:

1	69,93 - 74,91 %	(Σ = 4)
2	74,92 - 79,88 %	(Σ = 1)
3	79,89 - 84,86 %	(Σ = 1)
4	84,87 - 88,73 %	(Σ = 9)
5	88,74 - 92,61 %	(Σ = 4)
6	92,62 - 96,49 %	(Σ = 2)

Distància Lingüística entre les varietats valencianes

Flexió nominal: Mapa de similitud

Algoritme d'intervalització MINMWMAX 6 Grups Per veure el codi de les localitats en el mapa



Localitat de referència escollida:

P. 62 Sagunt

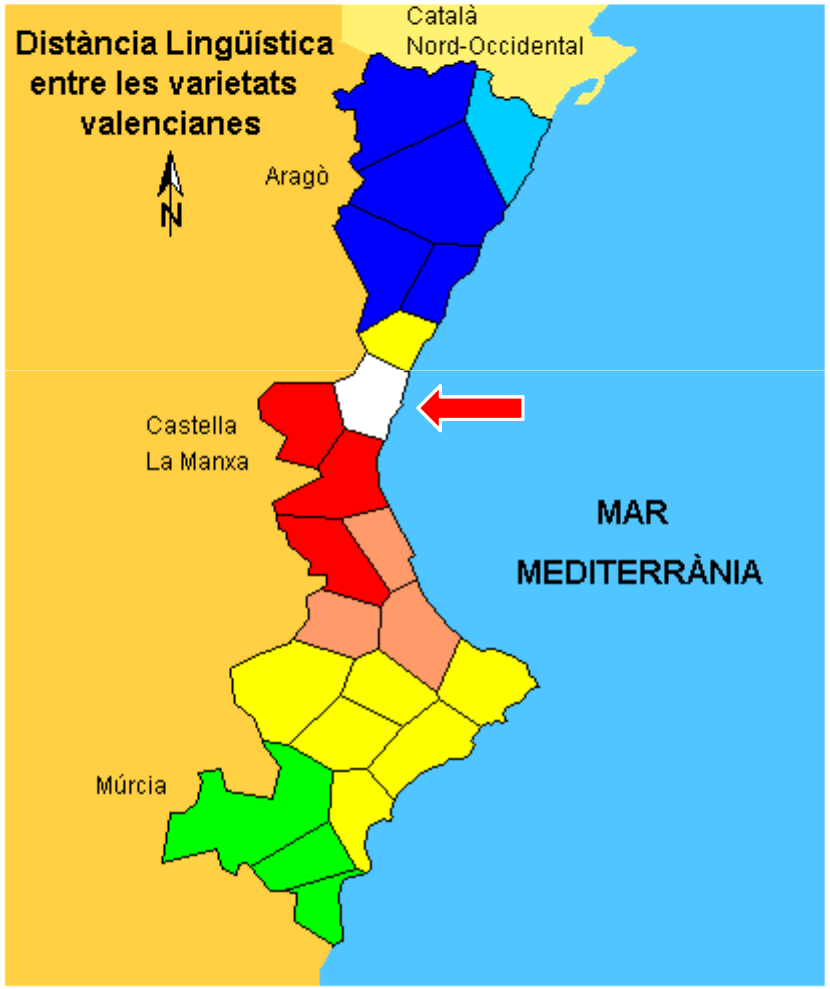
Distribució de similituds:

1	58,08 - 63,38 %	(Σ = 3)
2	63,39 - 68,69 %	(Σ = 1)
3	68,70 - 73,99 %	(Σ = 10)
4	74,00 - 81,06 %	(Σ = 4)
5	81,07 - 88,14 %	(Σ = 0)
6	88,15 - 95,21 %	(Σ = 3)

Distància Lingüística entre les varietats valencianes

Processos fonològics: Mapa de similitud

Algoritme d'intervalització MINMWMAX 6 Grups Per veure el codi de les localitats en el mapa



Localitat de referència escollida:

P. 62 Sagunt

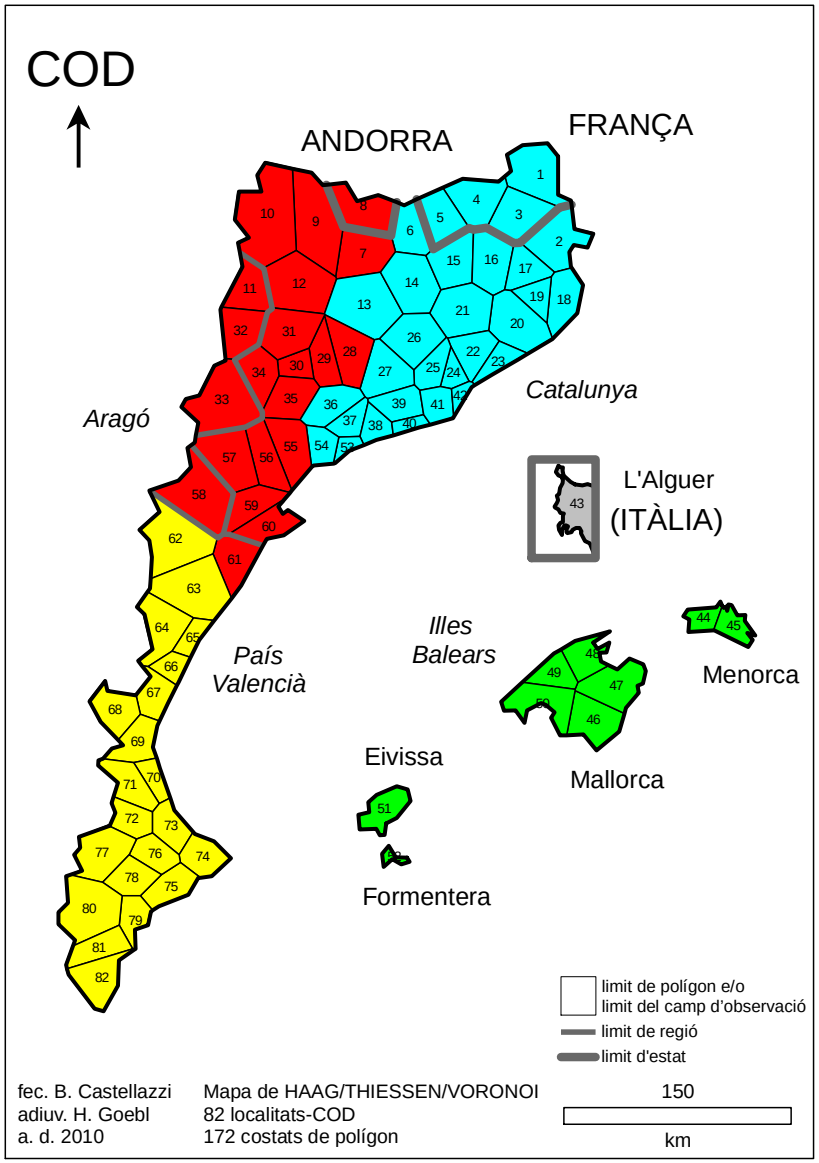
Distribució de similituds:

1	58,31 - 65,32 %	(Σ = 4)
2	65,33 - 72,33 %	(Σ = 1)
3	72,34 - 79,34 %	(Σ = 3)
4	79,35 - 85,46 %	(Σ = 7)
5	85,47 - 91,59 %	(Σ = 3)
6	91,60 - 97,71 %	(Σ = 3)

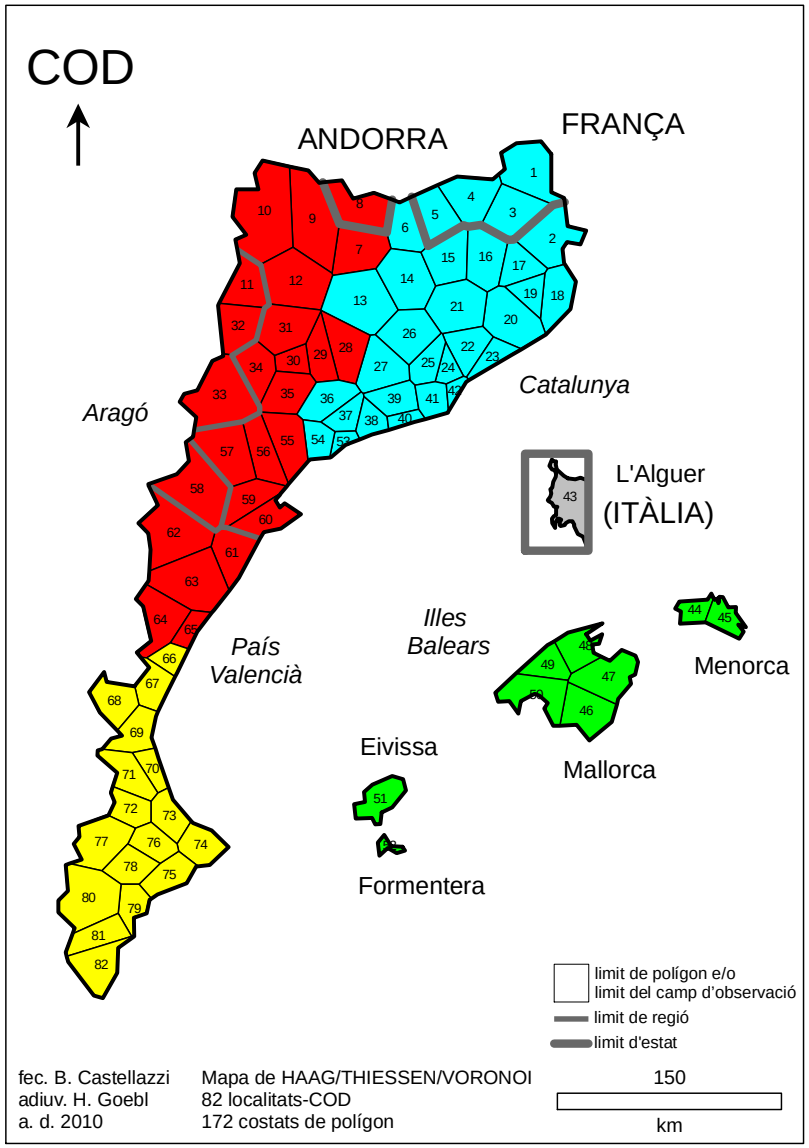
www.iula.upf.edu/rec/vdm_val/

o a través p. web del proyecto COD:
www.ub.edu/lincat

Análisis dialectométrico del COD (datos fonéticos). Representación cartográfica. VDM

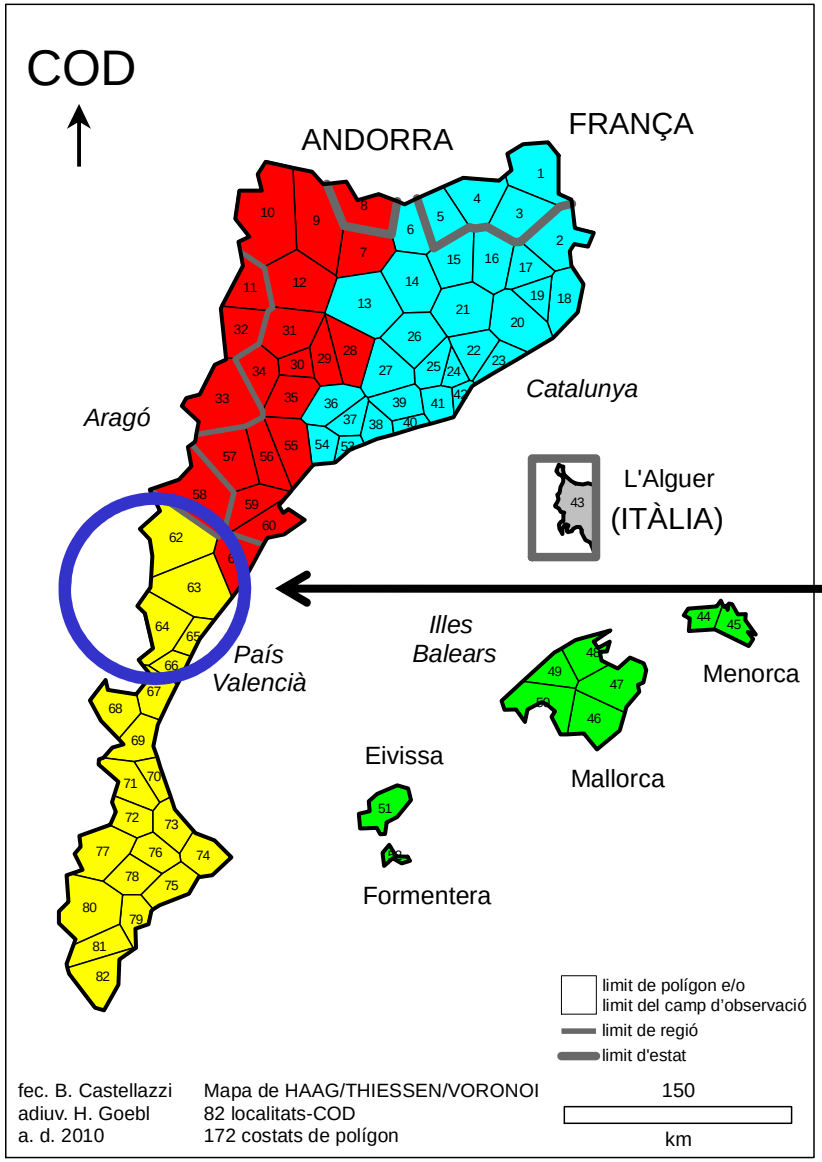


Análisis dialectométrico del COD (análisis fonológico previo de los datos). Representación cartográfica. VDM

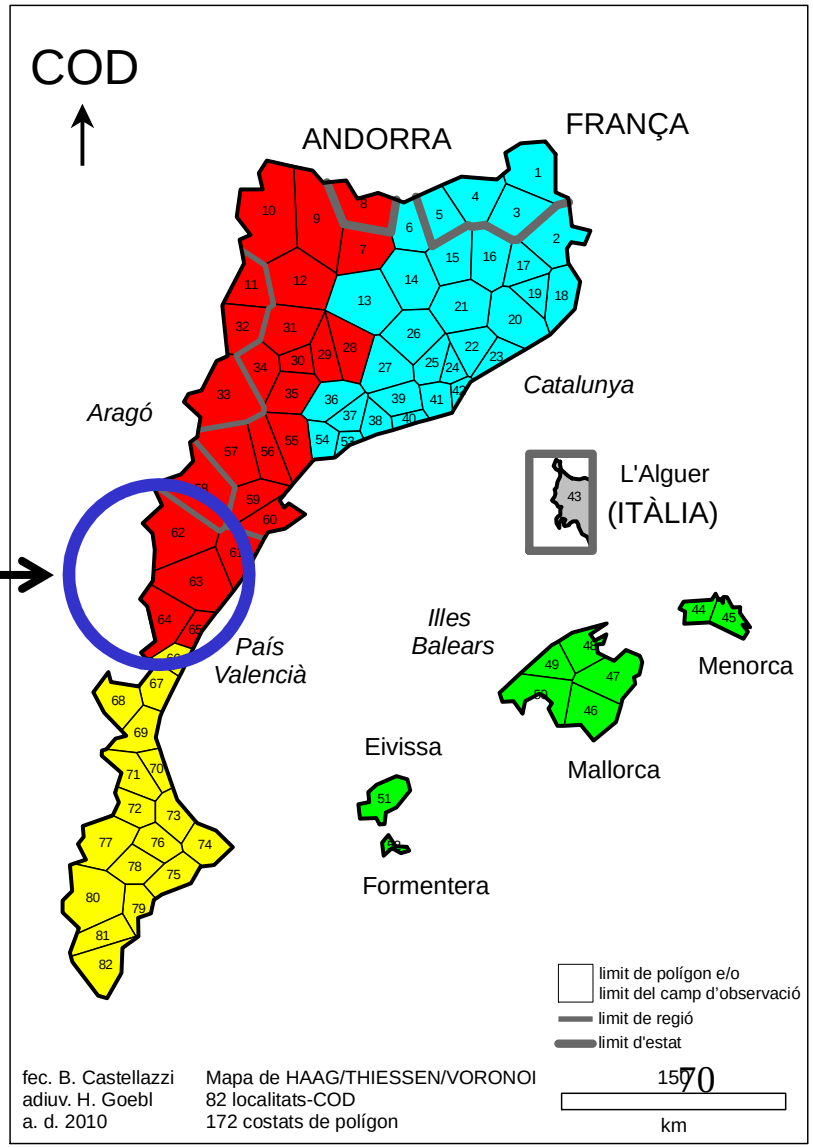


Representación cartográfica. VDM

Datos fonéticos

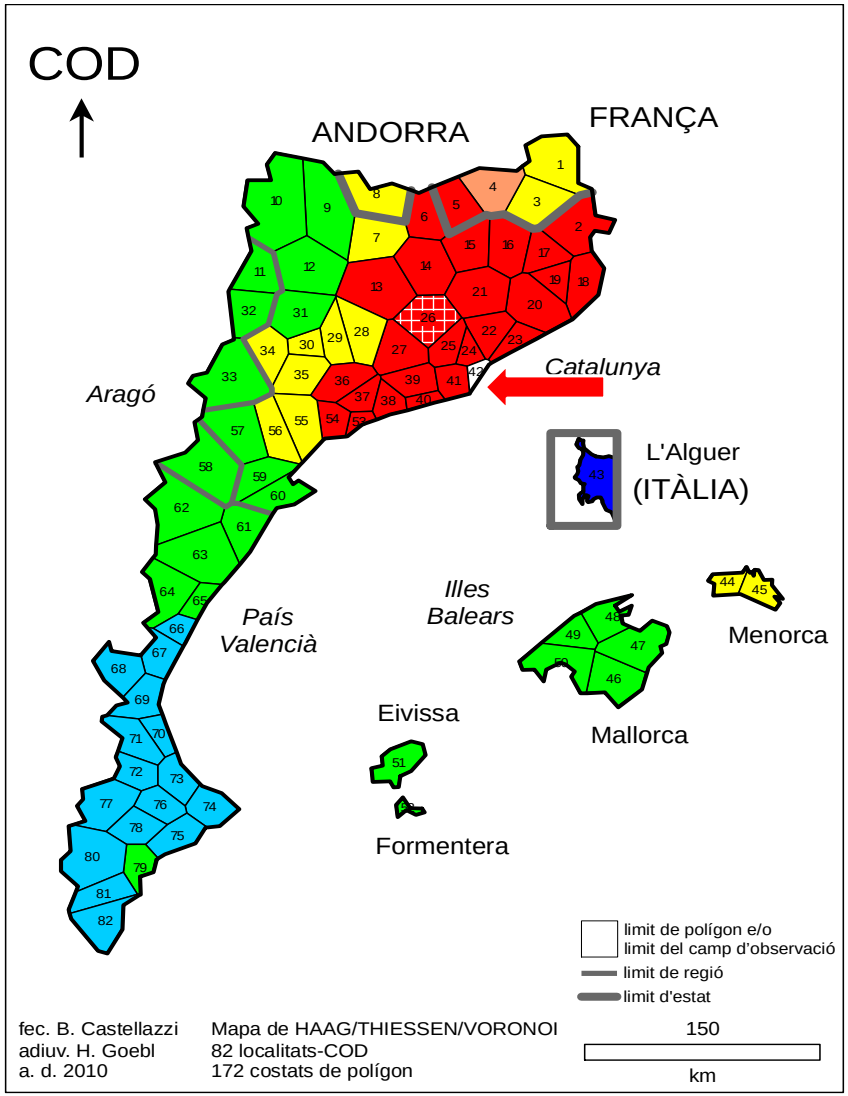


Datos fonológicos



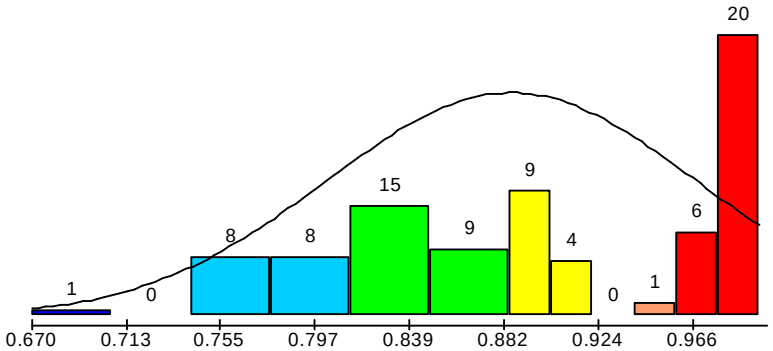
Mapa de similitudes

Localidad de referencia: 42. Barcelona



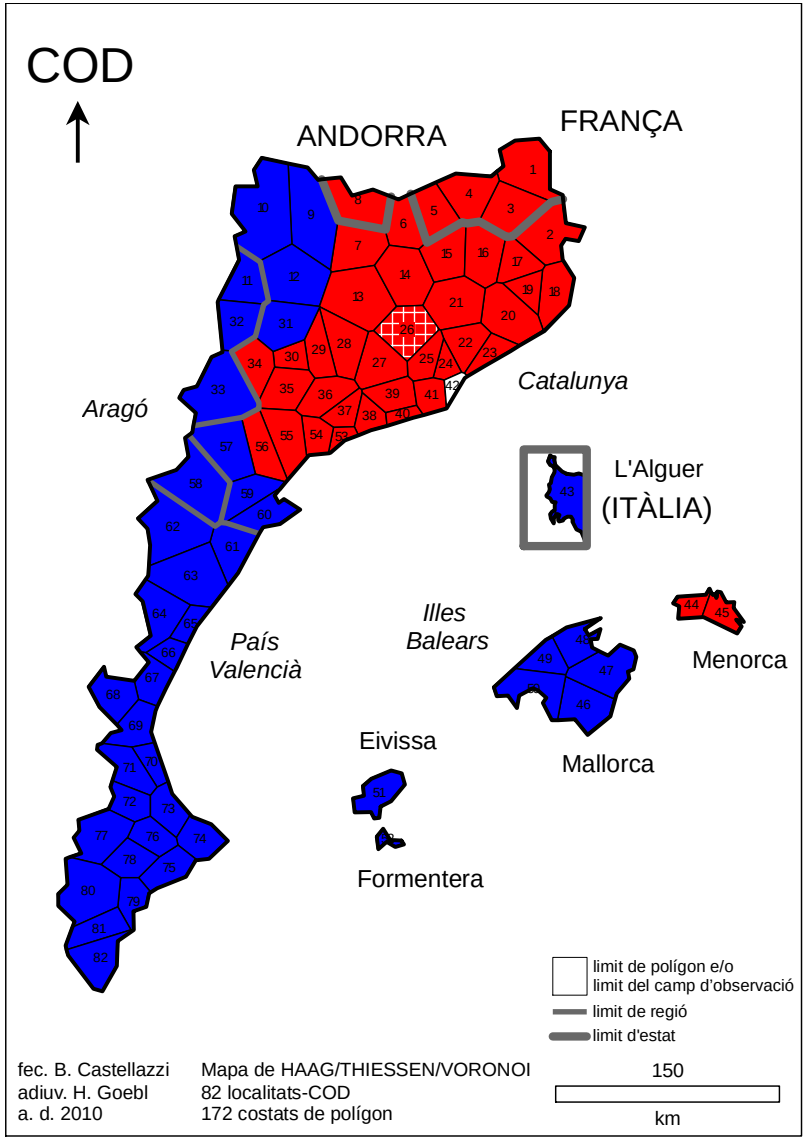
Algoritmo de intervalización: MINMWMAX
Grupos: 6

- [1] von 0.67 bis 0.74 (n = 1)
- [2] bis 0.81 (n = 16)
- [3] bis 0.88 (n = 24)
- [4] bis 0.92 (n = 13)
- [5] bis 0.96 (n = 1)
- [6] bis 1.00 (n = 26)
- Summe der Elemente: 81



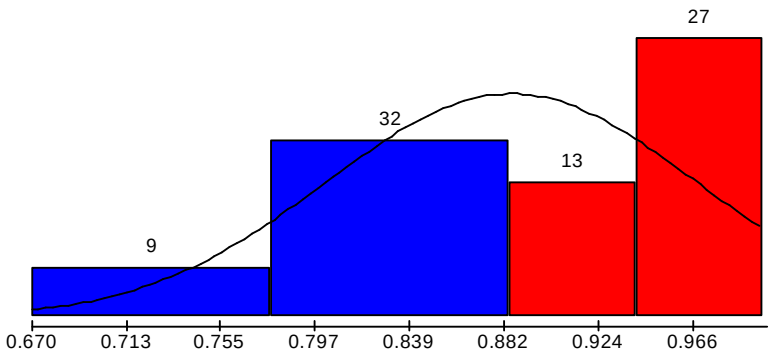
Mapa de similitudes

Localidad de referencia: 42. Barcelona



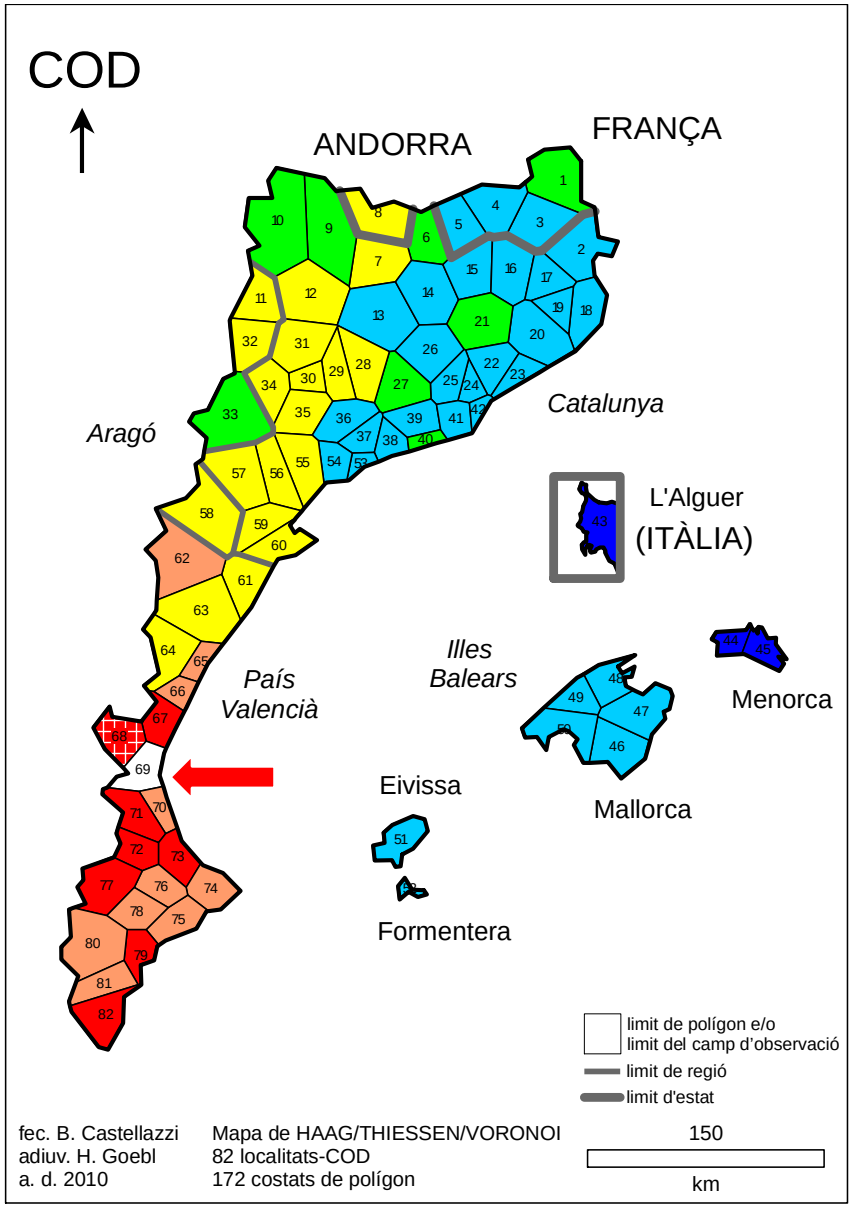
Algoritmo de intervalización: MINMWMAX
Grupos: 2

[1] von 0.67 bis 0.88 (n = 4:
[2] bis 1.00 (n = 40)
Summe der Elemente: 81



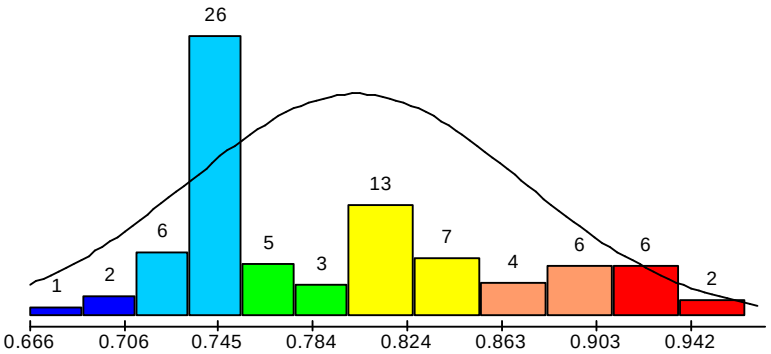
Mapa de similitudes

Localidad de referencia: 63. València



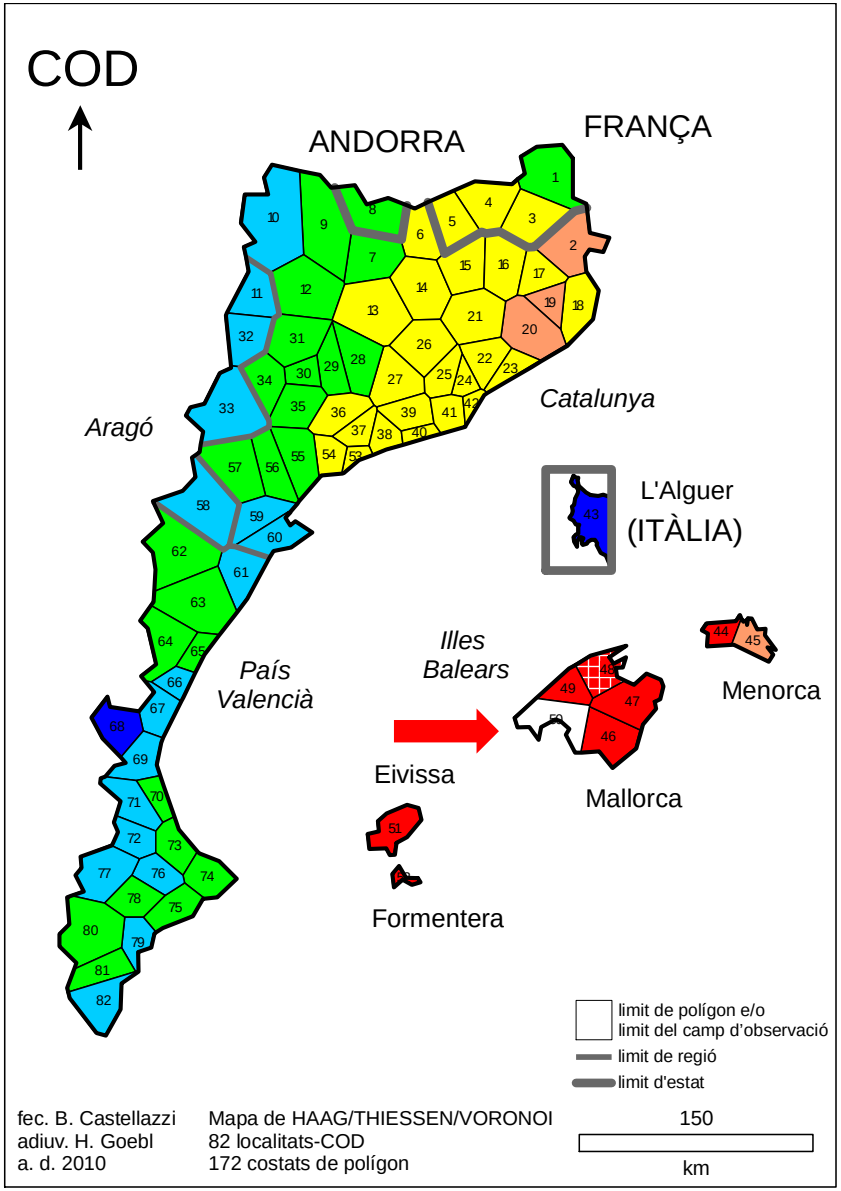
Algoritmo de intervalización: MINMWMAX
Grupos: 6

- [1] von 0.67 bis 0.71 (n = 3)
- [2] bis 0.76 (n = 32)
- [3] bis 0.80 (n = 8)
- [4] bis 0.86 (n = 20)
- [5] bis 0.92 (n = 10)
- [6] bis 0.97 (n = 8)
- Summe der Elemente: 81



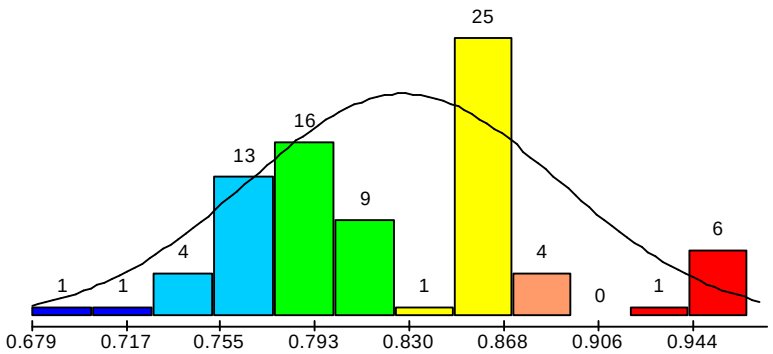
Mapa de similitudes

Localidad de referencia: 50. Palma (Mallorca)

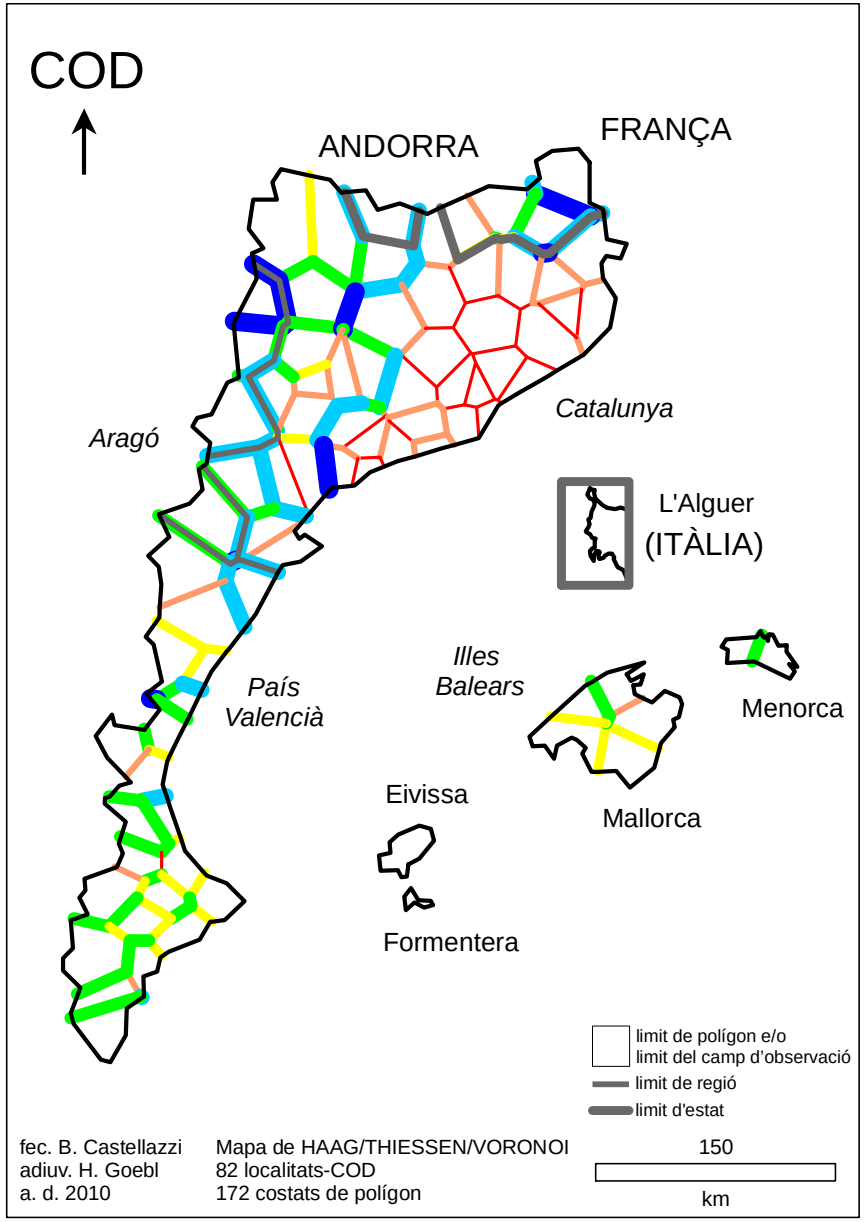


Algoritmo de intervalización: MINMWMAX
Grupos: 6

- [1] von 0.68 bis 0.73 (n = 2)
- [2] bis 0.78 (n = 17)
- [3] bis 0.83 (n = 25)
- [4] bis 0.88 (n = 26)
- [5] bis 0.92 (n = 4)
- [6] bis 0.97 (n = 7)
- Summe der Elemente: 81

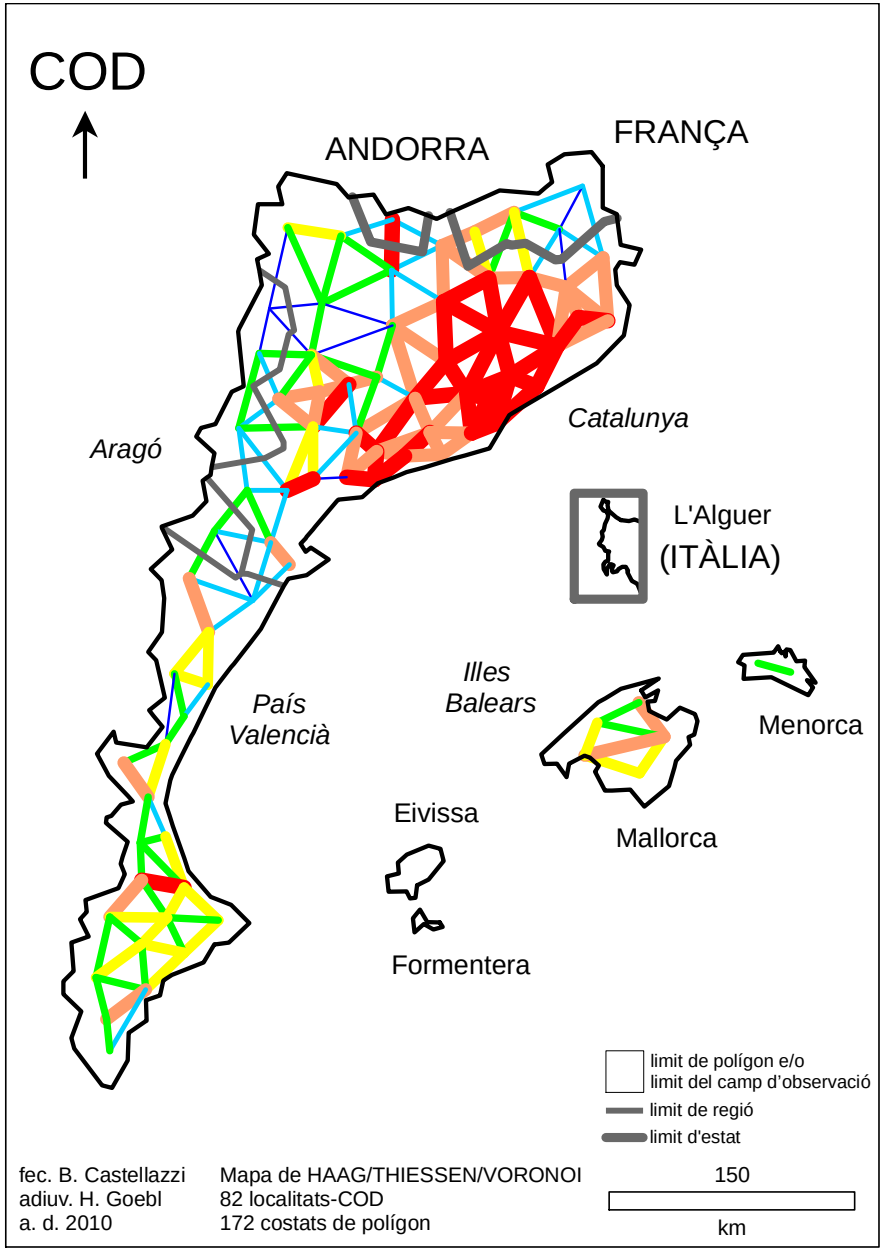


Mapa de isoglosas



- [1] von 0.00 bis 0.02 (n = 3
- [2] bis 0.03 (n = 39)
- [3] bis 0.05 (n = 22)
- [4] bis 0.08 (n = 36)
- [5] bis 0.11 (n = 27)
- [6] bis 0.14 (n = 11)
- Summe der Elemente: 172
- MIN = 0.00
- MW = 0.05
- MED = 0.04
- MAX = 0.14

Mapa de rayos



[1] von 0.86 bis 0.89 (n = 1:
[2] bis 0.92 (n = 27)
[3] bis 0.95 (n = 36)
[4] bis 0.97 (n = 22)
[5] bis 0.98 (n = 39)
[6] bis 1.00 (n = 37)
Summe der Elemente: 172
MIN = 0.86
MW = 0.95
MED = 0.96
MAX = 1.00

El trabajo analítico continua en el proyecto coordinado:

- FFI2010-22181-C03-03 (IP: Esteve Clua, U Pompeu Fabra)

  Distancia lingüística en los ejes espacial y temporal: aspectos fonológicos y morfológicos del catalán

- FFI2010-22181-C03-02 (IP: Maria-Rosa Lloret, U Barcelona)

  Descripción e interpretación de la variación dialectal: aspectos fonológicos y morfológicos del catalán

- FFI2010-22181-C03-01 (IP: Joan Mascaró, U Autónoma Barcelona)

  Análisis teórico de los procesos segmentales y morfofonológicos del catalán

FFI2010-22181-C03:

Estudio de la fonología y la morfología del catalán: descripción, teoría y variación

Muchas gracias por su atención

mrosa.lloret@ub.edu
esteve.clua@upf.edu

Qatar
Foundation
DRI-FIT