

# Revisiting Southern Gallo-Romance from a Complexity Theory viewpoint: Occitan

Complexity in Language Variation and Change (COMPILA 2023)  
Room Collaboration 1 (4th floor), Building 18, Komaba I campus, University of Tokyo, Japan  
Satellite meeting for the 28th International Conference on Statistical Physics (StatPhys28)  
August 4th (Fri.) 2023

Jean Léo Léonard (EA 739: Dipralang, UPV: Université Paul Valéry, Montpellier 3), Guylaine Brun-Trigaud (UMR 7320: Bases, Corpus, Langage, Université Côte d'Aur & CNRS), Flore Picard (Sorbonne Université) & Louise Albinet (UPV)

# Map of the talk

- 1) Introduction: preliminary remarks
- 2) Context, data, method and purpose: Occitan as a dialect network
- 3) Case Study 1: a central dialect ( ALLOc data, THESOC): Western Languedocian
- 4) Case study 2: a peripheral bow (Golfe du Lion)
- 5) Prospects

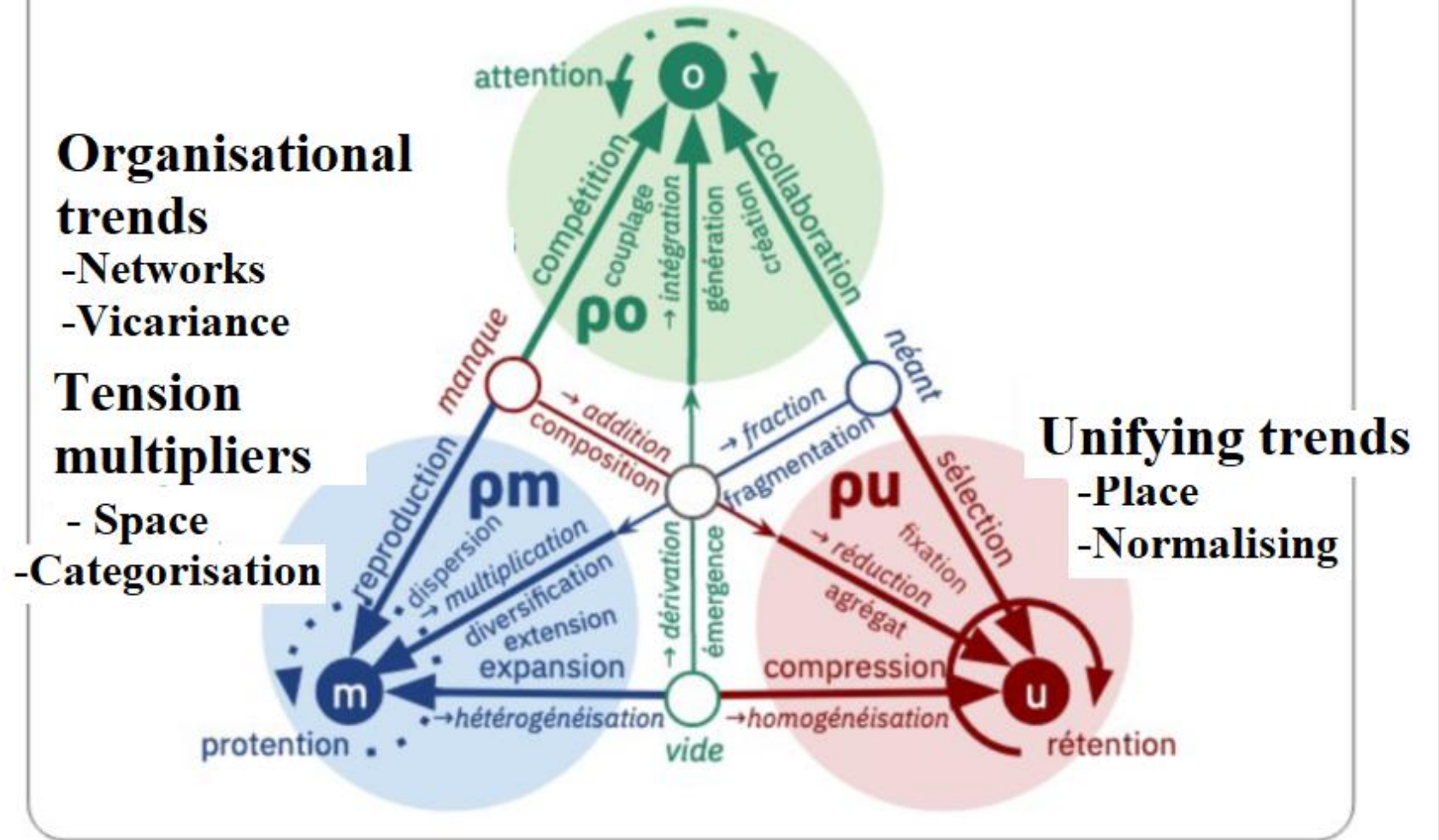
## 1) Some preliminary remarks

- **First**, we want to make clear that our approach differs from the *Automated Similarity Judgment Program (ASJP)*: we think dialectometry, especially LD, should be applied to *dialect continua*, or at least, inner diversity of well identified linguistic domains. Not to long range phylogenetic kinship – e.g. cladistics seems to be a more *challenging* tool for this purpose.
- **Second**, we assume LD is optimally heuristic for fathoming *linguistic landscapes*, or 'language dynamics, history and the social fabric from a quantitative viewpoint', rather than phylogenetic clusters and *longue durée*. Above all, dialectometry grasps typological trends, rather than phylogenetical evidence.
- **Third**, we suggest LD does not only highlight (synchronic) dialectical patterns, but also, to some extent, sociolinguistic/sociolectal patterns (see the Golfe du Lion component of this talk).
- **Fourth**: we evaluate Gabmap as an outstanding heuristic tool, as the team who has been engineering it for over 20 years already (Nerbonne, Heeringa, Bolognesi, Prokić, Weiling, etc.) has shown remarkable epistemological awareness, and has developed and integrated into its general design many strategic mathematical and computational devices to enhance its heuristic accuracy in processing dialectical complexity (see abstract).

# The Interplay of Tensions in a Field of Transformation (ITFiT),

according to Lacombe, Eric  
 2021. *Transformation numérique des organisations en réseaux : les potentiels d'une schématisation dynamique de l'information*.  
 Thèse de doctorat en Sciences de l'Information et de la communication, PhD dissertation, dir. Lise Vieira, Université de Bordeaux 3, p. 214.

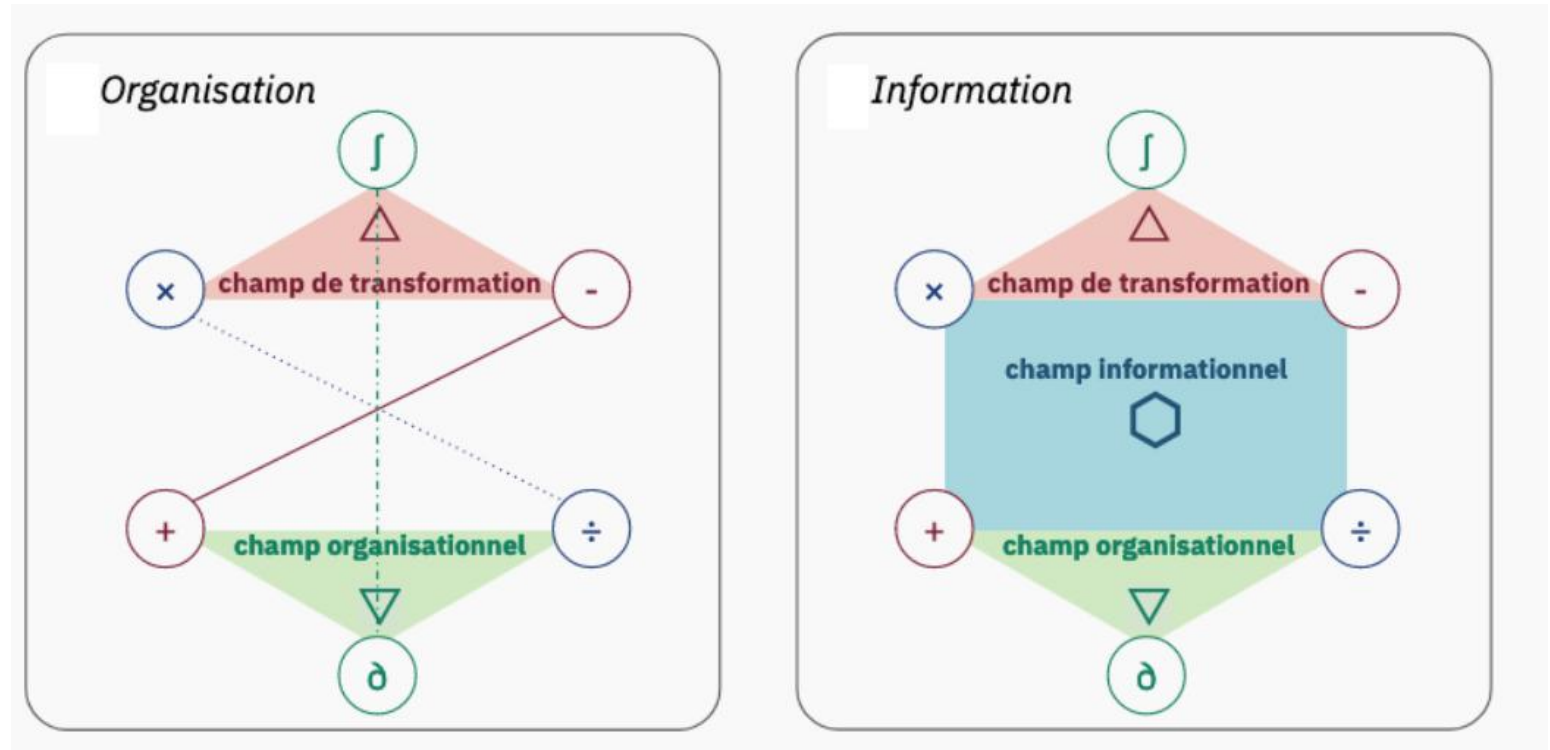
## Interplay of tensions in the field of transformations <sup>94</sup>





Organisation & Information Fields  
(x) = *multiplicans* etc.  
Source: Lacombe, E. (2023). Vers un langage pour exprimer la complexité.  
LHUMAINE, (2).  
<https://lhumaine.numerev.com/articles/revue-2/2959-vers-un-langage-pour-exprimer-la-complexite>

(x) = *Multiplicans*, (-) = *Subtrahens*, (+) = *Addens*, (./.) = *Dividens*  
& two symbols ( $\int$ ,  $\partial$ ) for *Tertium Inclusum*  
in the vertices of the graphs



2) Context,  
data, method  
and purpose:  
Occitan as a  
dialect  
network



# OCCITAN

- 14. Croissant's (transitional)
- 15. Limousin
- 16. Auvergnat
- 17. Languedocian
  - a. guyennese
  - b. cevennol
  - c. languedocian-N.
  - d. languedocian-C.
  - e. languedocian-E.
  - f. ariegese
- 18. Vivaro-alpin
  - a. vivarese
  - b. dromese
  - c. pre-alpine gavot
  - d. alpine gavot
  - e. cisalpine oc
- 19. Provençal
  - a. rhodanian
  - b. central
  - c. maritime
  - d. niçard

## BOUNDARIES

- national
- language
- dialectal

- LANGUE D'OIL**
- 1. Norman-N. (oil-NW)
  - a. insular norman
- 2. Picard (oil-N.)
- 3. Walloon (oil-NE.)
- 4. Lorrain (oil-CE.)
- 5. Weitchs (oil-E.)
- 6. Franc-Comtois-N. (oil-SE.)
- 7. Oïl-CS.
  - a. bourgundian
  - b. morvandian
- 8. Champagnese (oil-CN.)
- 9. Oïl-S.
  - a. berrian
  - b. bourbonnais
  - c. nivernais
  - d. auxerrois
- 10. Parlange (oil-SW.)
  - a. poitevin
  - b. saintongeais
  - c. angoumois
- 11. Gallo (oil-W.)
- 12. Oïl-CW.
  - a. normand-S.
  - b. lower maine
  - c. anjou
  - d. upper maine
  - e. tourain
  - f. orleanese
- 13. Francien (oil-C.)

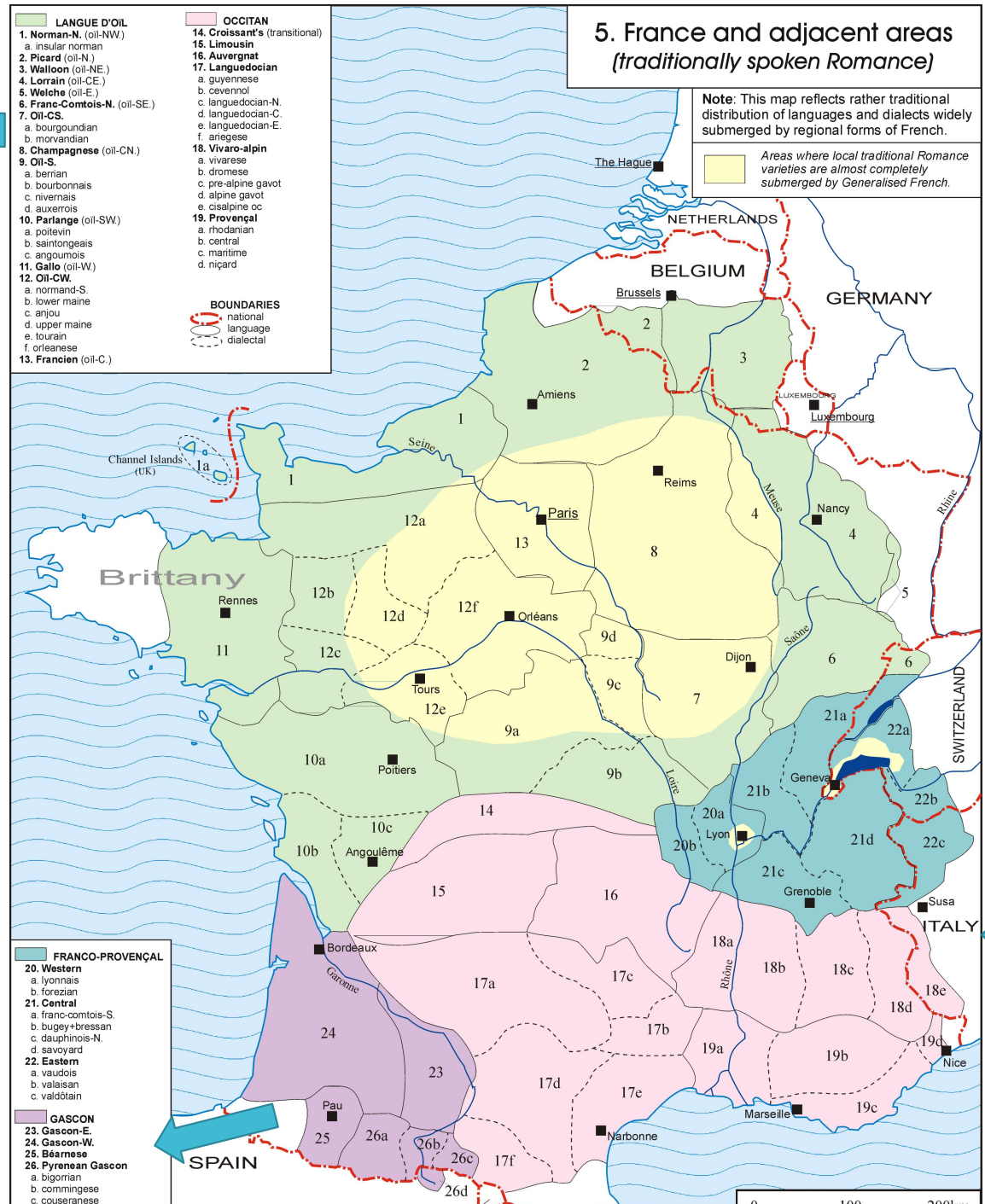
## BOUNDARIES

- national
- language
- dialectal

## 5. France and adjacent areas (traditionally spoken Romance)

Note: This map reflects rather traditional distribution of languages and dialects widely submerged by regional forms of French.

Areas where local traditional Romance varieties are almost completely submerged by Generalised French.



Source: Koryakov Yu.B.  
*Atlas of the Languages of the World.  
Romance languages.* Moscow, 2001.

Yuri B. Koryakov's map  
of the three Gallo-Romance  
Dialect networks  
-a canonical view.

# The canonical divisions and subdivisions of the Occitan dialect network (in dark and light purple)

(i.e. « attributive » dialects)

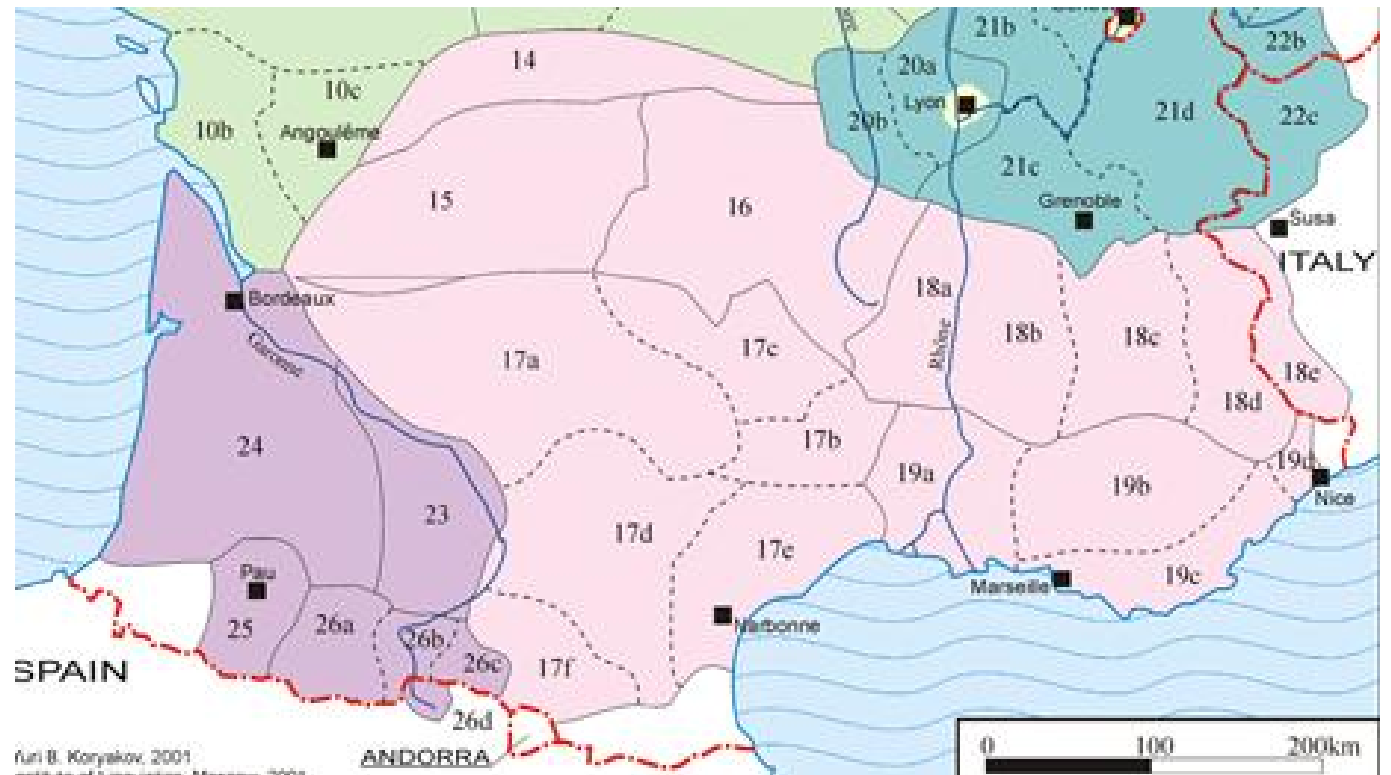
Source: Koryakov Yu.B.

*Atlas of the Languages of the World.*  
*Romance languages.* Moscow, 2001

**OCCITAN**

- 14. Croissant's (transitional)
- 15. Limousin
- 16. Auvergnat
- 17. Languedocian
  - a. guyennese
  - b. cevennol
  - c. languedocian-N.
  - d. languedocian-C.
  - e. languedocian-E.
  - f. ariegese
- 18. Vivaro-alpin
  - a. vivarese
  - b. dromese
  - c. pre-alpine gavot
  - d. alpine gavot
  - e. cisalpine oc
- 19. Provençal
  - a. rhodanian
  - b. central
  - c. maritime
  - d. niçard

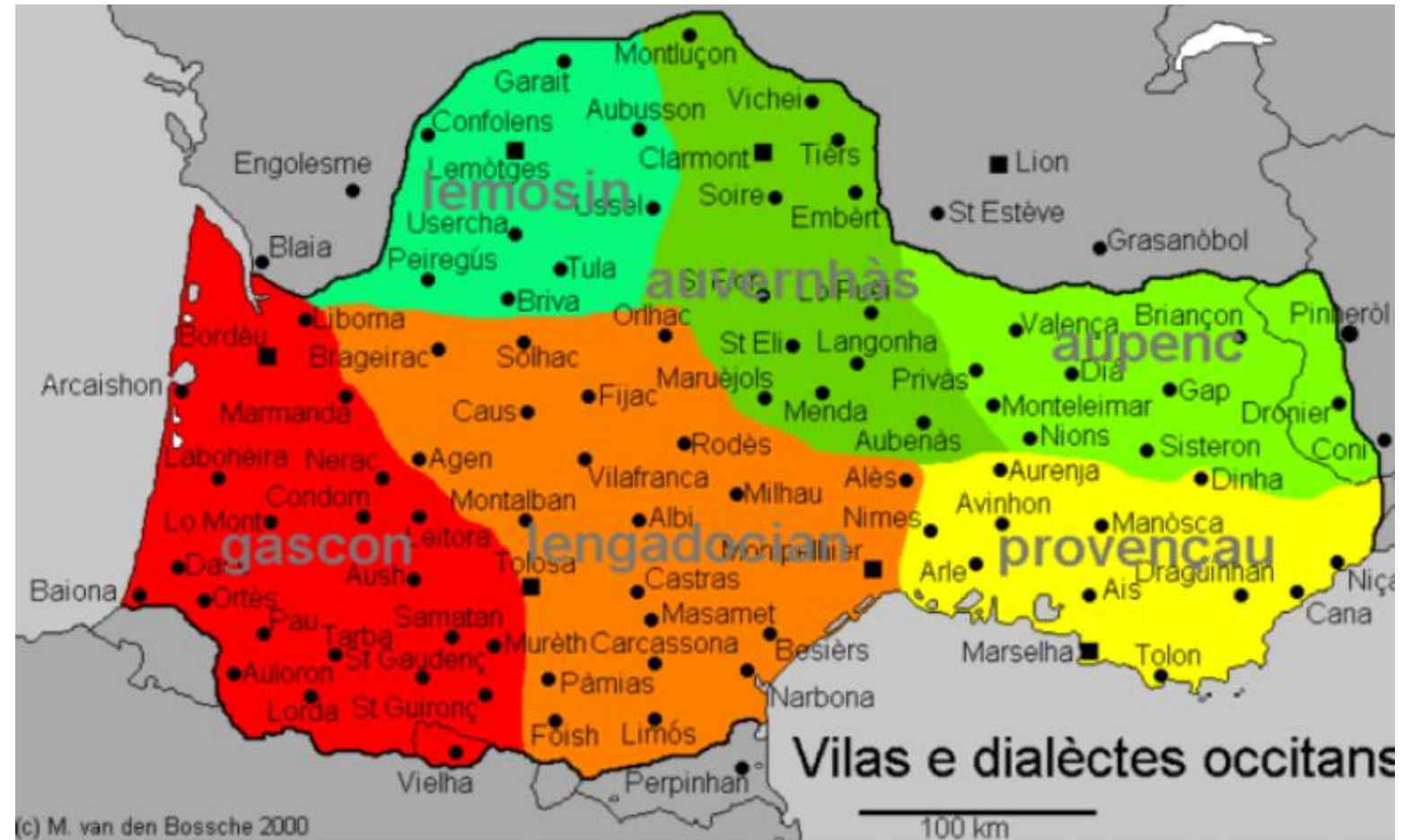
- GASCON**
- 23. Gascon-E.
  - 24. Gascon-W.
  - 25. Béarnese
  - 26. Pyrenean Gascon
    - a. bigorran
    - b. commingese
    - c. Couseransese
    - d. aranese



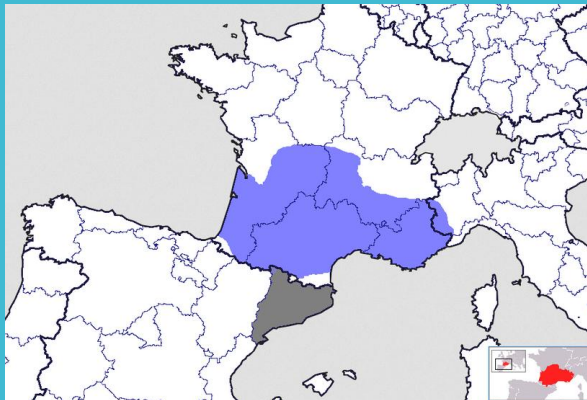


These are but ...  
« Attributive » dialects

**Complexity Theory**  
should help us to revisit  
these patterns, and to  
grasp *covert patterns* of  
emergence and  
**interaction** between  
*choremes*



# Why Occitan?



- A few advantages, for a survey of geolinguistic complexity, of the Occitan dialect network:
- **Huge** territory
- **Compact** area
- **Critical mass of reliable data**, thanks to the THESOC
- **71 phonological maps**
- **232 lexical maps**
- **Well studied**, stable (though sometimes challenging) « attributive dialects »
- **Geographic and historical complexity**
- **Polycentric**, former urban dialects (or not), strong entities (Provençal, Gascon, Languedocien, Limousin)

LOG IN WITH GENERIC

## Outils pour l'analyse dialectométrique par distance d'édition : Gabmap Libre d'accès en ligne



CLAPOP

Data ▾

Centres

Find data ▾

Find Services

CLARIN Recipes ▾

SEARCH

[Home](#)

[Services](#)

# Gabmap



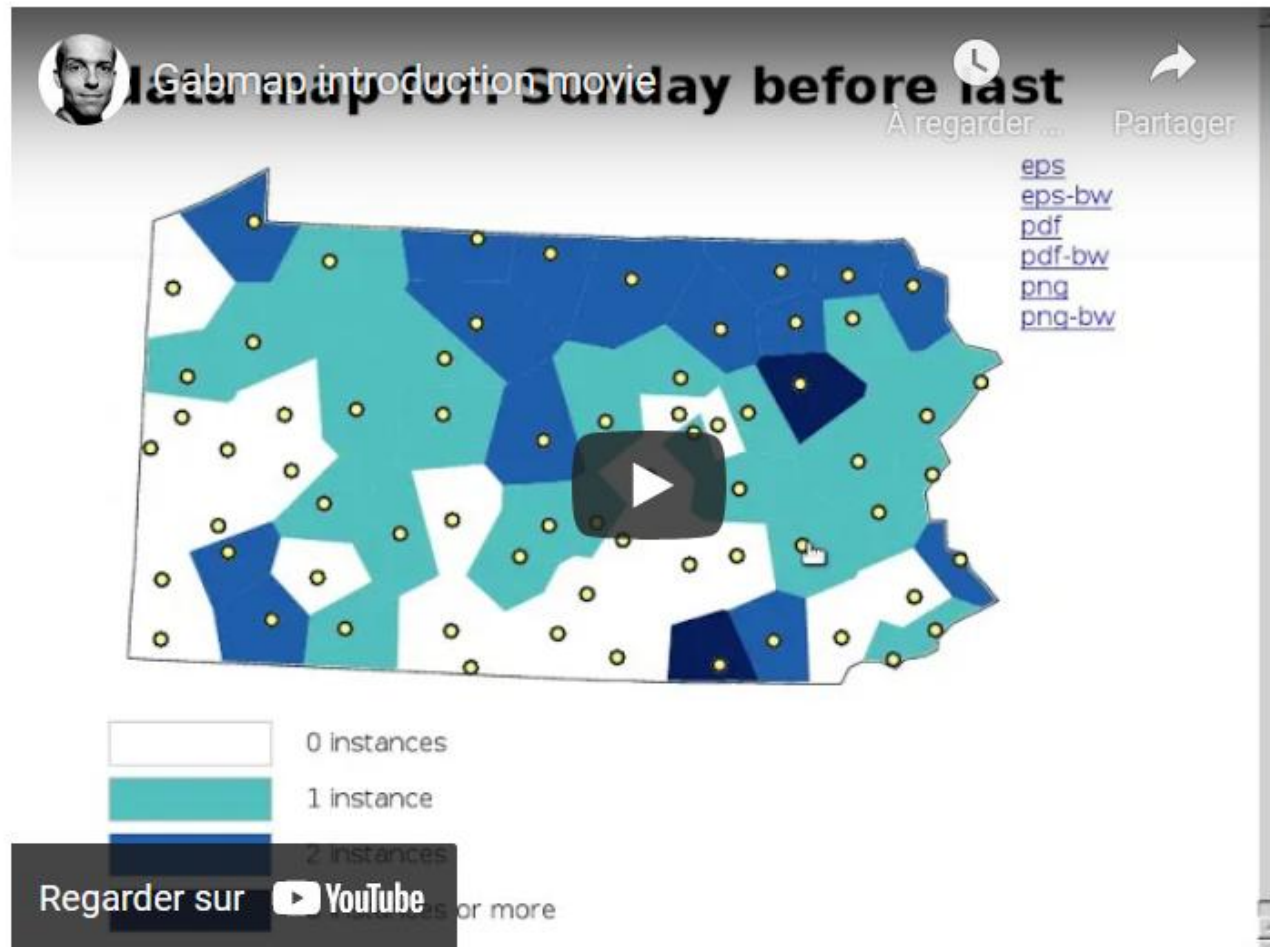


## This is Gabmap

Gabmap is a web application that visualizes dialect variations. It's made for dialectologists and students.

start application

try the demo

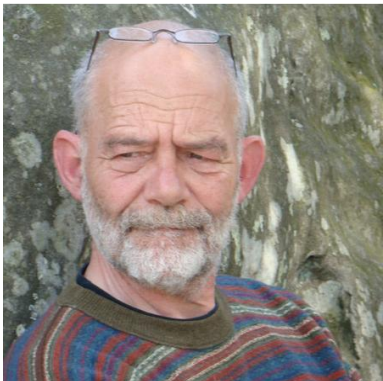


# Levenshtein algorithm (edit distance) procedure:

Adding, dropping, changing (phonology)  
 => ADDITION (+), SUBTRACTION (-)  
 TRANSFORMATION (x)

Operating on **CV templates** (string chains of characters)

Cf. CVCV  
 Templatic phonology  
 (Scheer, Lowenstamm & al.)



Method: String Edit Distance — Tokenized

Item: abeille

Place: ALAL 23

Show alignments

abeille

Page: « 1 2 3 4 5 6 7 8 9 10 11 »

a	b	'	œ	j	
a	v	'	ɛ	ʎ	ə
	1		1	1	1

ALAL 23 — ALP 2

a	b	'	œ	j	
a	b	'	e	ʎ	ə
			1	1	1

(-) = *Subtrahens* V#  
 (& *Substituens* V, vCv)

Place: ALAL 23

Show alignments

cheval

Page: « 1 2 3 4 5 6 7 8 9 10 11 »

ALAL 23 — ALP 1

ʃ	œ	v	'	o
ʃ	i	v	'	a
	1			1

ALAL 23 — ALP 2

ʃ	œ	v	'	o
ʃ	y	v	'	a
	1			1

ALAL 23 — ALP 3

ʃ	œ	v	'	o
ʃ	i	v	'	a
	1			1

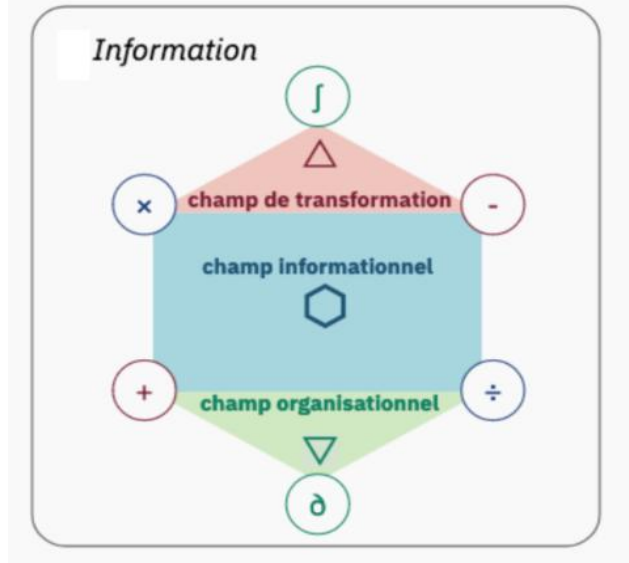
ALAL 23 — ALP 6

ʃ	œ	v	'	o
ʃ	i	v	'	a
ʃ	1			1

(Gentle) Reminder:

flow of **information**, self-**organization**

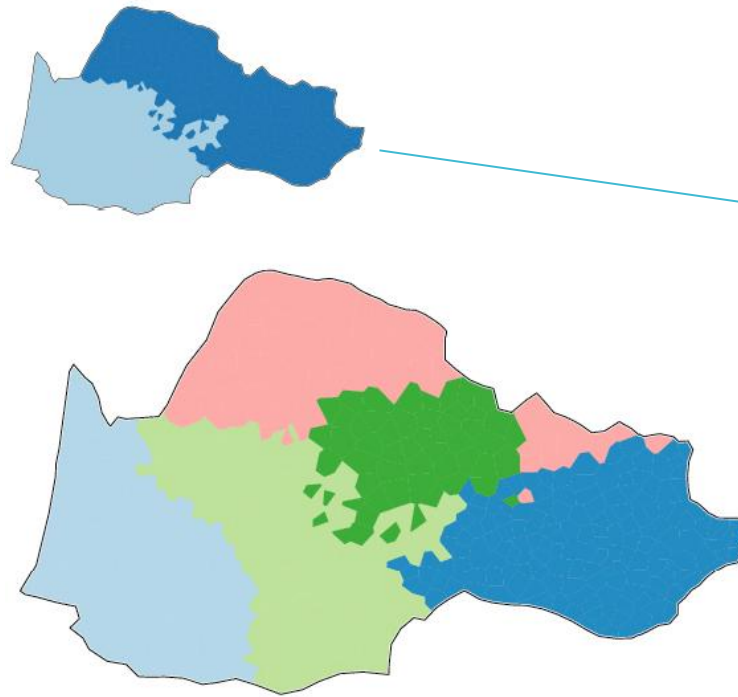
(x) = *Multiplicans*, (-) = *Subtrahens*,  
 (+) = *Addens*, (./.) = *Dividens*



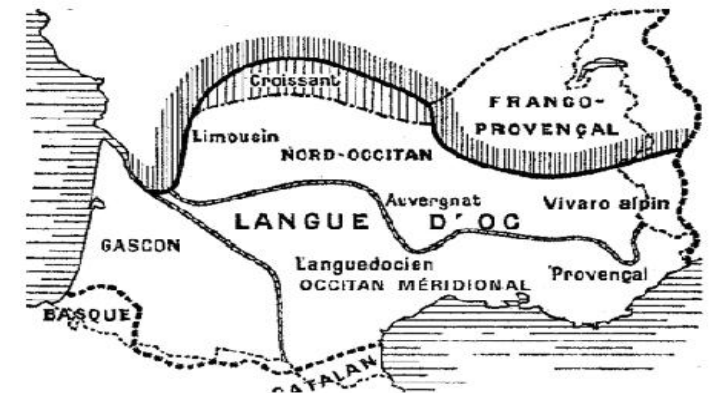
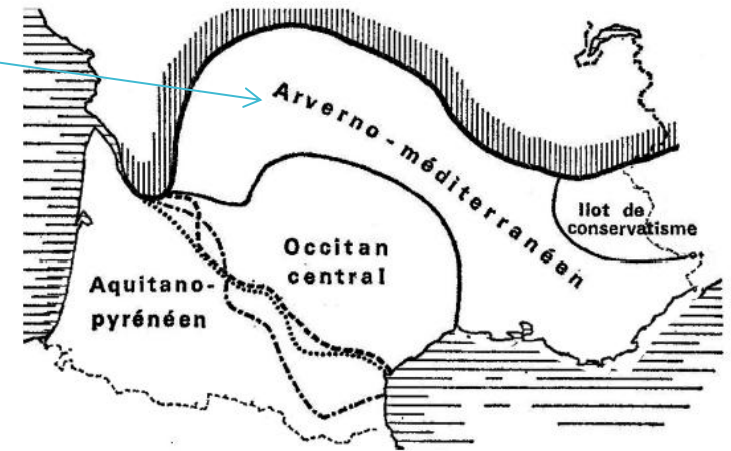
*Substituens* ↔ (x) = *Multiplicans*

## 2.1. Geolinguistic complexity: Occitan areas and choremes

Right: Pierre Bec  
Occitan Macrodialects  
& Dialects  
(isoglosses-driven),  
based on  
*betacism*  
*palatalizations*  
(of onset and coda)  
vs  
Ward's method,  
Phonology (71 items,  
THESOC), 2-5 classes



Gabmap, carte et dendrogramme :  
fonction *Ward's Method*, 5 classes,  
données phonologiques  
(71 cartes)

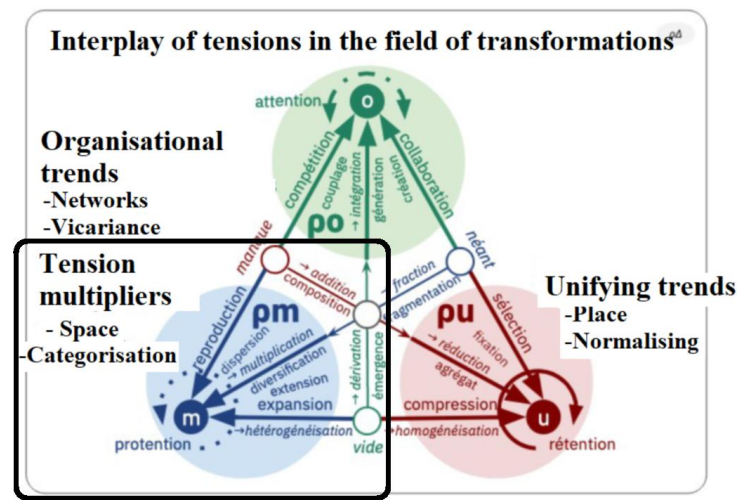
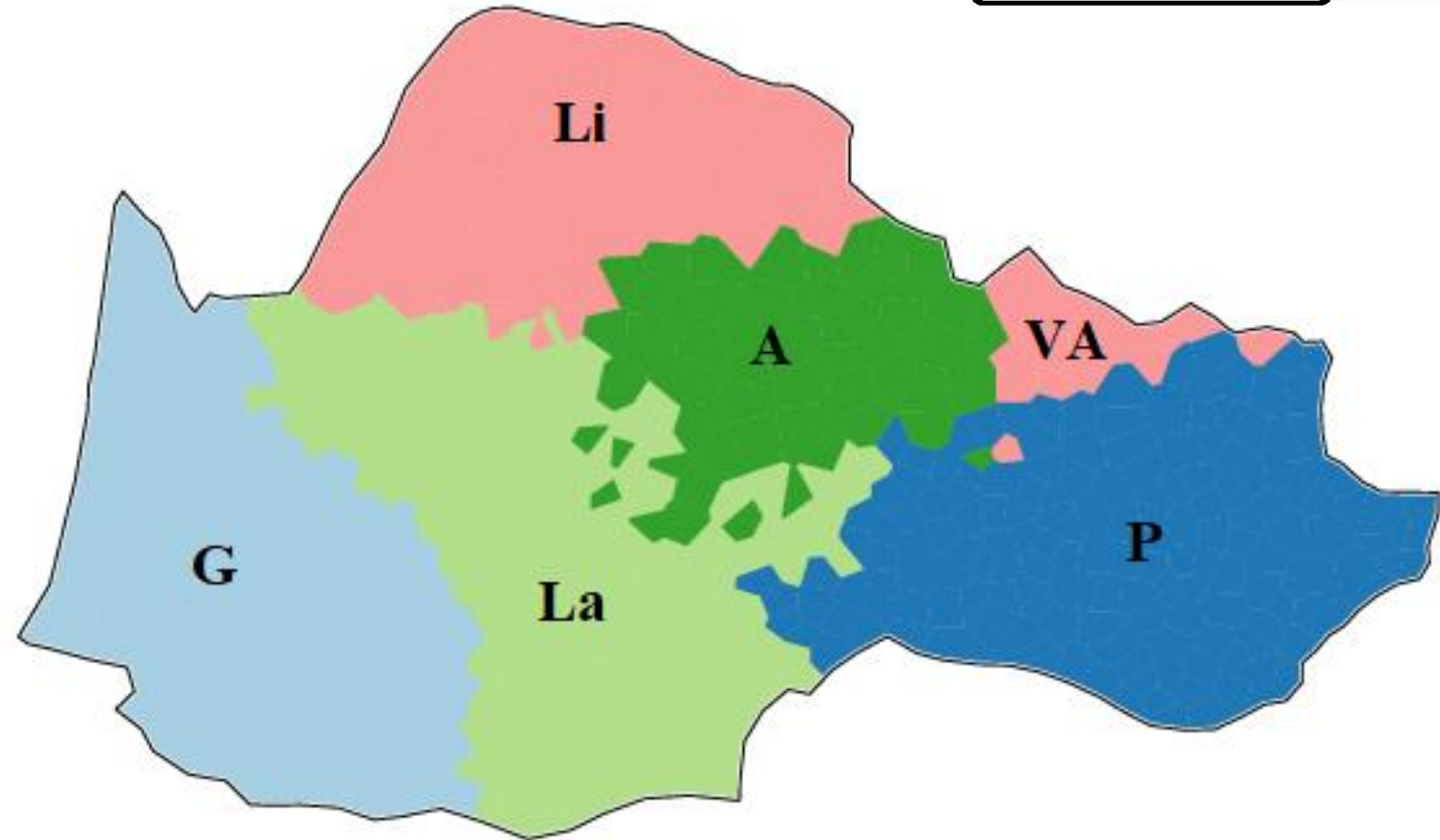


Bec (1973)

## 2.1. Geolinguistic complexity: Occitan areas and choremes

Phonology,  
Ward's method,  
5 classes

Li = Limousin  
VA = Vivaro-Alpin  
A = Auvergnat  
P = Provençal  
La = Languedocien  
G = Gascon

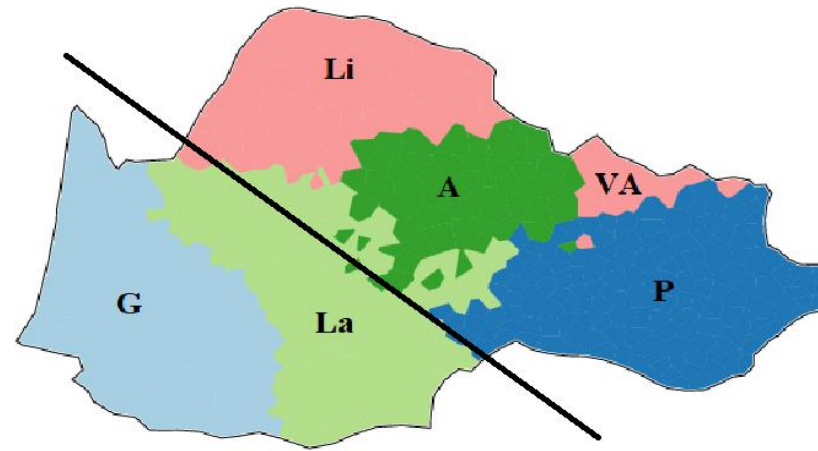




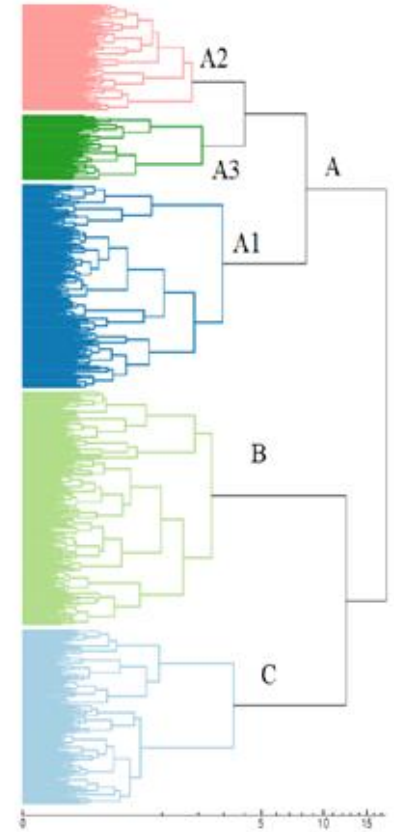
## 2) Geolinguistic complexity: Occitan areas and choremes

Ward's  
method,  
Phonology, 2-5  
classes

Gabmap, carte et dendrogramme : fonction *Ward's Method*, données phonologiques (71 cartes)



**Display, composition**

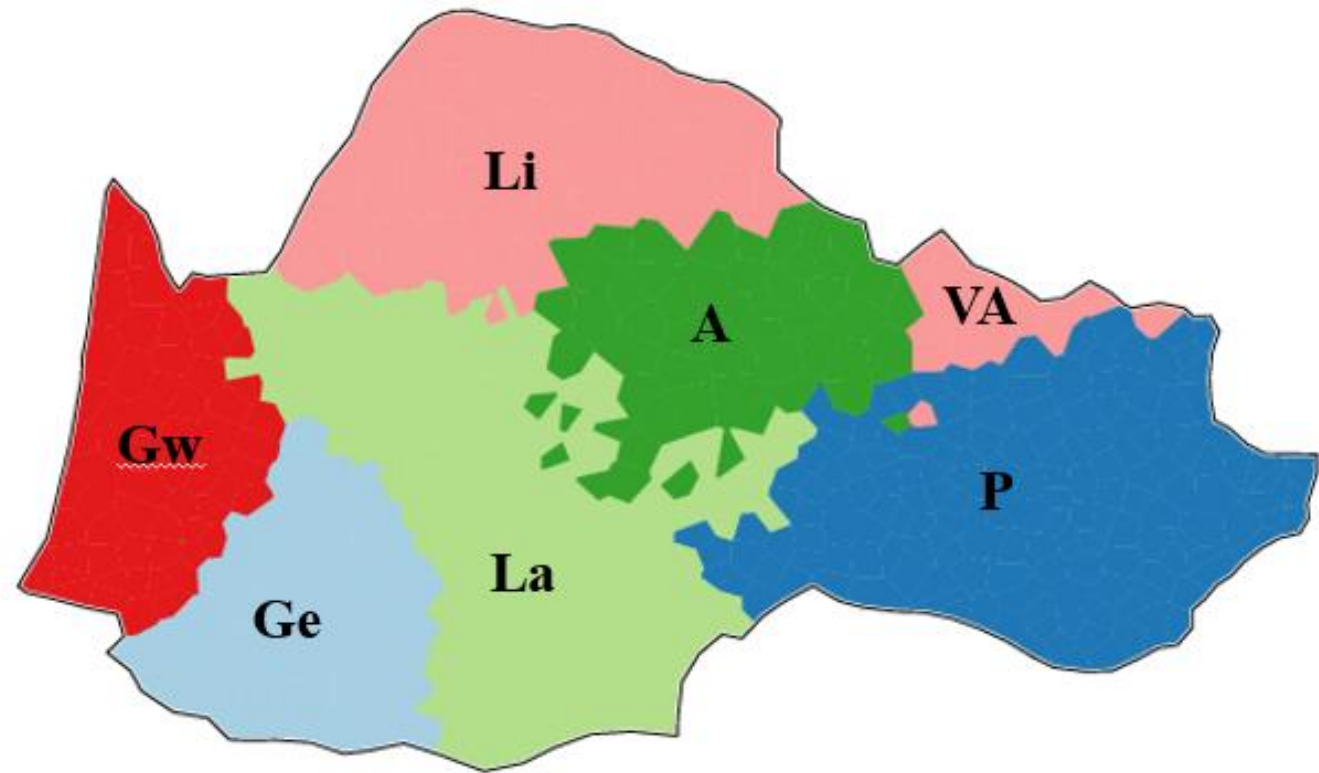


**Hierarchy, configuration**

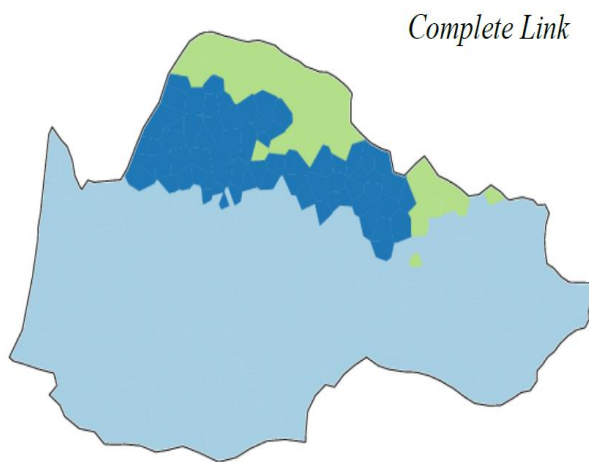
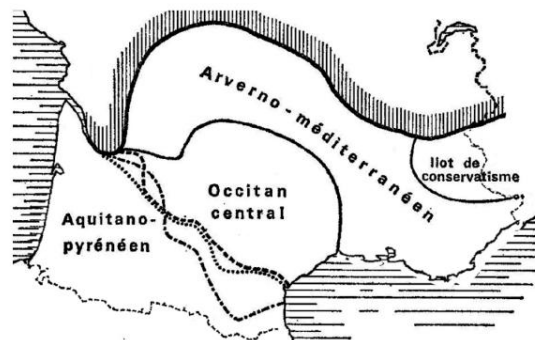
## 2) Geolinguistic complexity: Occitan areas and choremes

Phonology, Ward's  
method,  
6 classes

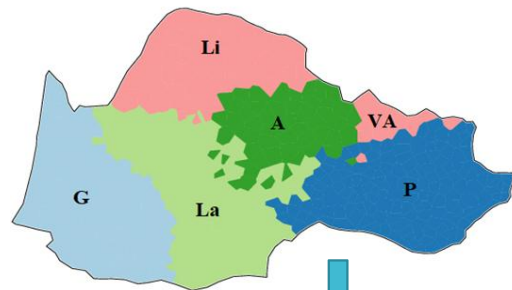
(& further abbreviations, if  
necessary  
=> Gw = Western Gascon,  
Ge = Eastern Gascon)



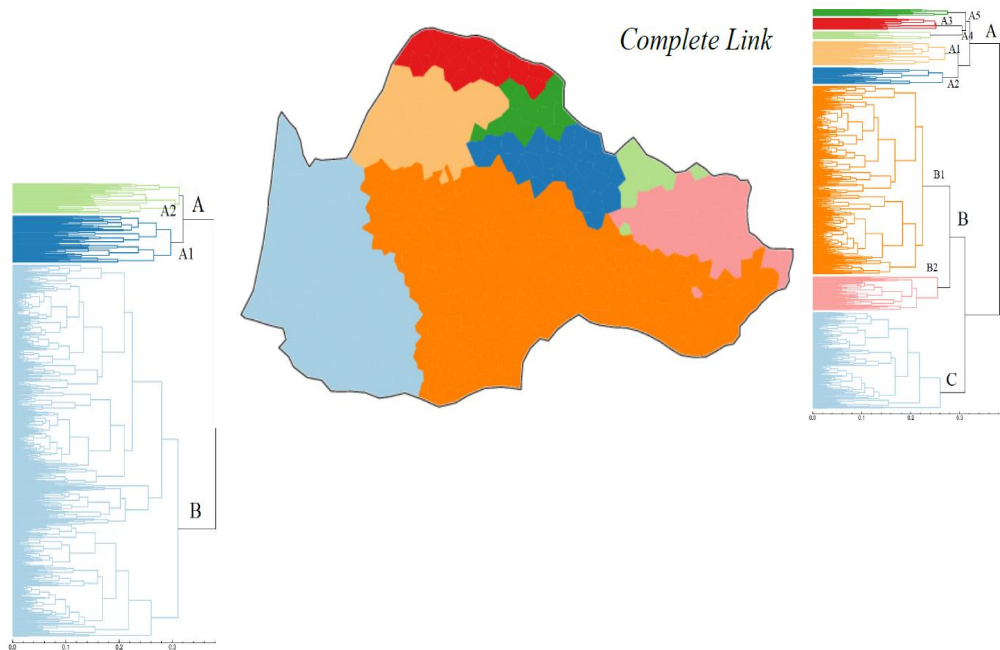
*Complete link,*  
Phonology  
3-8 classes



Ward's method



Compare *patterns*

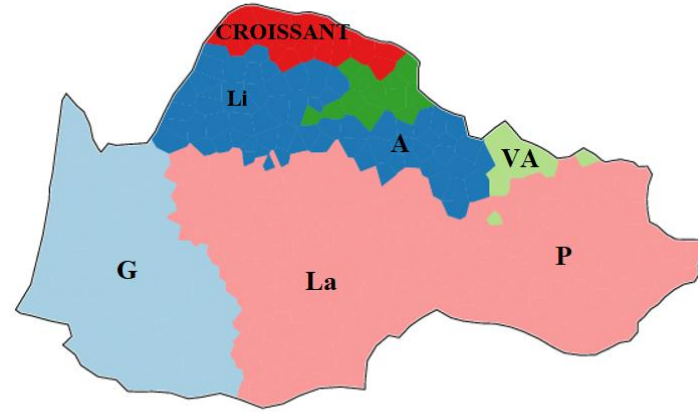




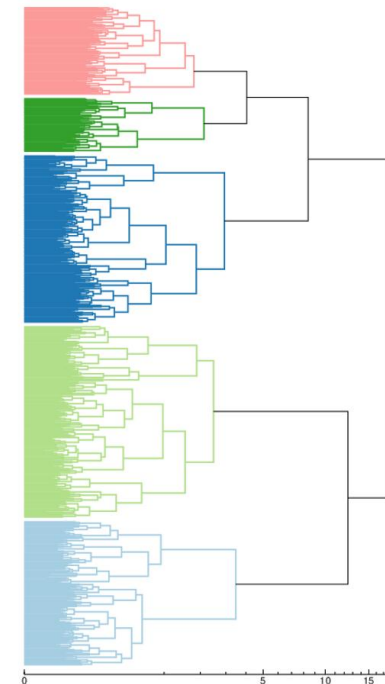
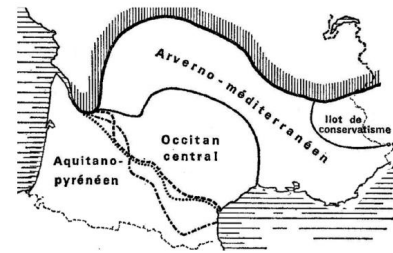
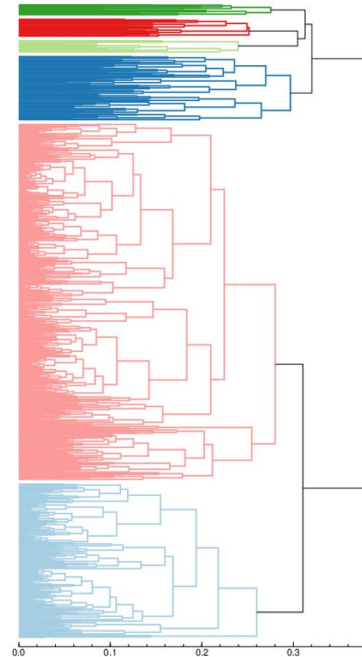
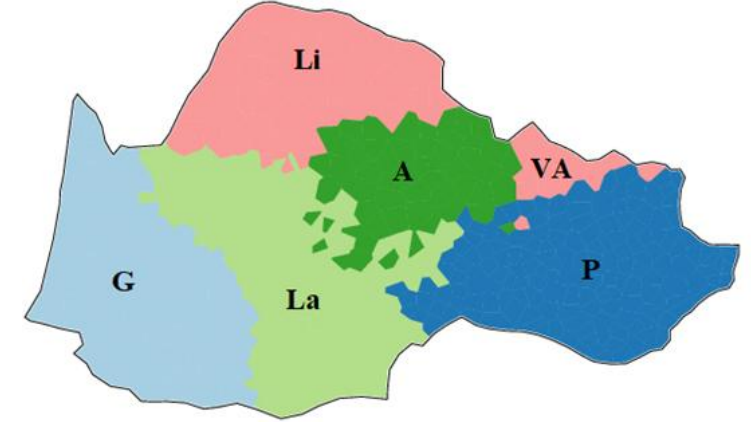
## 2) Geolinguistic complexity: Occitan areas and choremes

Compare a reductionist or *lumping* method (Complete Link) with splitting method (Ward's M)

Complete link, 6 classes



Ward's method



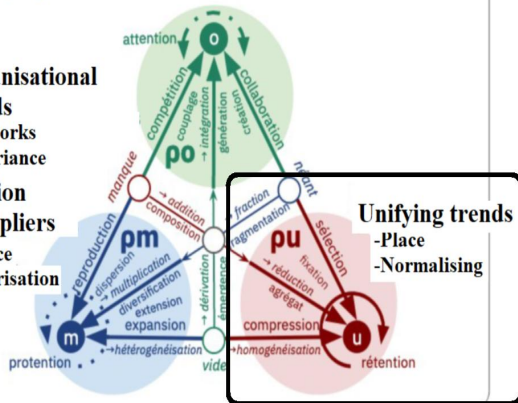
## 2.1. Geolinguistic complexity: Occitan areas and chormes

More  
reductionism:  
Group average  
Phonology,  
2-4 classes

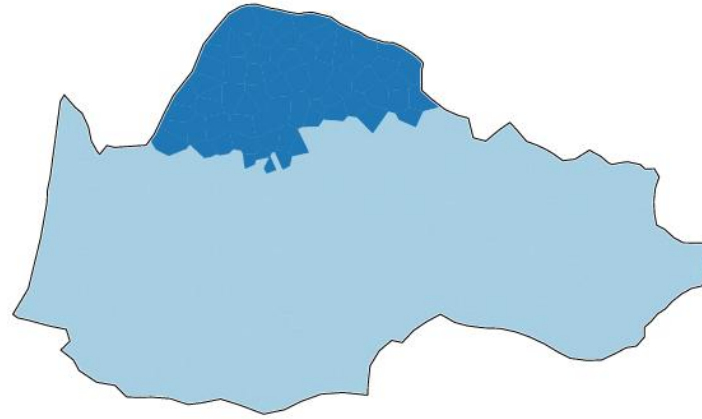
Interplay of tensions in the field of transformations<sup>4</sup>

Organisational  
trends  
- Networks  
- Vicariance

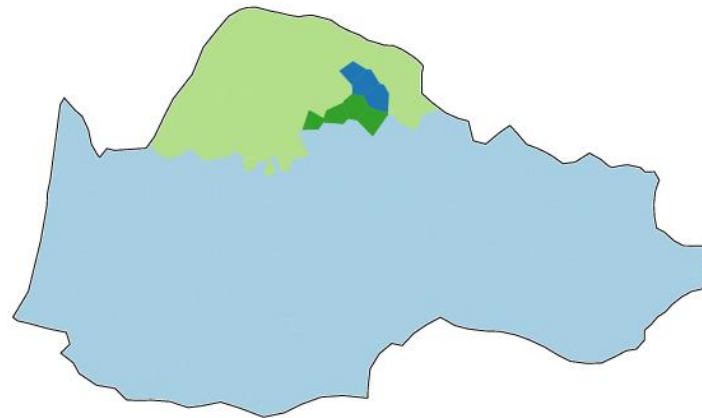
Tension  
multipliers  
- Space  
- Categorisation



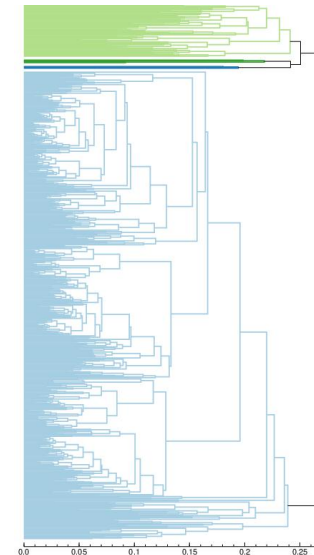
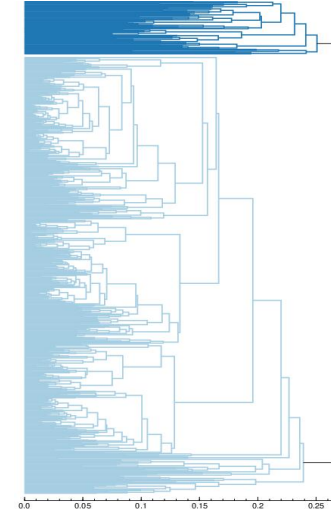
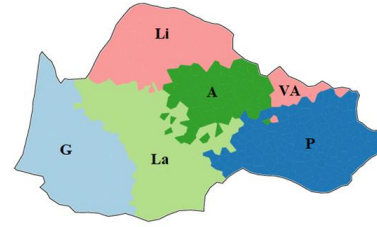
Group average 2 classes



Group average 4 classes

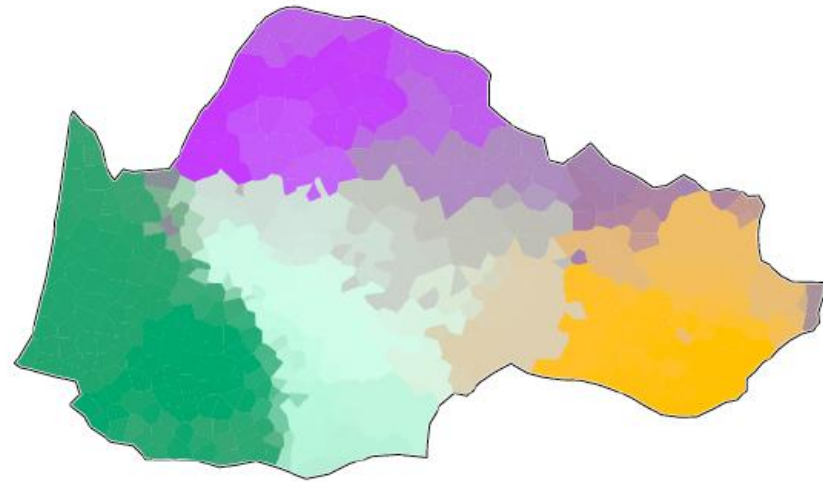


Ward's method

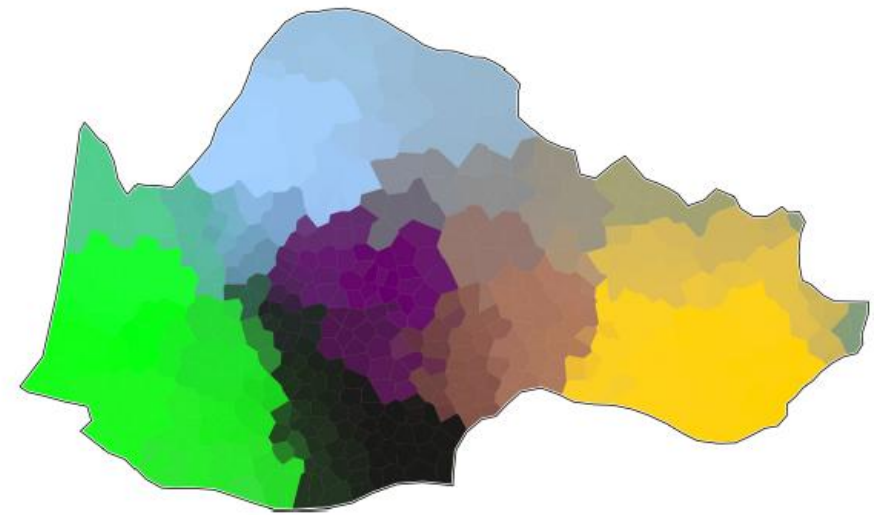


## 2.1. Geolinguistic complexity: Occitan areas and choremes

Fuzzy Cluster  
maps:  
phonology vs  
lexicon  
(THESOC)



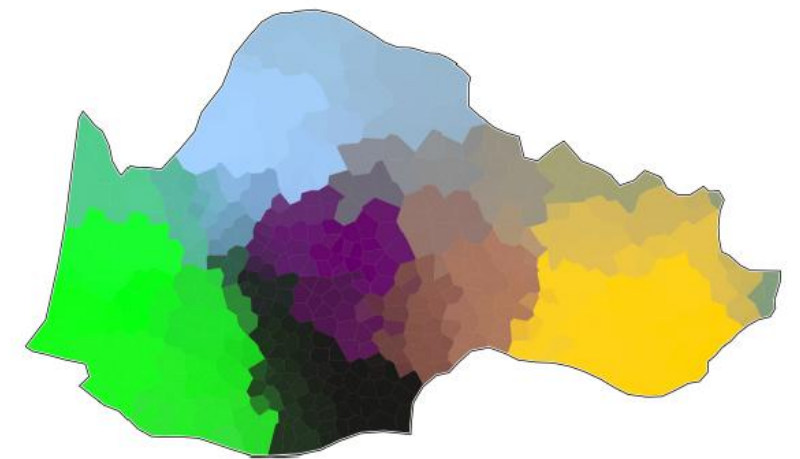
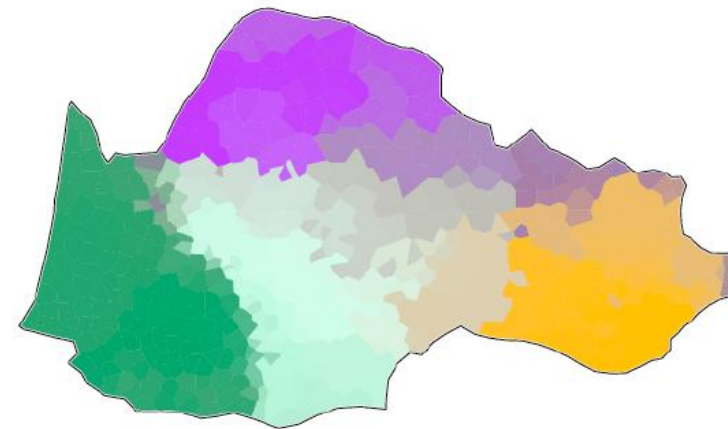
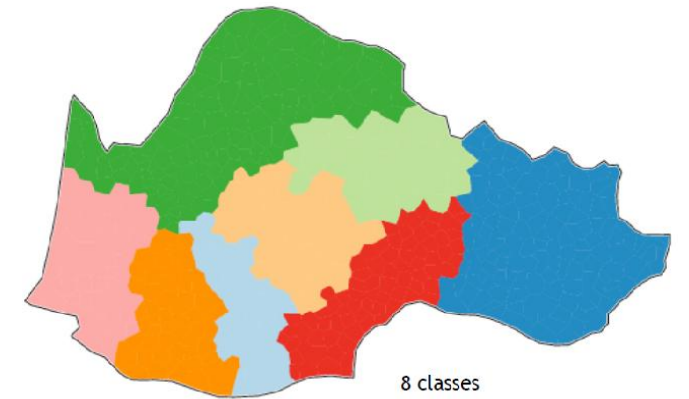
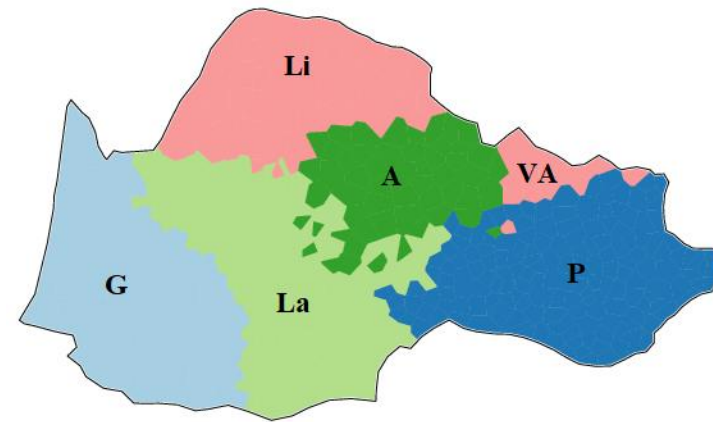
données phonologiques  
(71 cartes)



données lexicales  
(232 cartes)

Gabmap, fuzzy cluster

## 2.1. Geolinguistic complexity: Occitan areas and choremes



Gabmap, fuzzy cluster

données phonologiques  
(71 cartes)

données lexicales  
(232 cartes)

Fuzzy Cluster  
maps:  
phonology (left  
side) vs lexicon  
(right), below

Compare with Ward's  
method (above)

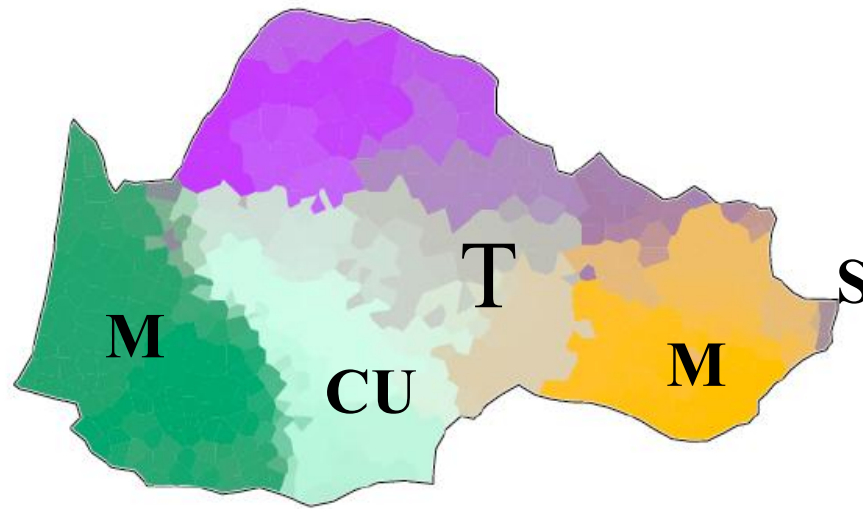


## 2.1. Geolinguistic complexity: Occitan areas and choremes

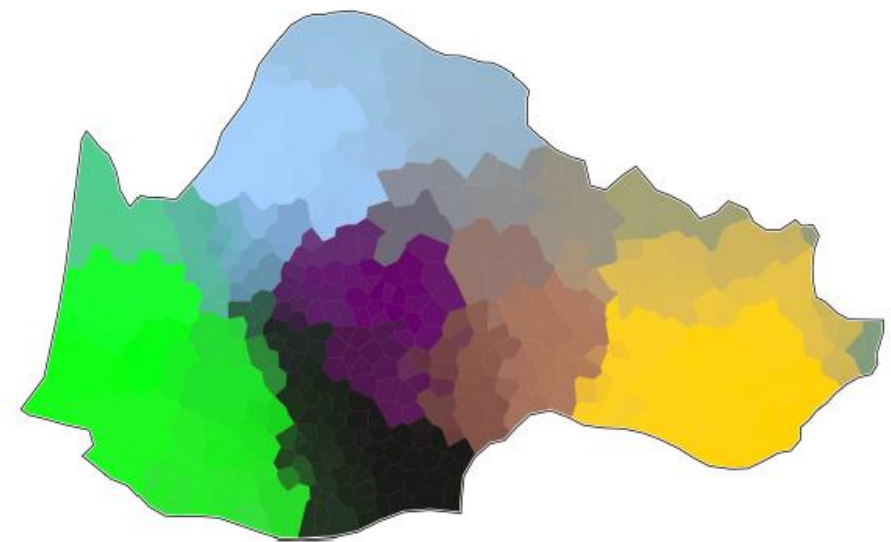
The 'don Ricardo's Model'  
(Modality vs Transition)

=> M vs T  
UC = Central Unit

See Dubert-Garcia (2021)  
=> **Modalities (M)**,  
**Central Unit (CU)**,  
**Transitions (T)** &  
**Outliers or Singletons (S)**  
ou micro-aïres alloglottes



données phonologiques  
(71 cartes)

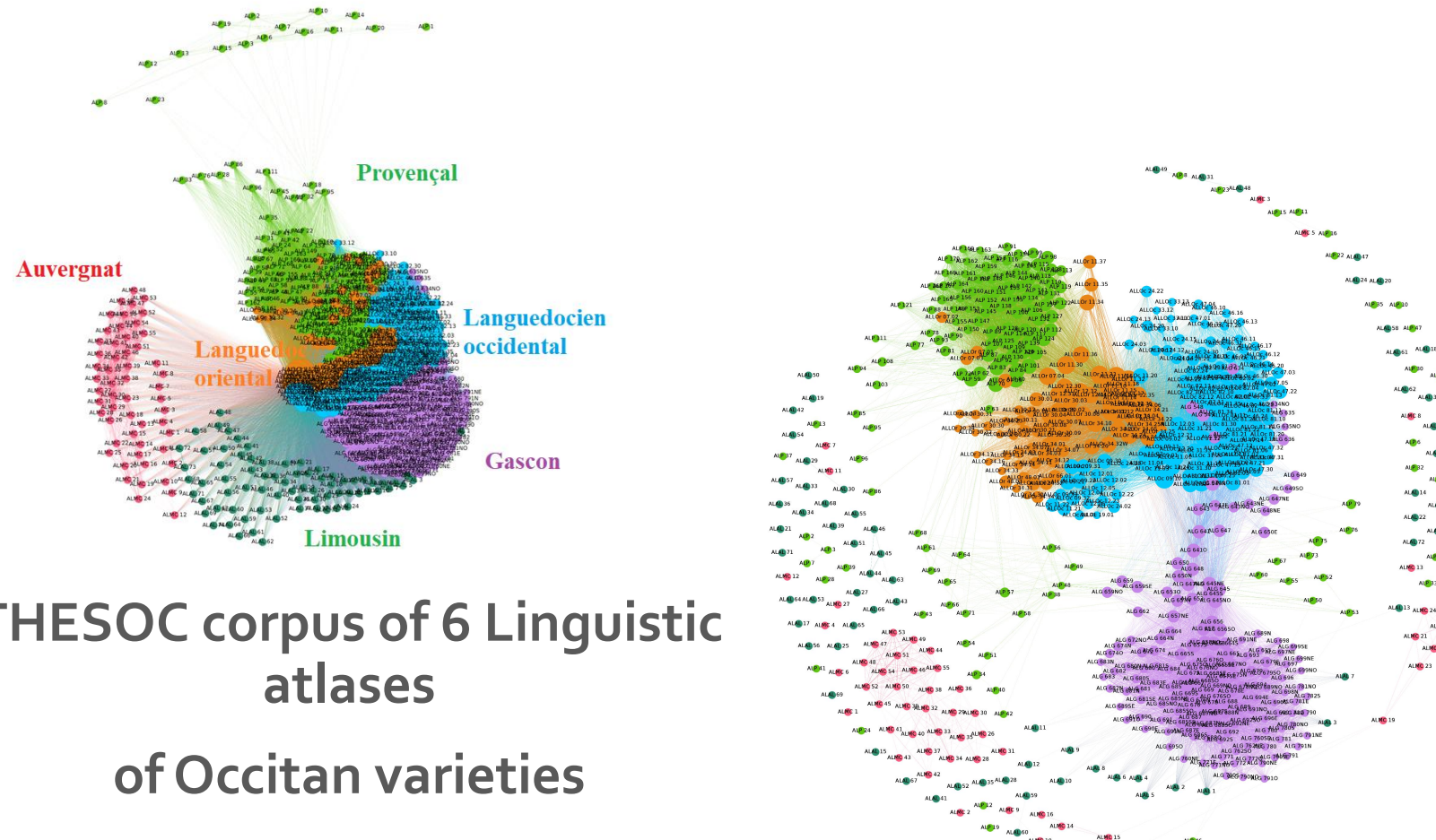


Gabmap, fuzzy cluster

données lexicales  
(232 cartes)

# 2.2. Dynamic networks (Gephi)

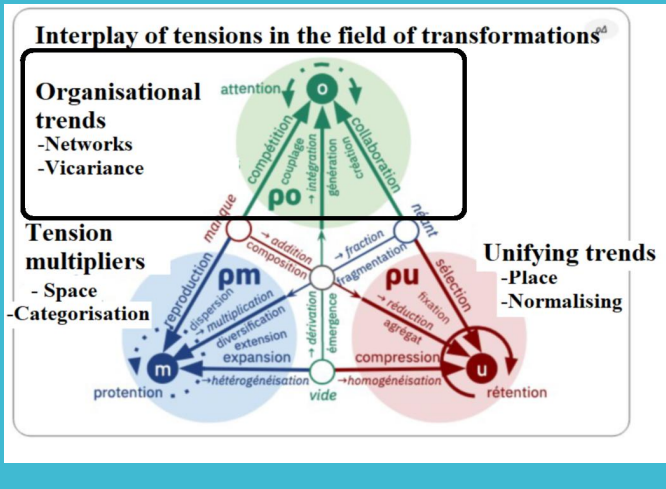
A network approach,  
using *Gephi*  
as a heuristic tool:  
streams of graphs,  
self-organization trends



THESOC corpus of 6 Linguistic  
atlases  
of Occitan varieties  
(ALG, ALP, ALLOc, ALAL, etc.)

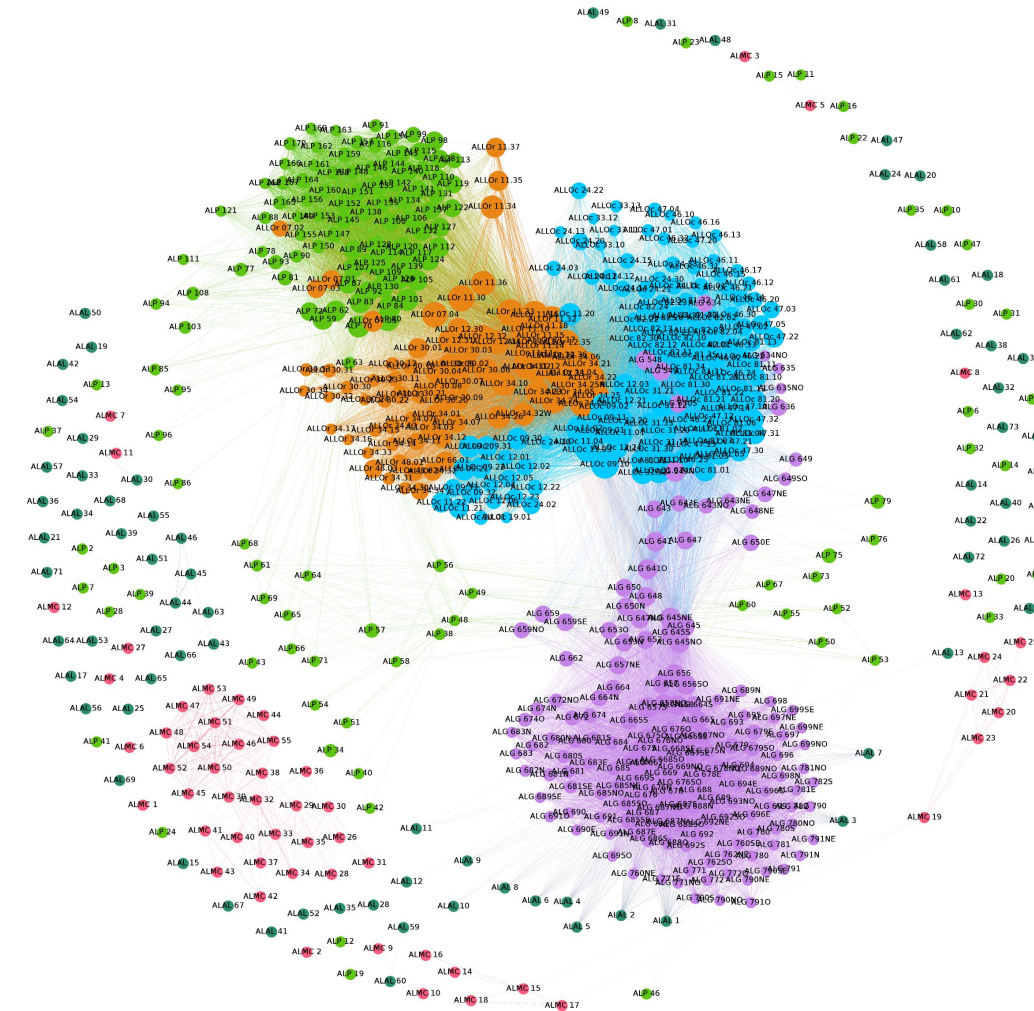
## 2.2) Dynamic networks (Gephi)

### 2.2. Reticular approach: normalized similarity threshold connectograms (Gephi)



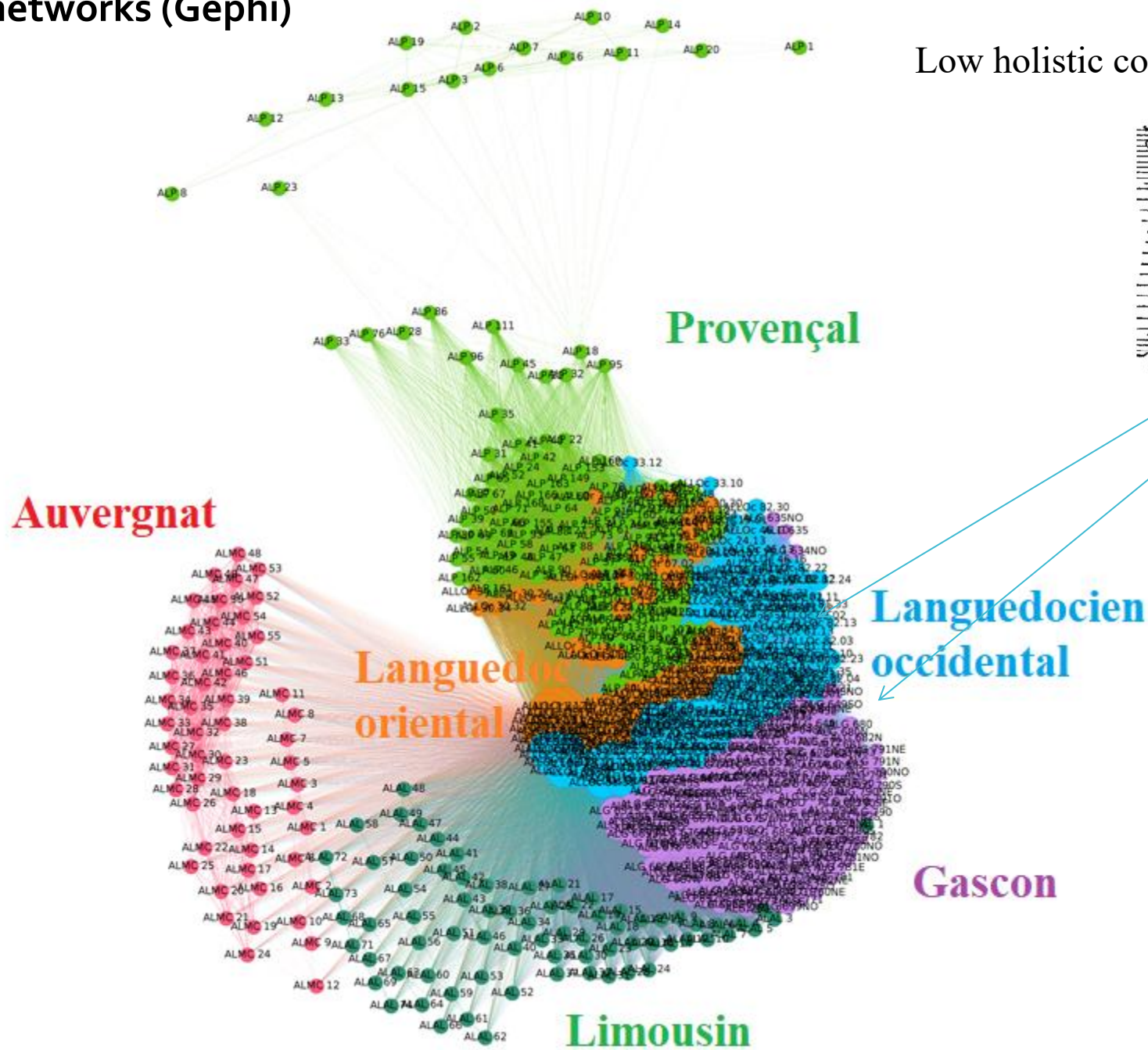
## Holistic convergence patterns

Revisiting interpointual *similarity* instead of *distance*  
(here: Index: Phonology 0.80)

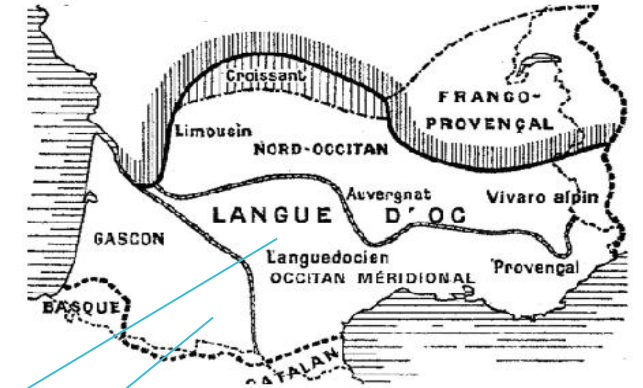




## 2.2. Dynamic networks (Gephi)



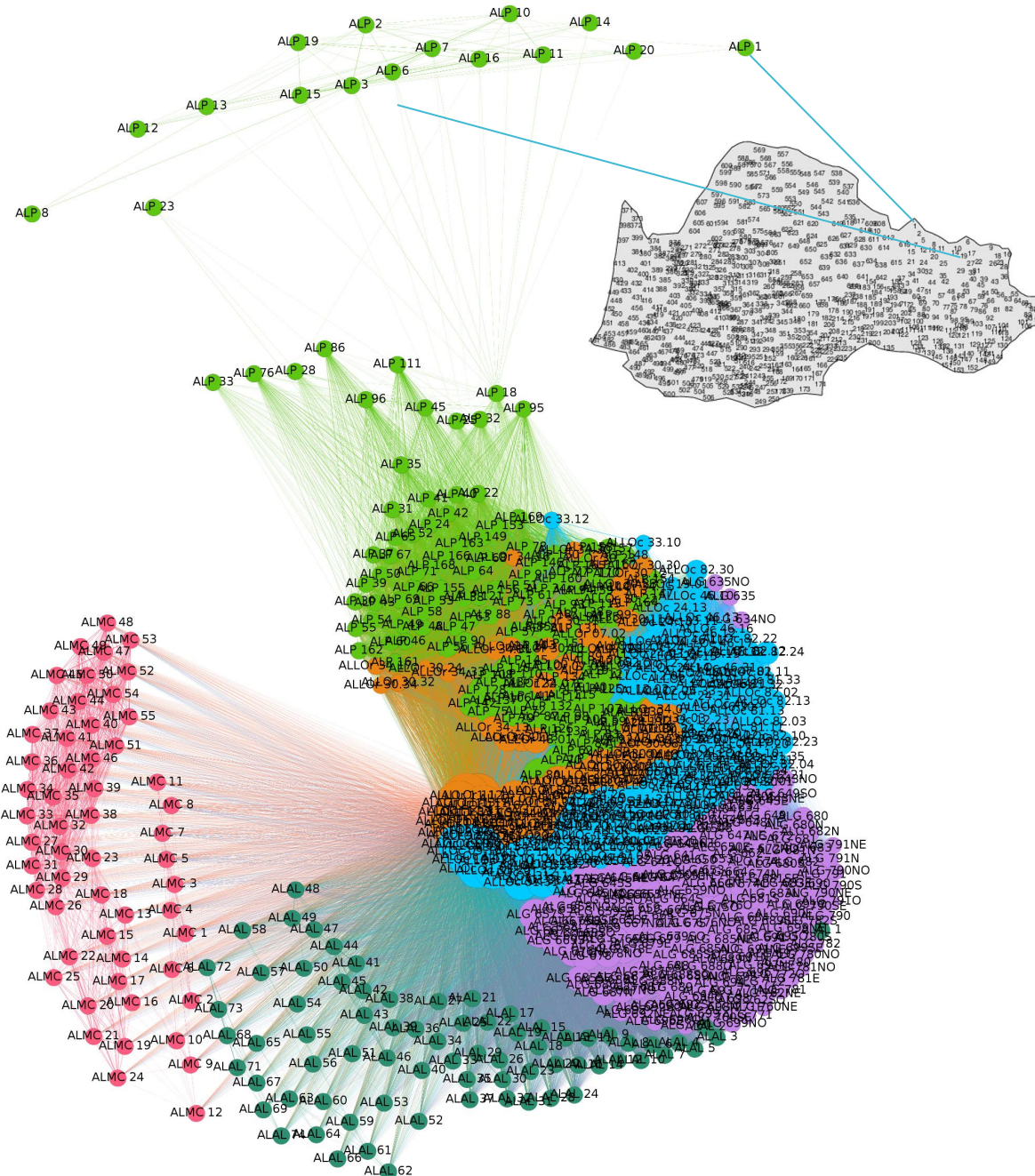
Low holistic convergence index



Bec (1973)

## 2.2. Dynamic networks (Gephi)

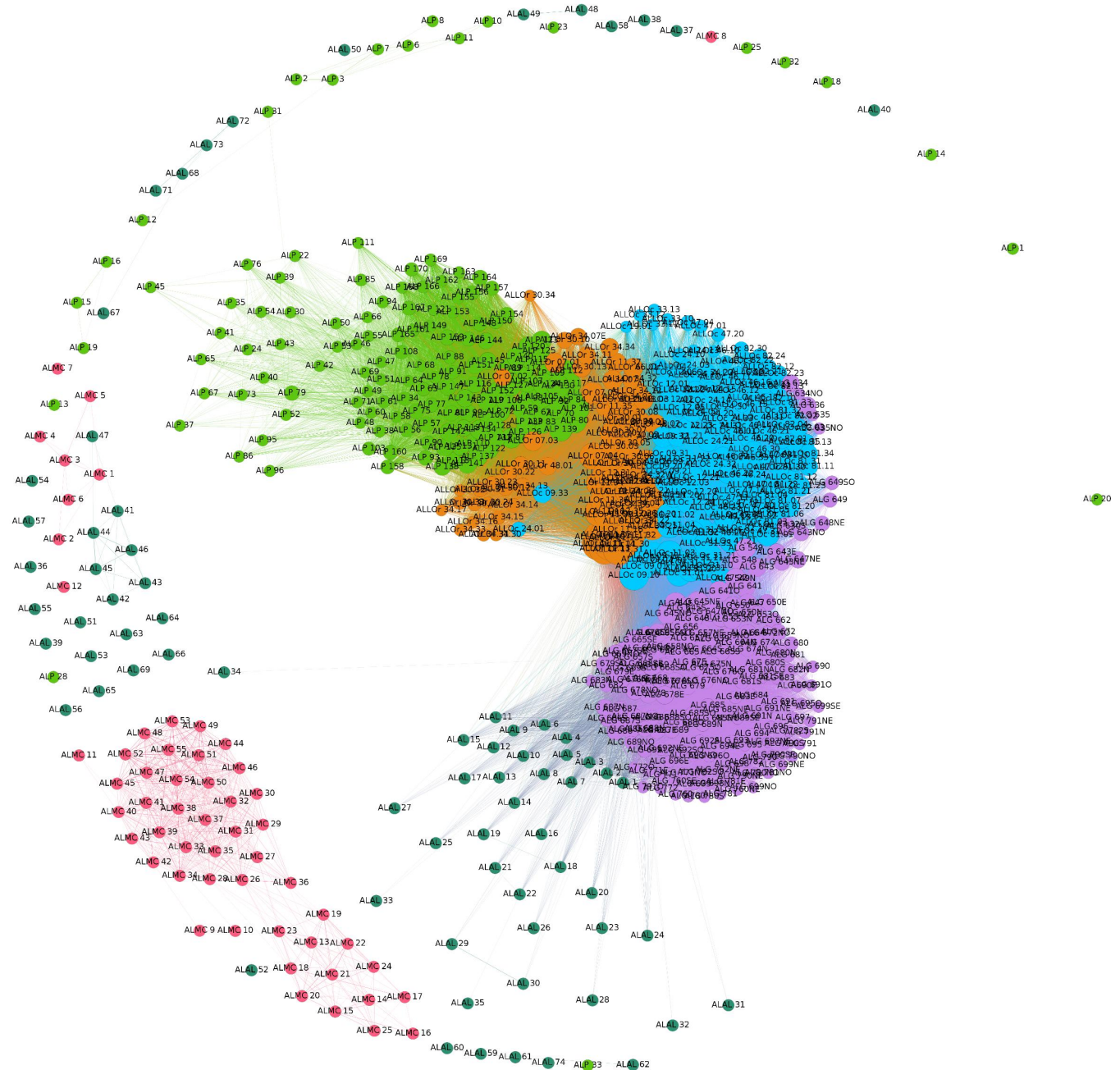
Gephi,  
phonological data,  
threshold of  
normalised  
similarity:  
0.70



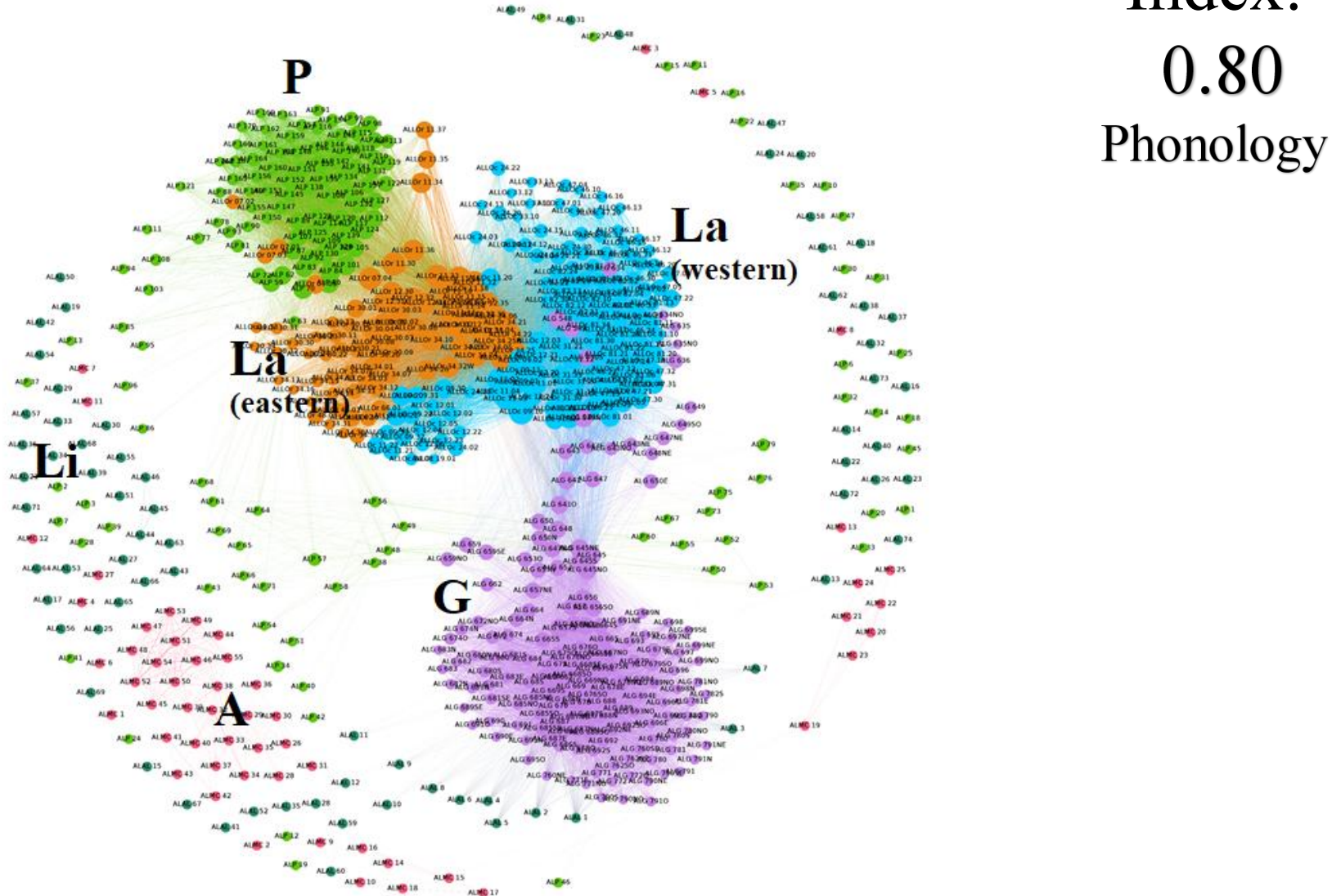
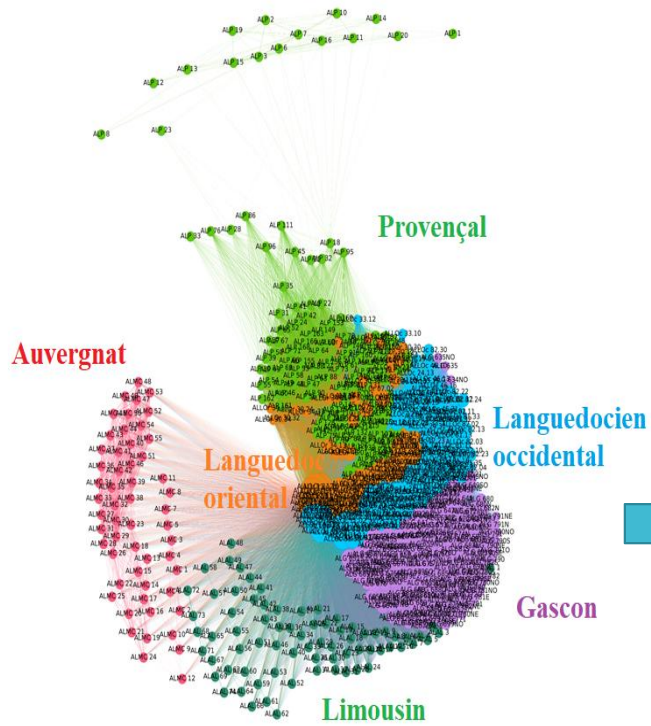


## 2.2. Dynamic networks (Gephi)

Phonology:  
0.75



## 2.2. Dynamic networks (Gephi)

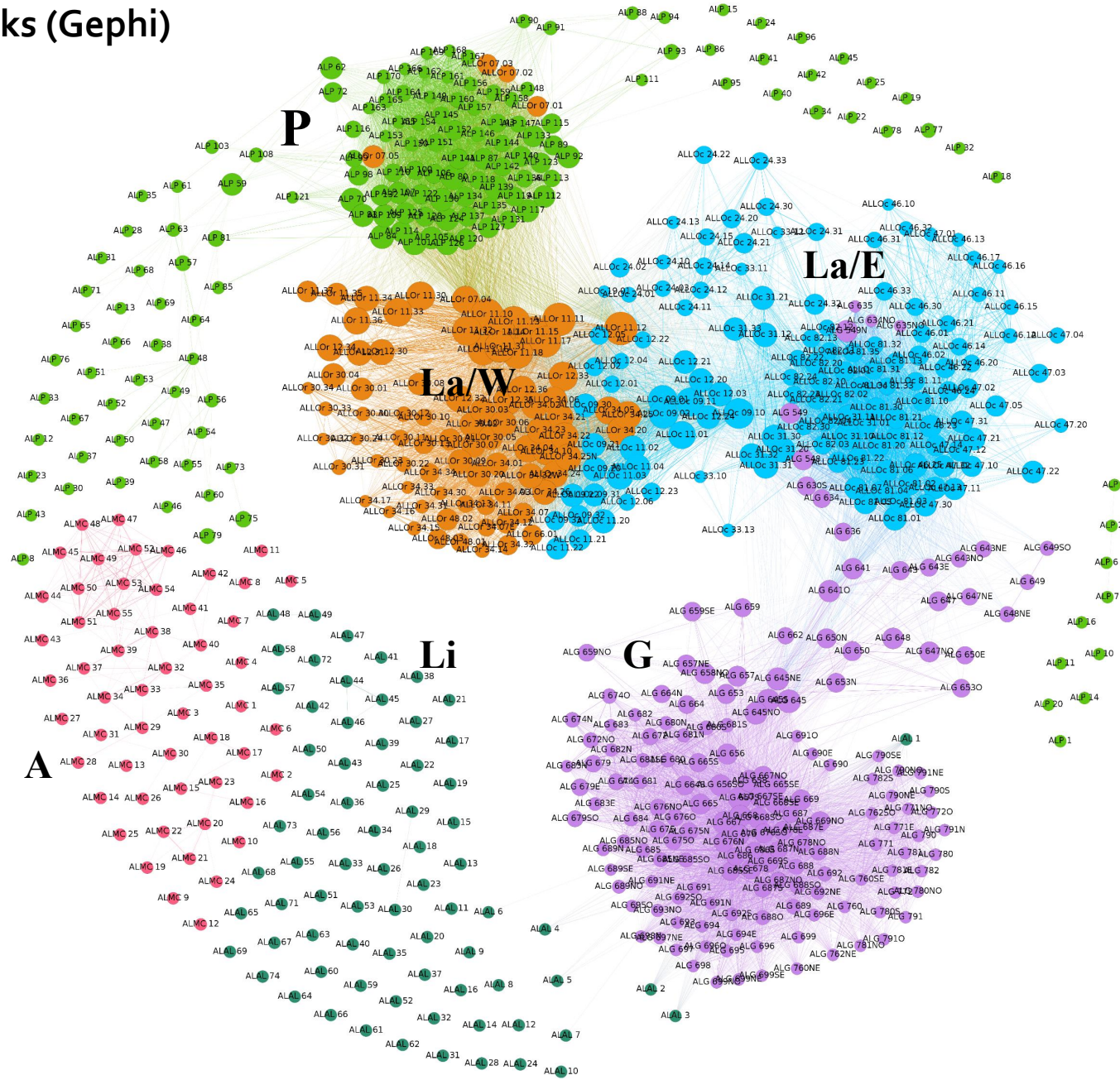


Index:  
0.80  
Phonology



## 2.2. Dynamic networks (Gephi)

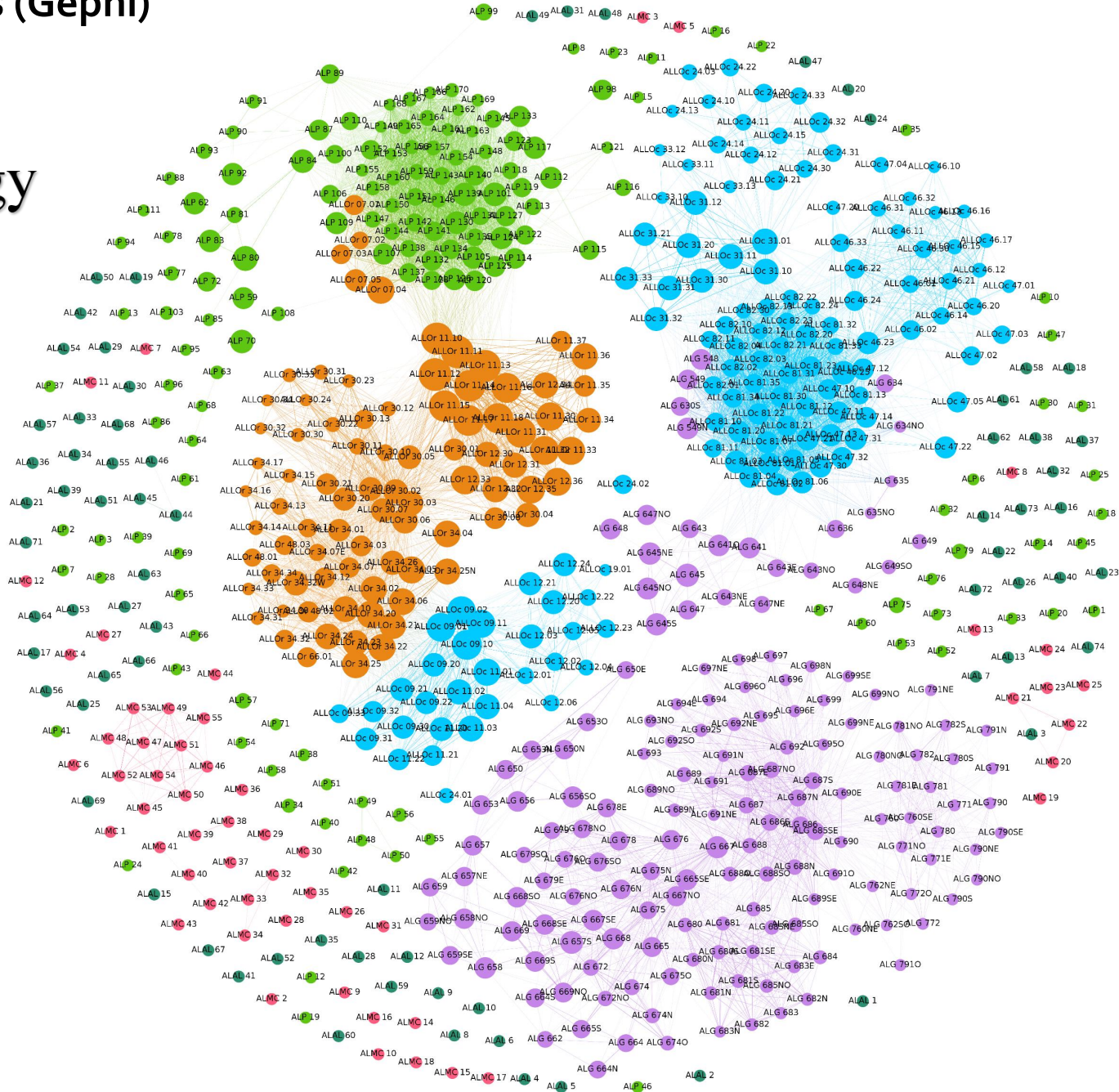
# Phonology



Structural  
convergence  
index:  
0.82

## 2.2. Dynamic networks (Gephi)

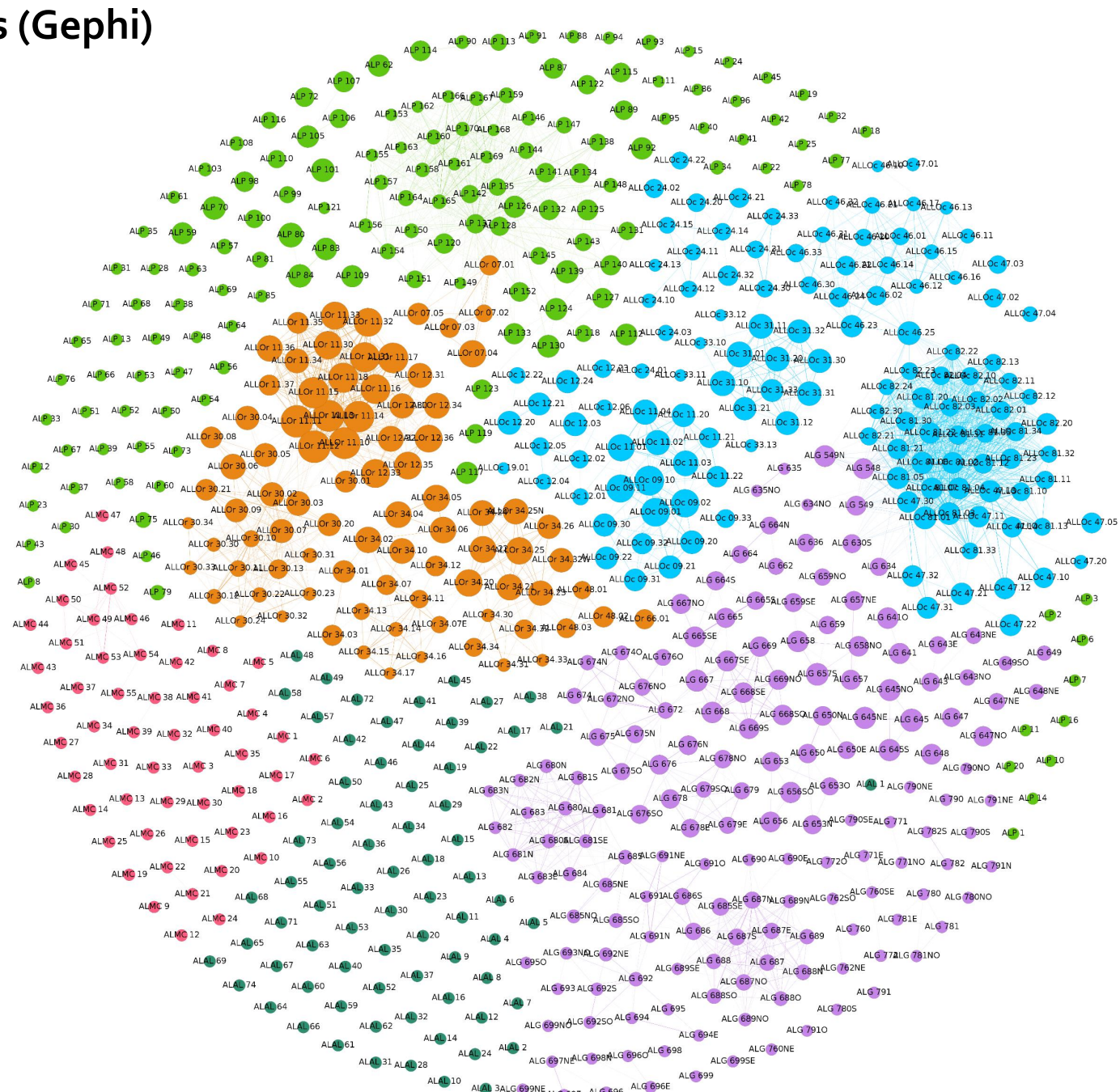
Phonology



Structural  
convergence  
index:  
0.85



## 2.2. Dynamic networks (Gephi)

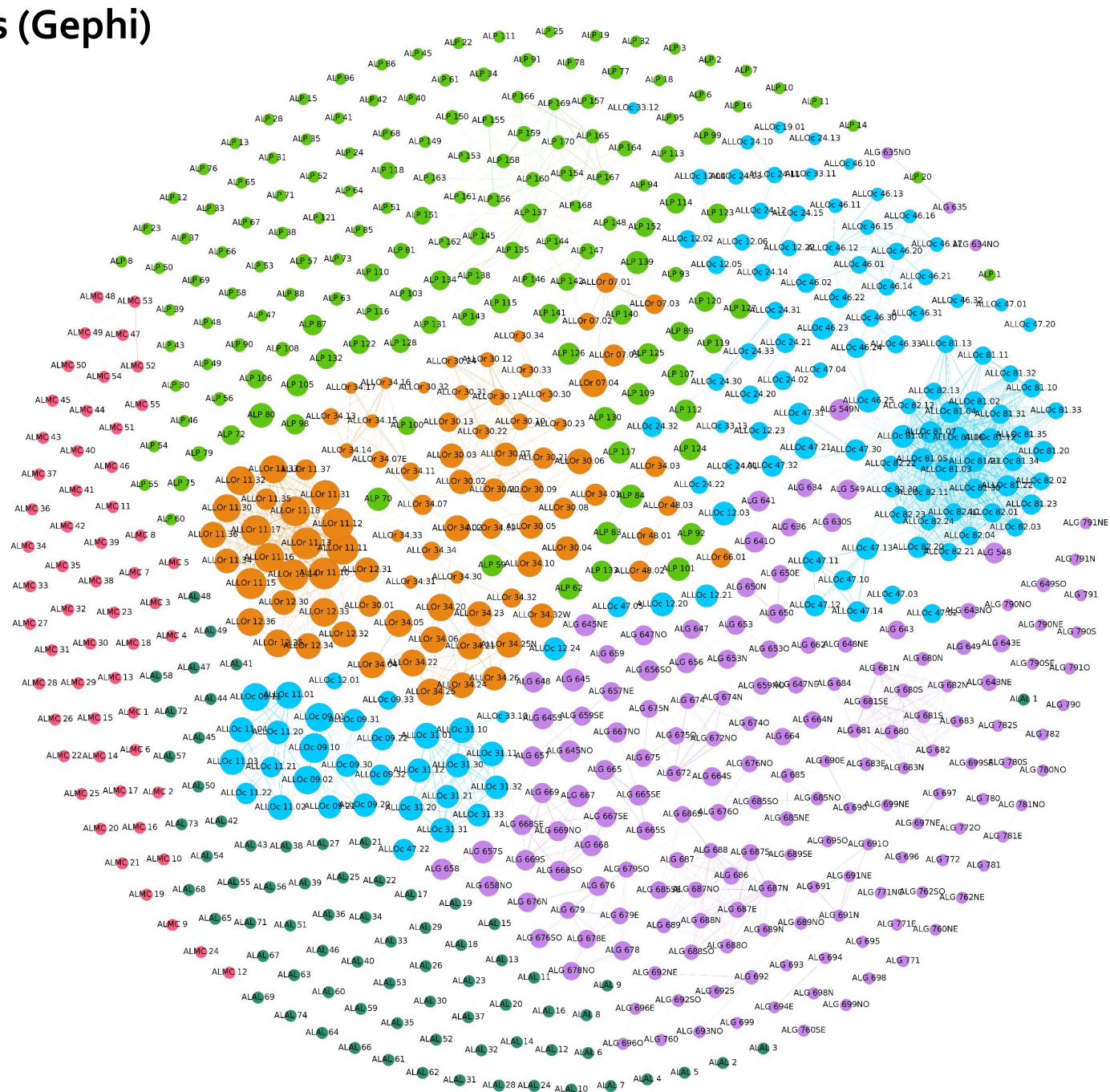


Structural  
convergence  
index:  
0.89

Phonology



## 2.2. Dynamic networks (Gephi)



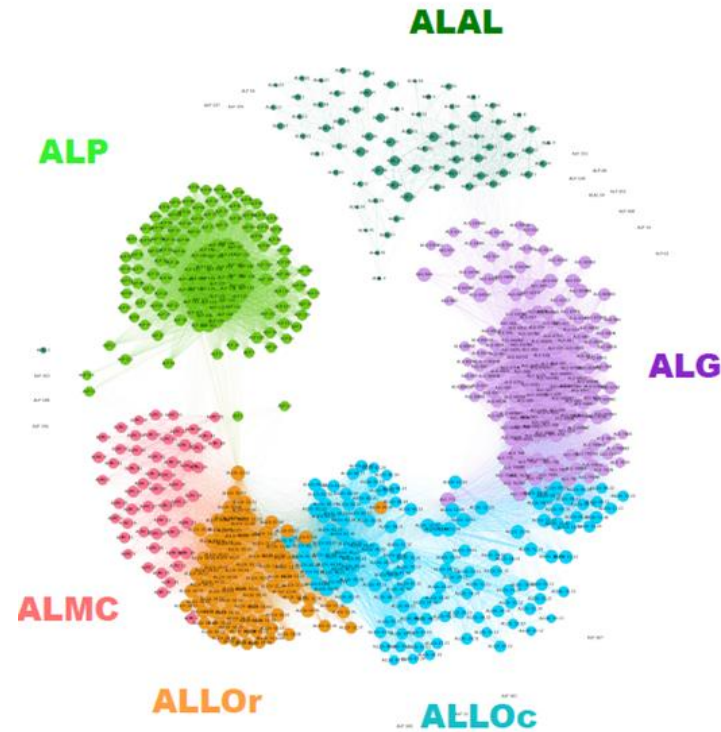
Structural  
convergence  
index:  
0.90

Phonology

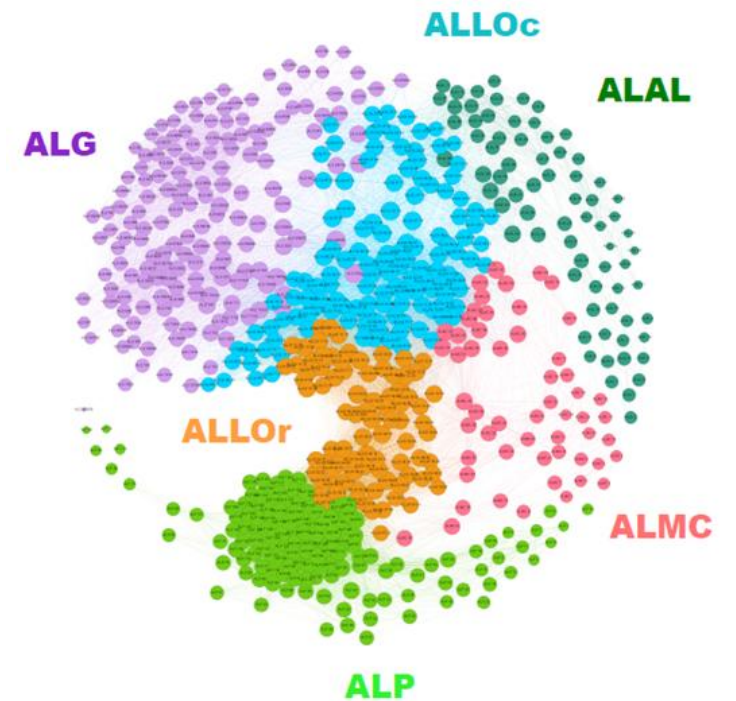
## 2.2. Dynamic networks (Gephi)

Lexicon, two  
*phases*: from  
choremes to  
areas and  
subnetworks

Lexicon: 0.65

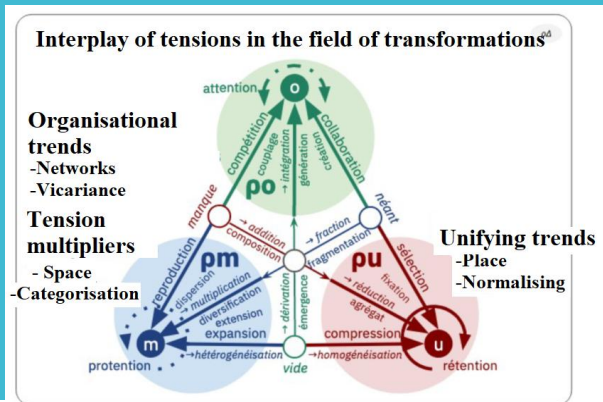


Lexicon: 0.875



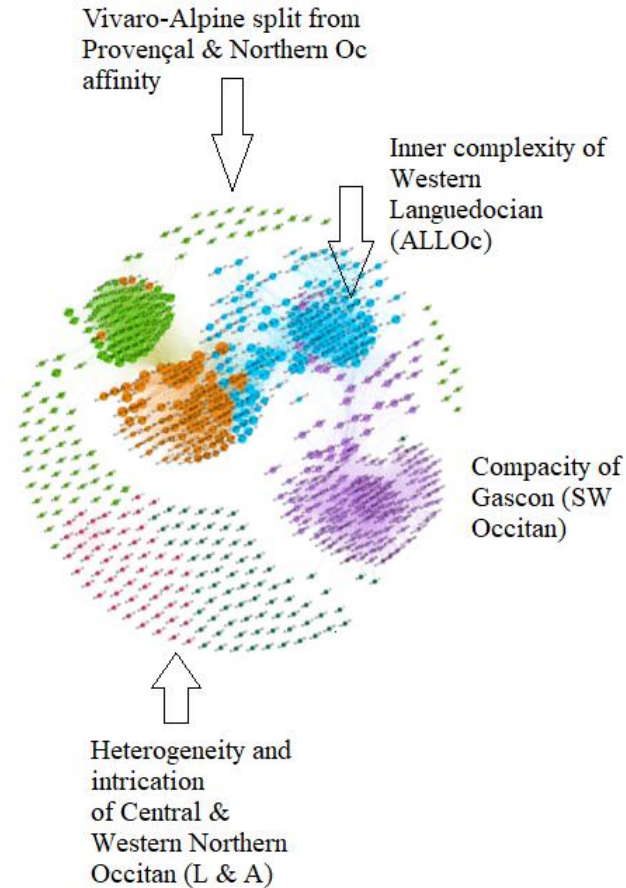
# What are the benefits of this method?

NB: gentle reminder of the Interplay of Tensions in the Field of Transformations



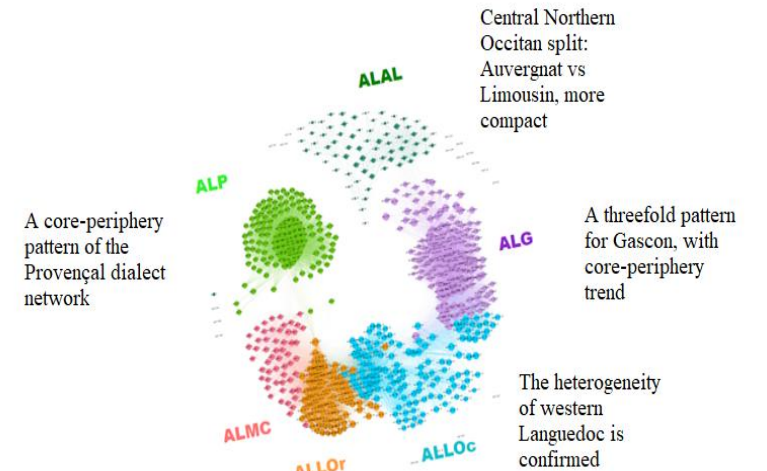
## Phonology, 71 items (THESOC)

Phonology, 0.82



## Lexicon, 232 items (THESOC)

Lexicon: 0.65





(3) A central  
dialect ( ALLOc  
data, THESOC)

Western  
Languedocian





ALLOc: a challenging linguistic atlas...

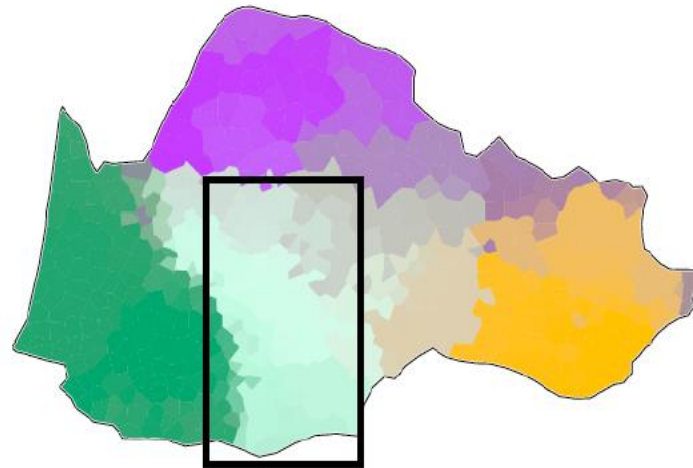
Why? a) The core of the «Central Unit»

b) **Modality** in the South, **Transitional** in the North

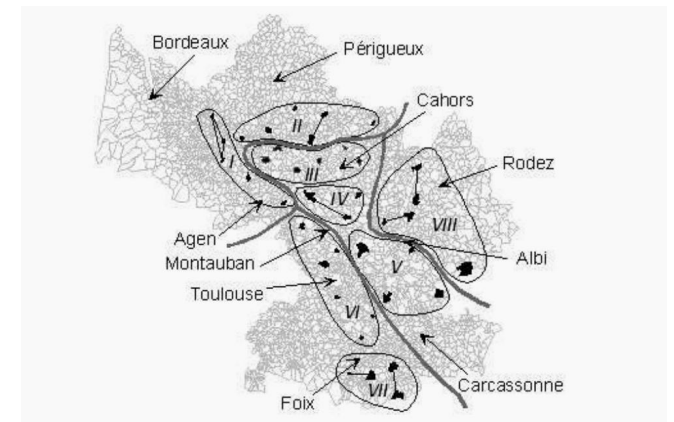
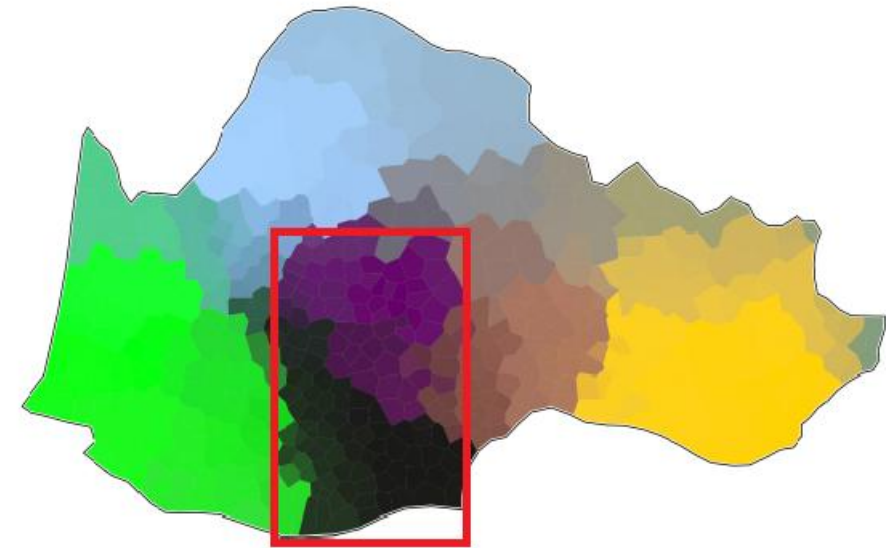
c) A fullhouse of **endemisms** in the transitional part

d) Several Town dialects in the past, among which...  
Toulouse, Albi, Cahors, Agen, Montauban, Rodez, Foix

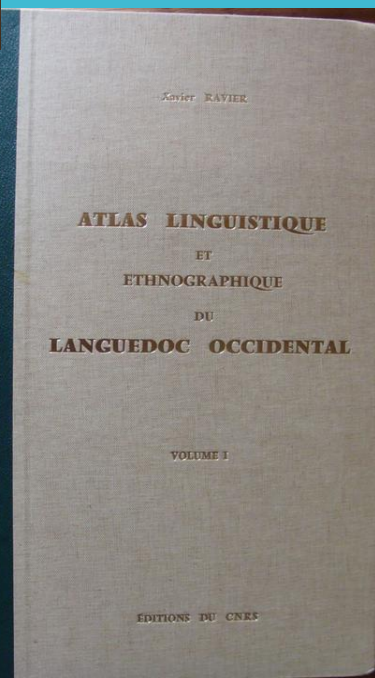
### Phonology, 71 maps/items



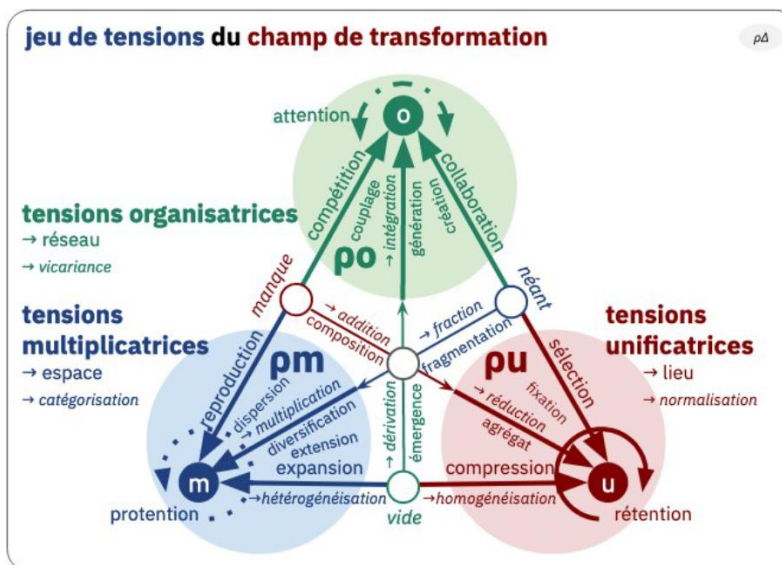
### Lexicon, 232 maps/items



A gentle reminder...  
The source (X. Ravier,  
ALLOc)

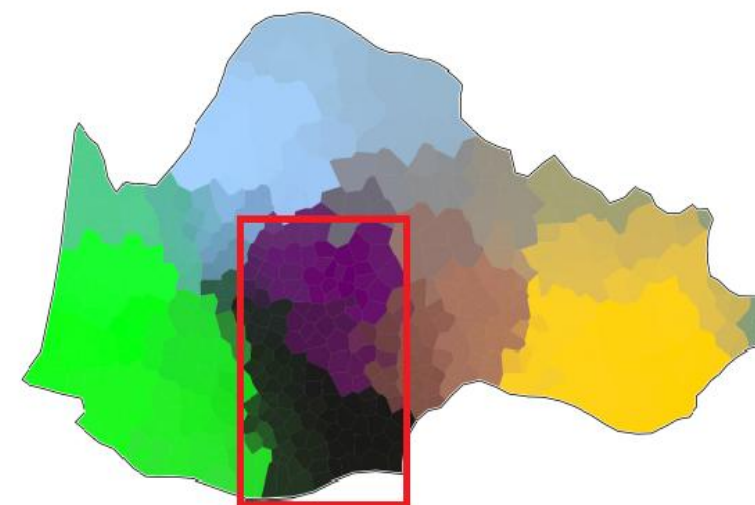
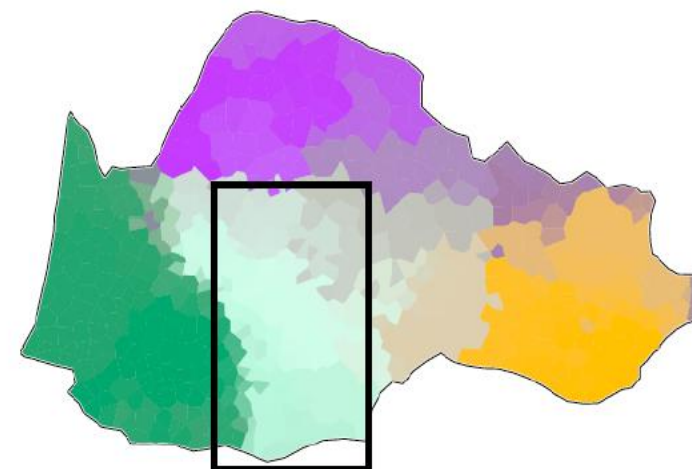


## Reminder: Lacombe's Model



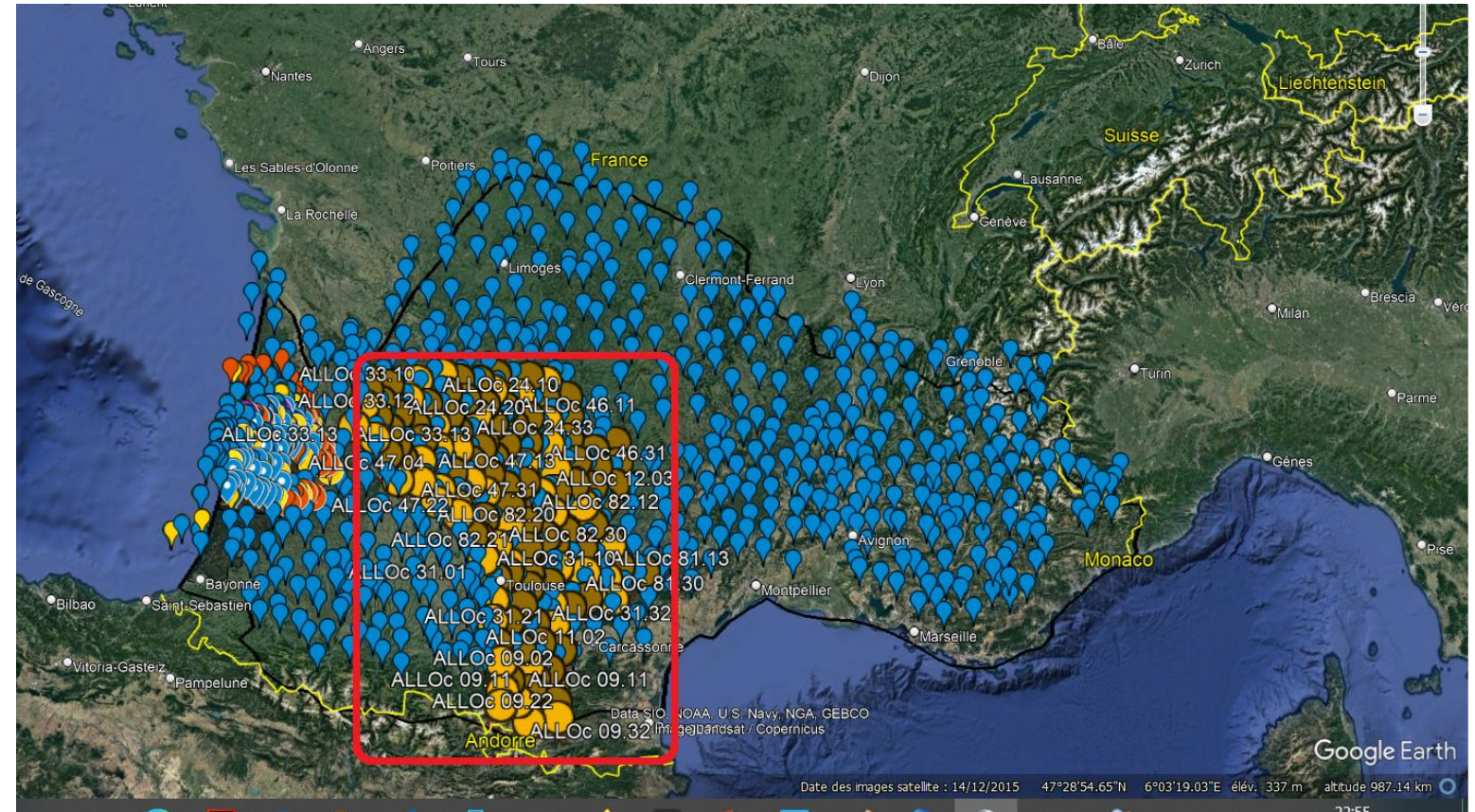
C59 — [pΔ] — jeu de tensions du champ de transformation

Phonology and lexicon in this distinguished area of ALLOc



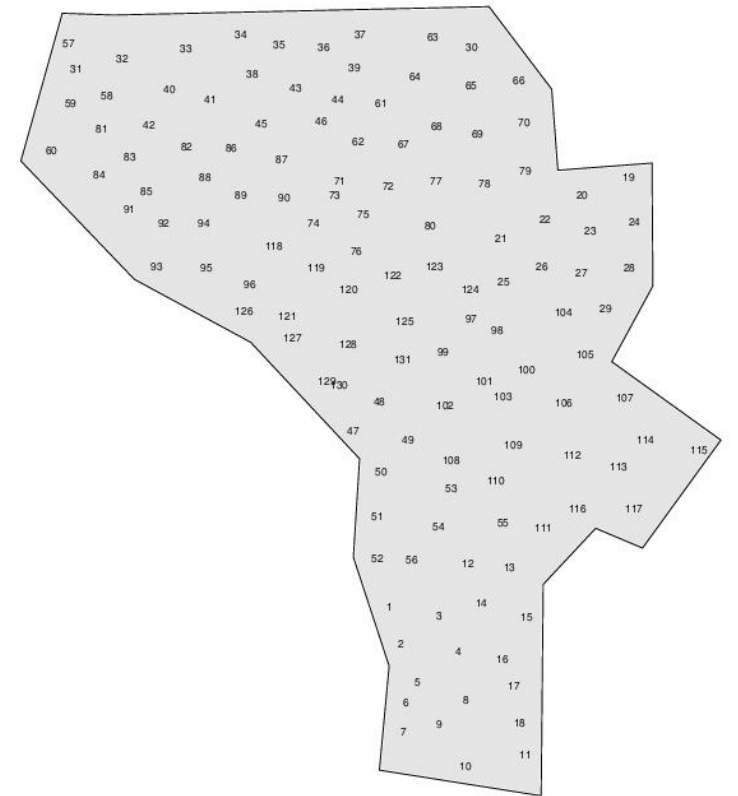
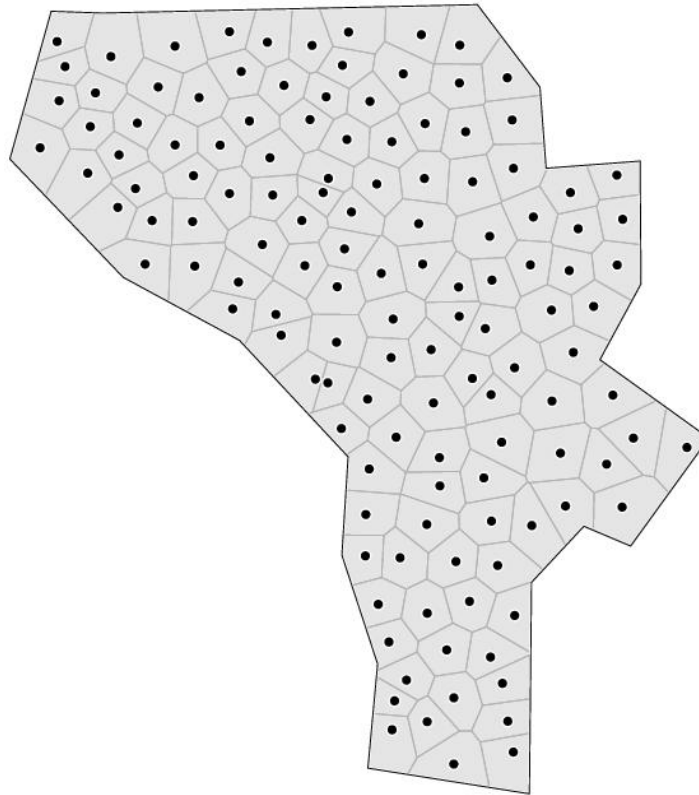
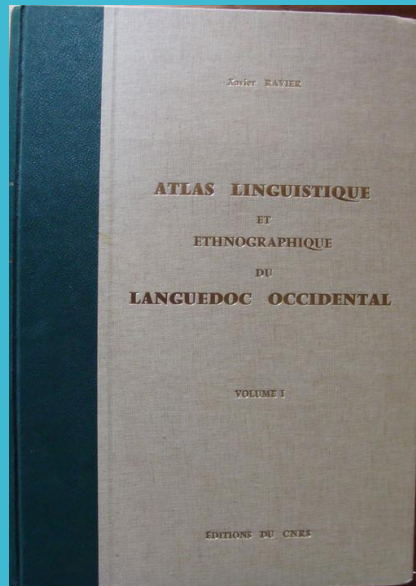


# Geodesic indexation of the ALLOc sites for Gabmap



# Dialectical *sites* (from ALLOc)

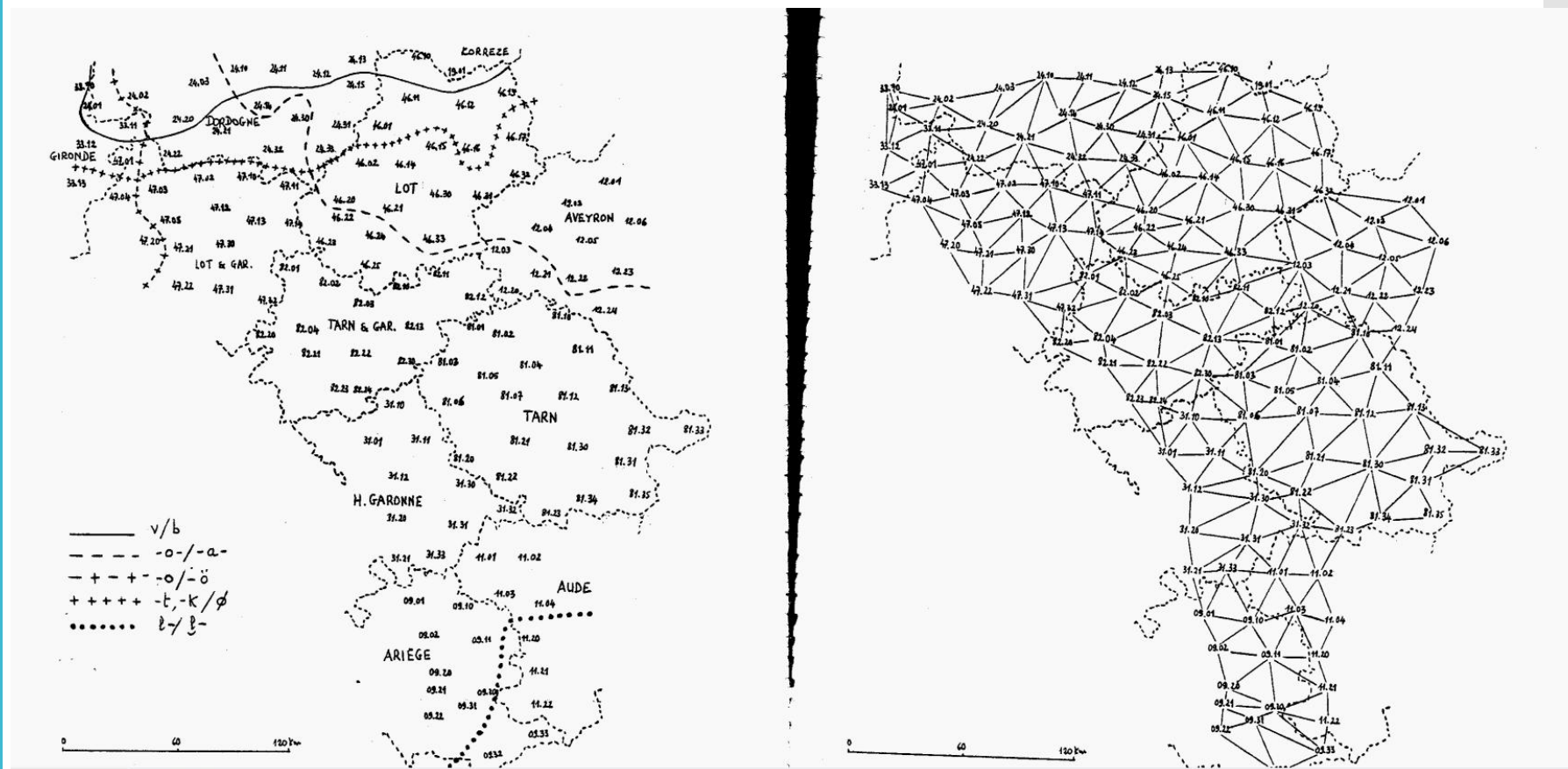
N = 131





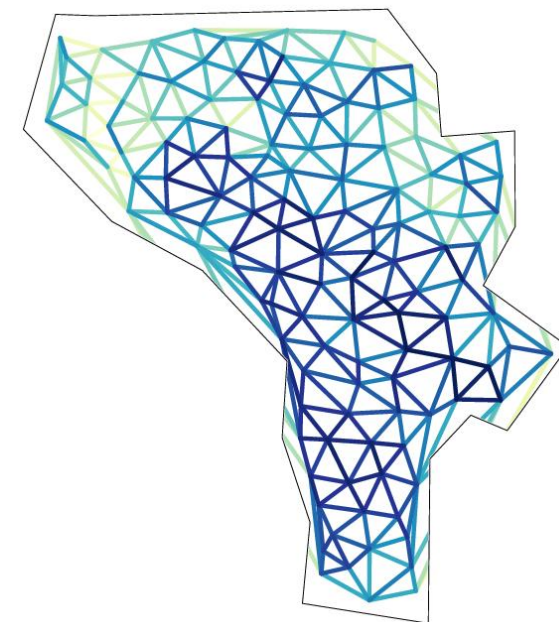
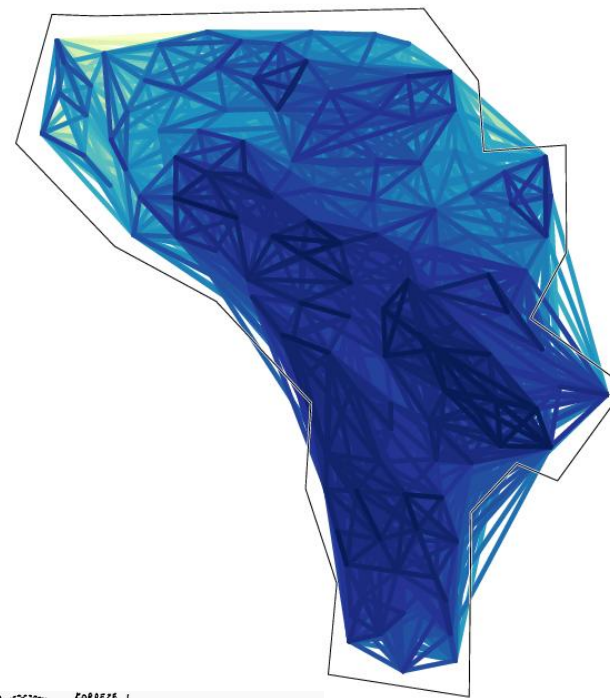
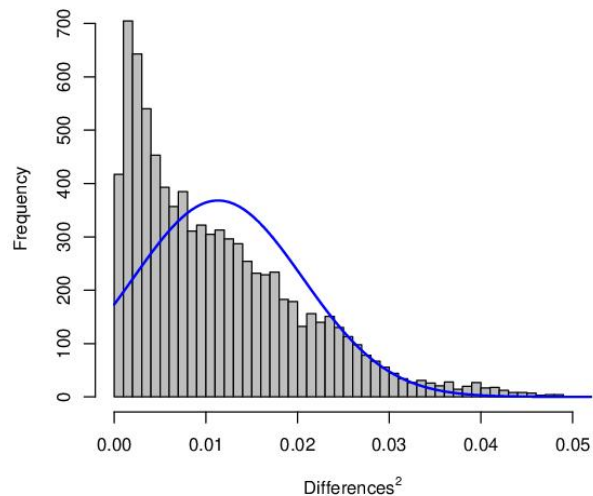
The «philological approach»  
(isoglosses) vs  
dialect network  
for DM  
(dialectometric)  
approach, by  
Christian Camps  
(1986)

[Guiter's «Global  
Method»]

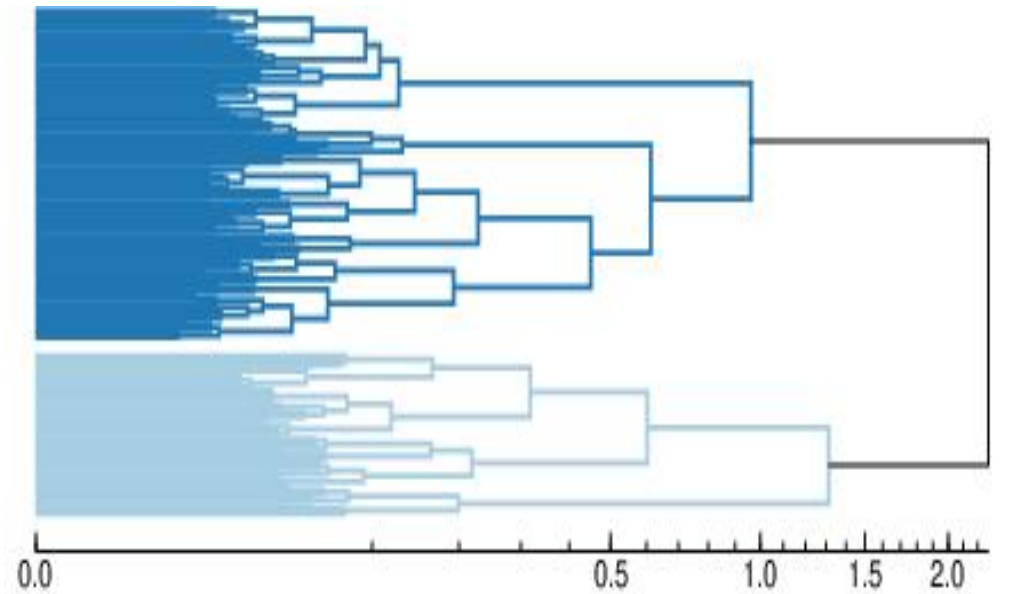
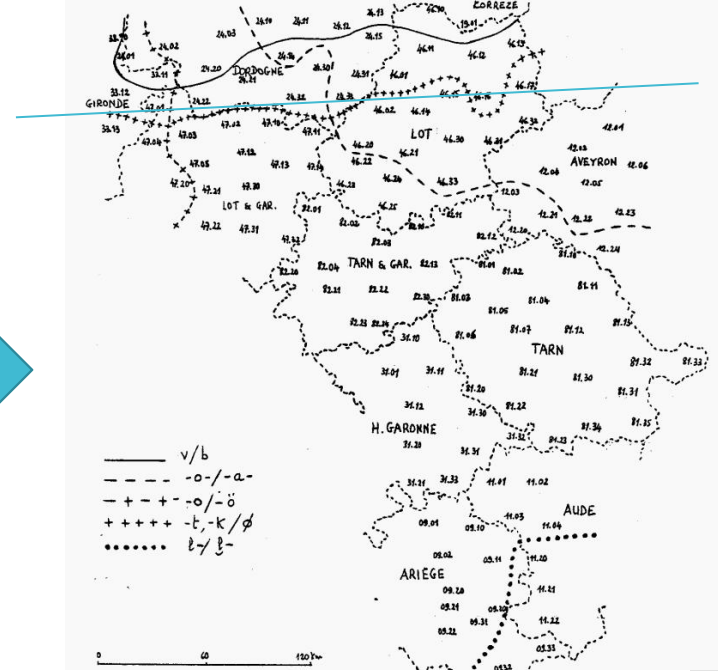
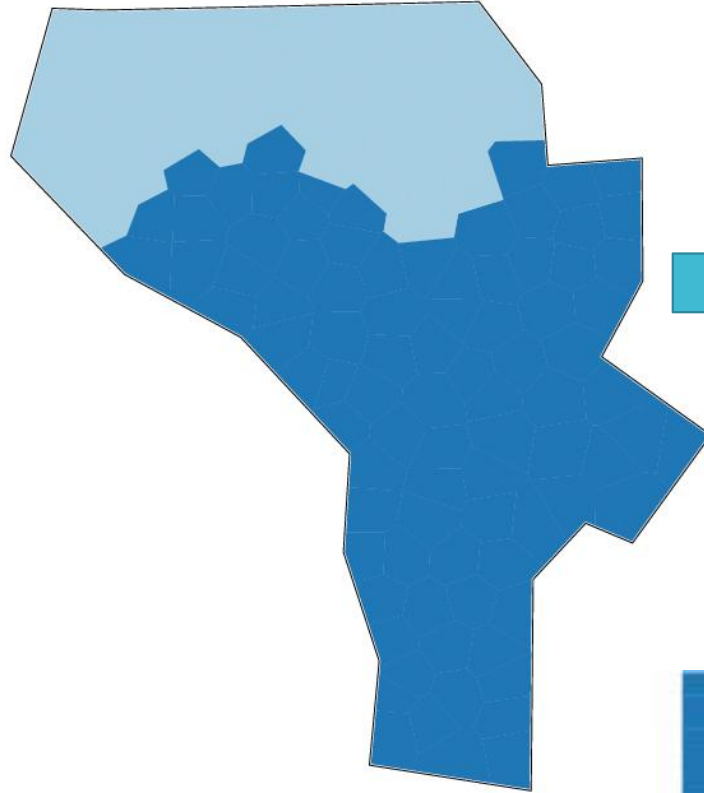


Phonological data  
 (string token analysis,  
 as above : CVCV  
 templates)  
 [Gabmap, edit distance]

**Places:** 131  
**Items:** 199  
**Instances:** 26276  
**Characters:** 155889  
**Unique characters:** 75  
**Tokens:** 151366  
**Unique tokens:** 107

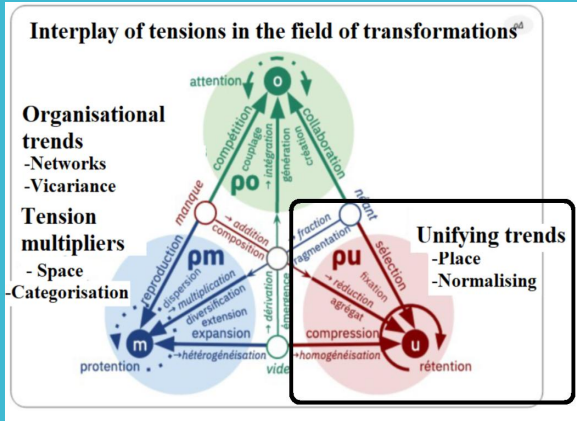
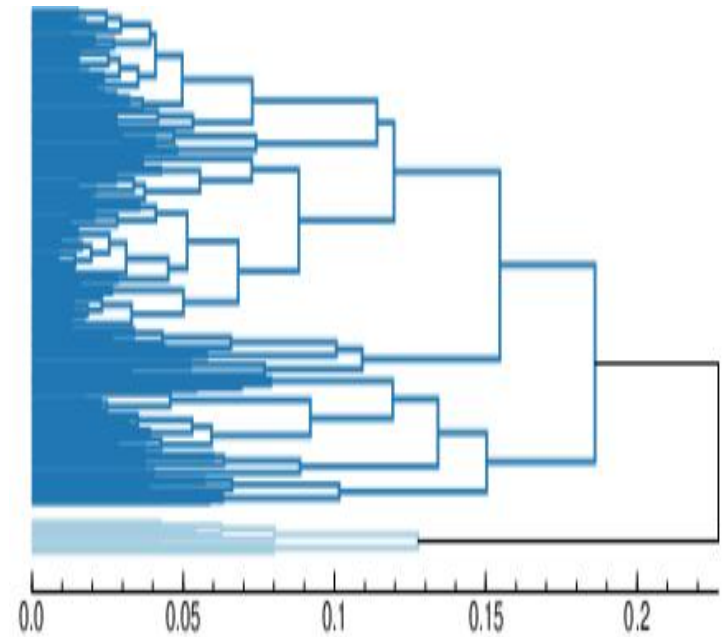
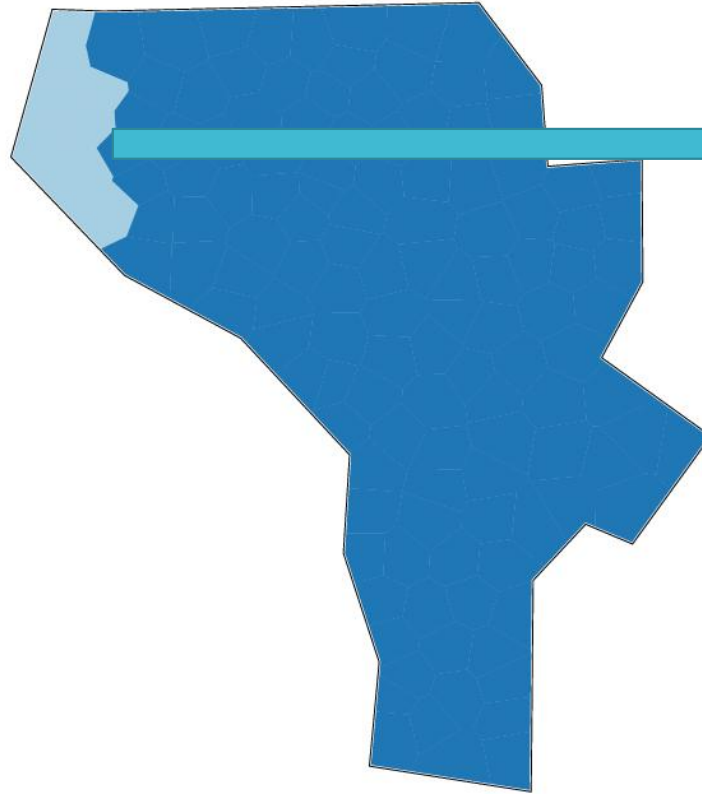


Ward's  
Method:  
2 classes  
(phonology)  
1st level of  
(taxonomic)  
organisation  
N vs C & S



Complete Link (& Group +/- pondered average):  
2 taxonomic classes  
(phonology)

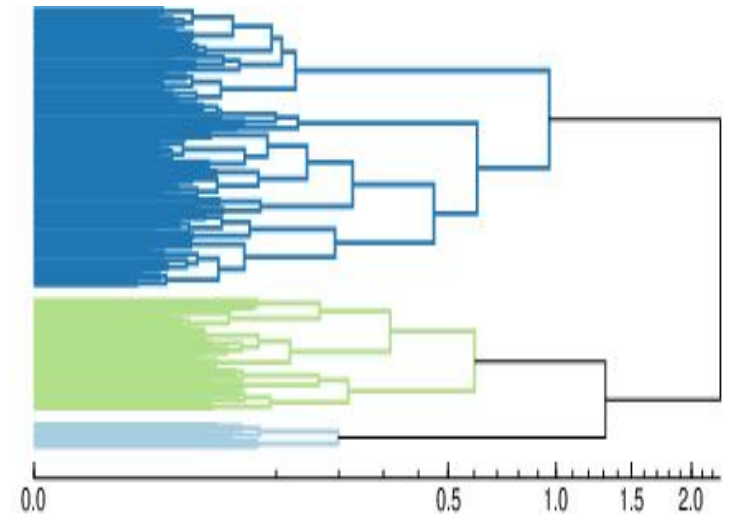
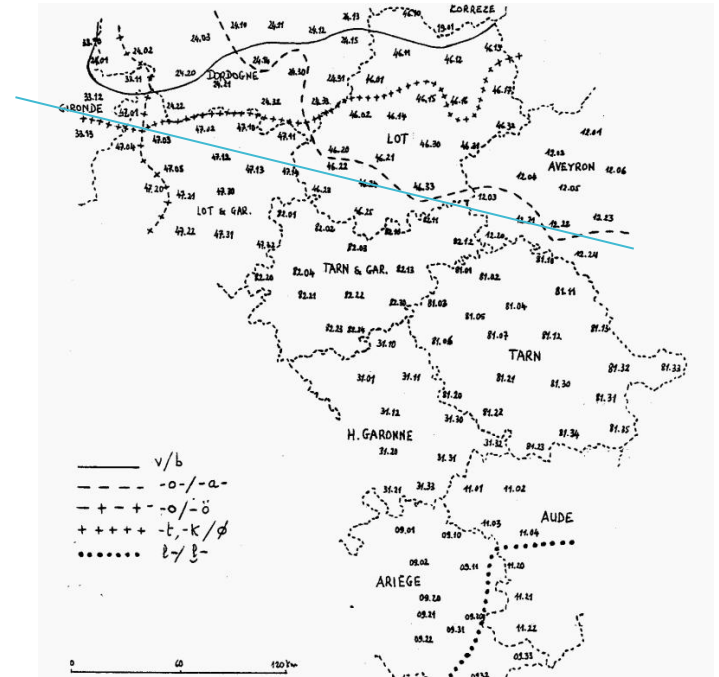
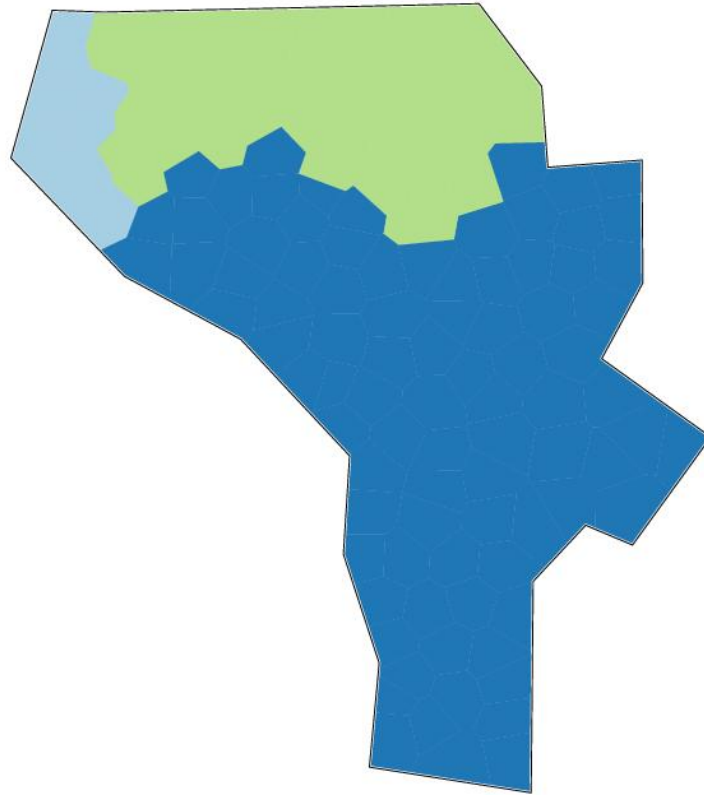
=> Unifying dynamics,  
reductionism (-)



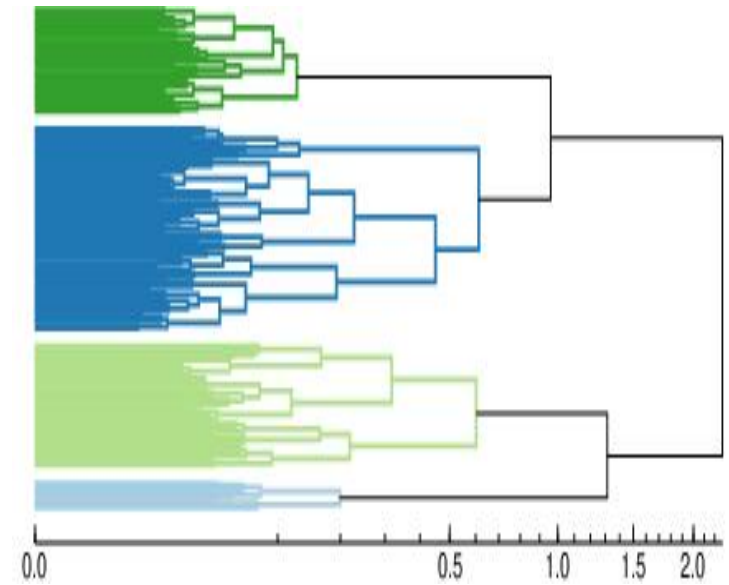
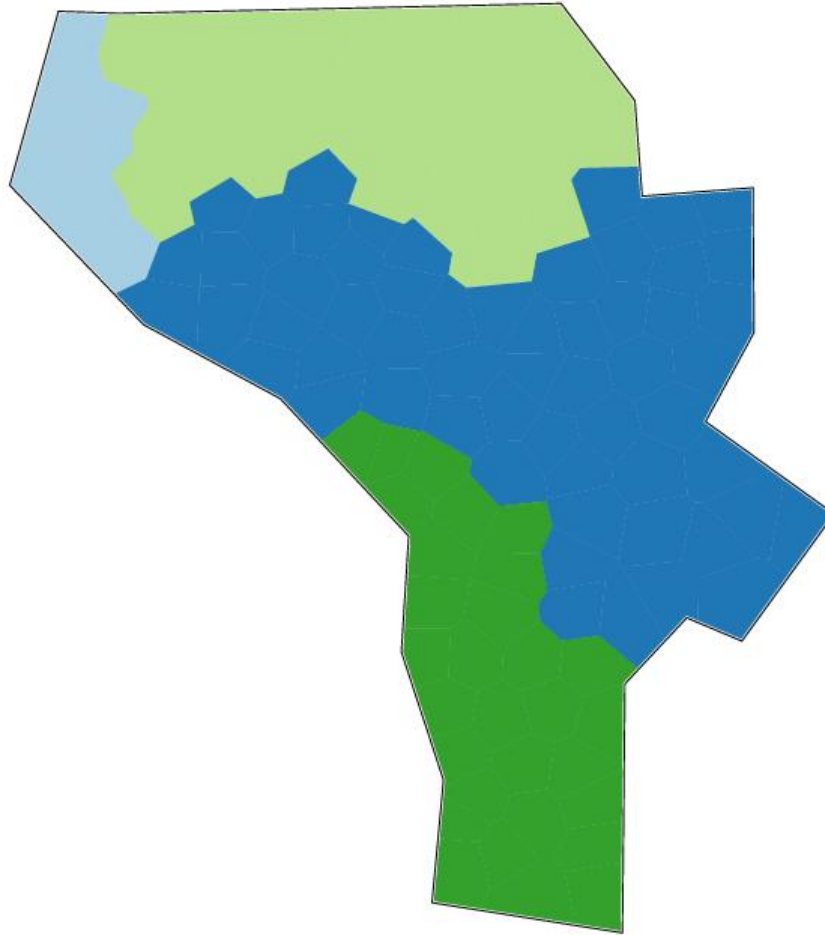


Ward's Method:  
3 classes  
(phonology)

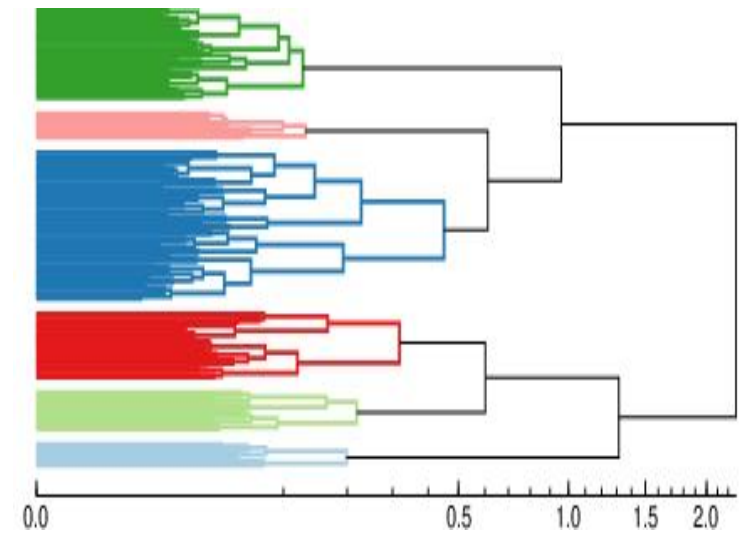
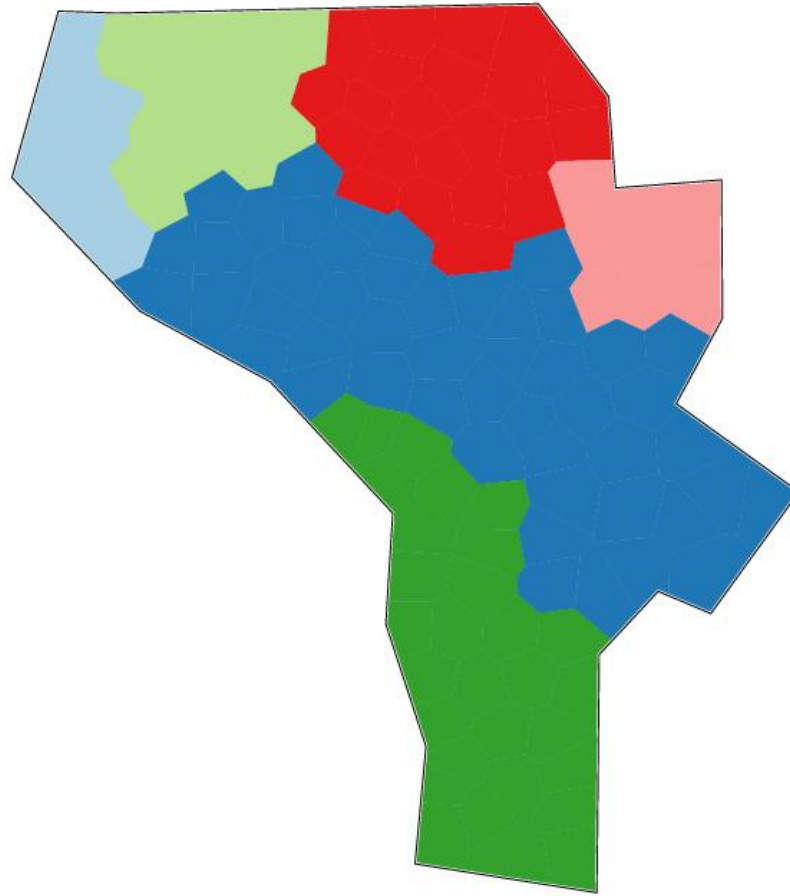
2nd step of  
taxonomic  
diversification  
& its isoglottic  
correlates  
according to  
Camps (*ibidem*)



Ward's Method:  
4 classes  
(phonology)  
A further step in  
taxonomic inner  
diversity

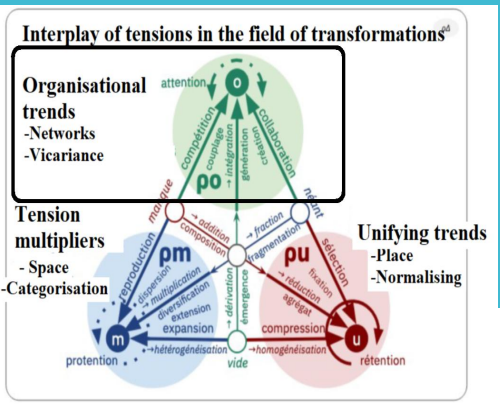
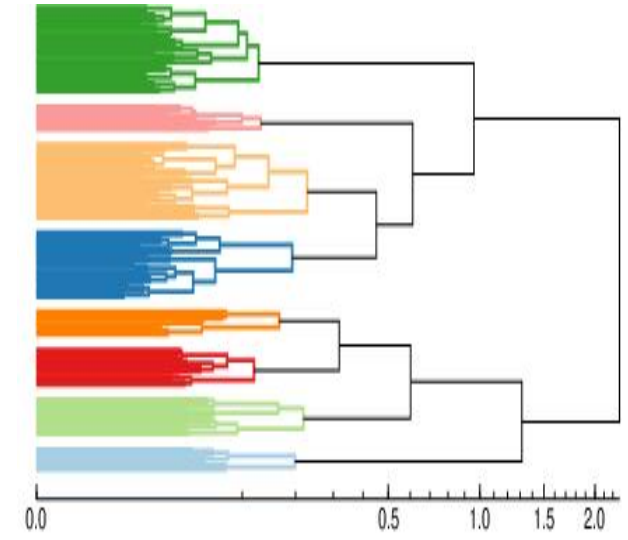
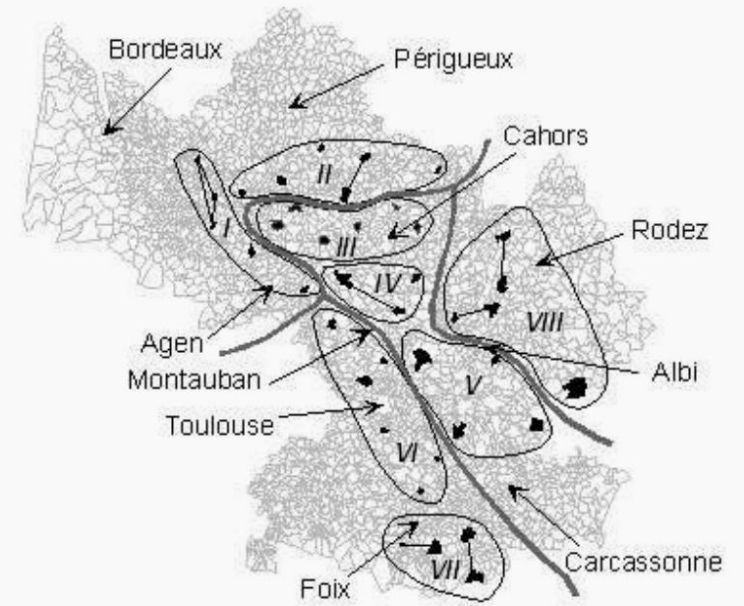
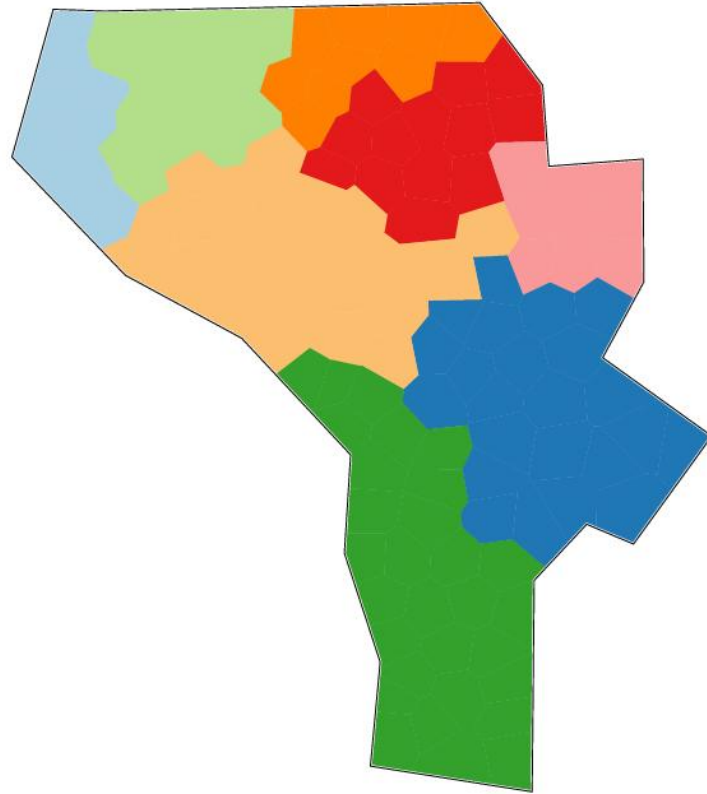


Ward's  
Method:  
6 classes  
(phonology)





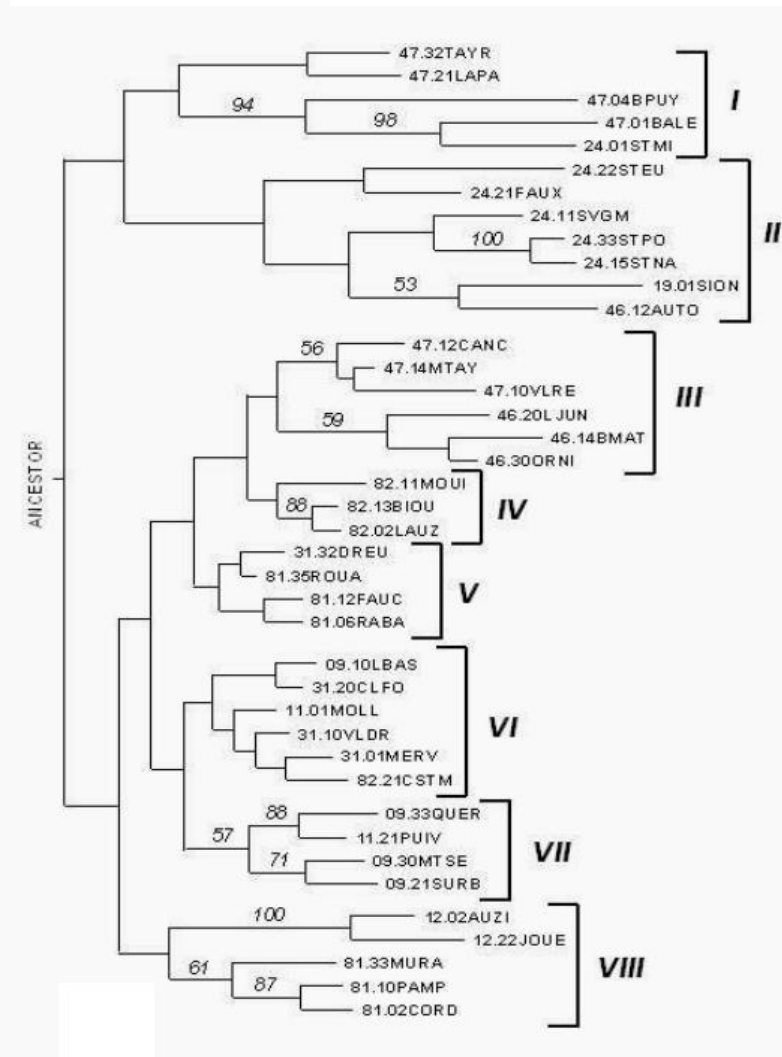
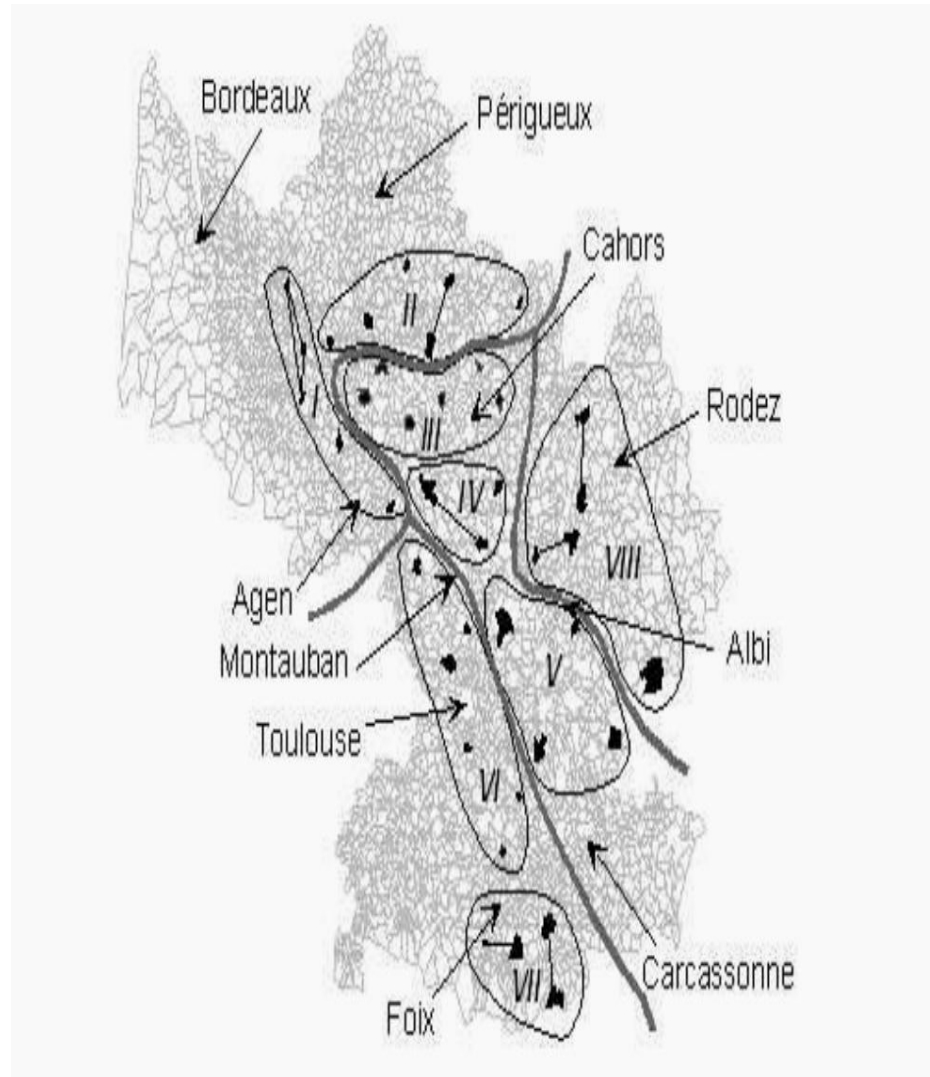
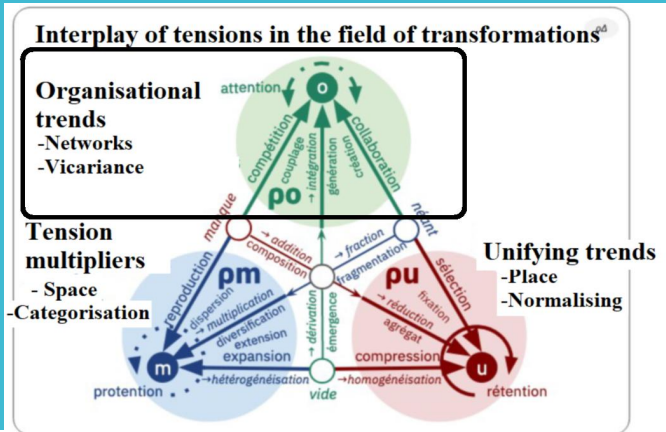
# Gabmap: Ward's Method: 8 classes (phonology) vs Cladistics (Brun-Trigaud & al. 2014)



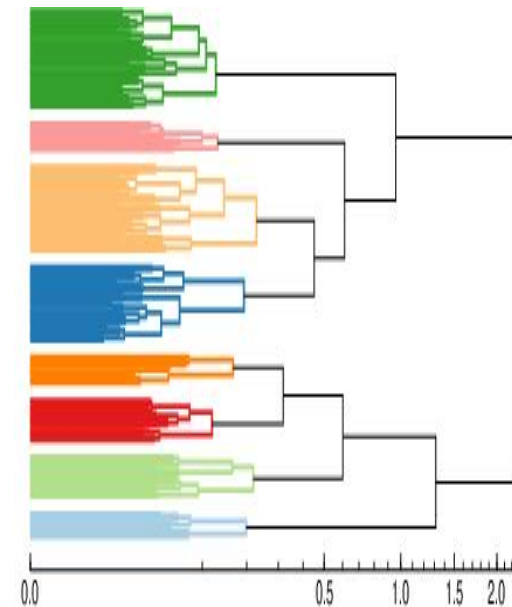
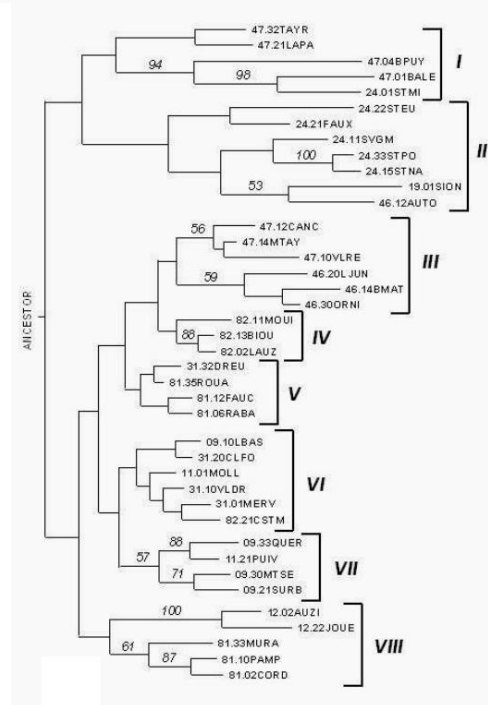
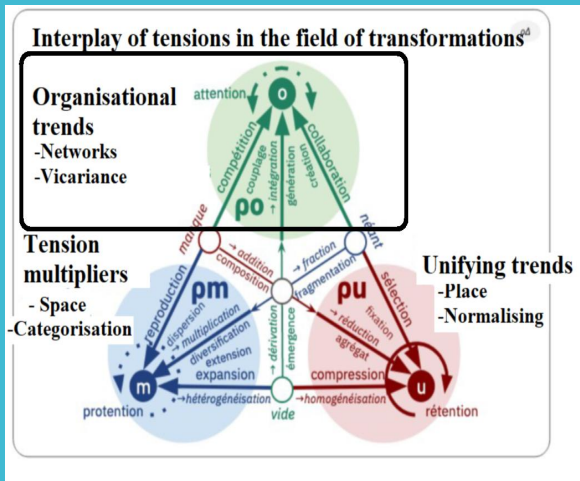
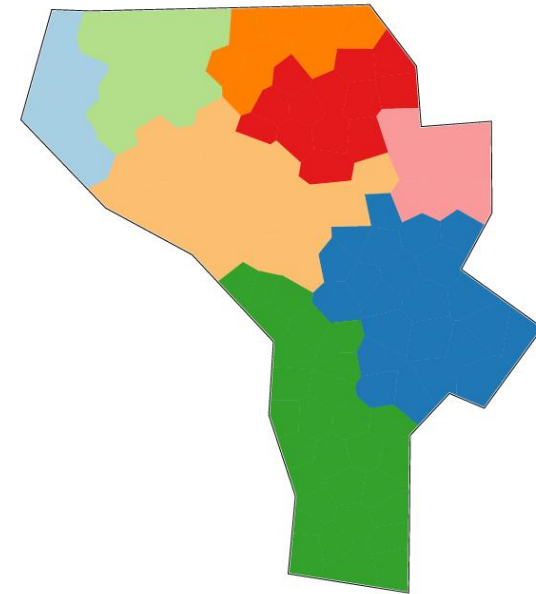
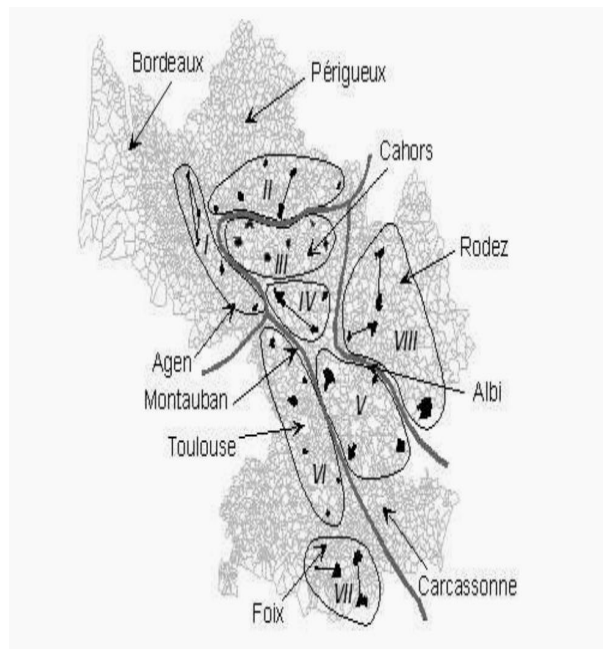
# Cladistic approach (ALLOc, phonology):

8 classes (Darlu & al. 2014)

=> Vicariance

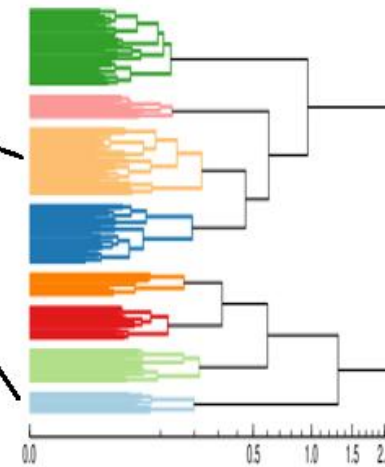
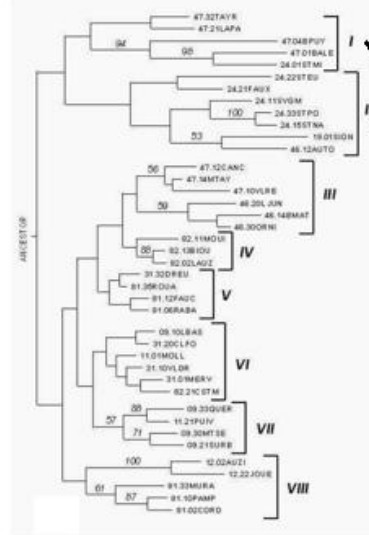
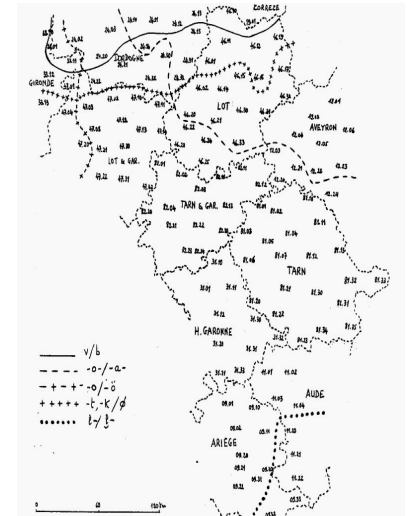
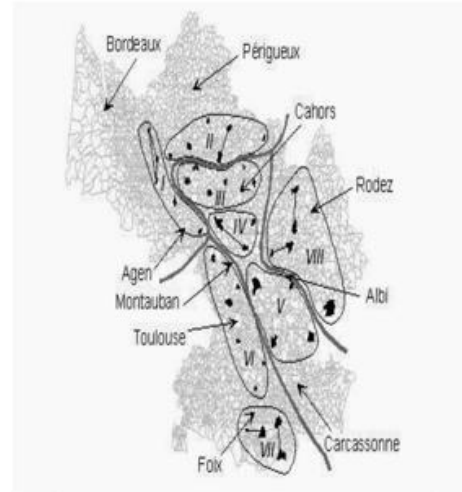


# Cladistic approach, 8 classes vs Ward's Method (Gabmap, 8 classes)

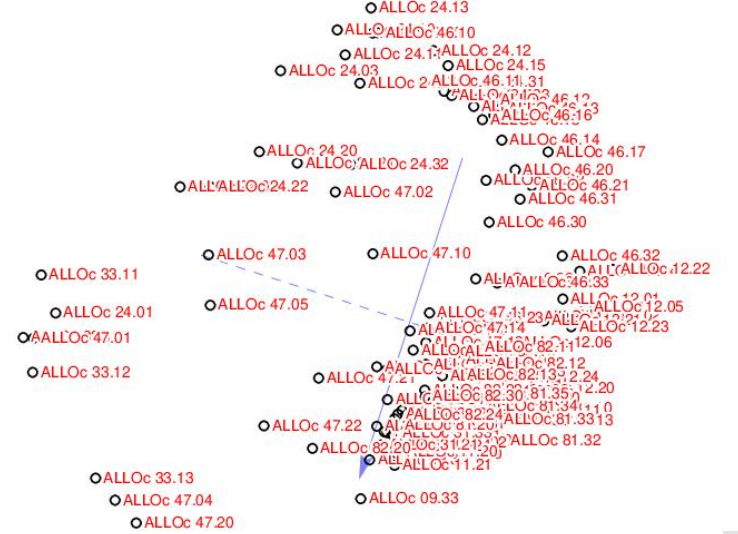
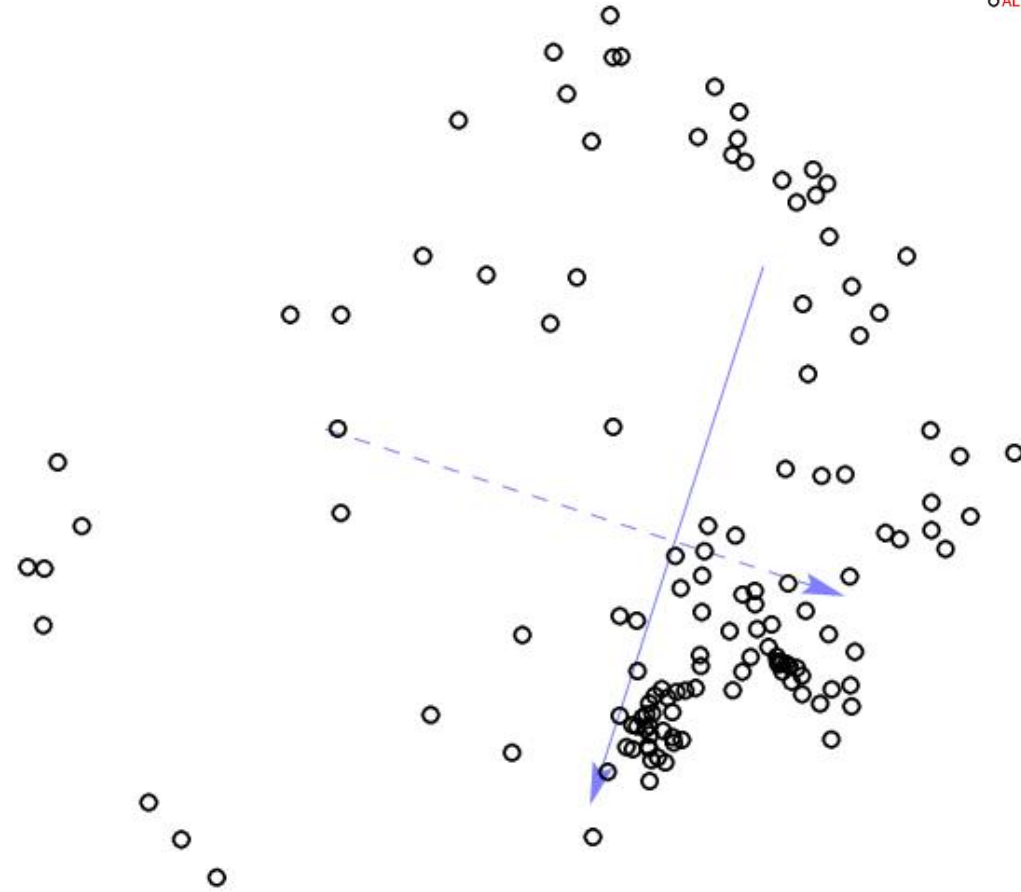




Cladistics vs  
LD  
geolinguistic  
patterns do  
correlate,  
whereas  
*taxonomy*  
slightly differs

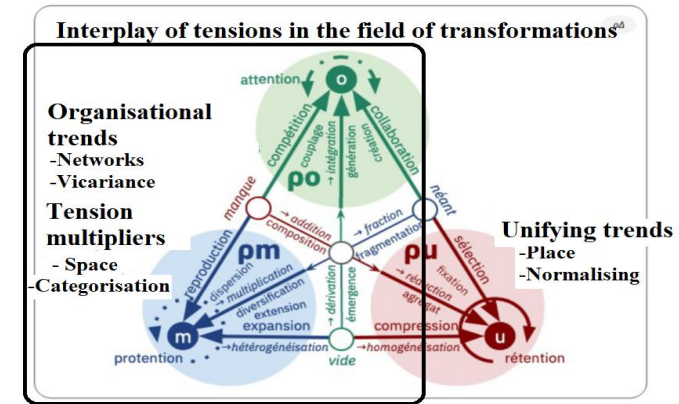
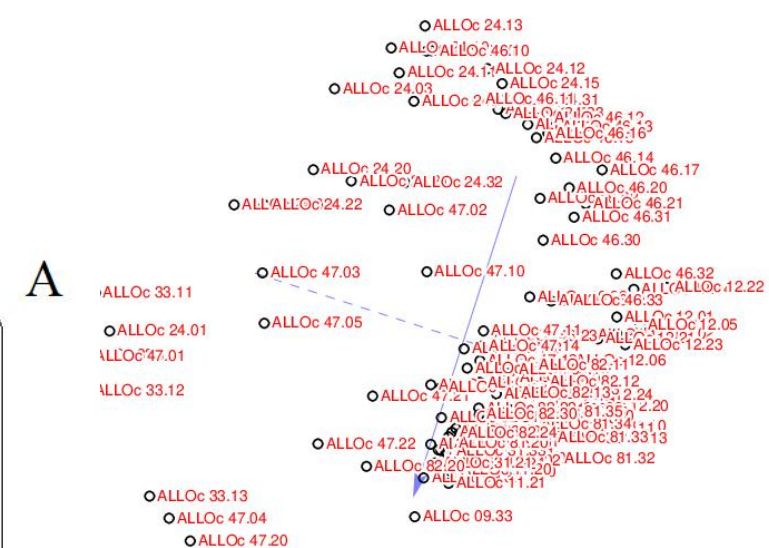
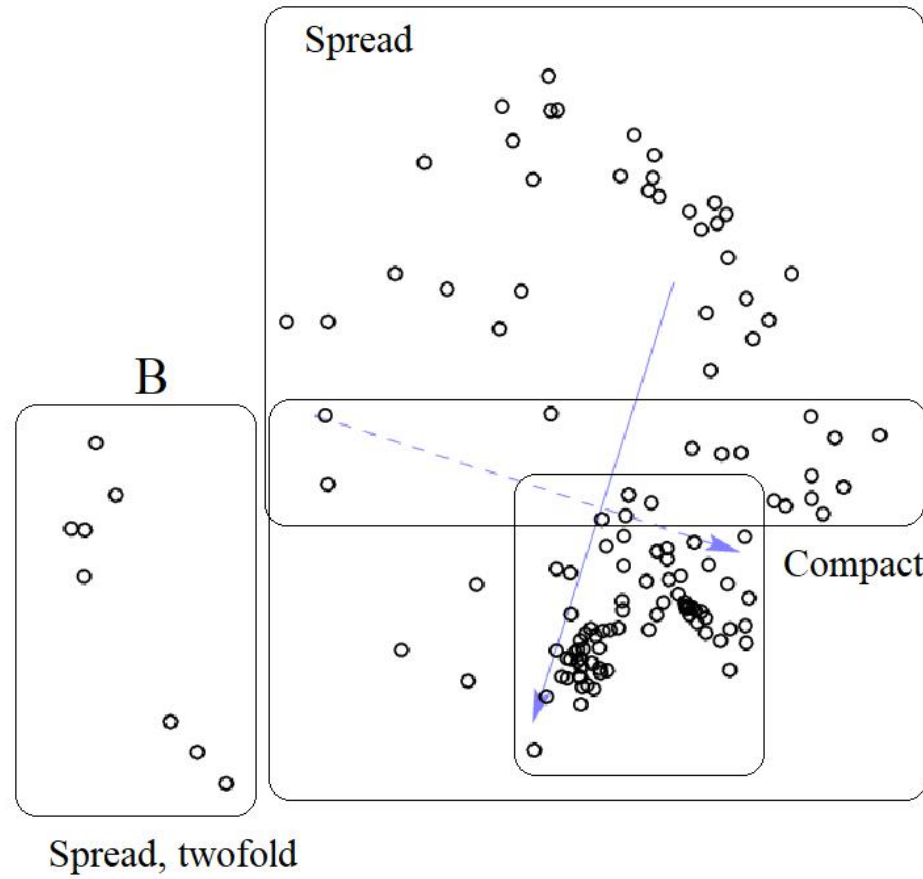


# MDS, ALLOc (Phonology)



# MDS, ALLOc (Phonology)

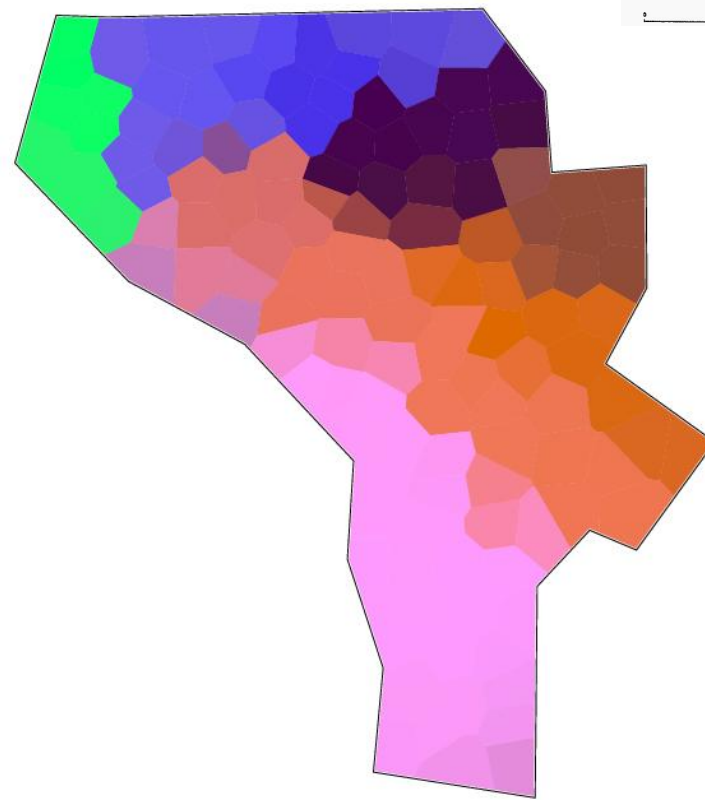
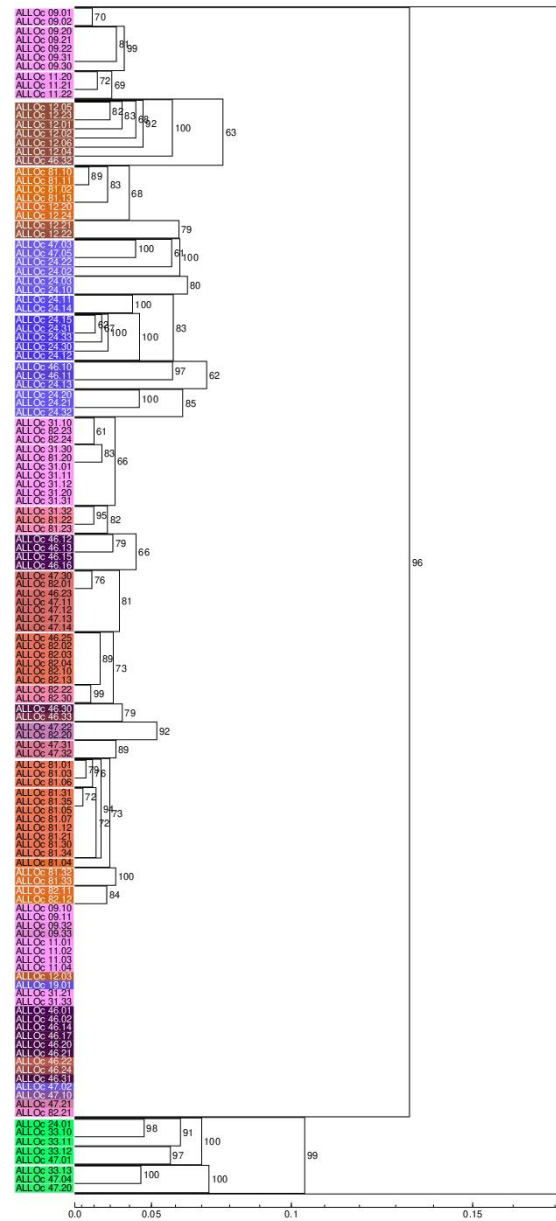
## Inducing aggregate properties & dichotomy



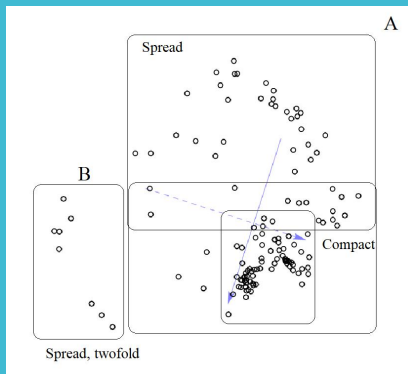


Readability,  
visualization,  
evaluation of  
cluster  
properties:

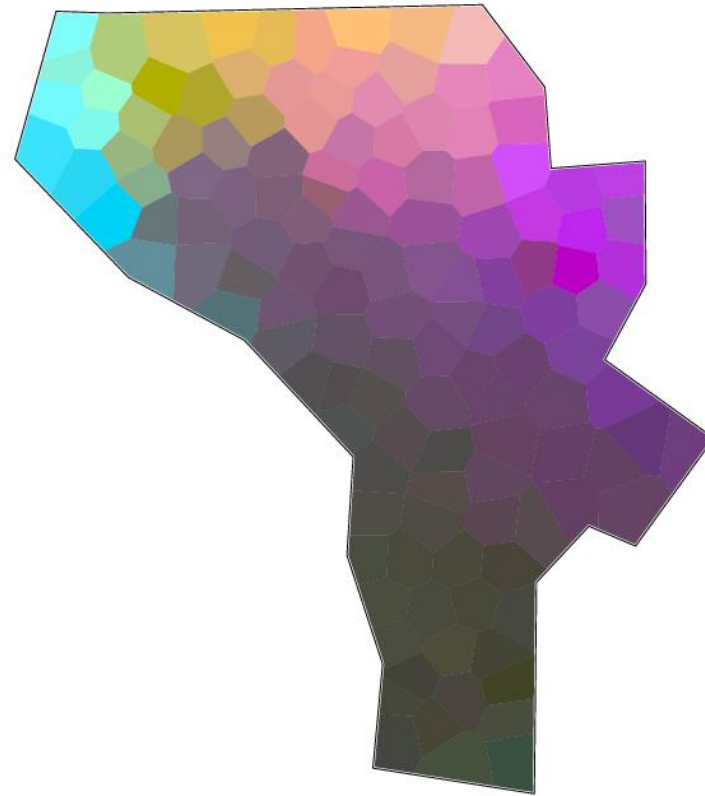
Probabilistic  
dendrogram &  
Fuzzy Cluster  
Map (Gabmap,  
ALLOc,  
phonology)



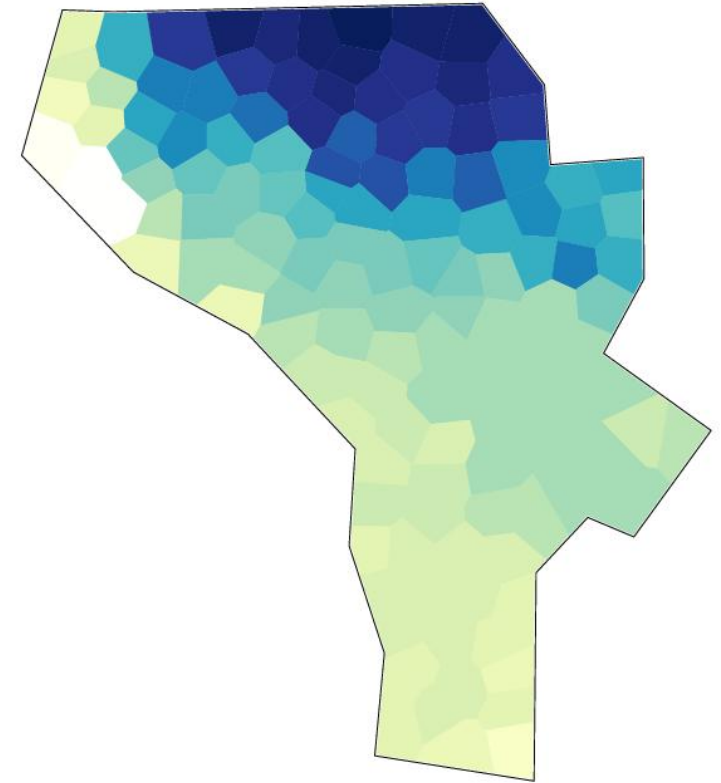
Gabmap  
(ALOC)  
199 items  
=> *Gradient* in a dialect  
continuum  
Classical MDS in three  
dimensions mapped  
onto RGB color space



ALOC phonology  
MDS maps; holistic ( $r = 0.98$ )



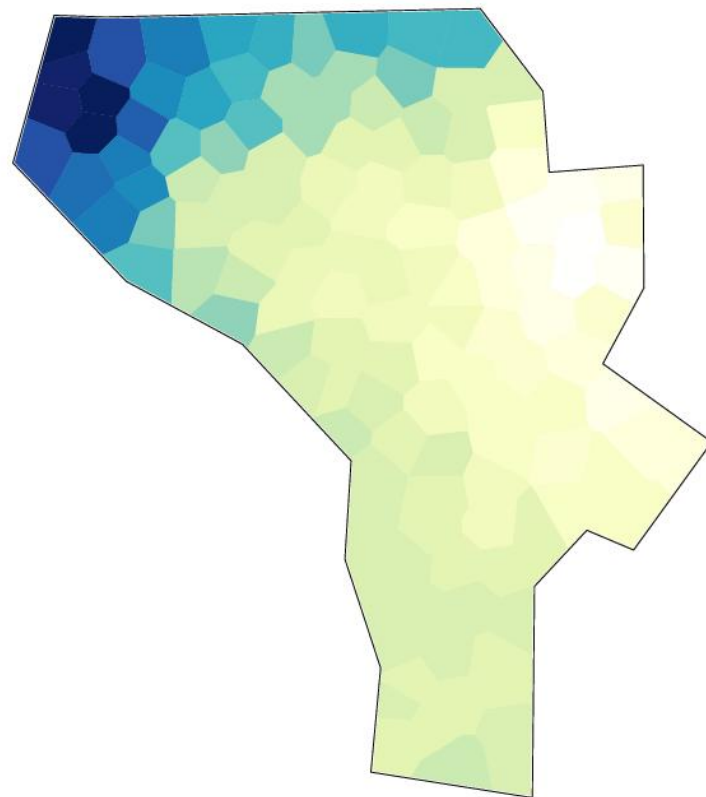
ALOC phonology MDS maps; 1st  
Dimension  
( $r = 0.71$ , stress = 0.77)  
Enhancing the Northern part of the  
*Dialect Continuum*



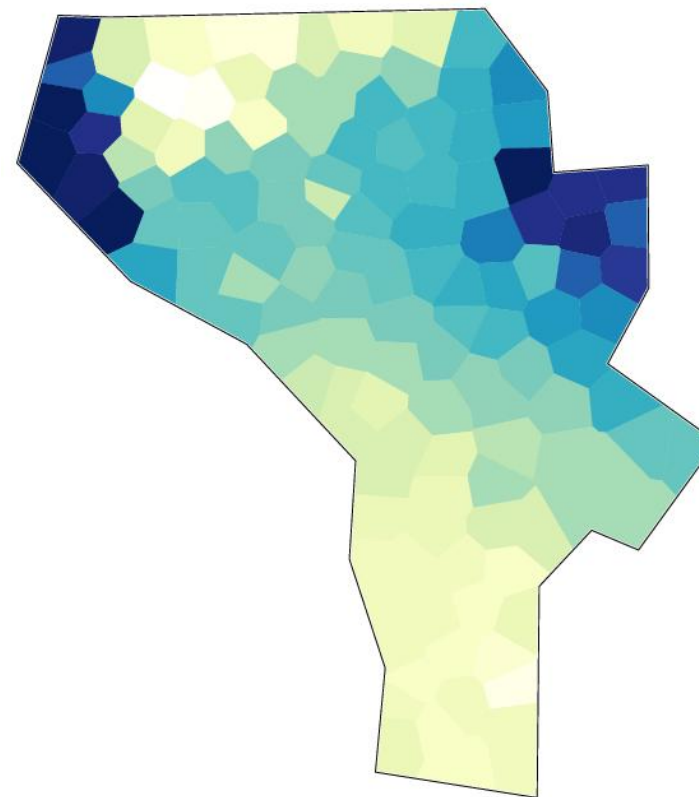
ALLOC  
phonology

2nd vs 3rd  
dimension

MDS maps: 2nd Dimension;  
 $r = 0.74-0.97$ , stress = 0.25  
North-Western segment of  
the dialect continuum



MDS maps: 3rd Dimension  
 $r = 0.49-0.98$ , stress = 0.19  
Lateral (i.e. peripheral areas)

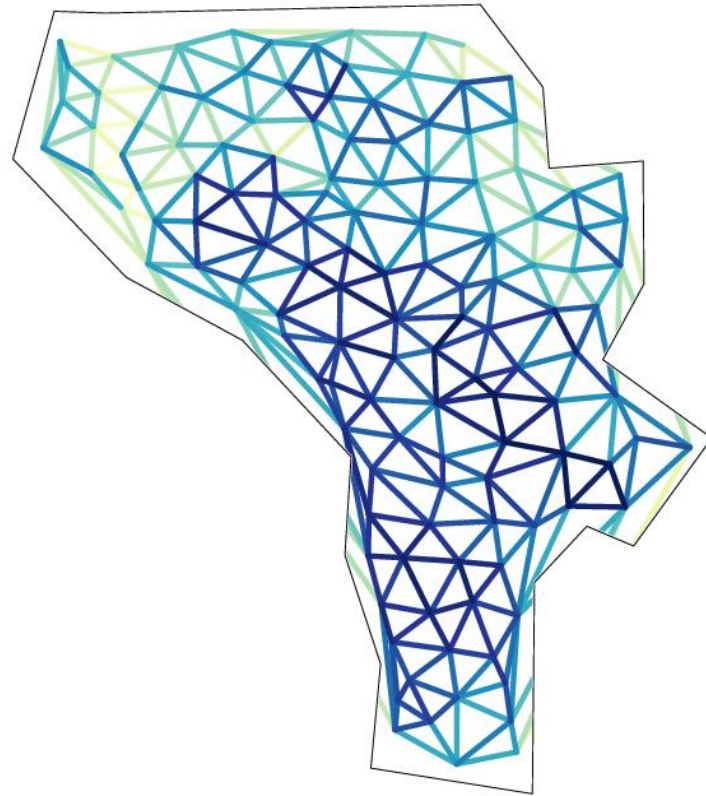




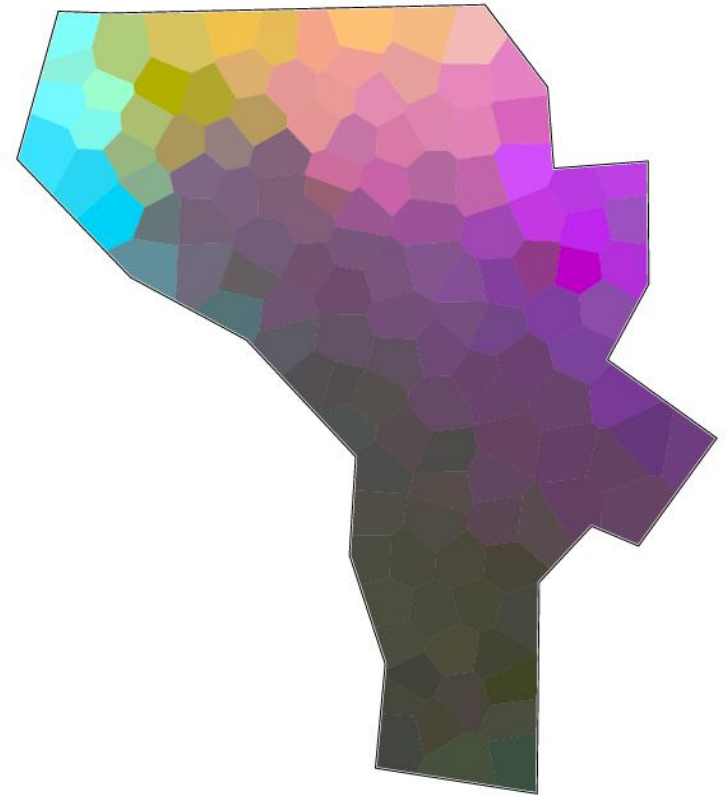
# Vicarious standpoints : Dialect network vs Continuum

(ALLOc, phonology,  
199 items)

Dialect Network inner  
structure



Dialect Continuum Gradient

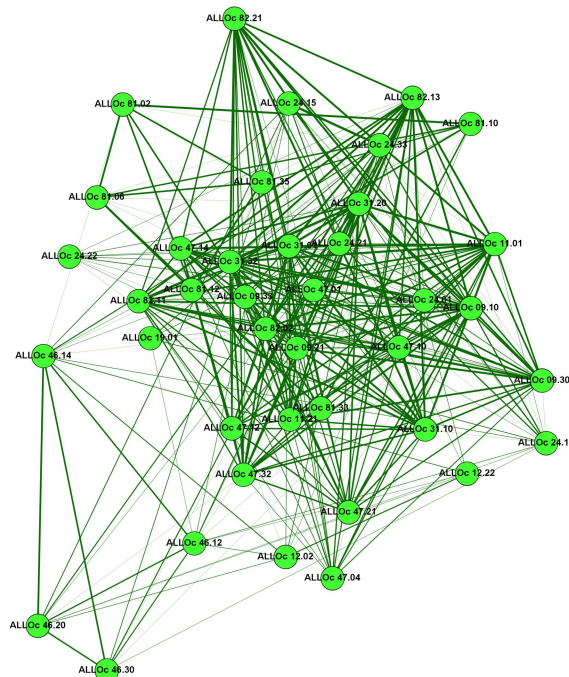


Gephi graphs  
(ALLOc, phonology) :  
threshold 84-85/100

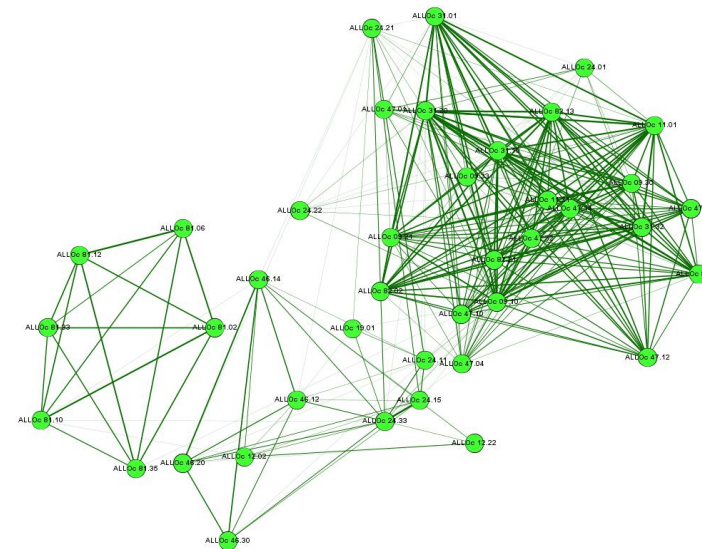
Emerging networks of  
communal aggregates  
(LD converted into  
similarities/ties)

=> a ball of yarn vs a  
triad

Gephi graph  
Threshold of similarity: 84/100  
ALLOc (phonology)

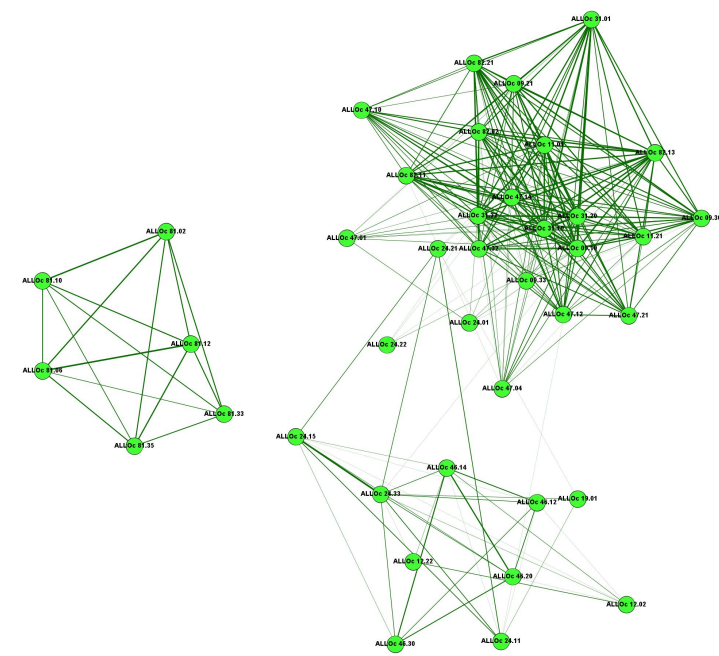


Gephi graph: threshold 85/100  
ALLOc (phonology)

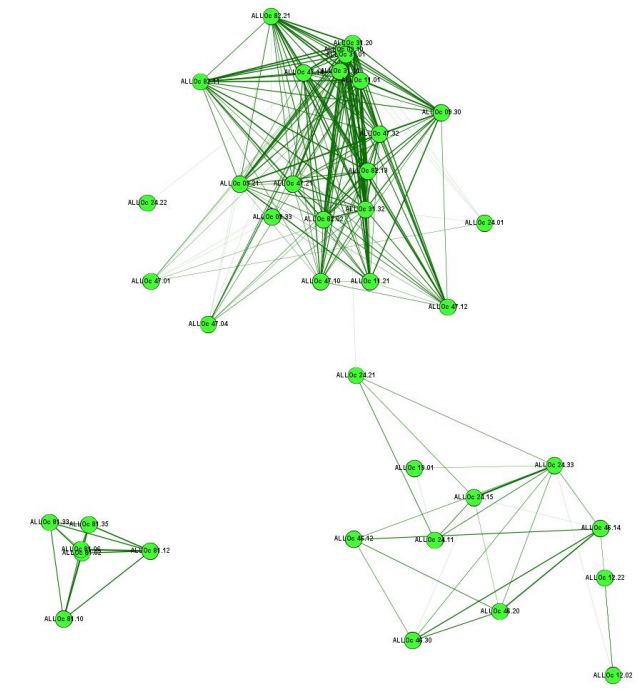


Gephi graph:  
threshold 86-  
87/100  
ALLOC  
(phonology)  
unfolding the  
triad...

Threshold of similarity:  
86/100



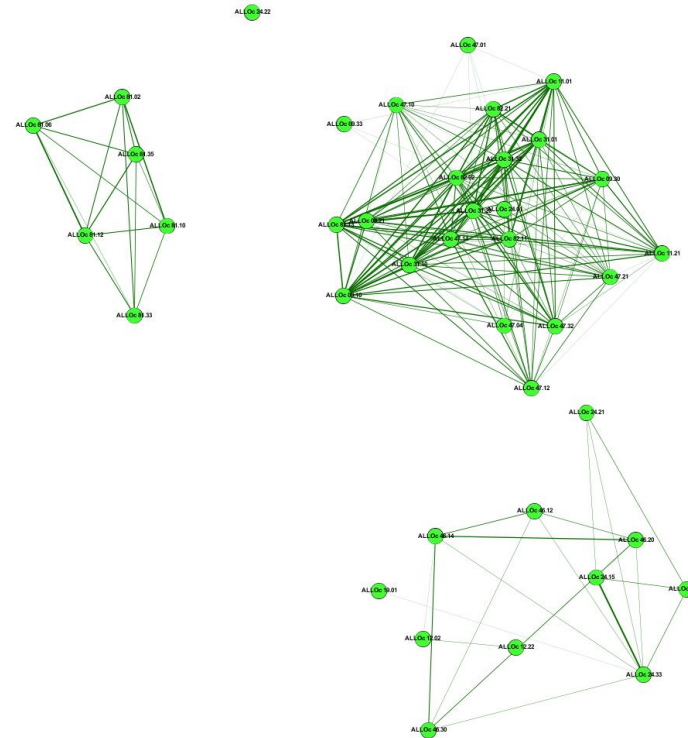
Threshold 87/100



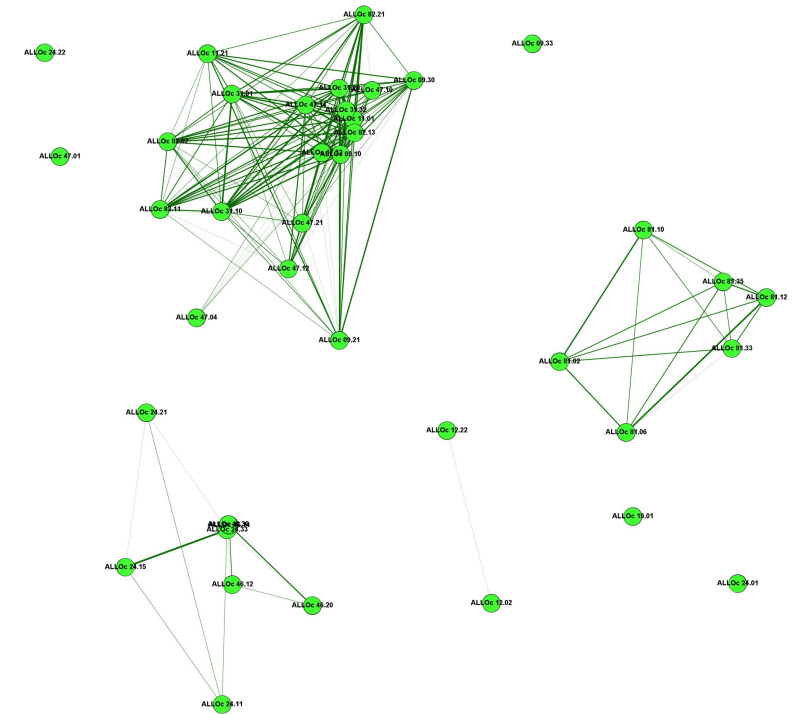


Gephi graph:  
threshold 88-  
89/100  
ALOC (phonology)  
=> emerging  
singletons/outliers  
vs robust  
cores/choremes  
within a  
disentangled triad

Threshold of similarity:  
88/100

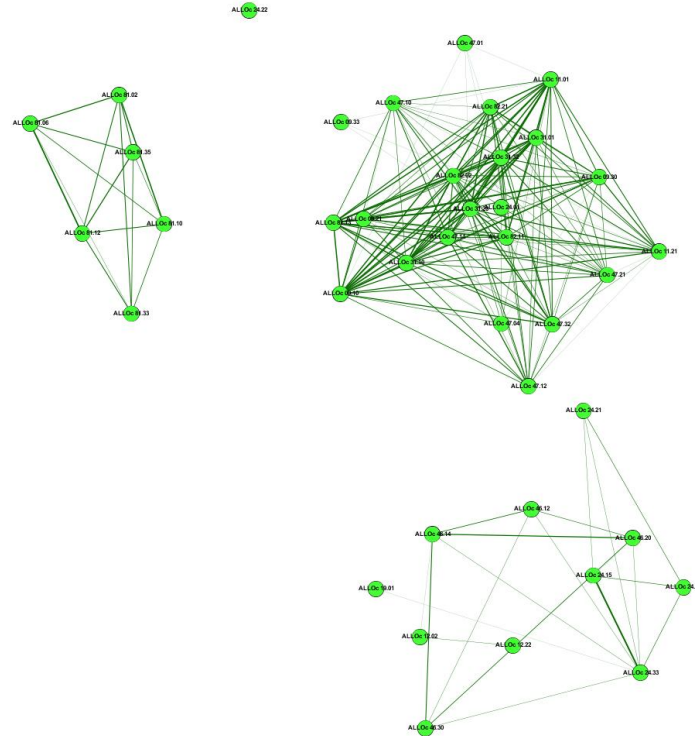


Threshold 89/100

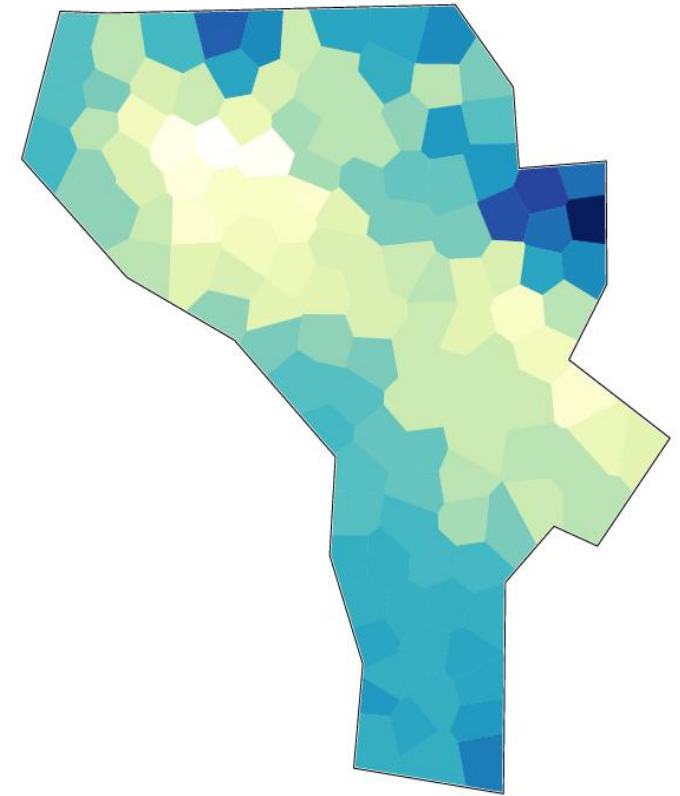


Gephi graph:  
threshold 88-  
89/100  
ALLOC (phonology)  
=> emerging  
**singletons/outliers**  
vs robust  
**cores/choremes**  
within a  
disentangled triad

Threshold of similarity:  
**88/100**

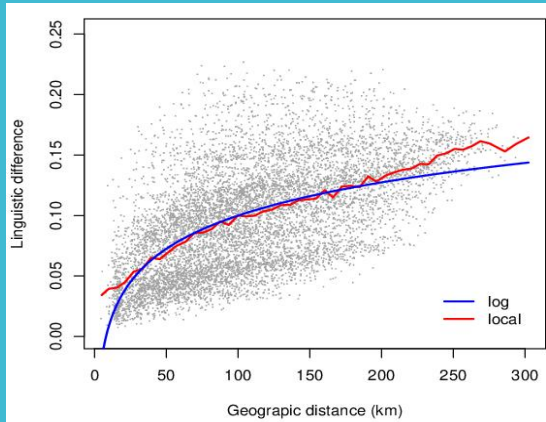


=> Close to MDS 4th dimension:  $r = 0.13-99$   
4 dimensions; stress = 0.17  
(but not only: the graph tells more about  
the network intricacy/inner structure)

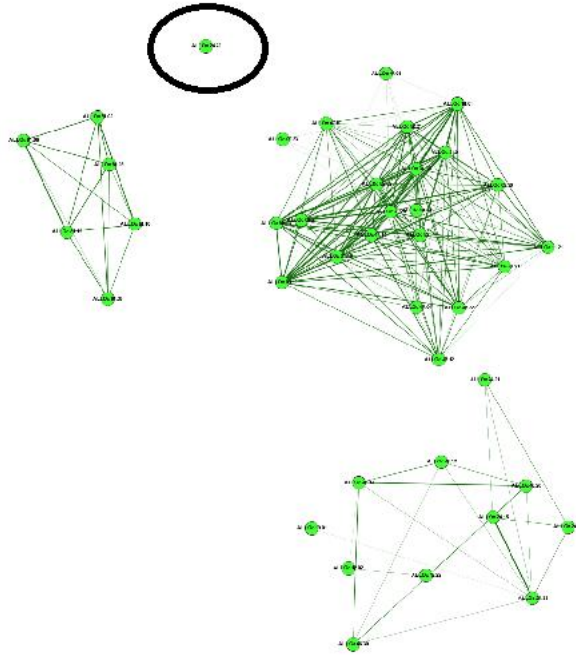


Heuristic hints from outliers:  
ALLOc 24.22 at 88/100  
threshold as compared to its  
linear distance with the  
thread of the dialect  
continuum/gradient

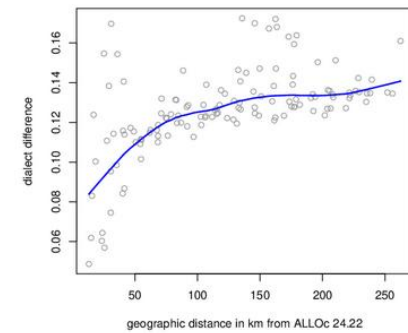
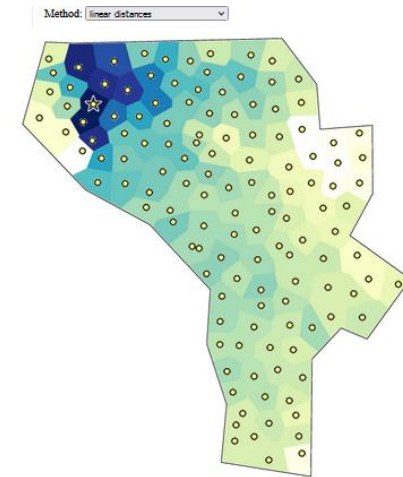
Below: overall pattern of  
geolinguistic distance in the  
ALLOc, according to Gabmap



A singleton  
at 88/100 threshold:  
locality ALLOc 24.22  
Reminder: Gephi graph

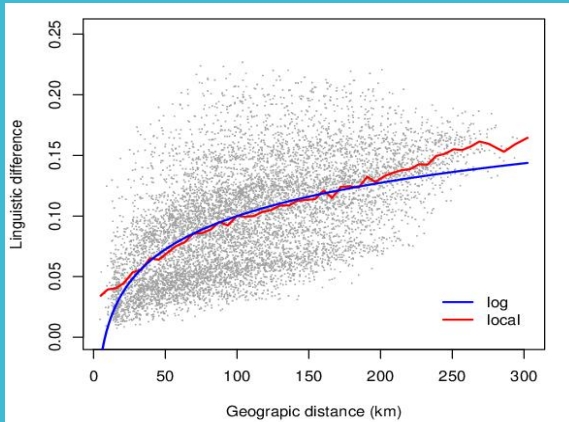


Site ALLOc 24.22: linear  
distance within the dialect  
continuum

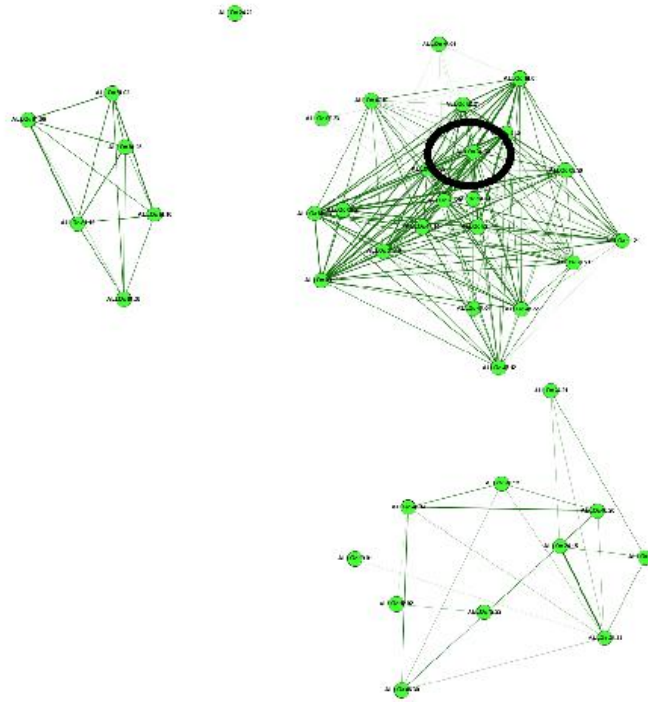




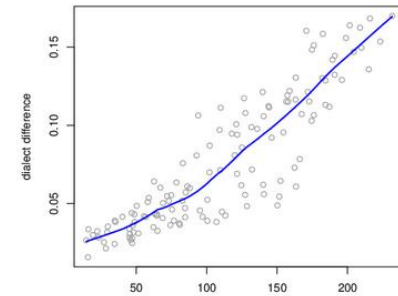
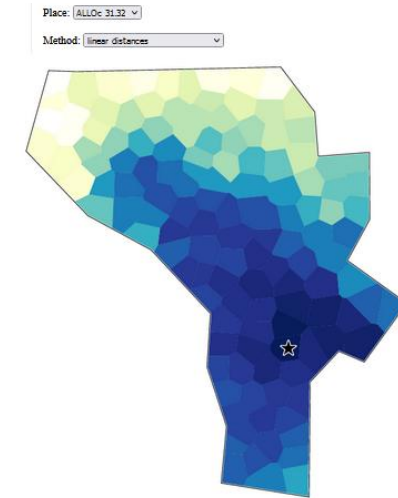
Heuristic hints from outliers: ALLOc 31.32 at 88/100 threshold as compared to its linear distance with the rest of the dialect continuum/gradient



A strongly embedded site,  
at 88/100 threshold:  
locality ALLOc 31.32  
Reminder: Gephi graph



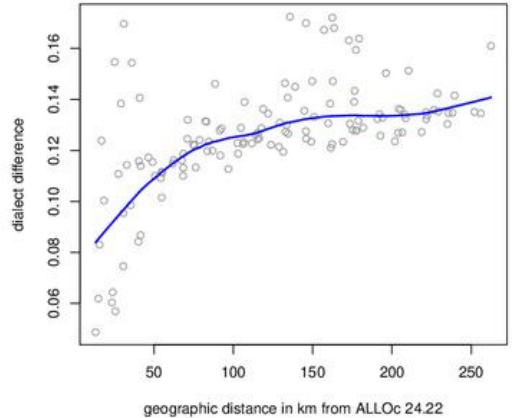
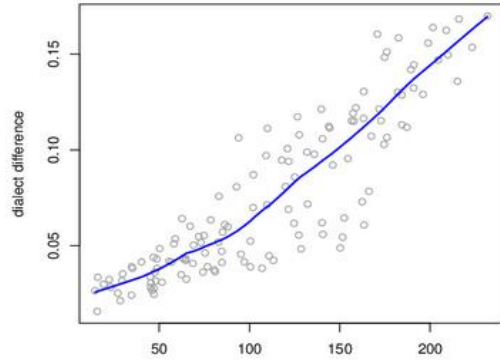
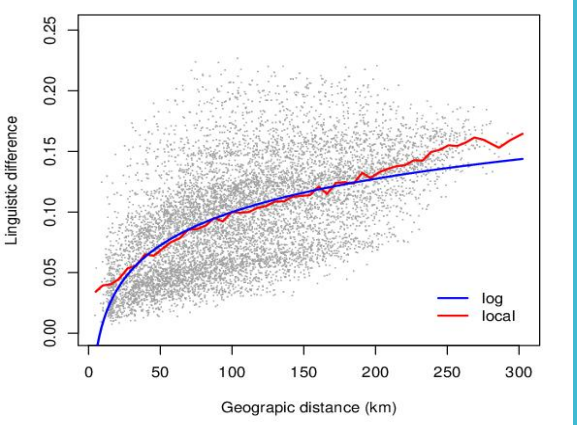
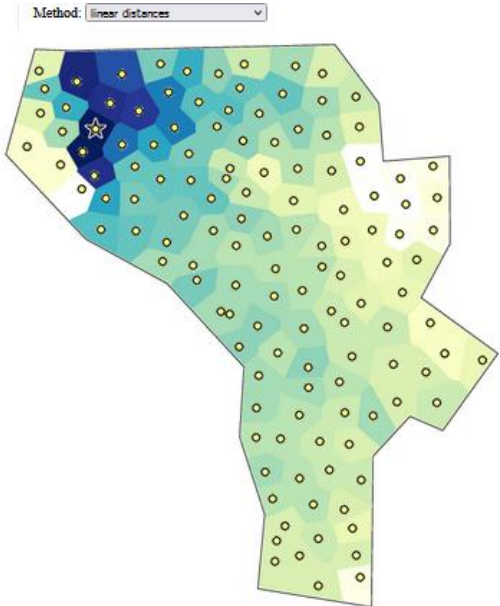
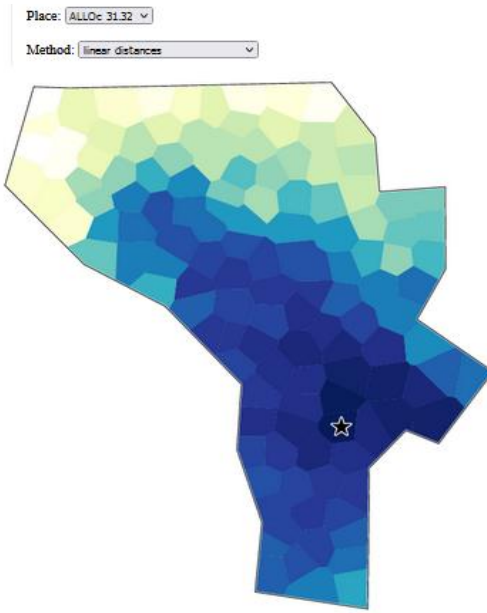
Linear distance (Gabmap)  
ALLOc 31.32



Heuristic hints from outliers: ALLOc 31.32 at 88/100 threshold as compared to its linear distance with the rest of the dialect continuum/gradient

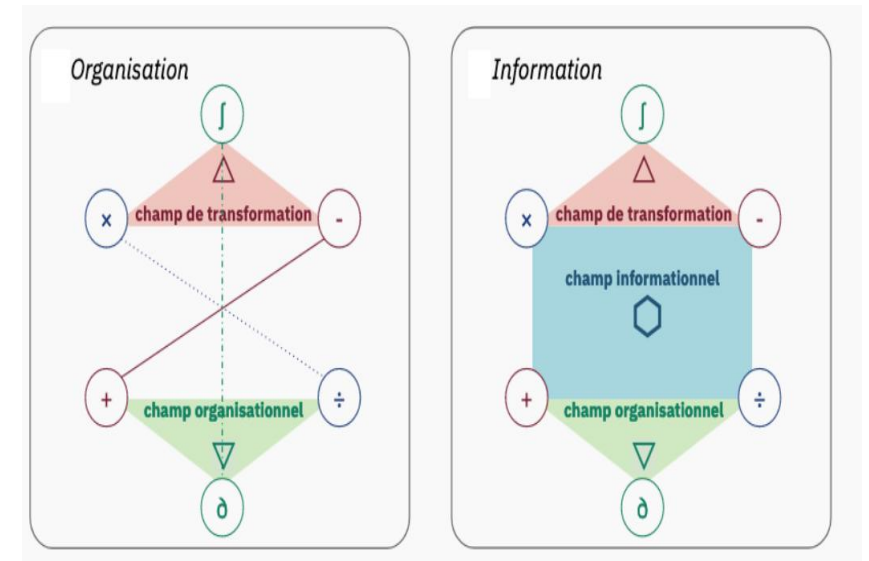
Core variety,  
at 88/100 threshold:  
locality ALLOc 31.32  
Reminder: Gephi graph

Outlier  
Linear distance (Gabmap)  
ALLOc 24.22



# What did this case study tell us? What benefits, from the standpoint of Complexity Theory?

- 1) **Vicariance**: embedding & comparison of several techniques/methods (LD vs Cladistics)
- 2) Dialect **Network** (Gabmap, Gephi) vs Dialect **Continuum** (MDS, from Gabmap)
- 3) **Finer grain ontology & agentivity**: the quest for heuristic outliers (*Tertium Comparationis* or *Inclusum*?)
- 4) **Multidimensional dialectology** through facts and various techniques of visualisation.





More heuristic hunting grounds...  
 Back to data:  
 Phonological variables  
 (1) unstressed -a# and  
 (2) -l#) in the  
 peripheral areas  
 => Function  
 «Shibboleth» in  
 Gabmap)

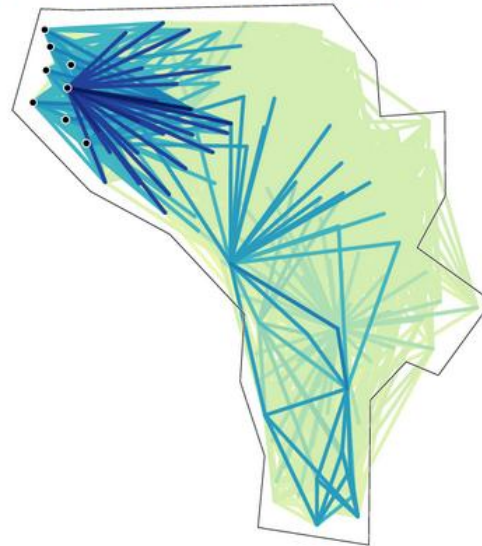
Reminder: patterns from  
 the MDS 3rd Dimension...  
 (right side at the bottom of  
 the slide)

Step 3: select item

Items sorted by value: 2.47 (-1.23 : 1.24) *diabé* Select item  
 The format of the select list is 'Score (Within difference : Between difference) item'.  
 → download as list

Current item: eau  
 Score: 2.124700  
 — Within difference: -0.464405  
 — Between difference: 1.660295

Difference map for *eau* (note: darker link color means higher difference)



eps  
pdf  
png

Step 4: show the distribution of relevant forms

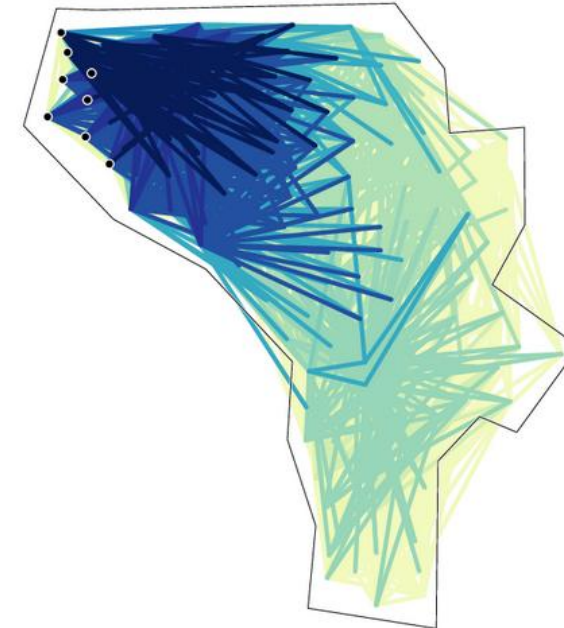
Patterns in the cluster	Patterns not in the cluster
<ul style="list-style-type: none"> <li>ɛjyɛ (6:6)</li> <li>ɛjyɛ (1:1)</li> <li>ɛjgɛ (1:1)</li> </ul>	<ul style="list-style-type: none"> <li>'ajgɔ (0:69)</li> <li>'ajɔ (0:38)</li> <li>'ajgɔ (0:2)</li> <li>'ajgɔ (0:2)</li> <li>ajyɔ (0:2)</li> <li>'ajyɔ (0:12)</li> <li>ajgɔ (0:1)</li> </ul>

Show distribution map

Current item: cheval

Score: 2.202964  
 — Within difference: -0.893600  
 — Between difference: 1.309364

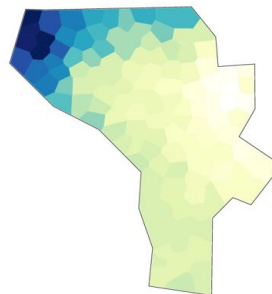
Difference map for *cheval* (note: darker link color means higher difference)



Step 4: show the distribution of relevant forms

Patterns in the cluster	Patterns not in the cluster
<ul style="list-style-type: none"> <li>fɪvɔw (3:3)</li> <li>fɪβɔw (5:8)</li> </ul>	<ul style="list-style-type: none"> <li>ʃaβ'al (0:9)</li> <li>tʃaβ'al (0:44)</li> <li>fɪβ'al (0:3)</li> <li>tʃɔβ'al (0:3)</li> <li>tsɔv'al (0:3)</li> <li>caβ'al (0:3)</li> <li>tʃaβ'al (0:24)</li> <li>tsɔβ'al (0:23)</li> </ul>

MDS maps: 2nd Dimension;  
 $r = 0.74-97$ , stress = 025  
 North-Western segment  
 of the dialect continuum



Back to data:  
Phonological endemisms  
& economy of phonological  
inventories (Clements)

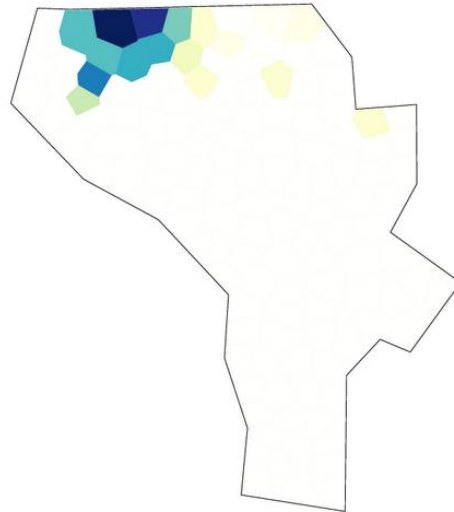
=> Markedness conditions in  
the diasystem

NB: the intricate correlation  
between the two  
phonotypes  
(implicational relation, and  
markedness hierarchy)

Alveopal\_S:  
Unvoiced >> Voiced)

=> Gabmap function  
«Characters»

## Hissing sibilants: unvoiced alveopalatal fricative (101 items)



found 101 items, showing all

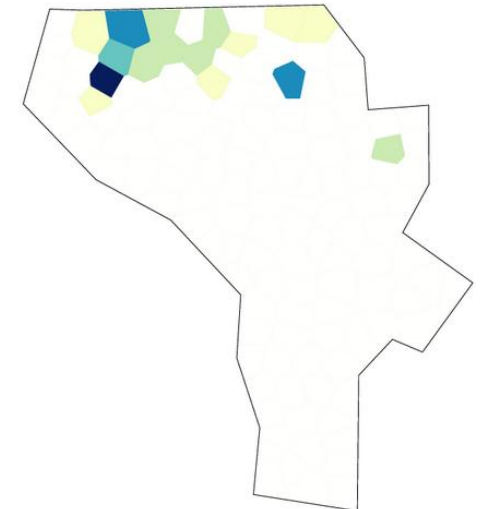
ɛɔwe'iko  
ALLOc 24.02 — saucisse

ɛjv'ɛɛke  
ALLOc 24.11 — eveque

ɛ'ur  
ALLOc 24.12 — sourd

ɛ'ɔ  
ALLOc 24.22 — sou

## Hissing sibilants: voiced alveopalatal fricative (35 items)



found 35 items, showing all

bez'ũ  
ALLOc 24.30 — besoin

ɛɔz'u  
ALLOc 46.15 — saisons

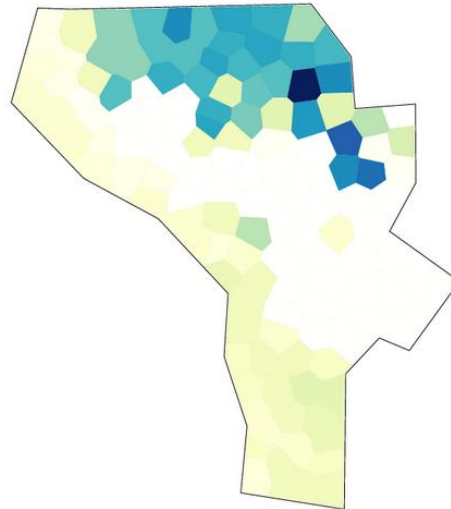
ɛaz'u  
ALLOc 24.03 — saisons

r'az'in  
ALLOc 24.20 — raisin

Back to data:  
Phonological endemisms  
« Local rules »  
=> Allophonical trends in  
the diasystem  
& Markedness  
hierarchy  
Palatoalv >> alveopal

Gabmap function:  
«Characters»

## Hushing sibilants: endemic unvoiced palatoalveolar



found 1263 items, showing 200

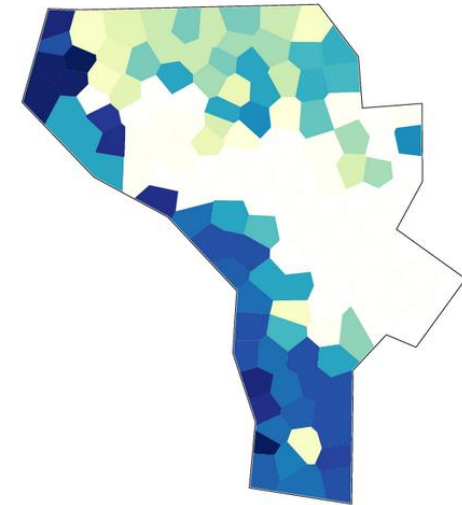
**furn'atʃe**  
ALLOc 11.02 — fromage

**k'ɛjʃ**  
ALLOc 46.12 — cuisse

**ʃiʒ'ew**  
ALLOc 24.14 — ciseaux

**k'ɛjʃ**  
ALLOc 12.21 — cuisse

## Hushing sibilants: endemic voiced palatoalveolar



found 884 items, showing 200

**furn'adʒe**  
ALLOc 09.22 — fromage

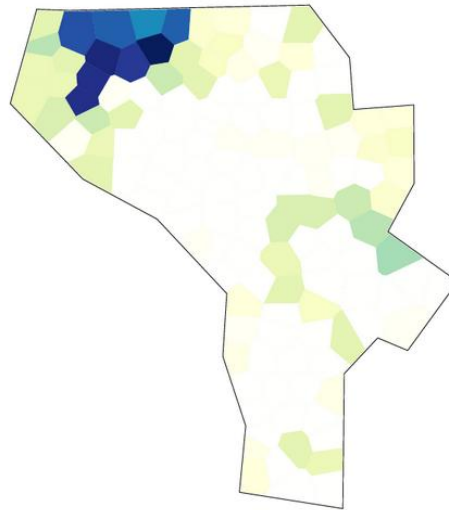
**ʃɔʒ'uʃ**  
ALLOc 46.31 — saisons

**kuʒ'i**  
ALLOc 24.10 — cousin

**r:aʒ'im**  
ALLOc 24.32 — raisin

«Rare sounds», as  
laminal fricatives,  
and reverse  
hierarchy of sonority  
(Lam[+cont.]:  
Voiced >> Unvoiced)  
=> Independent  
processes?  
=> Questions of  
General phonology

## Voiced laminal fricative (here, 200 items)



found 486 items, showing 200

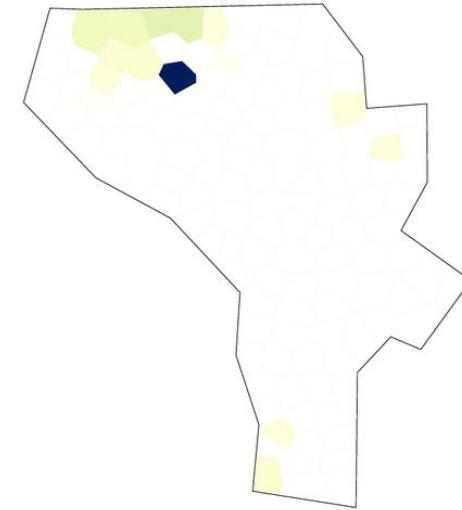
**ðɛ'd'armo**  
ALLOc 24.10 — gendarme

**tsum'ado**  
ALLOc 46.12 — journee

**zum'adoe**  
ALLOc 33.12 — journee

**medce'i**  
ALLOc 24.20 — medecin

## Unvoiced laminal fricative (59 items)



found 59 items, showing all

**boθ'i**  
ALLOc 12.05 — batir

**θav'al**  
ALLOc 24.20 — cheval

**θaβ'aw**  
ALLOc 24.22 — cheval

**θamin'ejo**  
ALLOc 24.21 — cheminee



## 4) Case study 2: a peripheral bow (Golfe du Lion)



## The benefit of this case study

- Analysis of interactions between vertices of a long, quasilinear graph (here, a dialect network):
- A coastline: shores and human groups, linked by a craft (fishing)
- A string of «graus» (gap in a lido), ponds and lidos at the foot of the Languedoc and Southern Provence hills («garrigues»).
- Empirically sound field data collected by Louis Michel in 1954-64 from fishing communities in the Golfe du Lion, in the languages of the corpus (Catalan and Occitan/Provençal).
- Digital capture of dialectal data from this documentary state on varieties of Roussillonese Catalan and Occitan, as a 'snapshot' at a given moment, in extremis before sociolinguistic assimilation.
- => high content of socio-historical indexicality of observable language dynamics, particularly through dialectometry (editing distance: Gabmap 2011- software).

A case study on  
the  
Northwestern  
Mediterranean  
coast:

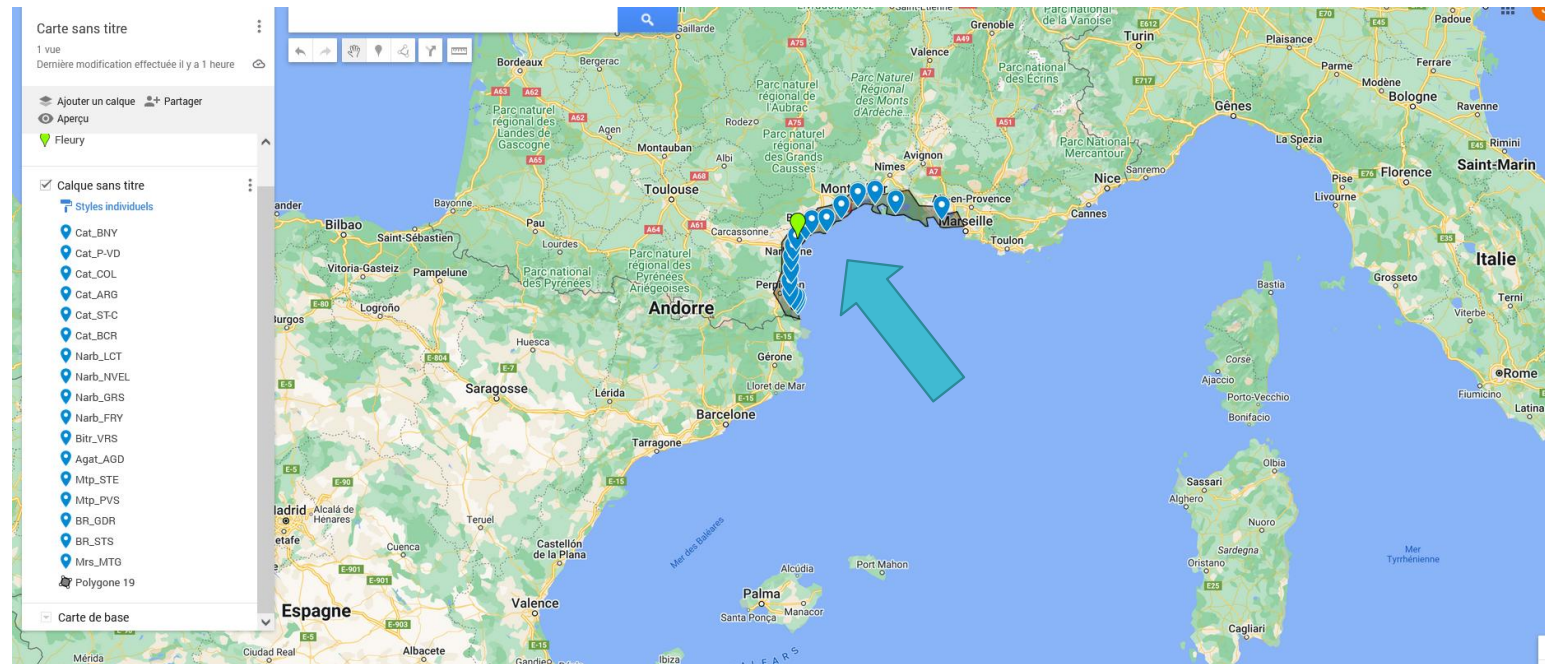
Golfe du Lion  
(source : Michel  
1964 : 142-186)



Source : [https://upload.wikimedia.org/wikipedia/commons/a/ad/Gulf\\_of\\_Lion\\_map.png](https://upload.wikimedia.org/wikipedia/commons/a/ad/Gulf_of_Lion_map.png)

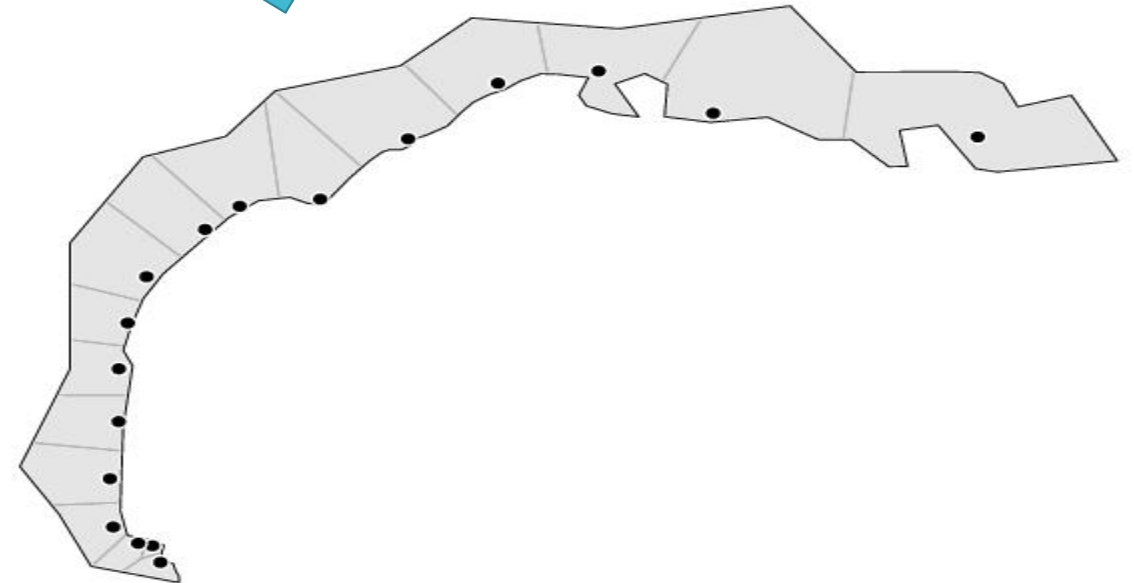
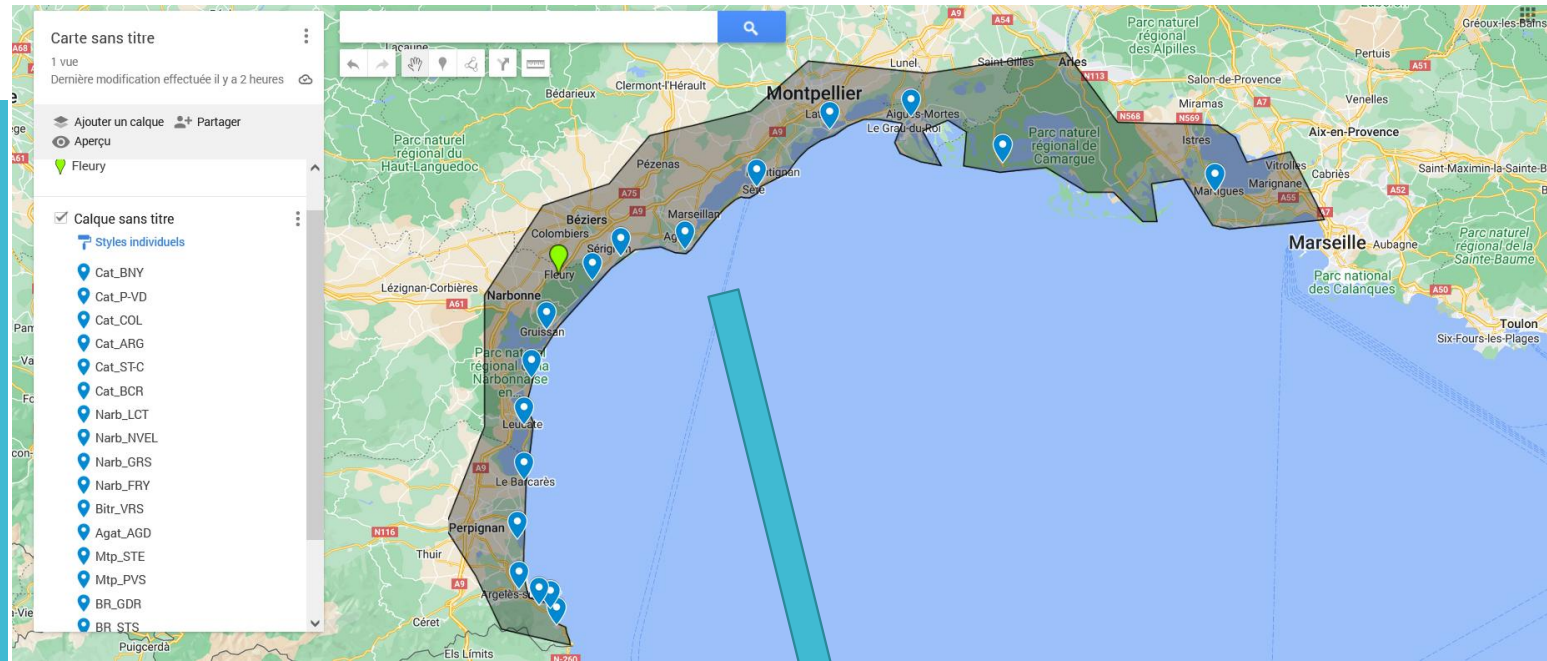


# Creating tags and polygons in Gabmap

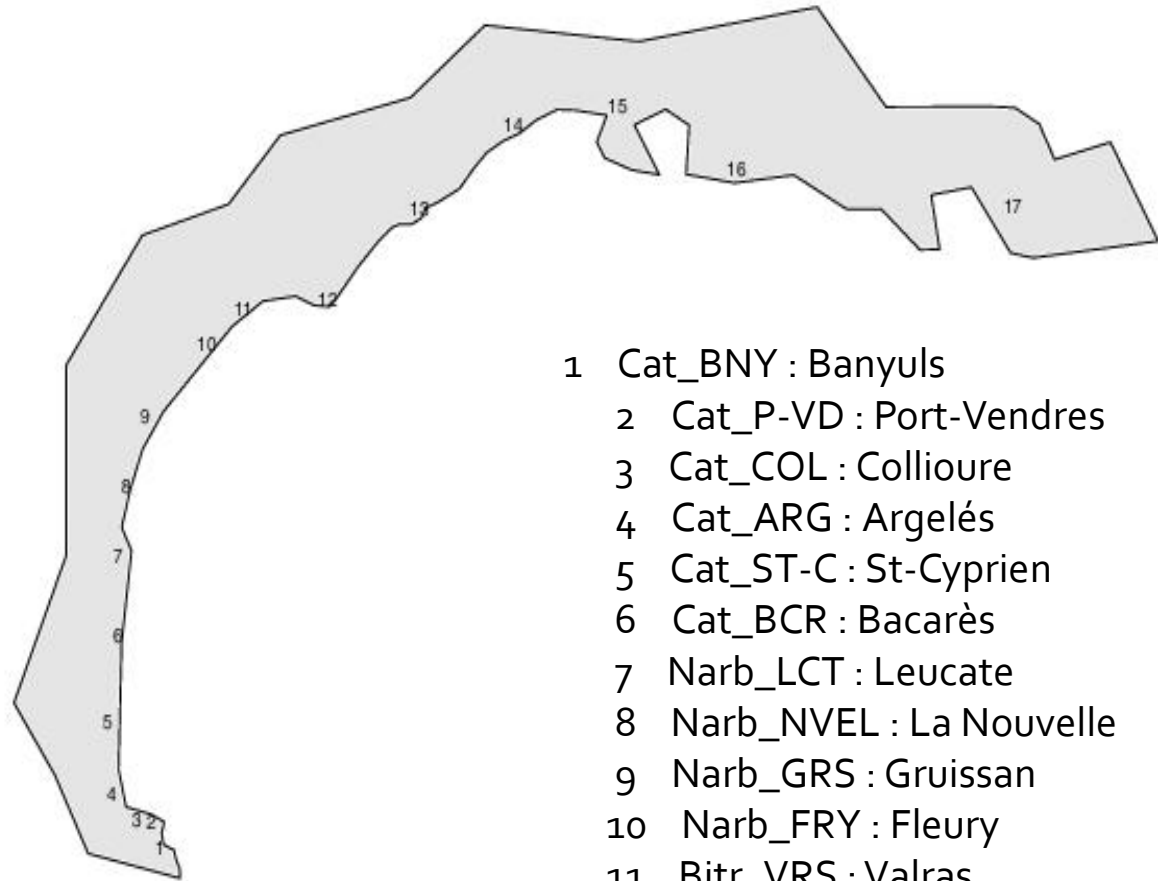




Plot for  
Gabmap of our  
coastal strip of  
the  
"fishermen's  
languages" of  
the Gulf of  
Lion (Michel,  
1964)



Localities between two languages (Catalan & Occitan, including Provençal) An arc of a maritime circle that is ecologically more complex than a simple "corniche"...



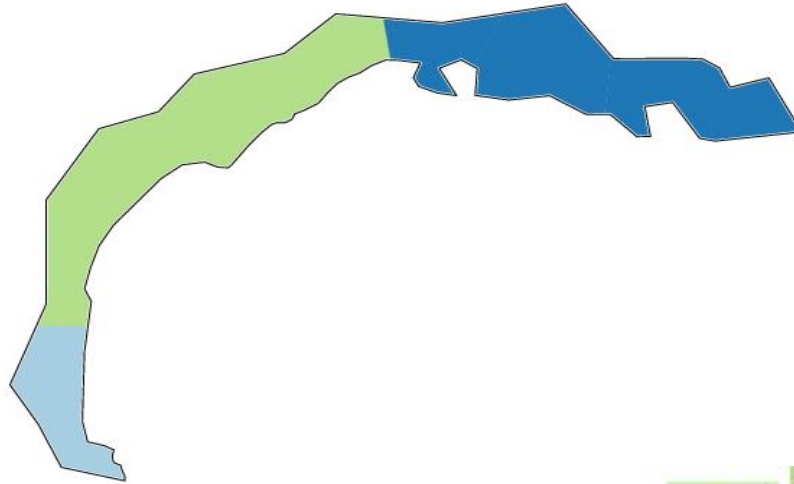
- 1 Cat\_BNY : Banyuls
- 2 Cat\_P-VD : Port-Vendres
- 3 Cat\_COL : Collioure
- 4 Cat\_ARG : Argelès
- 5 Cat\_ST-C : St-Cyprien
- 6 Cat\_BCR : Bacarès
- 7 Narb\_LCT : Leucate
- 8 Narb\_NVEL : La Nouvelle
- 9 Narb\_GRS : Gruissan
- 10 Narb\_FRY : Fleury
- 11 Bitr\_VRS : Valras
- 12 Agat\_AGD : Agde
- 13 Mtp\_STE : Sète
- 14 Mtp\_PVS : Palavas-les-Flots
- 15 BR\_GDR : Grau-du-Roi
- 16 BR\_STS : Les Saintes (Maries de la Mer)
- 17 Mrs\_MTG : Les Martigues

# Extraction of phonological data (& inflectional morphology)

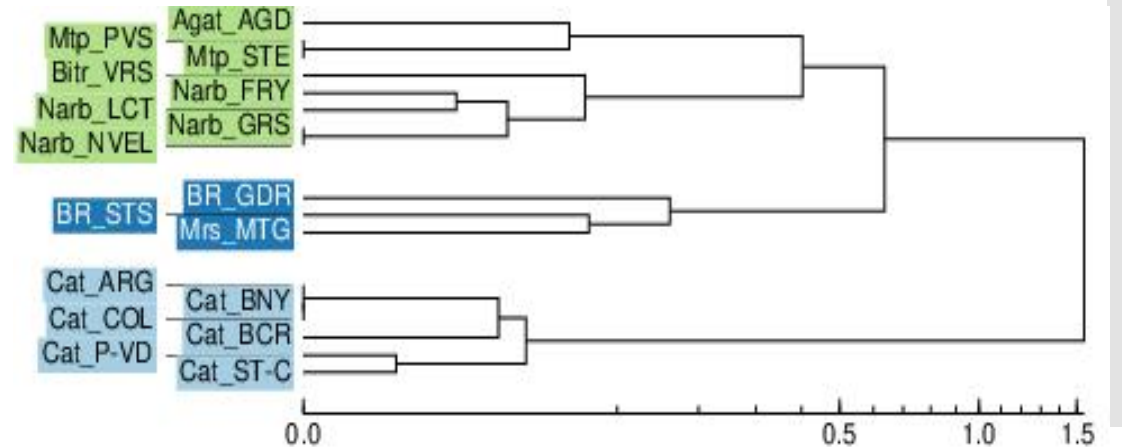
The blocks are very clearly grouped: Catalan (in light blue, to the south) vs Languedoc Occitan (in green, in the centre) & Provençal (in dark blue, to the north-east).

Places:	17
Items:	55
Instances:	935
Characters:	4245
Unique characters:	43
Tokens:	4242
Unique tokens:	43

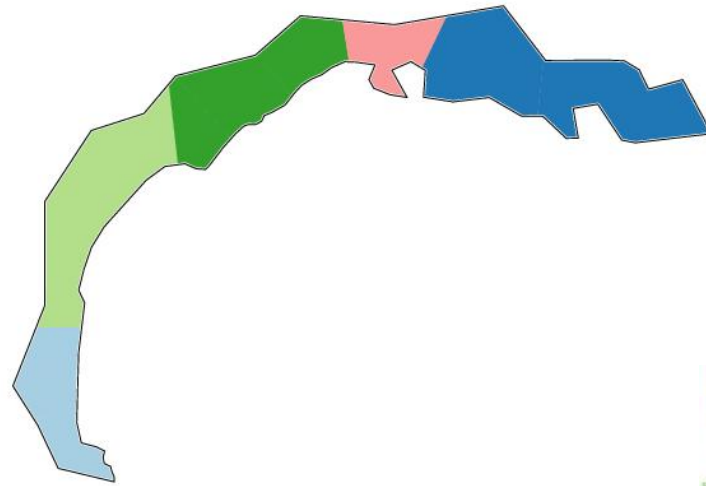
Ward's Method, phono & morpho flex, 55 items; 3 classes/intervals  
=> 3 "modalities" in the sense of don Ricardo (Dubert-Garcia 2021)



*Idem* : dendrogram :



CAH, phono & morpho flex,  
55 items ; 5 classes/intervalles

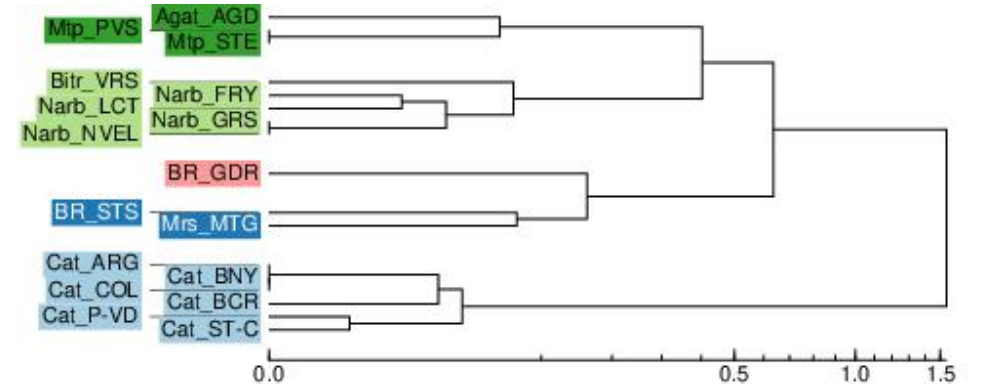


Données phonologiques  
(& morphologie  
flexionnelle)  
CAH, 5 classes

Le Grau du Roi apparaît  
bien comme variété  
transitionnelle du  
dendème/clade saintois-  
marseillais

Le catalan apparaît  
comme bloc structurant  
(outlier)

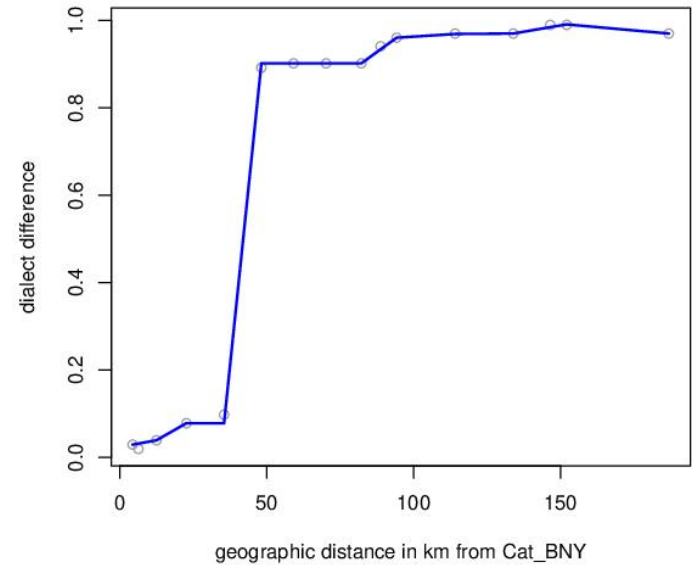
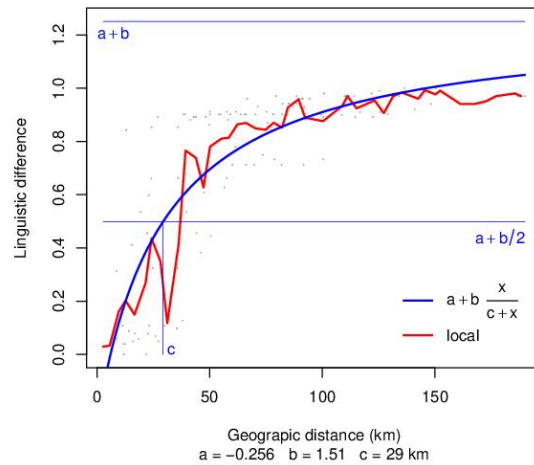
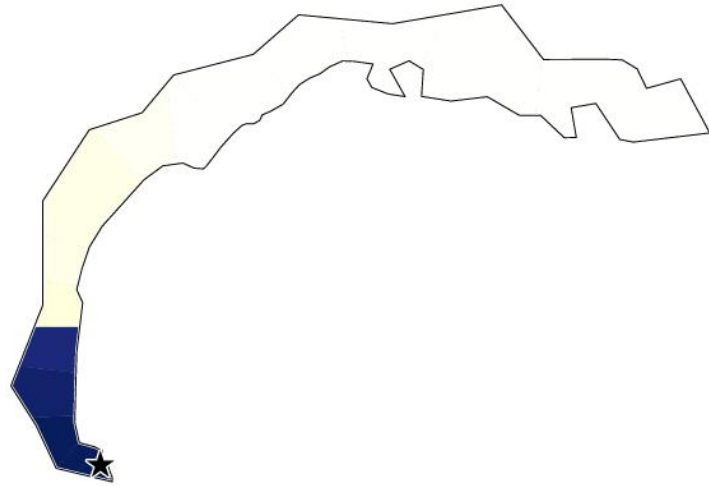
Idem : dendrogramme



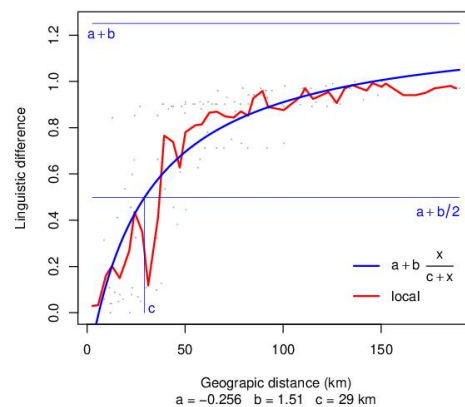
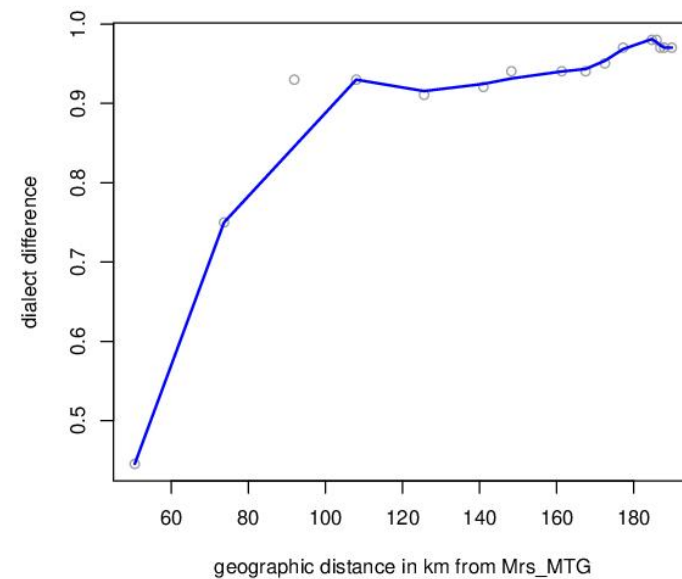
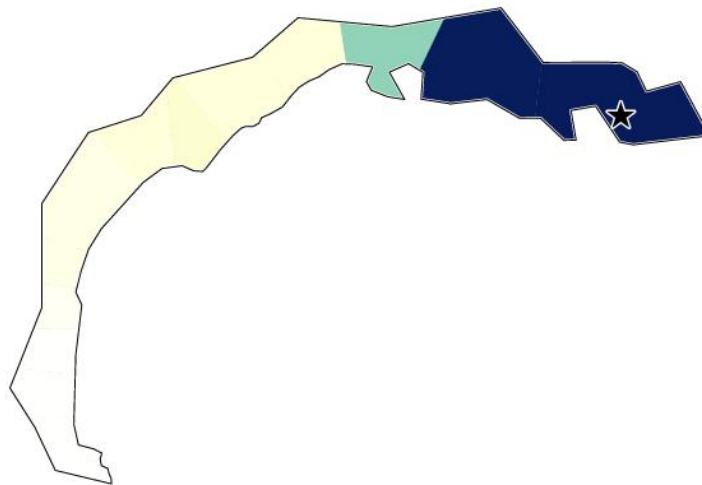


Linear distance  
(whole corpus),  
reference Point:  
Banuyls Catalan, at  
the southern end of  
the domain

=> language to  
language  
difference (Catalan  
vs Occitan)



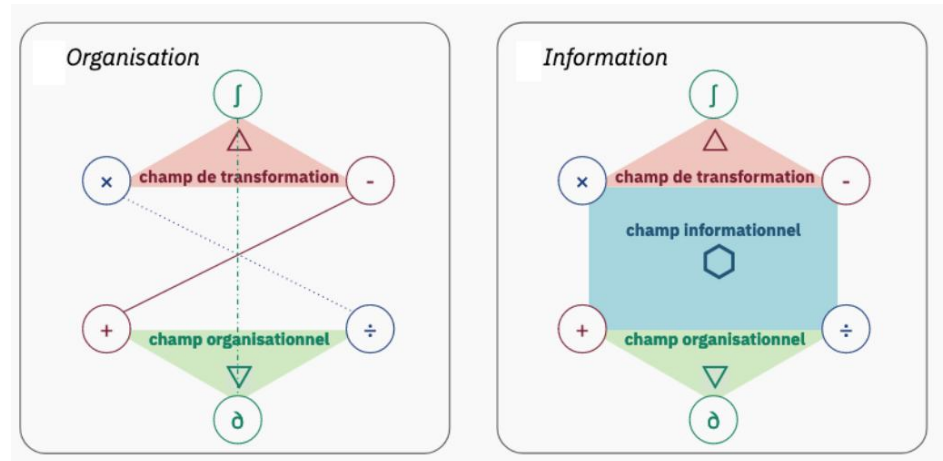
Maximum resonance of a Provençal variety: linear distance of Martigues from the rest of the dialect network under scrutiny



# Insights from the Golfe du Lion case study

- Two languages (Catalan vs Occitan), 3 varieties: gradient and thresholds of linguistic distance
- Sociolectal dimension (fishermen's varieties), to compare with hinterland varieties used by peasants («terriens»)

=> The challenge of  
Multidimensional dialectology for  
dialectometric tools and for  
Complexity Theory...



## 5) Prospects

- How these three sets of macro-, meso- and micro-geolinguistic empirical scales illustrate the relevance of quantitative geolinguistics for Complexity Theory?
- By providing the means to answer the two vexatae quaestiones which can be summed up as
- (1) the question of the ontology of dialects, and of the measurement instruments that allow us to classify language vs. dialect vs. sub-dialect vs. variety, and
- (2) beyond the dilemma of the dialectic between continuity vs discontinuity, a central question in science -that of the relationship between central modalities vs transitions (Dubert-Garcia 2021), which inform us about the fabric of interactive patterns in spatial dialectology, from different scales of observation (geographic as much as social, i.e. dialects vs sociolects) in all respect, a crucially relevant issue for CT.



# Merci

- ご親切にありがとうございました。
- thank you very much for your kind attention
- mercé plan per vòstra aimabla atencion
- mercés hèra entà la vòsta aimabla atencion
- moltes gràcies per la vostra amable atenció
- moltes gràcies per la seua amable atenció
- muchas gracias por su amable atención
- merci beaucoup pour votre aimable attention



# References

- Brun-Trigaud, G., Gaillard-Corvaglia, A., Léonard, J.-L., Darlu, P. (2014). Exploration cladistique de l'ALLOc (Atlas Linguistique du Languedoc Occidental), in C. Alén Garbato, Torreilles, C. & Verny, M.-J. (eds) (2014). Los que fan viure e tresluisir l'occitan, Actes du Xe Congrès de l'AIEO, Besièrs 12-19 juin 2011), Lemòtges : Lambert-Lucas, 437-448/963.
- Dubert-Garcia, F. (2021). A clasificación das variedades dialectais do galego de Ricardo Carvalho Calero, Madrygal. Revista de Estudios Galegos, 24, 69-80.
- Lacombe, Eric, 2023, Vers un langage pour exprimer la complexité. LHUMAINE, (2). <https://lhumaine.numerev.com/articles/revue-2/2959-vers-un-langage-pour-exprimer-la-complexite>
- Lacombe, Eric, 2021, *Transformation numérique des organisations en réseaux : les potentiels d'une schématisation dynamique de l'information*. Thèse de doctorat en Sciences de l'Information et de la communication, PhD dissertation, dir. Lise Vieira, Université de Bordeaux 3
- Léonard, J. L., Brun-Trigaud G. & Picard F. (2023, forthcoming). Atlas linguistiques et perspectives dialectométriques. In Sibille Jean & Louise Esher (éds). Manuel de linguistique occitane, Berlin, Mouton.
- Léonard J. L. & Brun-Trigaud G. (2022). " Théories et pratiques en dialectologie générale : empirisme critique et intégration macro-systémique des endémismes. Master et doctorat. Séminaire Théories et pratiques linguistiques dirigé par Samir Bajrić, le 8 novembre 2022, Université de Dijon., Dijon, France. 2022. hal-03916232.
- Léonard, J. L., Gélébert, G. (2023). Gnose de Théobald Lalanne, dialectométrie et complexité. LHUMAINE, (2), accessible sur <https://lhumaine.numerev.com/articles/revue-2/2893-gnose-de-theobald-lalanne-dialectometrie-et-complexite>.
- Léonard, J. L., Brun-Trigaud G., Gélébert, G. (2023). Retour de la Dauna à... Le « paysage dialectal » des landes gasconnes revisitées : de Millardet et Lalanne à Lartigue. De patois à éradiquer à langue de création artistique: avatars et représentations de l'occitan dans l'histoire, Agresti, Giovanni, Feb 2023, Bordeaux (Gironde), France. ( hal-04054509)
- Michel, L. (1964). La langue des pêcheurs du Golfe du Lion. Introduction : I. L'homme et la mer, II. Dialectologie côtière, Paris, d'Artrey.

# Références en dialectométrie

- Camps, Christian, 1986, Limites linguistiques d'après l'Atlas linguistique et ethnographique du Languedoc occidental. *Actes du XVIIe Congrès international de Linguistique et Philologie Romanes. Aix-en-Provence, 1983. Université de Provence - Jeanne Laffitte, Marseille, t. 6, 120-135. (VOIR à la maison).* Nerbonne John & al. 2011, Gabmap - a web application for dialectology, *Dialectologia: revista electrònica*, p. 65-89.
- Nerbonne John & Wilbert Heeringa, 2001, Computational Comparison and Classification of Dialects, *Dialectologia et Geolinguistica* 9, 69-83.
- Nerbonne John, Ilse van Gemert & Wilbert Heeringa 2005, A Dialectometric View of Linguistic "Gravity", Alpha Informatica report, University of Groningen, accessible sur <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.530.5957>, consulté le 24/08/2022.
- Nerbonne, John & William Kretzschmar 2003, Introducing Computational Techniques in Dialectometry, *Computers and the Humanities* 37, 245-255.
- Séguy, Jean, 1971, "La relation entre la distance spatiale et la distance lexicale", *Revue de Linguistique Romane*, pp. 335-357.
- Séguy, Jean, 1973, "La fonction minimale du dialecte", in Gardette, Pierre (éd.) : *Les dialectes romans de France à la lumière des atlas linguistiques régionaux*, Paris, éd. CNRS, pp. 21-42.

**SYNOPSIS of the TALK**

Jean Léo Léonard, Guylaine Brun-Trigaud, Flore Picard &amp; Louise Albinet, 2023

Talk COMPILA 23 4th of August 2023, Tokyo, *Webinaire***Complexity in Language Variation and Change (COMPILA 2023)**

Room Collaboration 1 (4th floor), Building 18, Komaba I campus, University of Tokyo, Japan

Satellite meeting for the 28th International Conference on Statistical Physics (StatPhys28)

August 4th (Fri.) 2023

**Revisiting Southern Gallo-Romance from a Complexity Theory viewpoint: Occitan (Southern Gallo-Romance)****Slide\_2**

- 1) Introduction: preliminary remarks
- 2) Context, data, method and purpose: Occitan as a dialect network
- 3) Case Study 1: a central dialect ( ALLOc data, THESOC): Western Languedocian
- 4) Case study 2: a peripheral bow (Golfe du Lion)
- 5) Prospects

**Slide\_3**

Our approach differs from ASJP. We advocate using dialectometry, particularly LD, for studying dialect continua and inner linguistic diversity, not long-range phylogenetic kinship like cladistics, which is more adequate for this purpose.

LD is optimally heuristic for understanding linguistic landscapes, language dynamics, history, and social fabric quantitatively, focusing on typological trends rather than phylogenetic evidence.

LD not only highlights synchronic dialectal patterns but also reveals sociolinguistic/sociolectal patterns to some extent (see the Golfe du Lion component of this talk).

We highly praise Gabmap as an outstanding heuristic tool. The team behind it (Nerbonne, Heeringa, Bolognesi, Prokić, Weiling, etc.) has shown remarkable epistemological awareness and integrated strategic mathematical and computational devices into its design, enhancing heuristic accuracy in processing dialectal complexity (see below).

**Slide\_4** : The Interplay of Tensions in a Field of Transformation (ITFiT) according to Eric Lacombe : If CT in a nutshell can be described as

- a) Flow of information
  - b) Self-organization
  - c) Emergence of (critical-state) patterns
- Then, the ITFiT Model grasps this set of processes as in this figure; in which three component interact :
- **Organisational** trends at the top, in green (networks, vicariance)
  - **Tension multipliers** at bottom, in blue (space & emic/etic categorisations)
  - **Unifying** trends (place & normalising action) at bottom, pink colour

We'll apply this model to geolinguistic space/topologies, thanks to DM.

Processes of reproduction, diversification, dispersion, heterogeneization compete with processes of unification, as selection, reduction, compression or conflation, homogeneisation (analogy, metatypes, etc.)

**Slide 5**

Organisation & Information Fields, according to Lacombe, E. (2023). *Vers un langage pour exprimer la complexité*. LHUMAINE, (2).

Four operations prevail on natural as much as on symbolic (semiotic) entities/objects :

(x) = Multiplicans, (-) = Subtrahens, (+) = Addens, (./) = Dividens

& two symbols (∫, ∂) for Tertium Inclusum



as to be seen in the vertices of the graphs. These basic operations combine within fields of information flows & organisation settings.

Now, some hints at linguistic geography of the Occitan domain :

Slide 6 : the main SIX dialects of Occitan

⇒ Let's get into section 2 of the talk.

**Slides 7-8** : Yuri B. Koryakov's map of the three Gallo-Romance Dialect networks -a canonical view.

i.e. a « Textbook categorisation » of the Gallo-Romance Dialect networks dialectical space

NB : North (Oïl dialects) in green has a twofold pattern : a *core* (Paris/île de France) vs *periphery*, out of strongly *unifying trends* since Middle-ages

South : in light and deep purple stands for the Occitan dialectical space, with a wide center and a lateral area in the SouthWest (Gascon).

We won't talk about the Francoprovençal area, in the Central-Eastern part of the territory (around Lyon and Savoie).

**Slide 9 :**

A canonical map as this one by Van den Bossche, from Wikipedia, shows only what Nerbonne & Kretschmar call « Attributive » dialects : emic entities or parts of a wider diasystem. But as such, it lacks intrication and hierarchy. Such a map does not tell us anything about the inner structure (IS) of the diasystem. We need taxonomies, or... Networks, unravelling the potential interactions within the diasystem as a Complex System. .

**Slide 10 :**

By the way, why Occitan?

A few advantages, for a survey of geolinguistic complexity, of the Occitan dialect network:

Huge territory

Compact area

Critical mass of reliable data, thanks to the THESOC (a big database, gathering data from at least 6 linguistic atlases)

**Parameters** : 71 phonological maps

232 lexical mapsThe area is well studied, stable (though sometimes challenging) « attributive dialects », as in van den Bosche's map of Occitan dialects.

Geographic and historical complexity

Polycentric, former urban dialects (or not), strong entities (Provençal, Gascon, Languedocien, Limousin)

**Slides 11-13**

How to get access to the Online application Gabmap.

**Slide 13** : how it works, on the basis of three elementary operations

Adding, dropping, changing (phonology)

⇒ ADDITION (+), SUBTRACTION (-)

TRANSFORMATION/SUBSTITUTION (x)

Operating on CV templates (string chains of characters)

⇒ Gabmap has a strong « realisational drive », as compared to classical dialectometry, which relies more on typological indexation of taxa (Goebl, Salzburg).

⇒ It matches to some extent CVCV and other Templatic dependency models in modern phonology (Scheer, Lowenstamm, etc.)

**Slide 14- 20**

### 2.1. Geolinguistic complexity: Occitan areas and choremes

Here we aim at defining the main entities (dialects) which make up the Occitan diasystem :

- 1) Results with Gabmap match the « canonical approach » on the basis of isoglosses, as Pierre Bec in his 1973 textbook « La llengua occitane »
- 2) We get more than basic display of geolinguistic patterns (entities, areas) : we can provide hierarchies, taxonomies, i.e. configurations.
- 3) We can get vicarious/alternative results, as in slide 19-20 : compare Complete link reductionism vs Ward's method cumulative clustering (and Bec's hyperdialects according to isoglosses, in the middle of slide 19).

#### Slides 21-23

Gabmap provides Fuzzy cluster maps, which enable to visualize the interplay between core-areas and transitions in a dialectic space : here for two diasystemic parameters : a « closed-system » *phonology* (71 items) vs an open-system *lexicon* (over 230 items)

- ⇒ This enhances complexity in the dialect continuum. It combines cores and gradients
- ⇒ It suggests parametric variation in transition dynamics -phonology being less intricated than the lexicon.
- ⇒ In slide 23, we oppose M (Modalities) or core entities to T (Transitions) and S (Singletons, Outliers). A triad we'll soon need when handling the same data from a *network perspective*.
- ⇒

#### Slides 24-33

The former maps and dendrograms (i.e. topologies and taxonomies), PLUS sets of functional entities, such as *core areas*, *transitions* and *singletons*. It tells us much more than categorisations available in textbooks. But now we need more than just spotting the « good old » components of the Occitan dialect continuum...

OK, we know much about the inner structure of the diasystem. But we still lack information about *inner consistency/cohesion* and *connectivity* of the various components of the diasystem and its local emerging patterns.

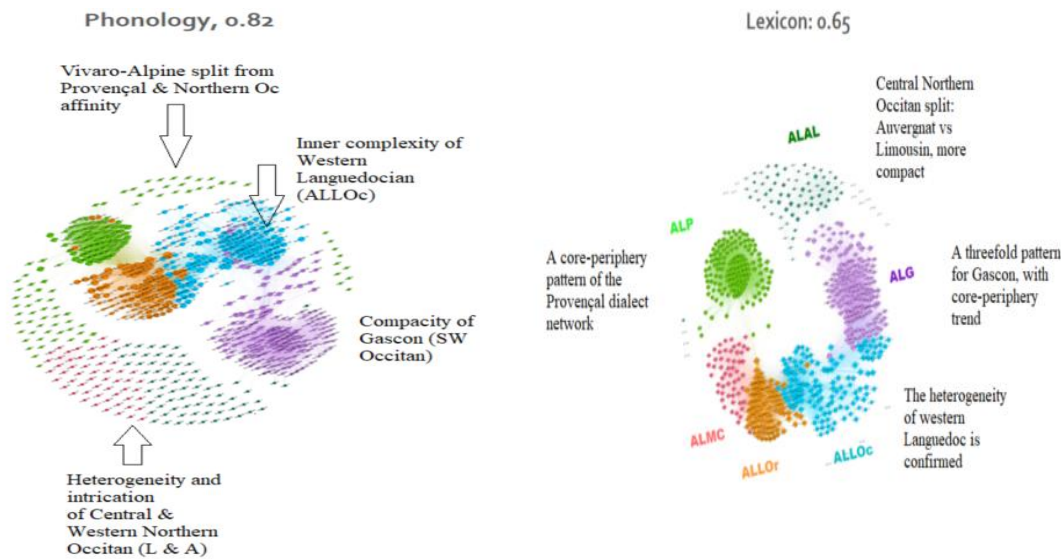
So, we'll use graph-streaming, using another tool : the Gephi application, also freely available Online.

These graphs show thresholds of emerging states of the unifying vs diversifying trends.

The lower the index (0.65 in lexicon), the denser the clouds of localities in the various atlases from the THESOC database (ALG, ALP, ALMC, etc.). The higher the index (0.90 in phonology, slide 90), the less compact the (core)clusters

#### Slides 34-35

What are the benefits of this method? Let's see what those clouds suggest to us, for further research on the strategic components within the overall and the local networks :



### (3) A central dialect ( ALLOc data, THESOC) Western Languedocian

What are we up to, with this case study ?

Well, it's supposed to be a strategic entity within the Central Unit of the Occitan diasystem (slide 37)  
It's time to have an eye on such a sensitive component of the diasystem, still keeping in mind Lacombe's model : we'll see the interplay of multiplying and unifying tensions is going on too, at this componential level -a recursive, « fractal » property of complex systems.

#### Slides 39-40

Localities : 131

#### Slides 41-47

Application of the geolinguistic or areal (maps) and taxonomic (dendrograms : trees) protocol to phonological data (199 items) on the ALLOc data, from THESOC

#### Slide 48

As we already said, most dialectological methods, especially based on linguistic distance, are more typology-driven than phylogeny-driven.

In this particular case, we previously handled the same ALLOc data from THESOC with a more phylogenetic approach, i.e. cladistics, thanks to Pierra Darlu, genetician at Inserm and CNRS, using PAUP (paper published in 2014).

The interesting thing with Cladistics is that, unlike LD or other similarity/distance based methods, Cladistics enhances derivational patterns between characters. It provides chains of interdependent phonological changes, ending up in « patristic distances ». It also uses bootstrap, to get an « ultimate », unique taxonomy, instead of the many vicariant algorithms in use in dialectometry (Ward'sMethod, Complete link, etc.).

So what ?

⇒ In slide 48, we see that the overall display of areas is similar, between our Gabmap WM results and Darlu & al.'s map (top right of the slide).

⇒

#### Slides 48-51

Nevertheless, the taxonomical hierarchy differs, as shows slide 51, although in a very predictable way : the cladistic approach conflates areas I and II in the peripheral north and NW in a separate (allogenic) cluster, whereas these peripheral areas are more embedded in the central diasystem.

One protocole (Cladistics) is more sensitive to *Initial Conditions*, as etyma are used to derive the characters one from another in the matrix, whereas the other protocol (LD, Gabmap) does not integrate this dimension in its computation of linguistic distance, working from a synchronic standpoint. Interestingly enough, both methods converge in detecting dialectical aggregates, but they differ in clustering these entities in their taxonomies.

### Slides 52-57

Nevertheless, Gabmap using LD includes MDS analysis of data.

This technique allows us to strip one by one the layers of the diasystem as if it would be an onion.

The MDS plot in slides **52-53** provide a more gradient overview of the display of varieties in the dialect continuum. To some extent, it vindicates the cladistic method, as block B for instance, in slide 53, is the same NW allogeous area we had in Darlu's map (area I in this taxonomy).

The various layers, according to stress on sets of discrete phonological variables responsible for the variegated dynamics of geolects, appear in slides 55 and 56.

**Slide 55** provides a holistic snapshot of the inner intricacy of the dialectical gradient, while stress hits at 0.77 in MDS contributes to vindicate the Cladistic approach, as Darlu's area II now also appears neatly in the upper North. We still have a gain, here : this MDS layer shows the fine-grain of structural convergence throughout the rest of the dialect continuum, in the centre and even down to the South.

This layer also suggest both Darlu's areas strongly differ between themselves-a fine-grained typological fact Darlu's cladogramme could not grasp : the NW area (Darlu's area I) is attracted by Gascon, whereas area II, in plain North is attracted to Northern Occitan Limousin dialect (on the road to Limoges).

### Slides 58-61

Can Gephi graphs help us to complete the picture ?

These graphs provide insights on connectivity within communal aggregates. LD score are converted into similarity scores, as previously with the bulk of the THESOC corpus. Here, as for the previous set of results, only phonology is under scrutiny.

Results clearly point at a threefold patterning of the dialect network, already emerging at threshold 85/100, in slide **58**.

In slide 61, we compare a clear-cut ternary complex of varieties embedded in subnetwork components to the 4th dimension of MDS, at stress 0.17. There is some correlation between the two artefacts, but the Gephi networks may be more heuristic to delve more into the fabric of the diasystem.

### Slides 62-64

Further research should be done of the structural properties of singletons or outliers, and their metonymic relation to the rest of the dialect network.

We therefore see very different patterns of linear distance (correlating linguistic distance to kilometric distance) for some singletons/outliers, as ALLOc 24.22 in the NW (slide 62) and 31.32 right in the core of the Central Unit i.e. CU (the CU within the Occitan CU...).

Each profile is fairly asymmetric : the former starts at higher level of differentiation, and is less geography-sensitive with the remainder of the dialect network, whereas the latter, embedded in the core area as it is, starts from a lower threshold of difference, which climbs more abruptly as it encounters northern varieties –under Gascon or Limousin influence, in the NW and North, respectively.

### Slides 65



What did this case study tell us?

What benefits, from the standpoint of Complexity Theory?

- 1) Vicariance: embedding & comparison of several techniques/methods (LD vs Cladistics)
- 2) Dialect Network (Gabmap, Gephi) vs Dialect Continuum (MDS, from Gabmap)
- 3) Finer grain ontology & agentivity: the quest for heuristic outliers (Tertium Comparationis or Inclusum?)
- 4) Multidimensional dialectology through facts and various techniques of visualisation.

### Slides 66-69

A survey of some heuristic hints at the data : shibboleths on the one hand, markedness conditions of reflexes on the other.

### Slides 70-80

The Golfe du Lion case study.

Slide 71 : Analysis of interactions between vertices of a long, quasilinear graph (here, a dialect network):

A coastline: shores and human groups, linked by a craft (fishing)

A string of «graus» (gap in a lido), ponds and lidos at the foot of the Languedoc and Southern Provence hills («garrigues»).

Empirically sound field data collected by Louis Michel in 1954-64 from fishing communities in the Golfe du Lion, in the languages of the corpus (Catalan and Occitan/Provençal).

Digital capture of dialectal data from this documentary state on varieties of Roussillonese Catalan and Occitan, as a 'snapshot' at a given moment, in extremis before sociolinguistic assimilation.

=> high content of socio-historical indexicality of observable language dynamics, particularly through dialectometry (editing distance: Gabmap 2011- software).

### Slide 80

Two languages (Catalan vs Occitan), 3 varieties: gradient and thresholds of linguistic distance

Sociolectal dimension (fishermen's varieties), to compare with hinterland varieties used by peasants («terriens»)

=> The challenge of Multidimensional dialectology for dialectometric tools and for Complexity Theory...

### Slide 81

How these three sets of macro-, meso- and micro-geolinguistic empirical scales illustrate the relevance of quantitative geolinguistics for Complexity Theory?

By providing the means to answer the two *vexatae quaestiones* of the existence of dialects and the continuity/discontinuity challenge to the ontology of complex systems which can be summed up as

- (1) the question of the ontology of dialects, and of the measurement instruments that allow us to classify language vs. dialect vs. sub-dialect vs. variety, and
- (2) beyond the dilemma of the dialectic between continuity vs discontinuity, a central question in science -that of the relationship between central modalities vs transitions (Dubert-Garcia 2021), which inform us about the fabric of interactive patterns in spatial dialectology, from different scales of observation (geographic as much as social, i.e. dialects vs sociolects) in all respect, a crucially relevant issue for CT.

