

Emergence forte *versus* émergence faible en écologie linguistique : le cas du mazatec

Colloque de l'Institut Universitaire de France : "Nature et Culture »

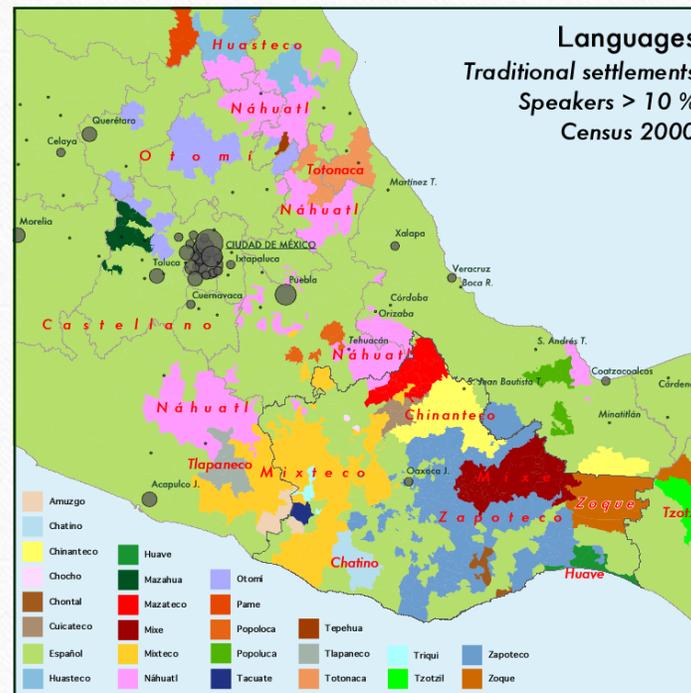
Dijon, 27-28 mai 2014

Jean Léo Léonard, Paris 3-CNRS, UMR 7018, en collaboration avec Vittorio dell'Aquila & Daniele Dalmaso (CELE, srl), Antonella Gaillard-Corvaglia (Paris 3/Inalco) & Stéphane Balev (RI2C, ISCN, Le Havre) & Pierre Darlu (CNRS).

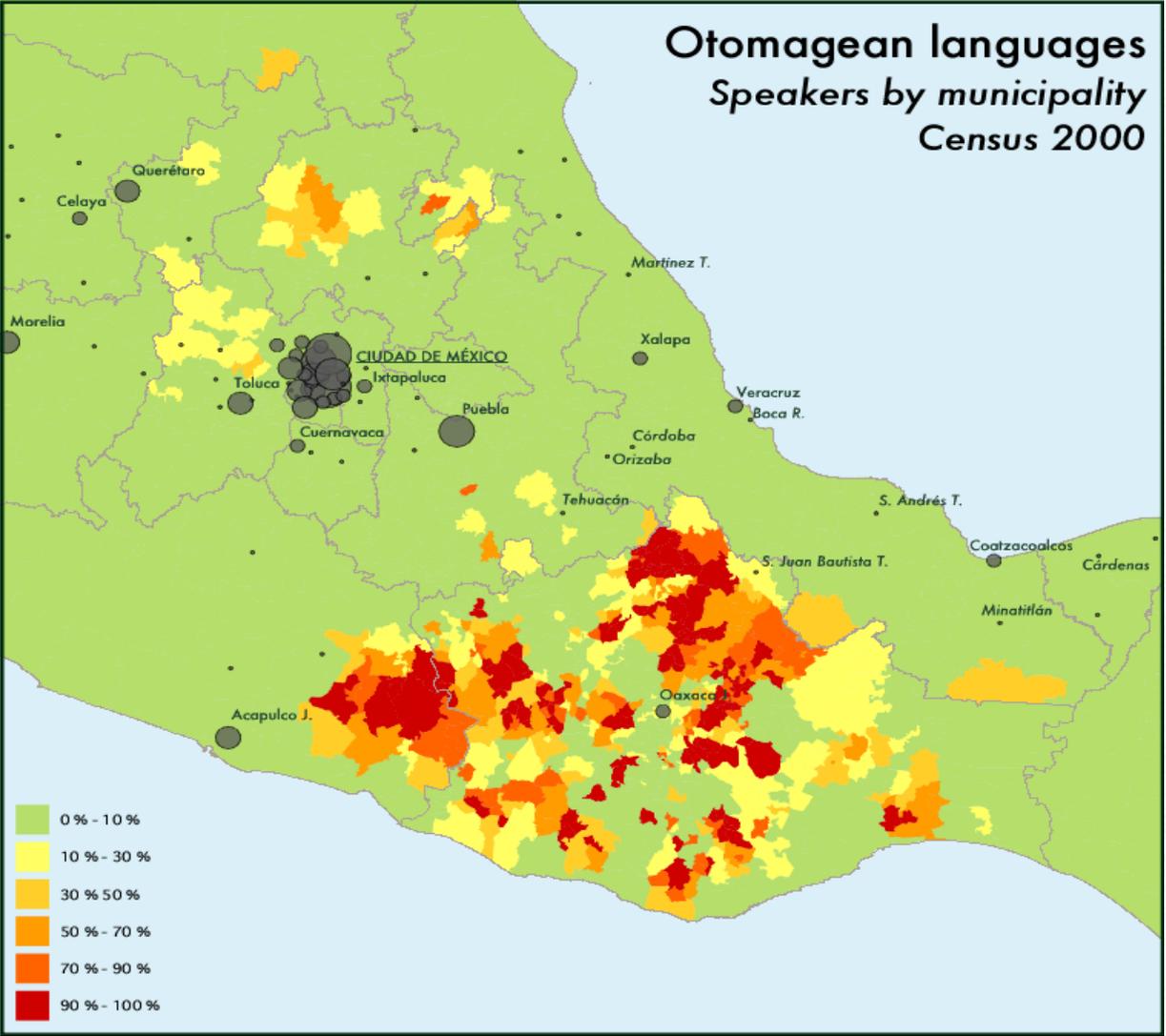
Plan

- 1. Contextualisation
- 2. Base empirique
- 3. Modélisation
 - 3.1. Facteurs externes
 - 3.2. Cladistique (émergence phylogénétique)
 - 3.3. Distance d'édition (algorithme de Levenshtein)
 - 3.4. Réseaux épigénétiques (intercompréhension)
- 4. Conclusion

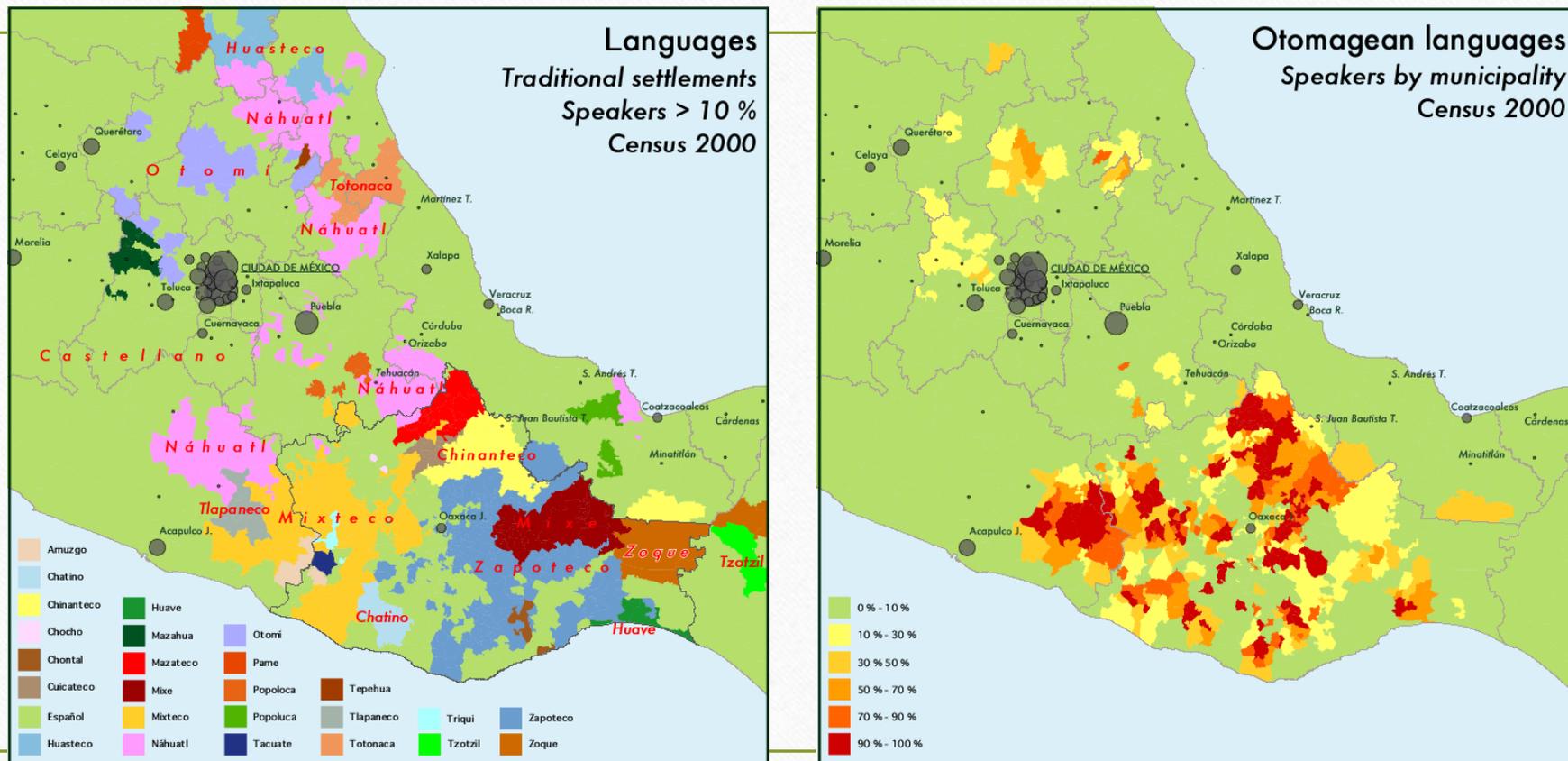
1. Contextualisation



Otomagean languages Speakers by municipality Census 2000



L'archipel vertical (cf. John Victor Murra), ou l'archipel des *versants interactifs* (cf. Nichols, 2004).



Langues Otomangues (Carte : Jossierand, 1983)

Map III-1. The Distribution of Otomanguean Languages in Mexico



Contextualisation

- Le mazatec : une langue otomangue du sud-est du Mexique
- Plus de 220 000 locuteurs
- Un biotope tropical fortement diversifié, sur trois paliers géographiques
- Trois systèmes de cultures postcoloniales :
 - basses terres (système **canne à sucre**),
 - terres moyennes (**élevage**)
 - hautes terres (système **café**, en crise depuis plus de 30 ans)
- Ces systèmes ont supplanté la polyculture autochtone - la *milpa* mésoaméricaine : maïs, courge, haricot et piment), qui reste cependant *résiliente*. Elle assure même la subsistance et la survie d'une grande partie de la population actuelle de la région de peuplement autochtone.

2. Base empirique

- Bases de données dialectologiques (Kirk, 1966, ALMaz 2010-13)
 - Données socioéconomiques officielles
- ⇒ Dynamiques d'émergence de la variation dialectale du mazatec au cours de l'histoire (cf. résultats cladistiques & algorithme de Levenshtein).
- ⇒ Complexité socioterritoriale de l'interaction entre nature et culture, à travers le prisme des configurations géolinguistiques.

Structures émergentes multiplexes

1) L'émergence forte d'une relation adaptative culture/nature (où la *culture* s'adapte à la *nature*)

=> auto-organisation d'un *flux* de variation dialectale des différentes communautés de peuplement *coordonné entre biotopes* en relation de *complémentarité*.

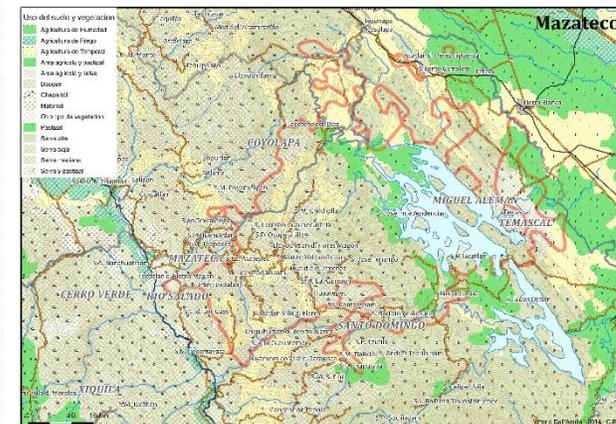
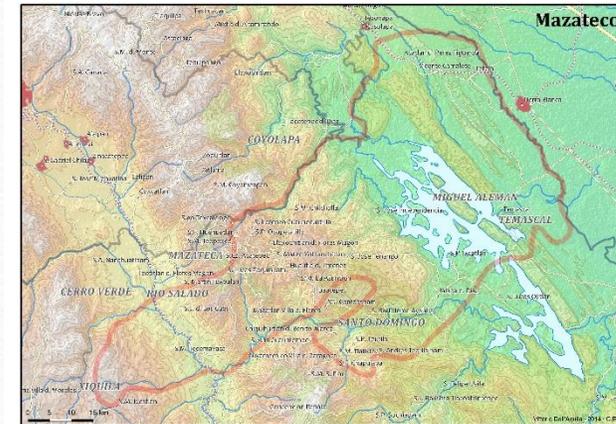
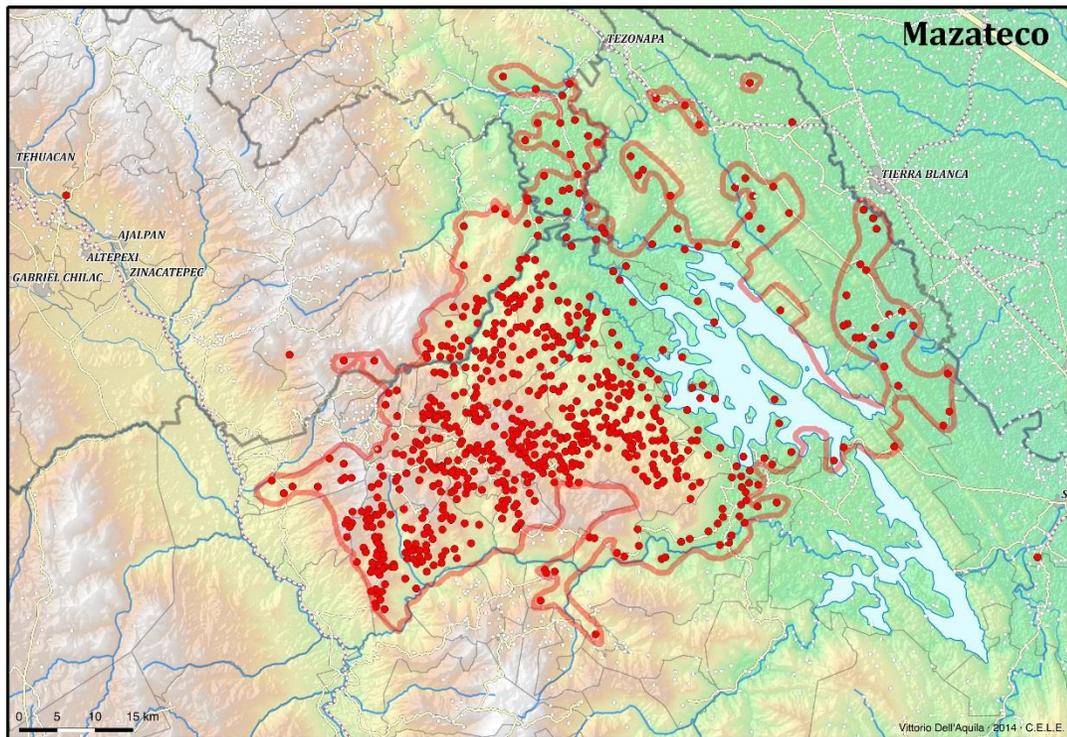
2) L'émergence de communautés écohumaines réparties sur un *éventail dialectal* reléguée à une dynamique d'émergence faible

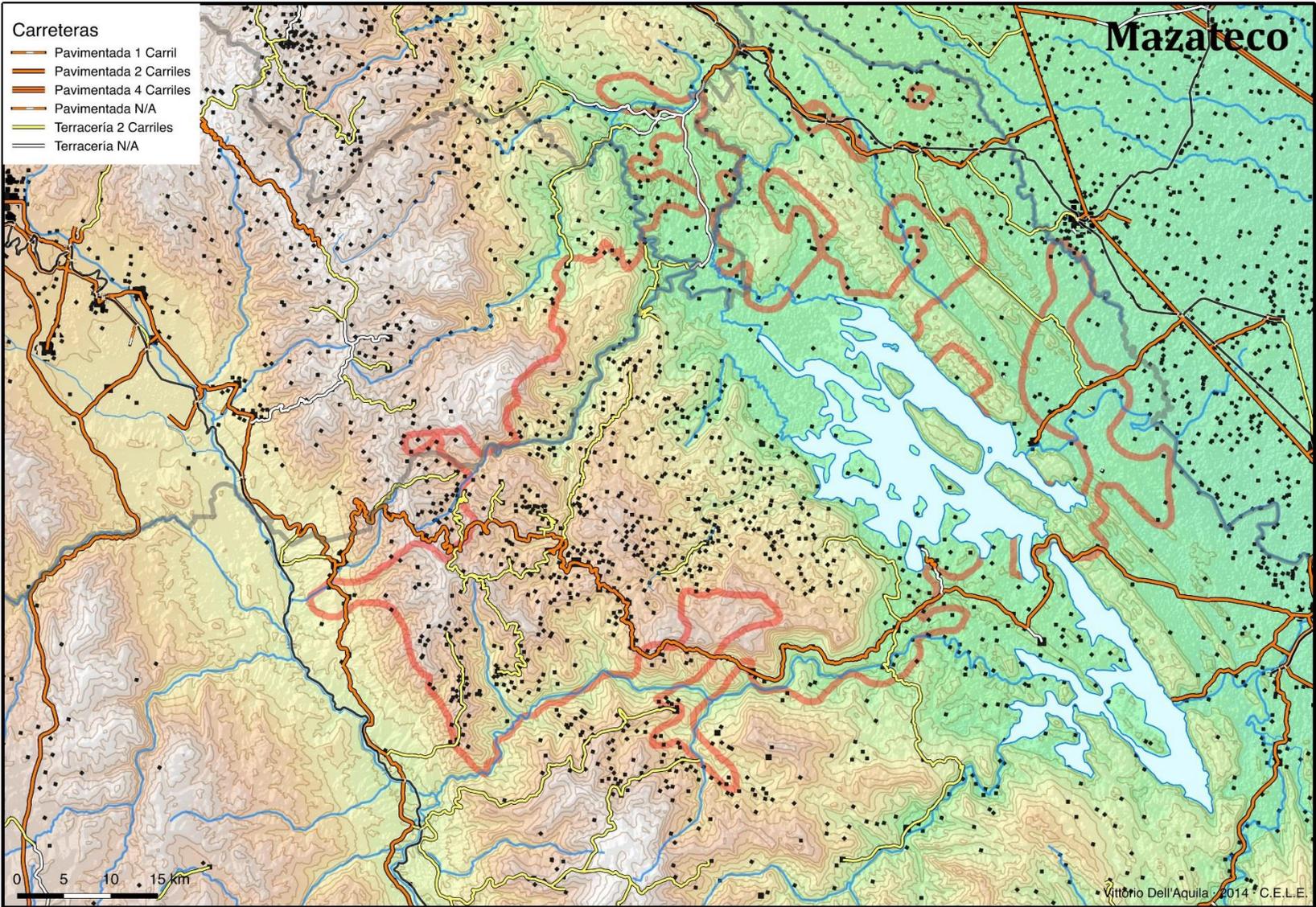
=> Relation nature/système où la *nature* doit s'adapter à des *systèmes productifs*, menant à une crise de la relation homme/environnement.

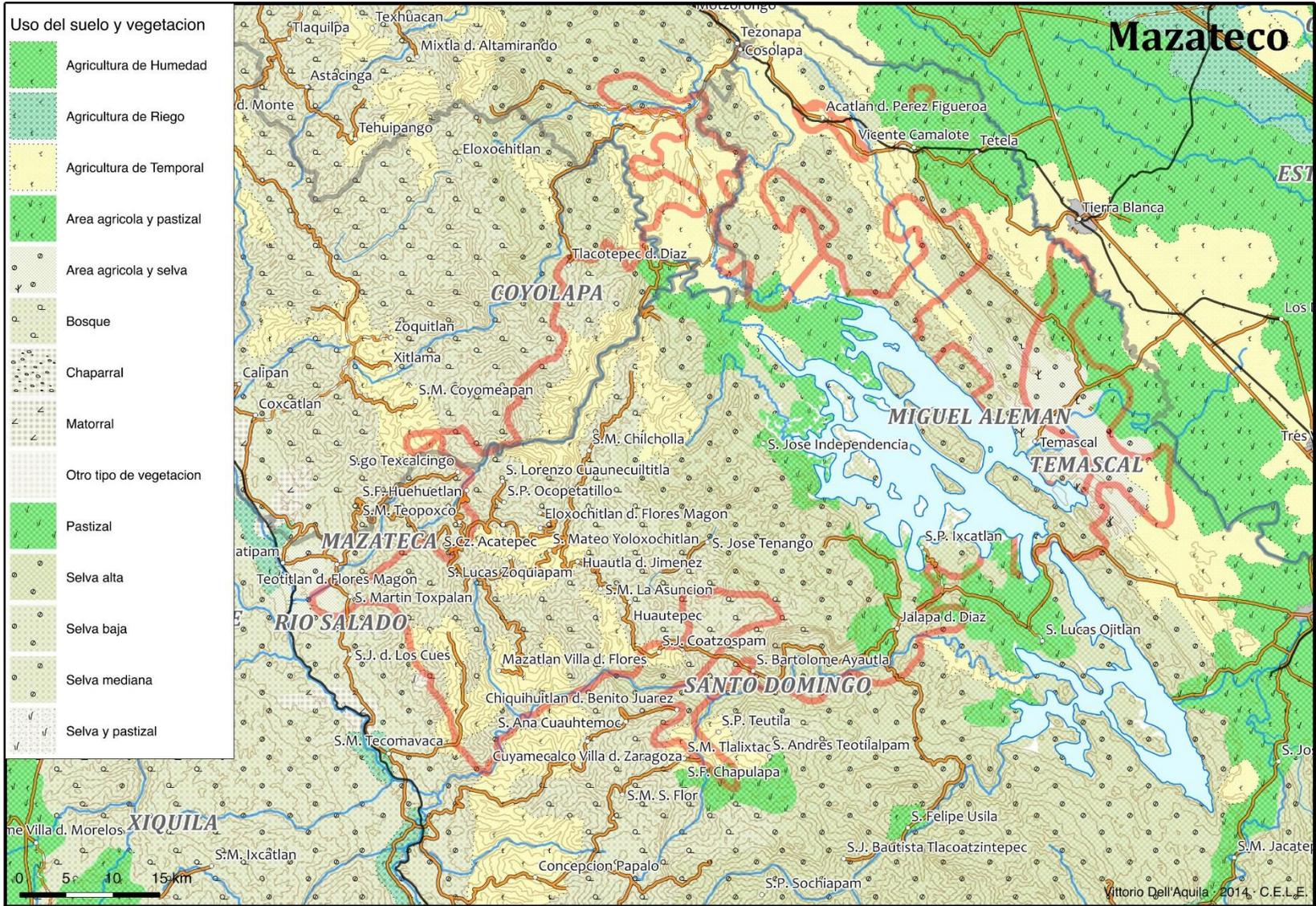
3. Modélisation

- Trois ordres de **complexité** et d'**émergence** (O'Sullivan, 2004) :
- **Phylogénétique** (complexité des *déterminismes*)
- **Ontologique** (complexité des *agrégats*)
- **Épigénétique** (complexité des *relations* et des *représentations*).

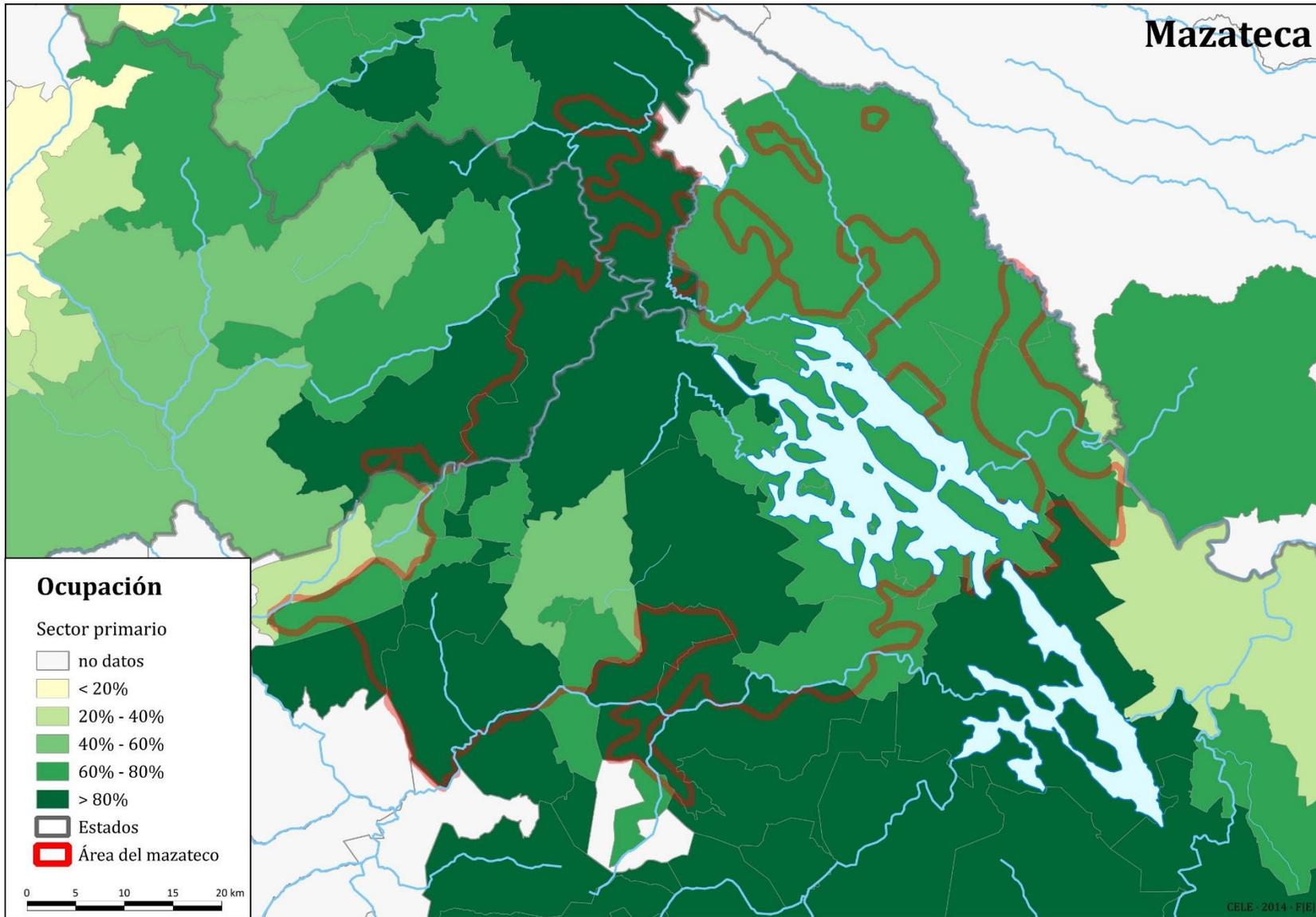
3.1. Facteurs externes (complexité des agrégats sociogéographiques)



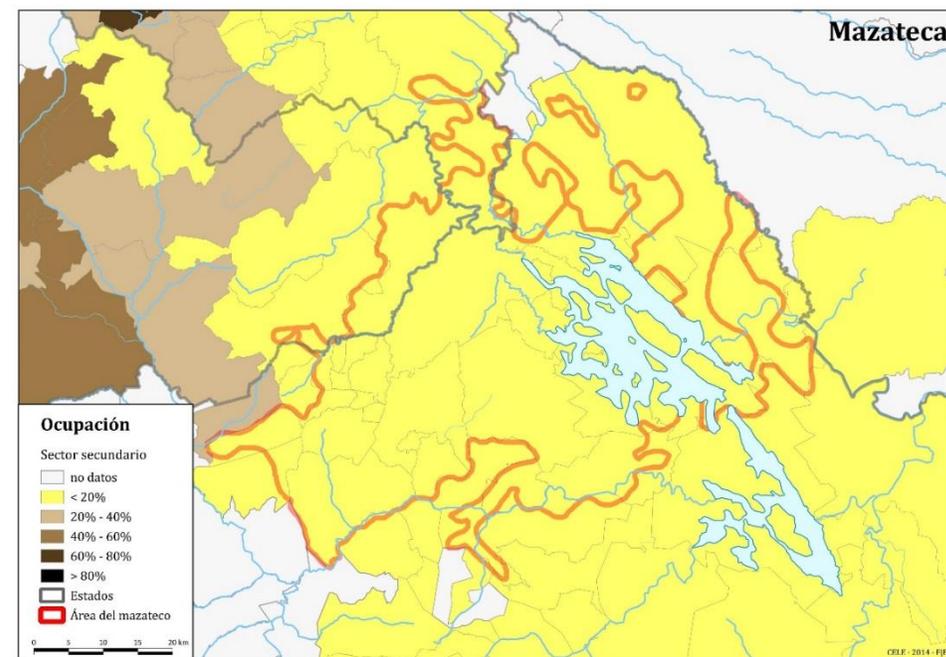
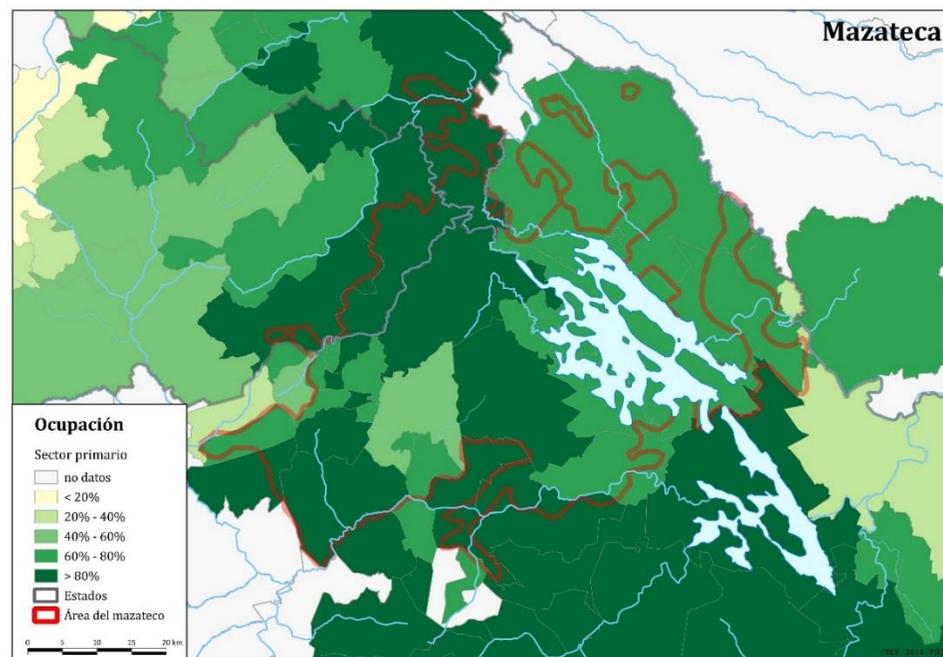




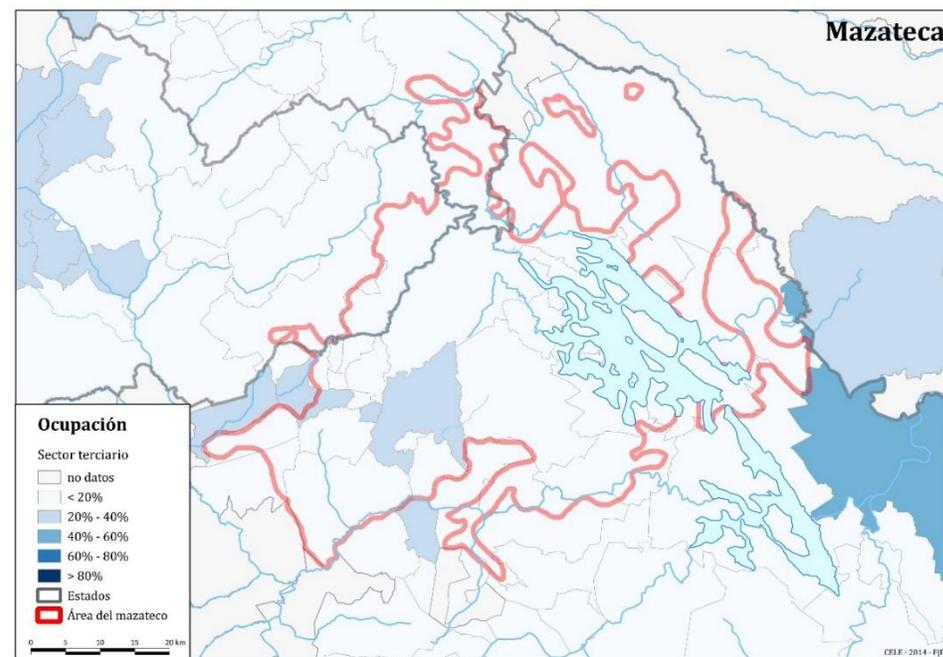
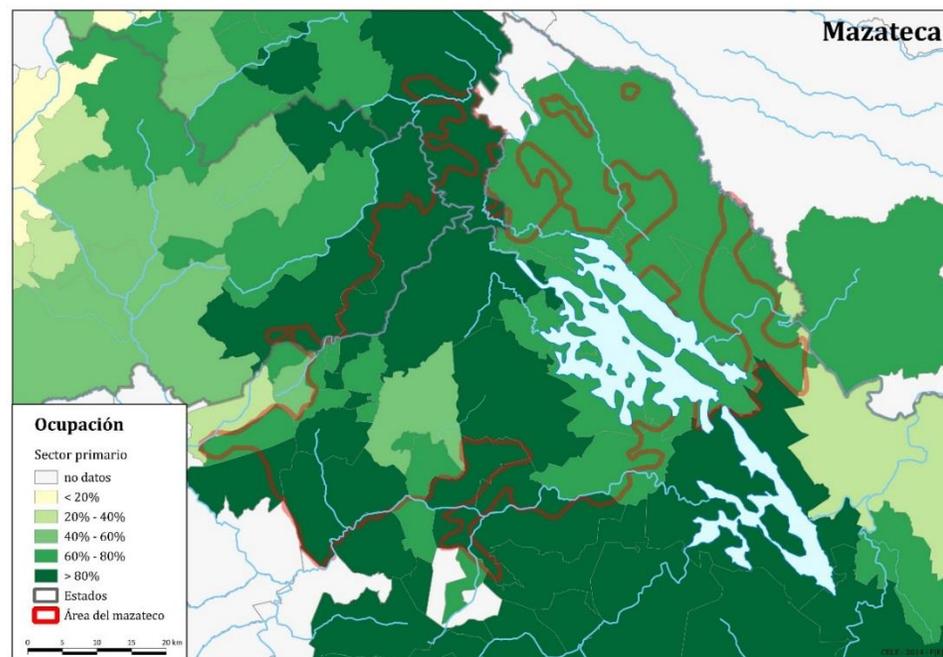
Mazateca



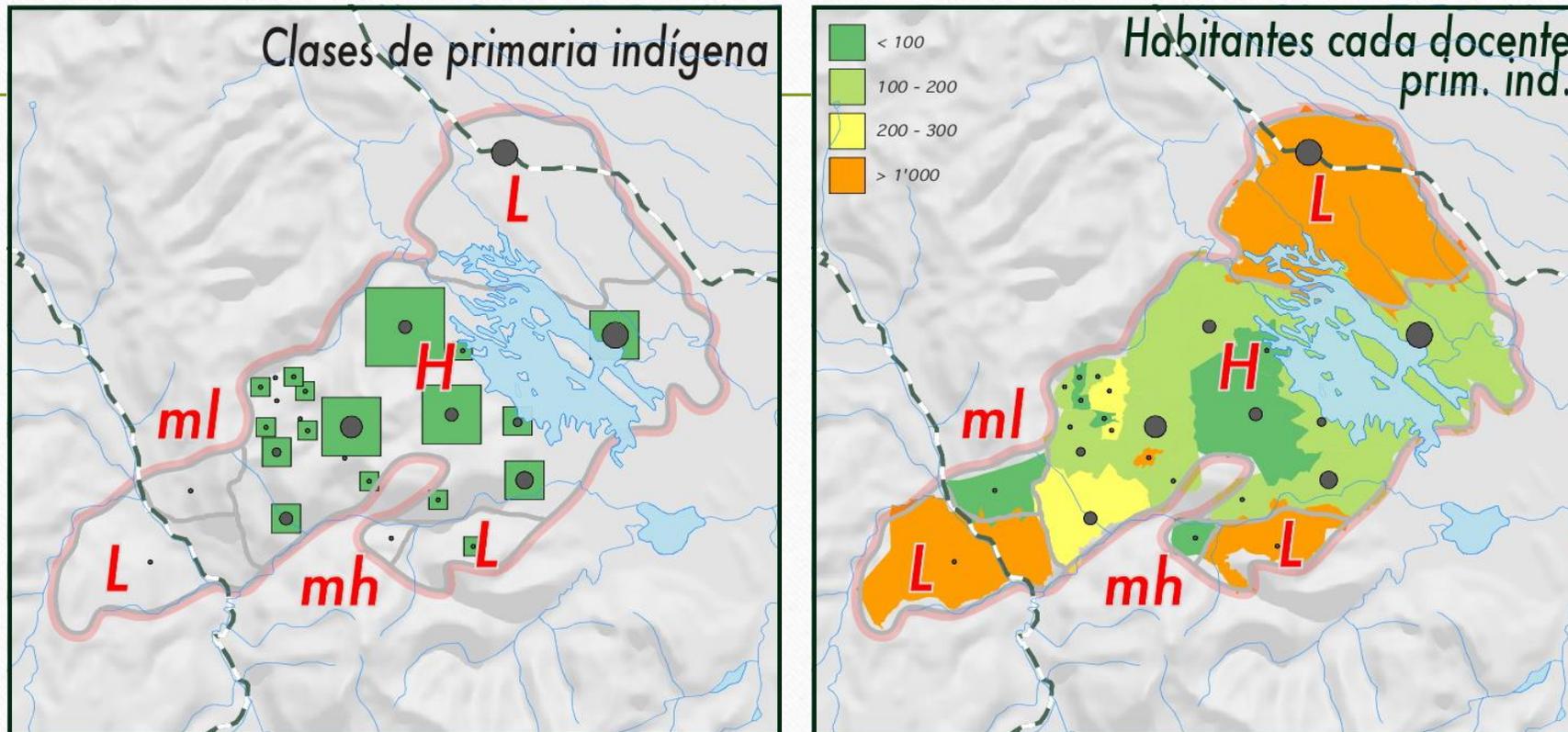
Secteurs primaire (à gauche), secondaire (à droite)



Secteurs primaire (à gauche), tertiaire (à droite)



Investissement public dans le tissu éducatif régional (années 2000) : infrastructures et ressources humaines



Hautes terres centrales : Huautla



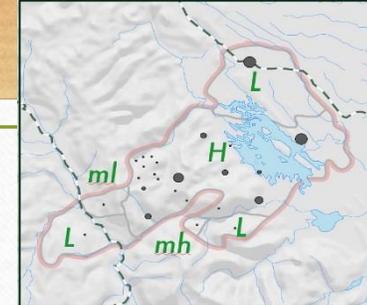
San Andres Hidalgo, « agencia » de Huautla



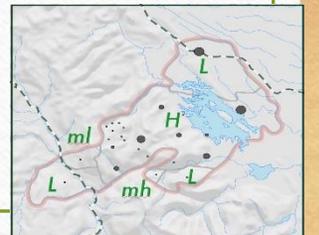




Basses terres mazatèques : zone de Soyaltepec



Le barrage et le lac artificiel Miguel Alemán, Une mosaïque de *minifundios* engloutis





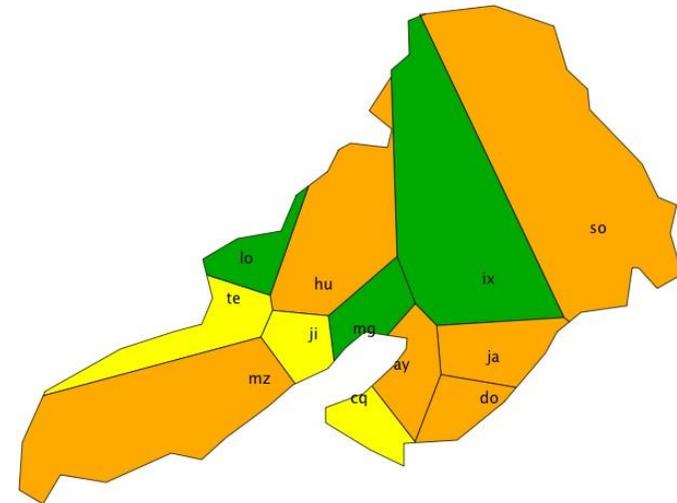
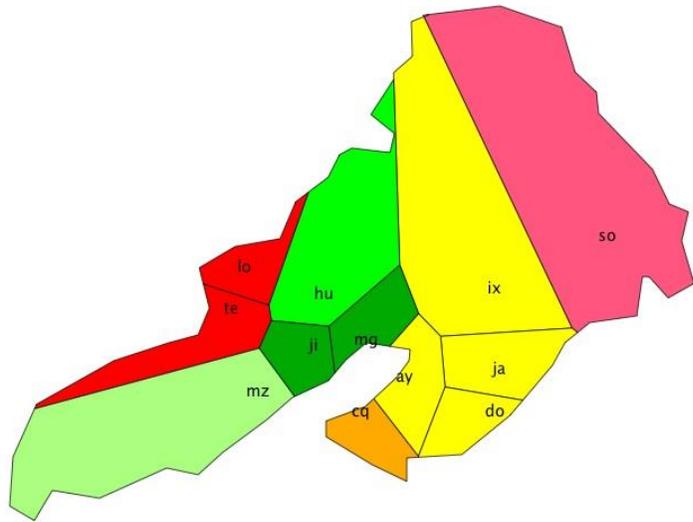


Une zone de résilience (et de résistance) :
l'île du *Vieux Soyaltepec*

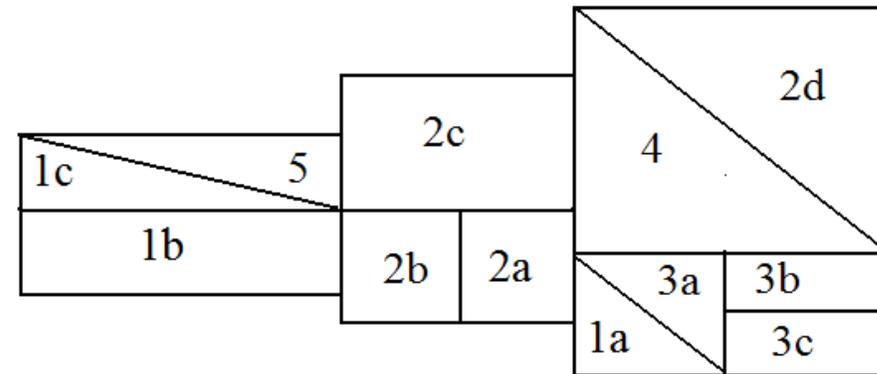
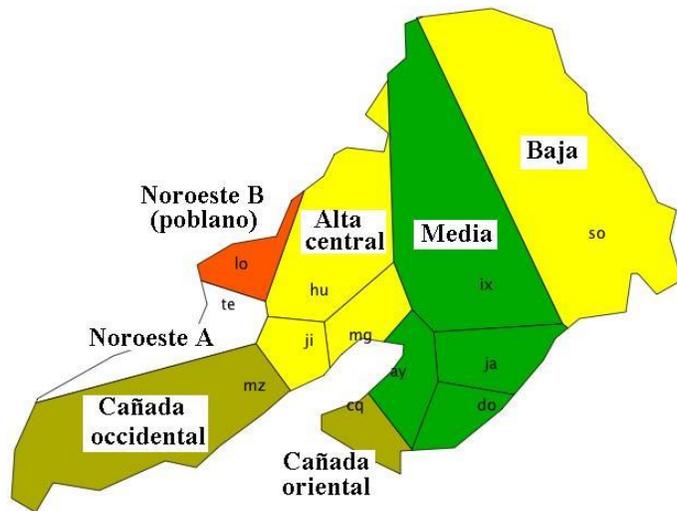


3.2. Cladistique (émergence *phylogénétique*)

Clades en phonologie diachronique
et en morphologie (pré)verbale



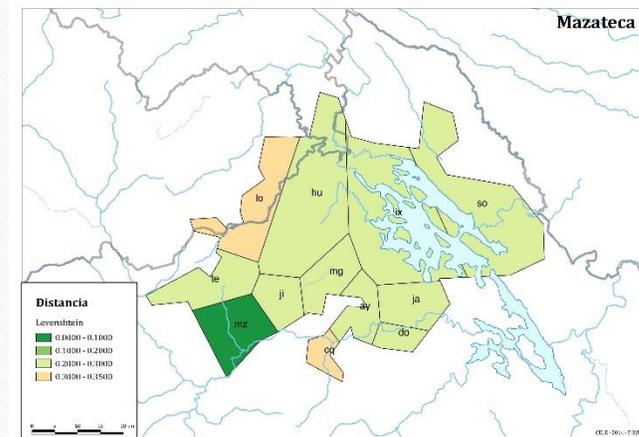
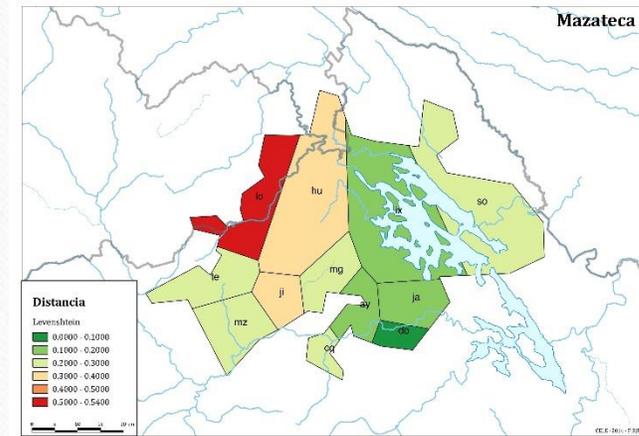
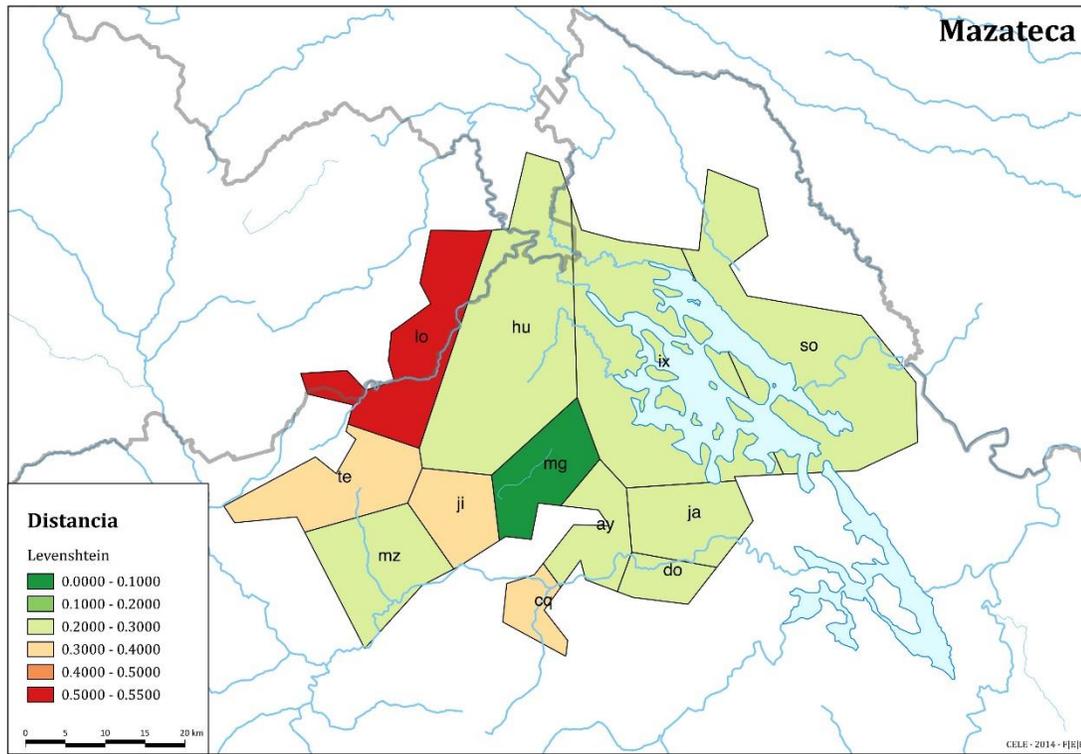
Modélisation d'après les résultats de l'analyse cladistique (données de Kirk, 1966)



	input	vis_hu	vis_te	vis_ix	vis_mg	vis_ji	vis_ja	vis_mz	vis_a)
2	able	hʃukhuaʔ ʃ	ʃikhua	ʃukhua	ʃukwʰa		ʃukwʰaʔ ʃ		ʃukwʰare
3	ables	mǎ-	mǎ-	mǎ			mǎʔ	akwimǎ-	mǎ-
4	above	nkasũʔ ʃ	isũnka	khihñāsũ		sũʔnka	khihñāsũʔ ʃ ʃ	khisũ	khihñāsũja
5	accustomed to	mǎnkaliʔ ʃ ʃ		kamǎnka	mǎnka		timǎnkaʔ ʃ ʃ	kamǎnka kamǎnka	kamǎnka
7	acid	sǎ	sǎ	sǎ	sǎ	sǎ	sǎʔ	sǎ	sǎ
8	across			nkuwakha			nqawakhaʔ ʃ ʃ		nkuwaha
10	afternoon	mǎʃũʔ ʃ		kamǎnkijũ		nkiʃũ	giʃũʔ ʃ	nkiʃura	kwankijũ
11	afterwards	haskǎʔ ʃ	iskǎ kiskǎ	haskǎ		askǎñi	iskǎʔ ʃ	haskǎ	haskǎ
12	alive	kũ	k tihñǎk	tikũ		kũ	tikũʔ ʃ		tihñǎkũ
13	alligator	ʃka	ʃka	ʃkha	ʃka		ʃkhaʔ		ʃkha
14	alone	kʔǎ suwa	kǎʔǎ suwa	kʔǎ suwa	sawa		kʔǎʔ suwaʔ ʃ	ʃutakǎ sawa	kʔa suwa
19	ant	niʔjuʔ ʃ	niʃju	niʔju	niʔju	niʔju	niʔjuʔ ʃ	niʔju	niʔju
22	appearance of	khiʔi	khi		khʔi		khiʔi	khi	khiʔi
23	arises	wisuthěʔ ʃ ʃ	tiwisithě	hwasuthě			wisithěʔ ʃ ʃ		ʔesethě
24	arm	ʃʒua	ʃʒa-	thianǎ	thia	ʃʒa	thiaʔ	thja	thia-
25	armadillo	ʃʒaʔaʔ ʃ	ʃʒa	ʃʒa		ʃʒaʔa	ʃʒaʔ	ʃʒaʔa	ʃʒeʔa
26	around	hantaiʔ ʃ	nkathanti	nkantai			n_daiʔ ʃ		ntai

3.3. Emergence *ontogénétique* ou *typologique*

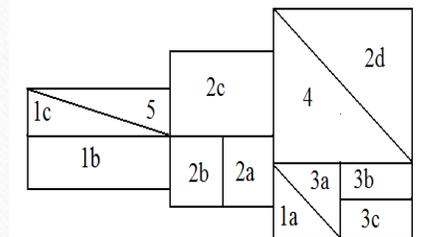
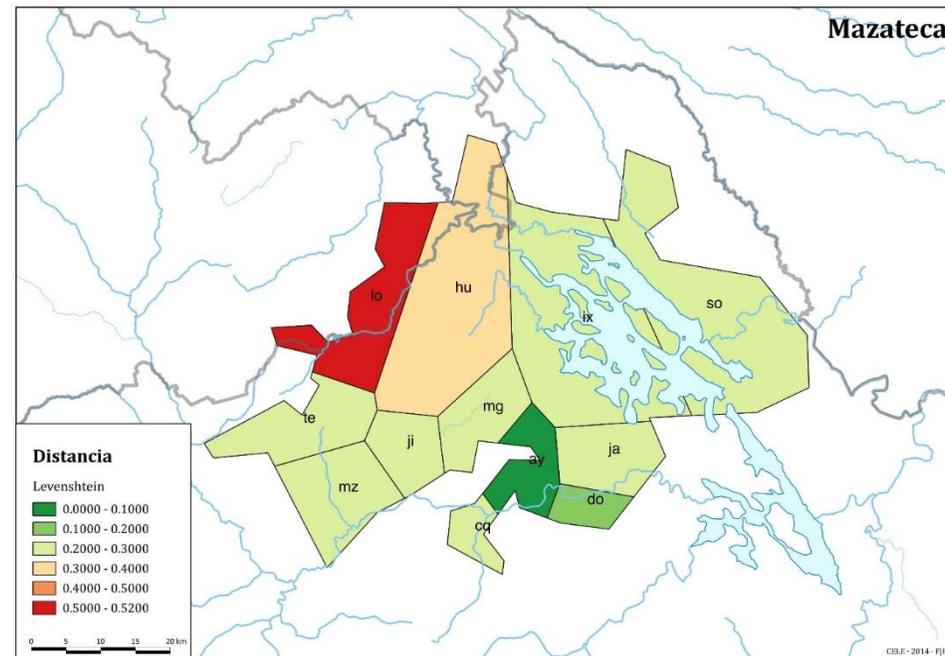
Algorithme de Levenshtein, cf. Beijering, Gooskens & Heeringa 2008 ; Bolognesi & Heeringa, 2002 ; Wichmann, Müller, Velupillai, Brown, & al.. 2010. Données de Kirk, 1966.



Ayautla : le piémont cafetier

NB : **vert** = identique ou proche, **rouge** = différence forte

NB : Réfraction des hautes terres centrales



Grille de lecture

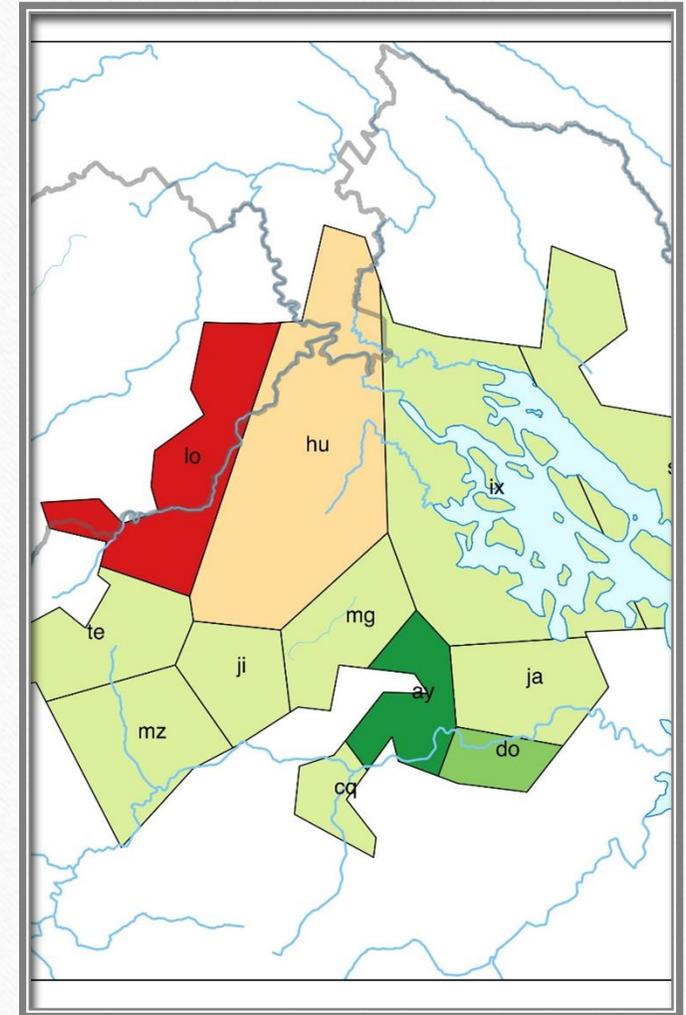
Vert sombre: locolecte

Vert foncé : régiolecte ou sous-dialecte

Vert clair : « arrière-pays », palier de convergences : *aire de résonance* structurale
(*effet moyen* de distance structurale)

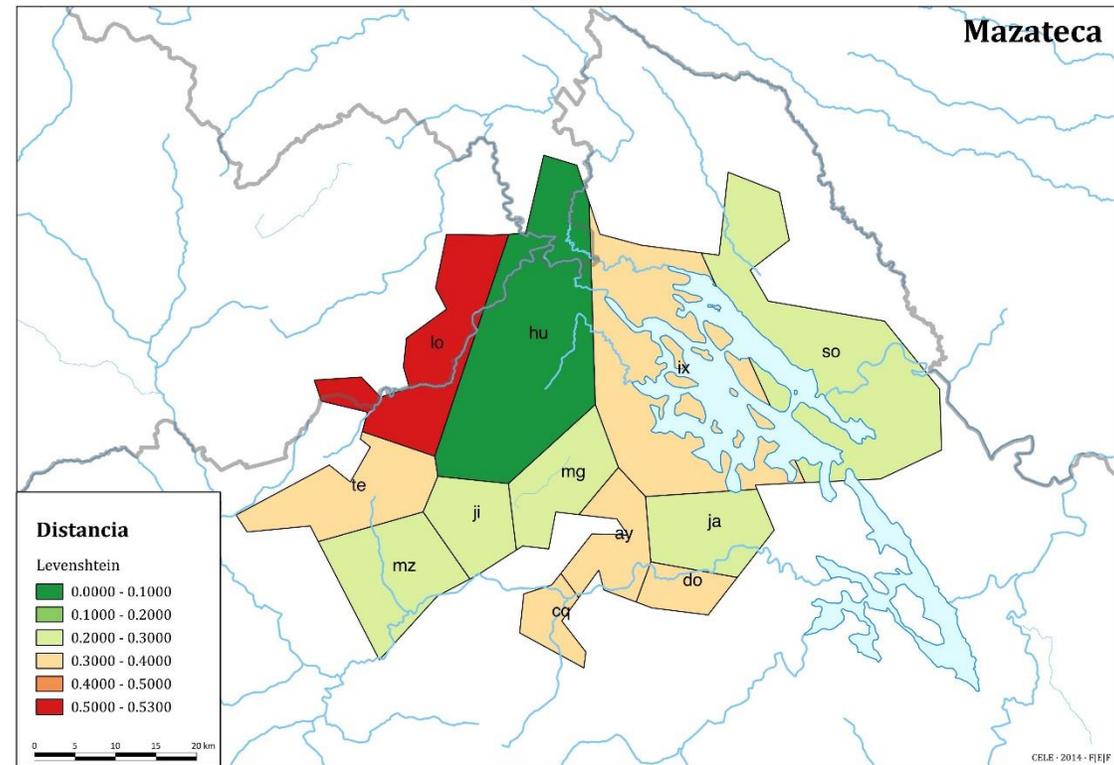
Orange : géolecte ou dialecte distinct

Rouge : antipode typologique, dialecte buttoir



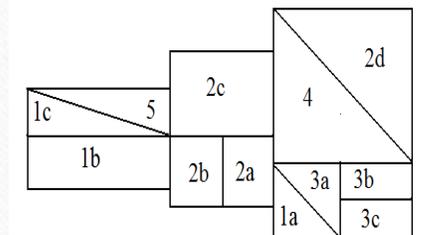
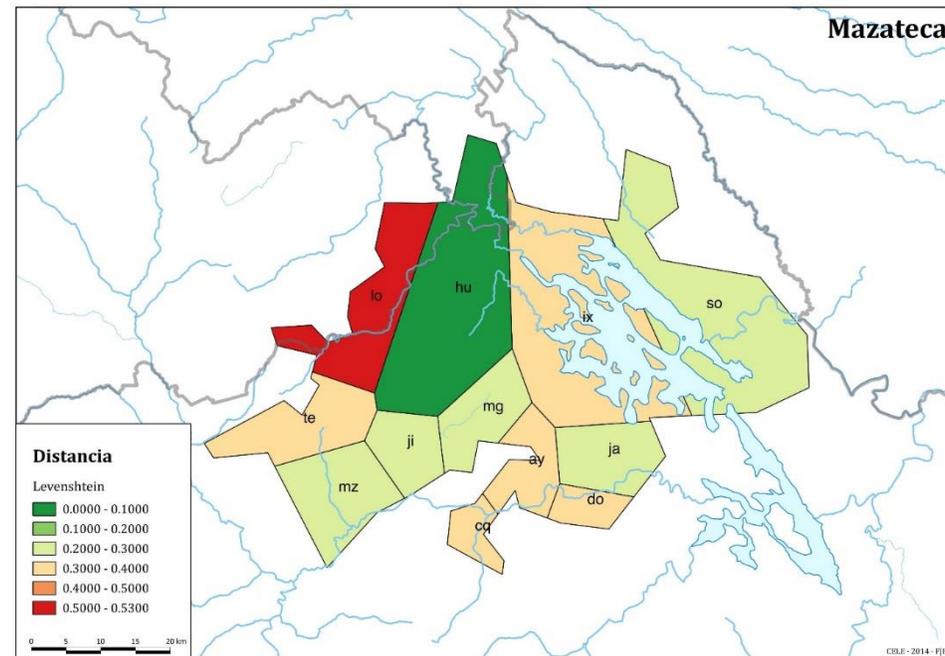
Grille de lecture de la *gamme chromatique*

La **gamme chromatique** peut se lire comme une *projection* de relations de *proximité* versus *distance structurale* (en phonologie principalement), de chaque **locolecte** comparé à tous les autres (12 au total, dans Kirk 1966).



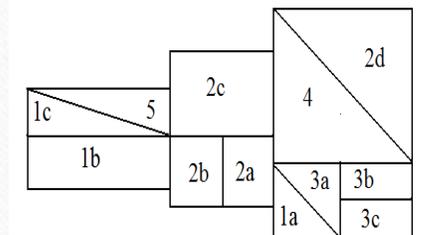
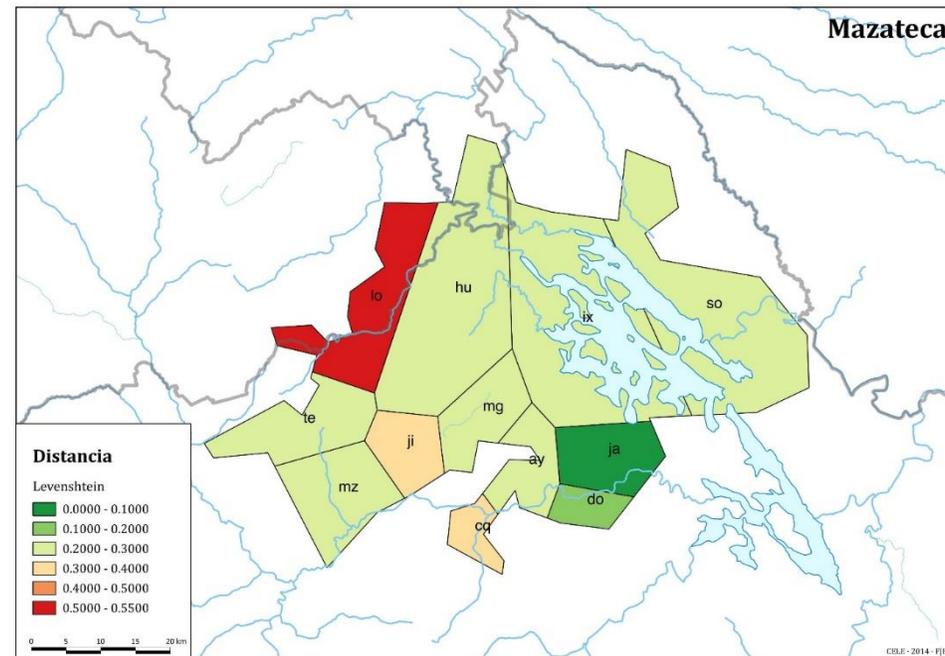
Huautla : centre cafetier des hautes terres

Rappel: **vert** = identique, proche, ou en résonance ; **rouge** = différence forte
=> Résonance périphérique, réfraction des versants occidentaux et orientaux



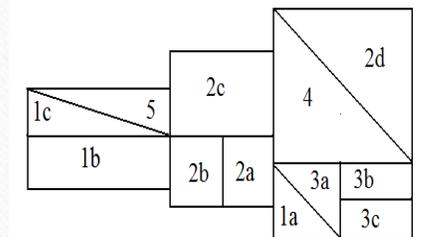
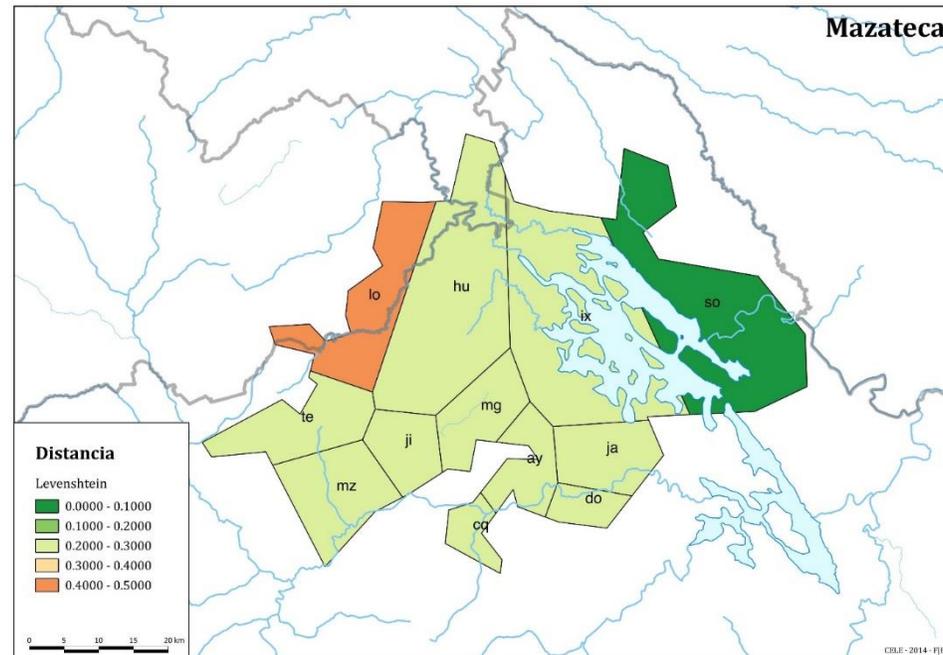
Jalapa : la plaine aux mille têtes (élevage)

=> Vaste aire de résonance depuis le sud-ouest du domaine, soit le *Piémont consensuel*. NB : *Points de réfraction* à Ji et Cq

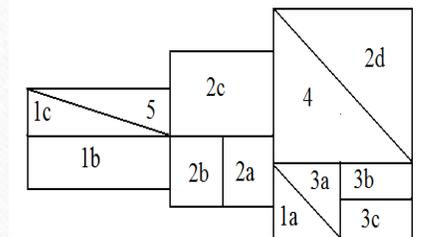
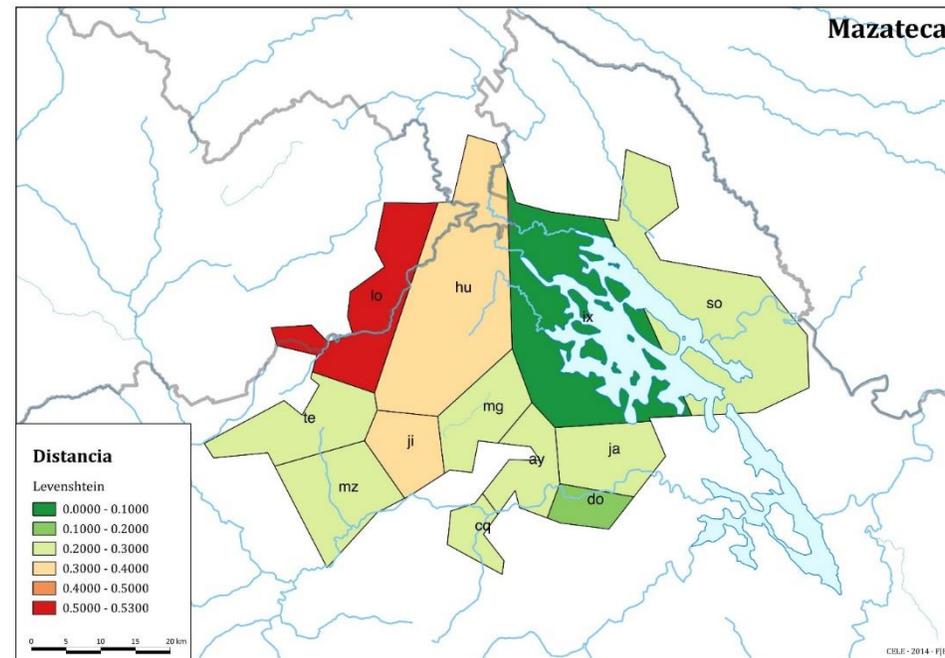


Soyaltepec : les basses terres sucrières, ancien versant
oriental de la *vallée encaissée* aux cent villages

=> Aire de résonance pan-mazatèque,
soit la *vallée consensuelle*.

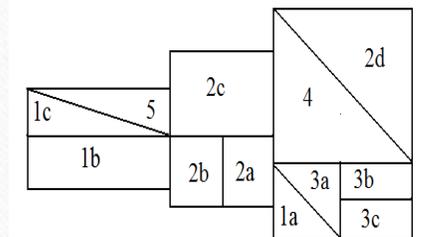
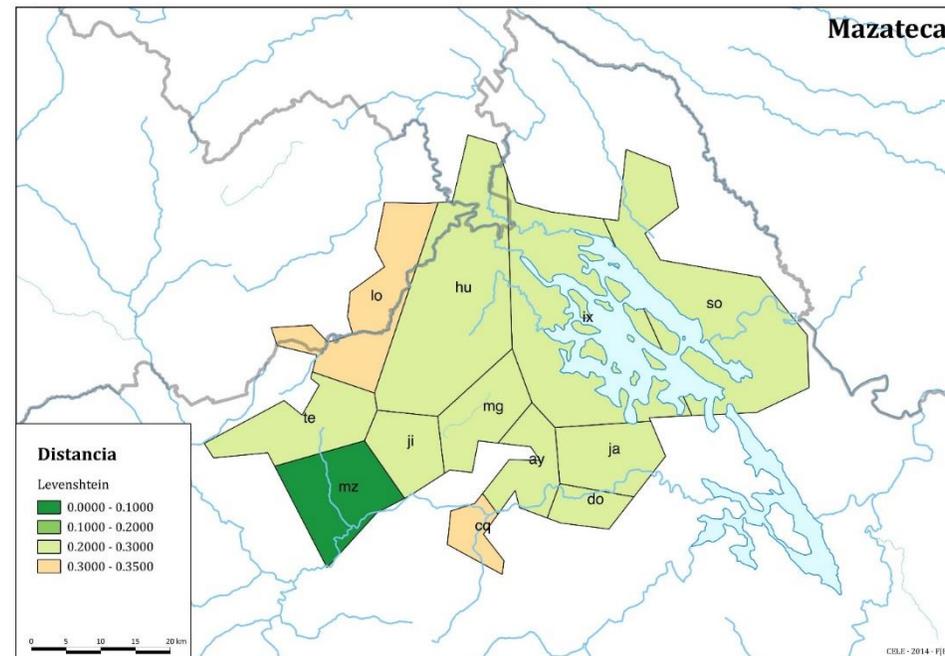


Ixcatlán : la vallée aux cent villages (engloutis)
=> Dichotomie *vallée centrale* vs. *Hautes terres centrales*
Résonance périphérique



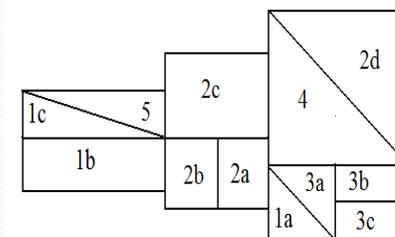
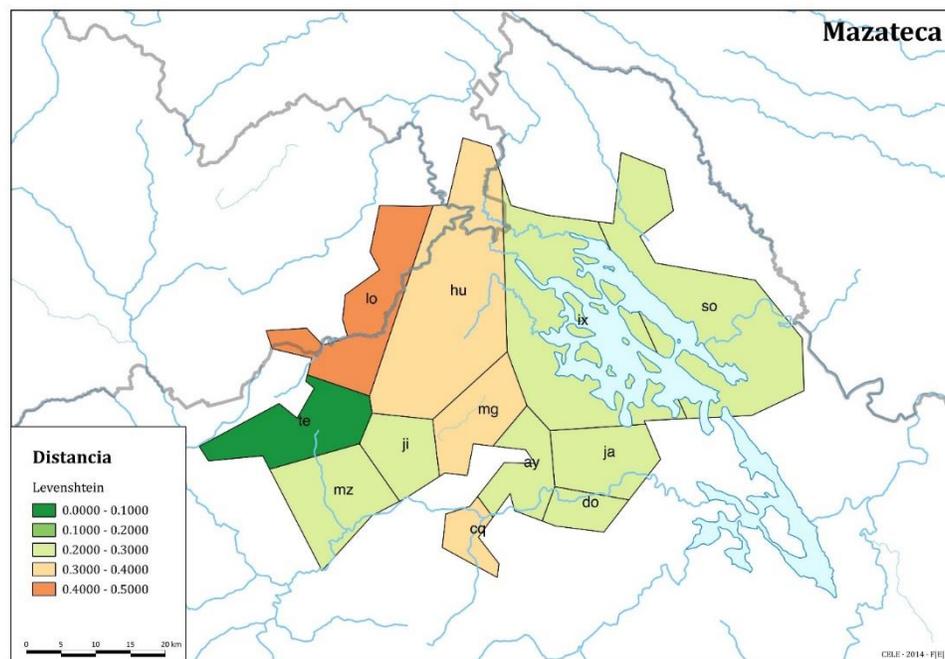
Mazatlán : le cañon aux jardins suspendus

=> résonance consensuelle panmazatèque, mais réfraction du versant oriental du cañon de Cuicatlán, et réfraction douce avec le versant poblano



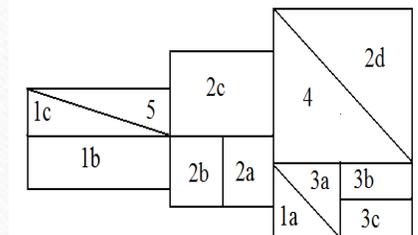
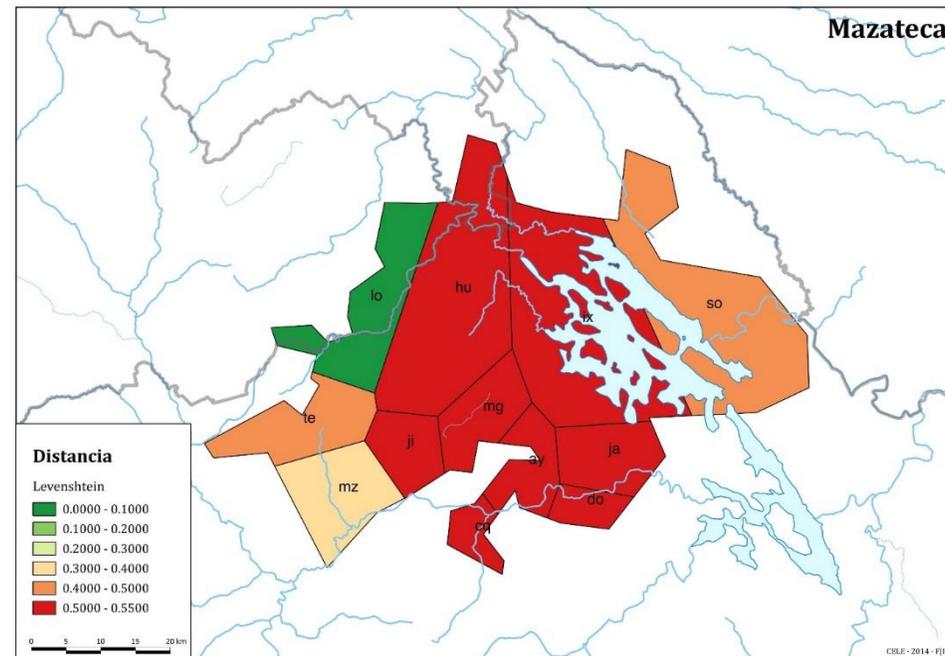
Tecoatl & Eloxochitlán : les versants cafetiers

=> résonance périphérique, mais réfraction avec les versants élevés de proximité (Hu, Mg)



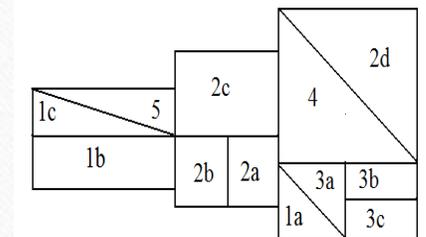
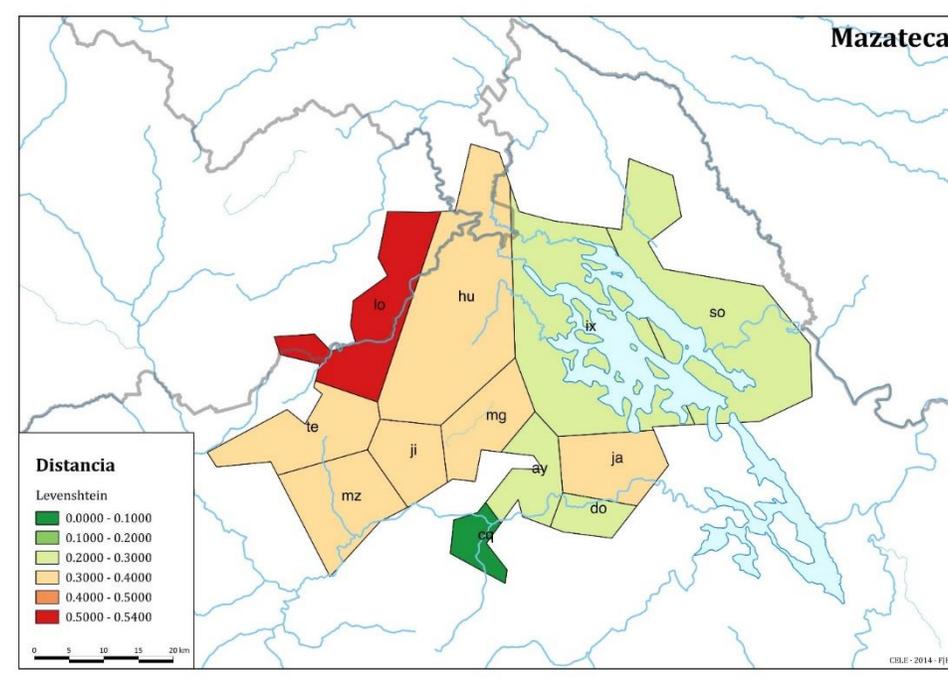
San Lorenzo Cuaunecuiltitla : le versant « poblano »

=> Locolecte innovant (chaînes de propulsion & de traction vocalique),
panréfractaire, avec résonance périphérique relative



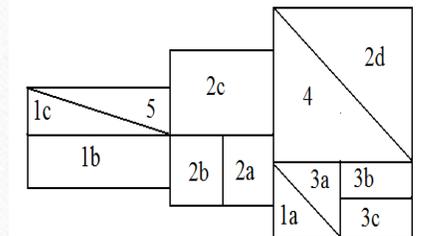
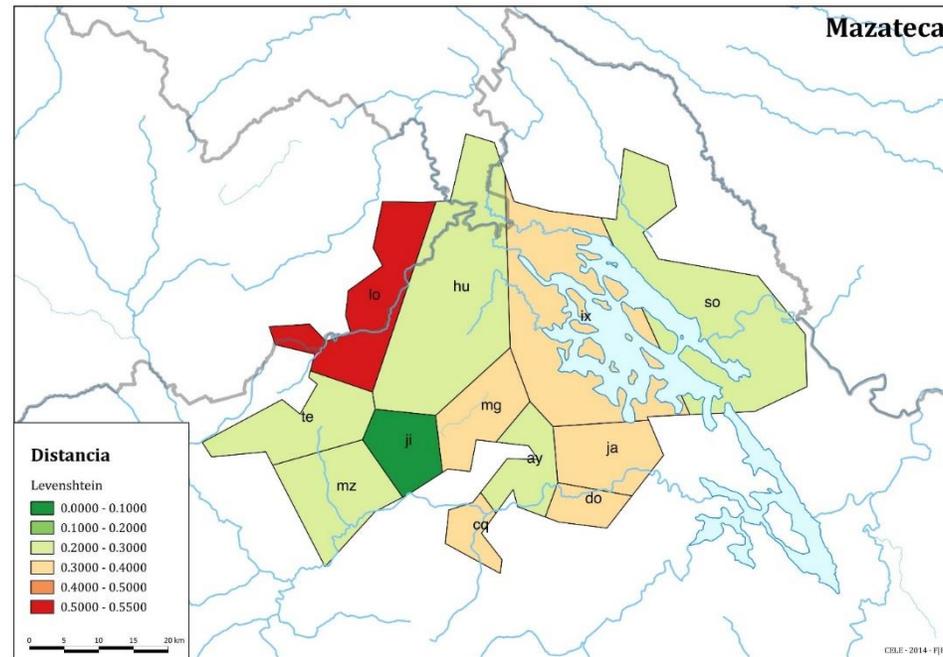
Chiquihuitlán : le cañon enclavé

=> résonance avec piémont et basses terres, réfraction avec hautes terres

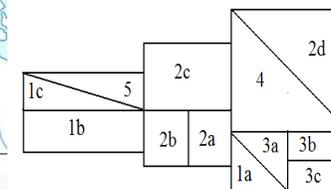
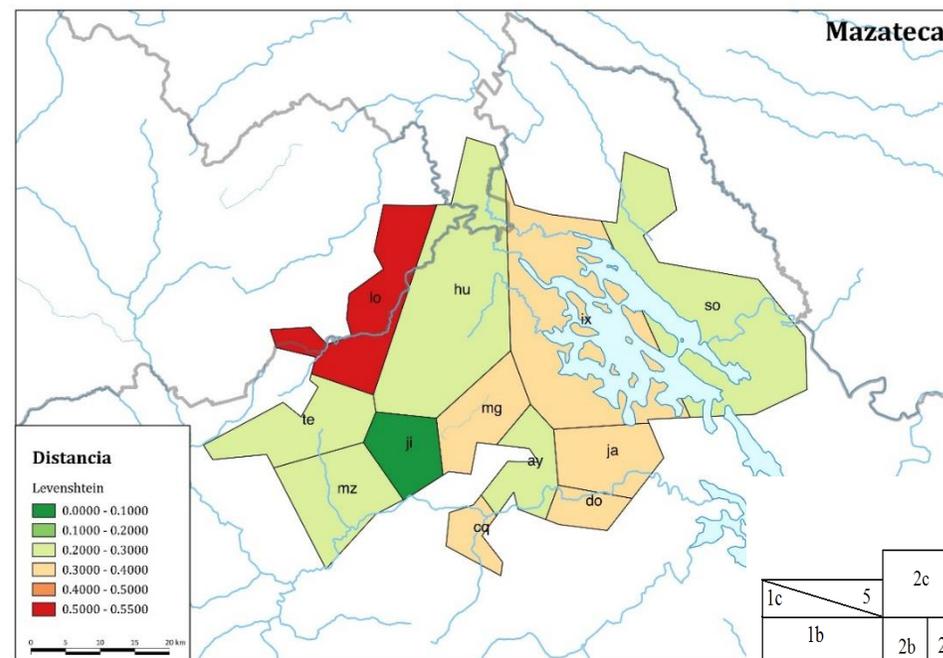
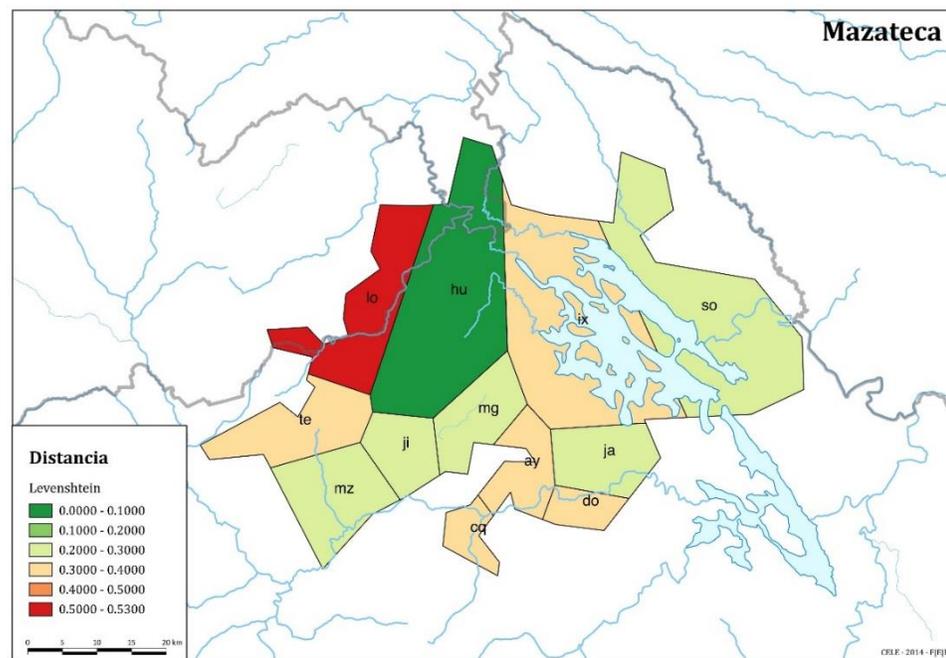


Santa Maria **Jiotes** : le carrefour des versants

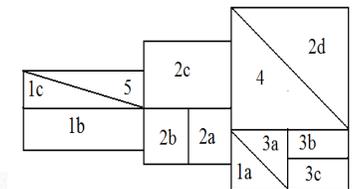
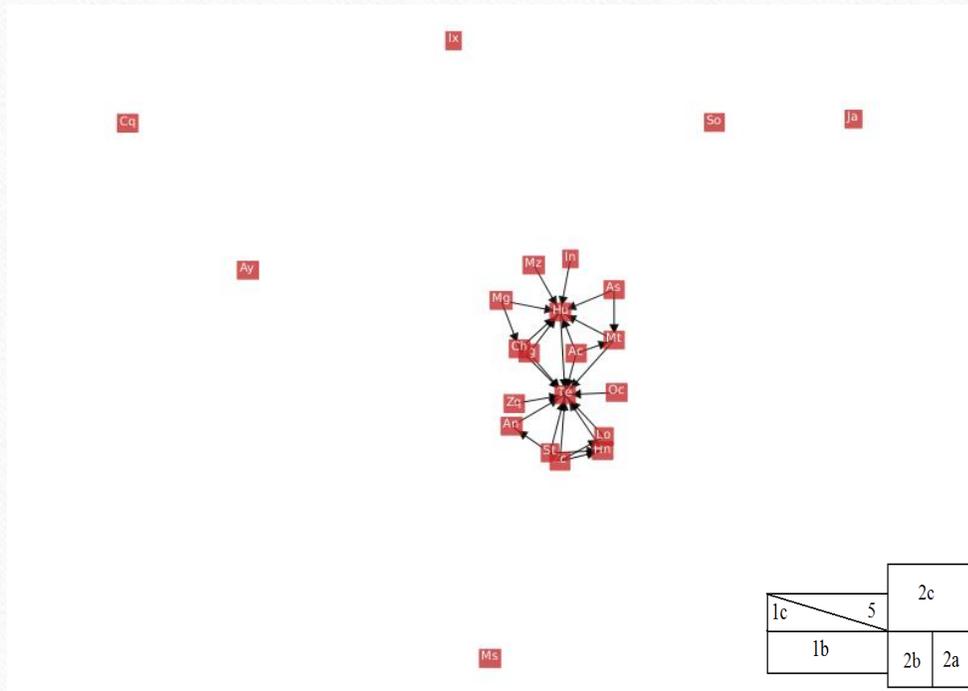
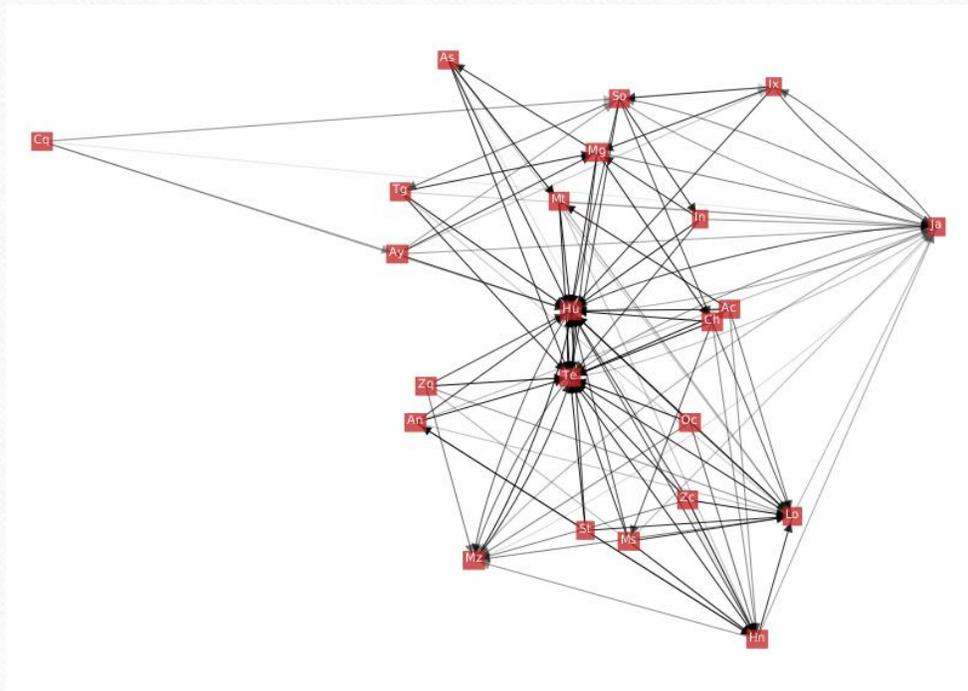
=> résonance nucléaire et périphérique, réfraction avec le piémont, l'ouest de la vallée encaissée et le cañon oriental



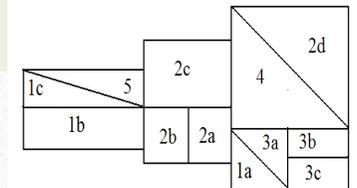
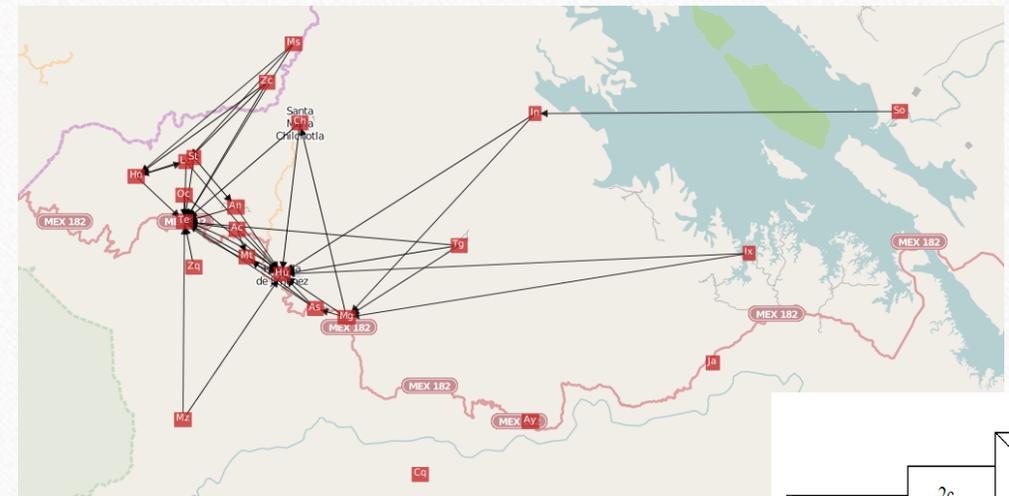
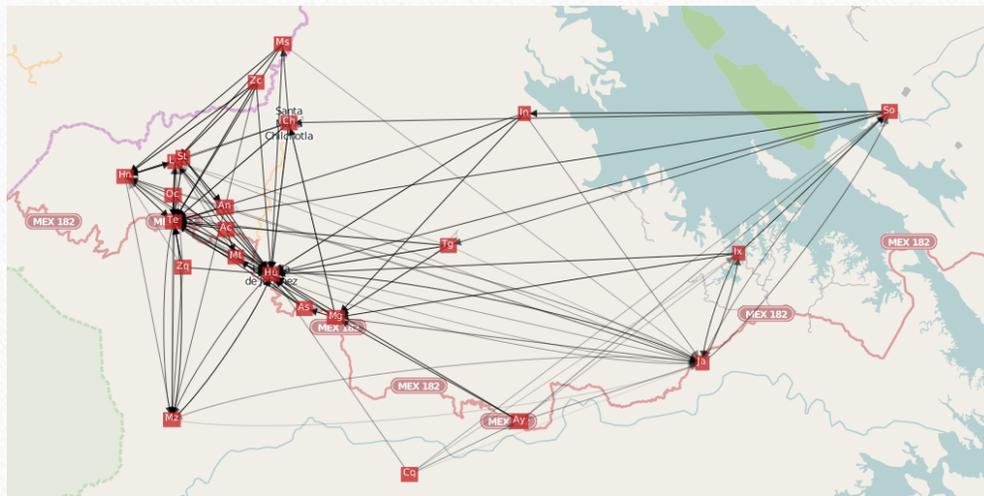
« Grain fin » typologique des ... Terres du café : Huautla *versus* Jiotes



3.4. Réseaux épigénétiques (intercompréhension), Kirk (1970)



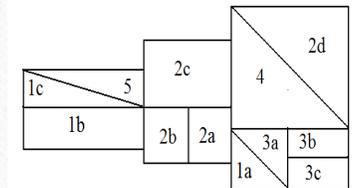
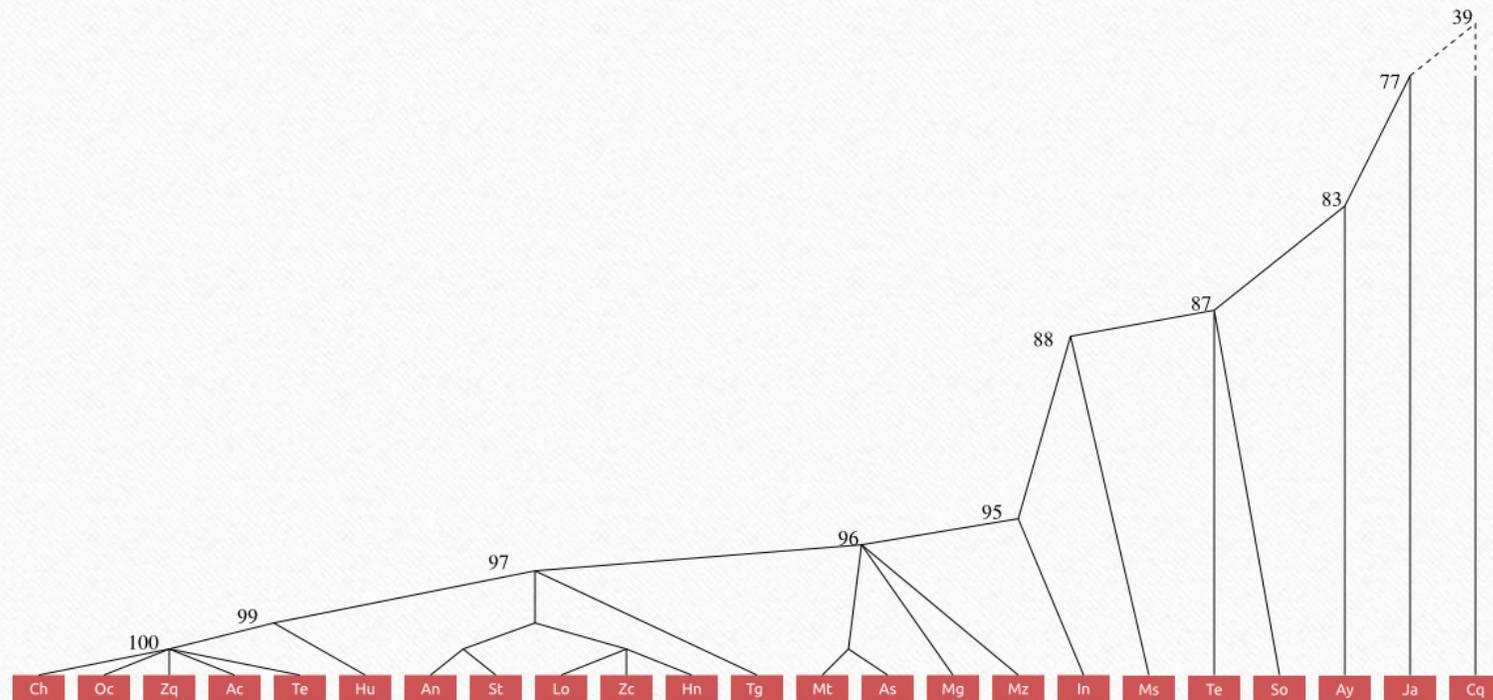
Intercompréhension : l'enquête de Paul L. Kirk (1970), traitement à l'aide de GraphStream (par Stefan Balev, ISCN)



Seuils d'intercompréhension (Kirk, 1970)

GraphStream (Stefan Balev, ISCN)

NB : 23 locoletes au lieu de 12



4. Conclusion

- La variation géographique des dialectes (la *géolinguistique*) offre un *prisme* sur l'émergence des *agrégats communaux* au cours de l'histoire d'un peuplement humain.
=> Un observatoire privilégié de l'*écologie linguistique*, et des relations homme/nature.
- Les structures qui émergent (les *aires dialectales*), sont multiplexes :
- **phylogénétique** (les phases d'expansion et de condensation : *géolectes*, *régiolectes*, *locolectes*),
- **ontogénétique** (les types, les dialectes et sous-dialectes)
- **épigénétique** (les représentations, les réseaux d'intercompréhension).

Nature / culture

- En termes de relations **homme/nature**, les données de notre approche multiplexe tendent à montrer que le peuplement de l'aire mazatèque s'est faite, au cours du dernier millénaire, par paliers successifs étagés sur l'éventail écologique de l'archipel vertical (hautes, moyennes et basses terres).
- La **complexité** (en termes de **diversité**) du **biotope** a contribué à une **diversification** dialectale opérant sur des **paliers de convergence** (les « arrières-pays » ou **géolectes profonds** obtenus par l'algorithme de Levensthein), ponctués de « niches ».
- Certaines « **niches** » ont formé des **isolats** (versant « poblano », Chiquihuitlan), certaines **versants** ont créé une dynamique de **zones d'expansion** et de **zones résiduelles**.

Impact *versus* Résilience

- **Impact** puissant et désagrégateur de **systèmes d'exploitation du milieu naturel** d'origine coloniale et postcoloniale (café, élevage, canne à sucre, implantations agroalimentaires)
- Mais le système *milpa* reste **résilient** et protège les **agrégats communaux**. De même nombre de « souches » dialectales, comme celles immergées sous la réserve d'eau du barrage Miguel Aleman.
- **L'approche multiplexe** permet d'apprécier la **pertinence** de ces phénomènes de **résilience**.

Retour vers l'empirisme

- **L'approche empirique** (terrain dialectologique) permet
- d'approfondir et d'enrichir cette *vision multiplexe*
- D'observer ou modéliser **l'auto-organisation** des agrégats communaux *versus* les facteurs favorisant la désagrégation et l'atomisation du milieu social et la *simplification* de la **relation culture/nature** par des systèmes de monoculture commerciale et les projets d'infrastructures (barrage Miguel Aleman).
- Les systèmes de production intensifs et à finalité agroindustrielle, émigration, désertification ou exode rural, acculturation ou assimilation sociolinguistique, etc. ont cependant un effet productif, de nouveaux cycles de variation et de diversification.

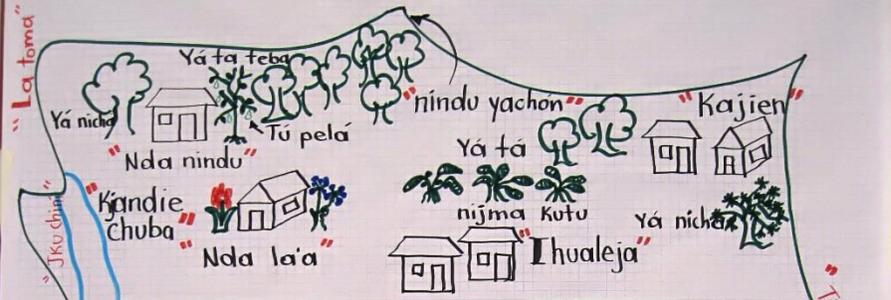
Perspectives

- Explorer la *complexité algorithmique* des *structures émergentes* (en phonologie, morphologie, syntaxe, lexicque) dans le nouveau contexte de relation homme/*milieu ressource* ou *milieu dépendant* ou *milieu périphérique/périphérisé*.
- Développer, à l'aide des outils issus du paradigme des systèmes complexes ou d'autres domaines scientifiques, de nouveaux *horizons de connaissance* (comme ici, la cladistique, la distance d'édition, GraphStream, etc.), rendant compte de la structure interne des agrégats géolectaux et sociolectaux.
- Comparer une gamme diversifiée d'*archipels verticaux* (Murra) comme celui que nous venons d'étudier, en tant qu'*observatoires de la diversification des langues du monde* à l'échelle de *microcosmes* (comparer à Nichols, 1992, pour approche au niveau macroscopique).

-
- Comprendre et valoriser les *processus résilients* des sociétés locales.
 - Plus d'informations sur <http://jll.smallcodes.com/home.page> (projet MAmP 2009-14, IUF, JLL)

ASIENRA NAXINANDA NAJNDIA.

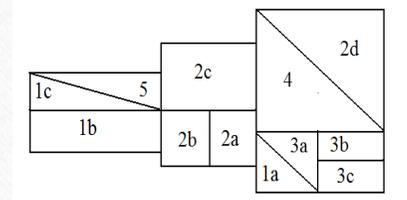
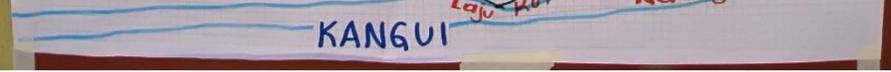
KAJUN



NGABASIEN



KANGUI



Références

- Abse Edward, 2007, « Don Patricio's Dream: Shamanism and the Torments of Secrecy in Fieldwork among the Mazatec Indians », dans Jean-Guy GOULET & Bruce MILLER (eds.), *Extraordinary Anthropology: Transformations in the Field*, University of Nebraska Press, p. 258-282.
- Beijering, K. Gooskens C. & Heeringa W. 2008. "Predicting intelligibility and perceived linguistic distance by means of the Levenshtein algorithm", Amsterdam, *Linguistics in the Netherlands*, 2008.), 13-24.
- Bolognesi, R. and W. Heeringa, 2002. "De invloed van dominante talen op het lexicon en de fonologie van Sardische dialecten". In: D. Bakker, T. Sanders, R. Schoonen and Per van der Wijst (eds.). *Gramma/TTT: tijdschrift voor taalwetenschap*. Nijmegen University Press, Nijmegen, 9 (1): 45-84.
- Blanton R., S. Kowalewski, G. Feinman, L. Finsten, [1981]-1993. *Ancient Mesoamerica. A Comparison of Change in Three Regions*, Cambridge, Cambridge University Press.
- Boege Eckart, 1988, *Los mazatecos ante la nación. Contradicciones de la identidad étnica en el México actual*, México, Siglo XXI.
- Bull Brian, 1984, « Aspect Formation of San Jeronimo Mazatec Verb », Summer Institute of Linguistics, *SIL-Mexico Workpapers*, vol 6, p. 93-117.
- Brown, Cecil H., Holman, Eric W., Wichmann, Søren and Velupillai, Viveka. 2008. "Automated classification of the world's languages: a description of the method and preliminary results". *STUF Language Typology and Universals* 61, 285-308.
- Casad Eugene, 1974. *Dialect intelligibility testing*, University of Oklahoma, the Summer Institute of Linguistics.
- Chierici, Frederica, 2012. *Dalla geografia linguistica alla dialettometria. Sviluppi e applicazioni*, Tesi di laurea, Università degli studi di Milano.

- Demanget Magali, 2006. *La chair des Dieux est-elle à vendre ? Chamanisme, tourisme et ethnicité en terre mazatèque (Mexique)*, Thèse NR, Université Paris X.
- Golston, Chris & Kehrein, Wolfgang 1998. "Mazatec onsets and nuclei", *International Journal of American Linguistics*, 64.4: 311-337.
- Golston, Chris & Kehrein, Wolfgang 2004. "A prosodic Theory of laryngeal contrasts", *Phonology* 21, 1-33. < <http://zimmer.csufresno.edu/~chrisg/>>
- Gudschinsky, Sarah 1958a. *Proto-Popotecan. A Comparative Study of Popolocan and Mixtecan*, *IJAL*, 25(2).
- Gudschinsky, Sarah C. 1958b. "Mazatec dialect history", *Language* 34, 469-481.
- Gribbin, J. 2004. *Deep Simplicity, Chaos, Complexity and the Emergence of Life*, London, Penguin.
- Josserand Kathryn, 1983. *Mixtec Dialect History*, Ph. D. Dissertation, New Orleans, Tulane University.
- Kaufman Terrence, 2006. "Early OtoManguean Homelands and Cultures:some premature hypothesis", accessible en ligne.
- Kirchhoff, P., 1943 « Mesoamérica. Sus límites geográficos, composición étnica y caracteres culturales », *Acta Americana. Revista de la Sociedad Interamericana de Antropología y Geografía*, México/Los Angeles : 92-107, reed. Par Garcia Mora & al., 2002. *Paul Kirchhoff. Escritos selectos. Estudios mesoamericanistas*, México, D.F., UNAM : 43-54.
- Kirk, Paul Livingston 1966. *Proto-Mazatec phonology*. PhD dissertation, University of Washington.
- Kirk, Paul Livingston, 1970. « Dialect Intelligibility Testing: The Mazatec Study », *International Journal of American Linguistics*, Vol. 36-3: 205-211.

- Léonard, Jean Léo 2010. "Enquêtes exploratoires pour l'ALMaz (*Atlas Lingüístico Mazateco*). Elicitation croisée, entre typologie et codification d'une langue otomangue", Grenoble, *Géolinguistique* 11 : 59-109.
- Léonard, Jean Léo 2012. *Eléments de dialectologie générale*, Paris, Michel Houdiard.
- Léonard, Jean Léo ; dell'Aquila, Vittorio & Gaillard-Corvaglia, Antonella 2012. "The ALMaz (*Atlas Lingüístico Mazateco*): from geolinguistic data processing to typological traits", *STUF*, Akademie Verlag, 65 : 78 -94.
- Léonard, Jean Léo & dell'Aquila, Vittorio 2013. "Temps et espace dans la réseau dialectal mixtec : variables multiplexes et indices aréologiques", in Snežana Gudurić, *Jezični i kulture u vremenu i prostoru*, I, Actes du colloque de Novi Sad, novembre 2011, Université de Novi Sad, pp. 75-86.
- Léonard, Jean Léo & Gilles Polian, 2013. "Herramientas cuantitativas en el proyecto ALTO (*Atlas Lingüístico del Tseltal Occidental*) y evaluación de la zona dialectal sur", à paraître dans Gerrero Galván, Orozco, Leonor [éds.] *Estudios de Variación geolingüística*, Colegio de México.
- Longacre, Robert 1957. *Proto-Mixtecan*, Indiana University Research Center in Anthropology, Folklore and Linguistics, 5.
- Massip-Bonet, Àngels & Albert Bastardas-Boada (eds.), 2013. *Complexity Perspectives on Language, Communication and Society*, Berlin & Heidelberg, Springer Verlag.
- Meinig, D.W. 1965. « The Mormon Culture Region: Strategies and patterns in the geography of the American West: 1847-1964 », *Annals of the Association of American Geographers*, 55: 191-220.
- Mufwene, Salikoko, 2001. *The Ecology of Language Evolution*, Cambridge Approaches to Language Contact, Cambridge, CUP.

- Nichols Johanna, 1992. *Linguistic Diversity in Space and Time*, Chicago & London, The University of Chicago Press.
- Nichols Johanna, 2004. "The Origin of the Chechen and Ingush: A Study in Alpine Linguistic and Ethnic Geography", *Anthropological Linguistics*, 46-2: 129-155.
- O'Sullivan, David, 2004. « Complexity Science and Human Geography », *Transactions of the Institute of British Geography*: 282-295.
- Pike, Kenneth 1948. *Tone Languages. A Technique for Determining the Number and Types of Pitch Contrasts in a Language, with Studies in Tonemic Substitution and Fusion*, Ann Arbor, University of Michigan Press. NB : Rééditions : 1949, 56, 57, 61.
- Rensch, Calvin [1966]-1976. *Comparative Otomanguean Phonology*, Ph.D. Dissertation, University of Pennsylvania, Indiana University Publications, Language Science Monographs 14, Bloomington: Indiana University Press.
- Silverman, Daniel 1997. "Laryngeal complexity in Otomanguean vowels", *Phonology* 14 : 235-261.
- Suárez, Jorge, 1983. *The Mesoamerican Indian languages*, Cambridge, Cambridge University Press, Cambridge Language Surveys.
- Trudgill, Peter, 2013. « Contact and Sociolinguistic Typology », in Hickey, Raymond (ed.), *The Handbook of Language Contact*, Oxford, Wiley-Blackwell: 299-317.
- Wichmann, Søren & Holman, Eric, 2009. *Assessing temporal stability for linguistic typological features*, Munich, Lincom¹.
- Wichmann, Søren, Müller, André, Velupillai, Viveka, Brown, Cecil H., Holman, Eric W., Brown, Pamela, Urban, Matthias, Sauppe, Sebastian, Belyaev, Oleg, Molochieva, Zarina, Wett, Annkathrin, Bakker, Dik, List, Johann-Mattis, Egorov, Dmitry, Mailhammer, Robert and Geyer, Helen. 2010. *The ASJP Database* (version 12)

Compléments

Compléments sur le traitement cladistique (Gaillard-Corvaglia, Darlu, Léonard, 2012)

1) Toutes les données, non pondérées

Quand les caractères ne sont pas pondérés, on trouve,

par bandb 3 arbres de 710 pas

Tree length = 710

Consistency index (CI) = 0.6521

Homoplasy index (HI) = 0.3479

CI excluding uninformative characters = 0.4374

HI excluding uninformative characters = 0.5626

Retention index (RI) = 0.5304

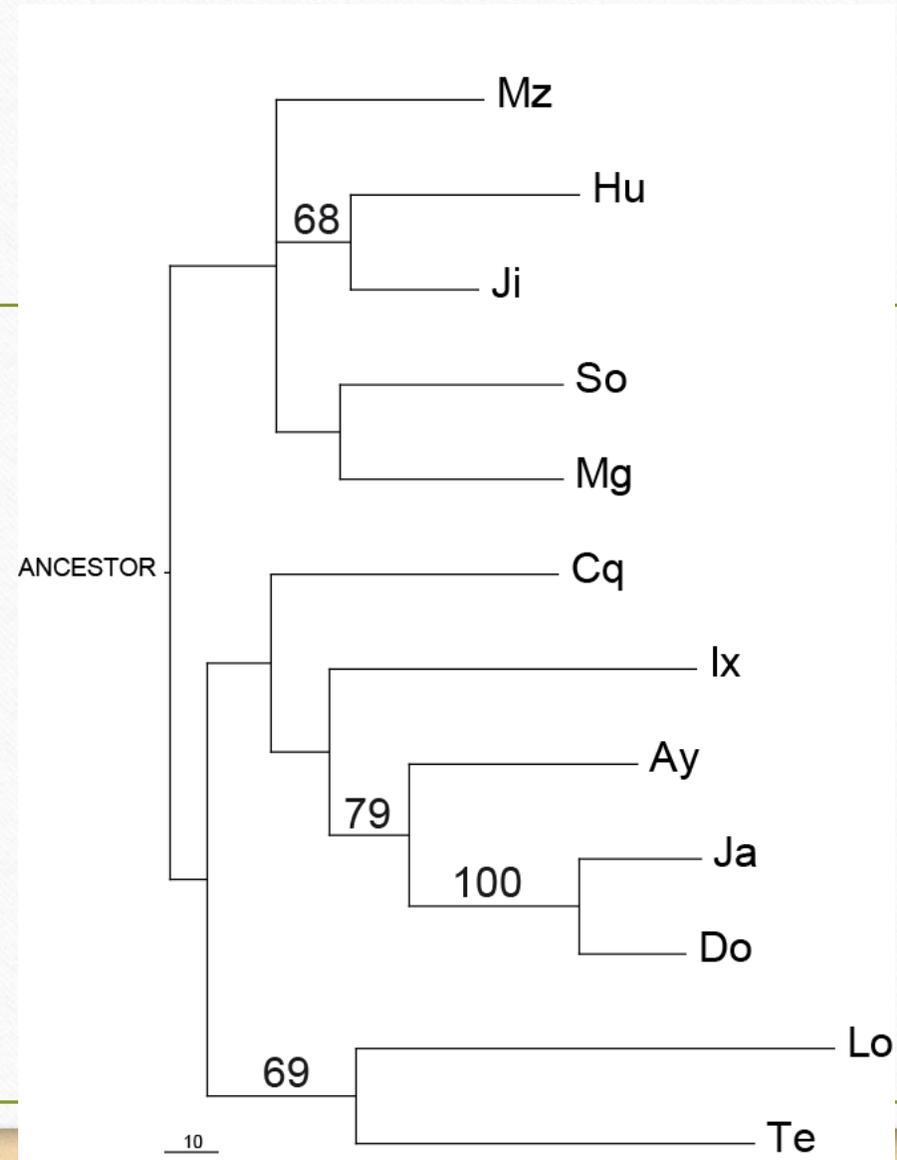
Rescaled consistency index (RC) = 0.3459

Ces trois arbres diffèrent au niveau de la trifurcation,
par la position de MZ :

- 1) (MZ,((HU,JI),(SO,MG)))
- 2) ((MZ,(SO,MG)),(HU,JI))
- 3) ((MZ,(HU,JI)),(SO,MG))

L'option d'optimisation est le « branch and bound », c.a.d qu'elle examine TOUS les arbres possibles pour choisir le plus parcimonieux.

Les valeurs portées sont les Jackknife à 50% (c'est mieux que le bootstrap pour ce genre de données, même si les résultats sont quasi identiques).



2) Toutes les données, pondérées

Branch-and-bound search settings: Optimality criterion = parsimony

Character-status summary:

Of 467 total characters:

All characters are of type 'irrev.up'

133 characters have weight 1

334 characters have weights other than 1

4 characters are constant

271 variable characters are parsimony-uninformative

Number of parsimony-informative characters = 192

Branch-and-bound search completed:

Score of best tree found = 1609

Number of trees retained = 1

Character-state optimization: Accelerated transformation (DELTRAN)

Consistency index (CI) = 0.6787

Homoplasy index (HI) = 0.3213

CI excluding uninformative characters = 0.4417

HI excluding uninformative characters = 0.5583

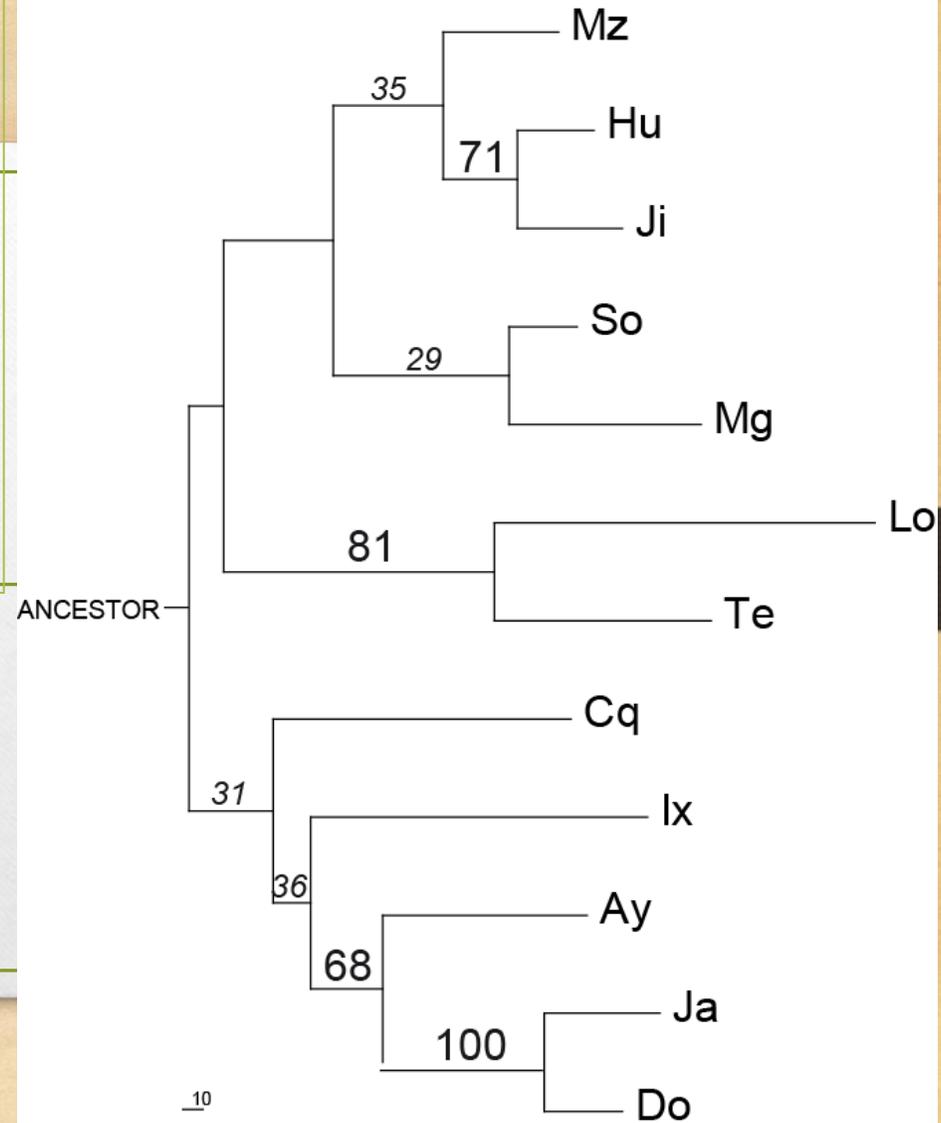
Retention index (RI) = 0.5313

Rescaled consistency index (RC) = 0.3606

L'option d'optimisation est le « branch and bound », c.a.d qu'elle examine TOUS les arbres possibles pour choisir le plus parcimonieux.

Les valeurs portées sont les Jakknjfe à 50% (c'est mieux que le bootstrap pour ce genre de données, même si les résultats sont quasi identiques. En italiques, ceux qui sont inférieurs à 50%, donc très peu robustes mais encore représentables dans l'arbre.

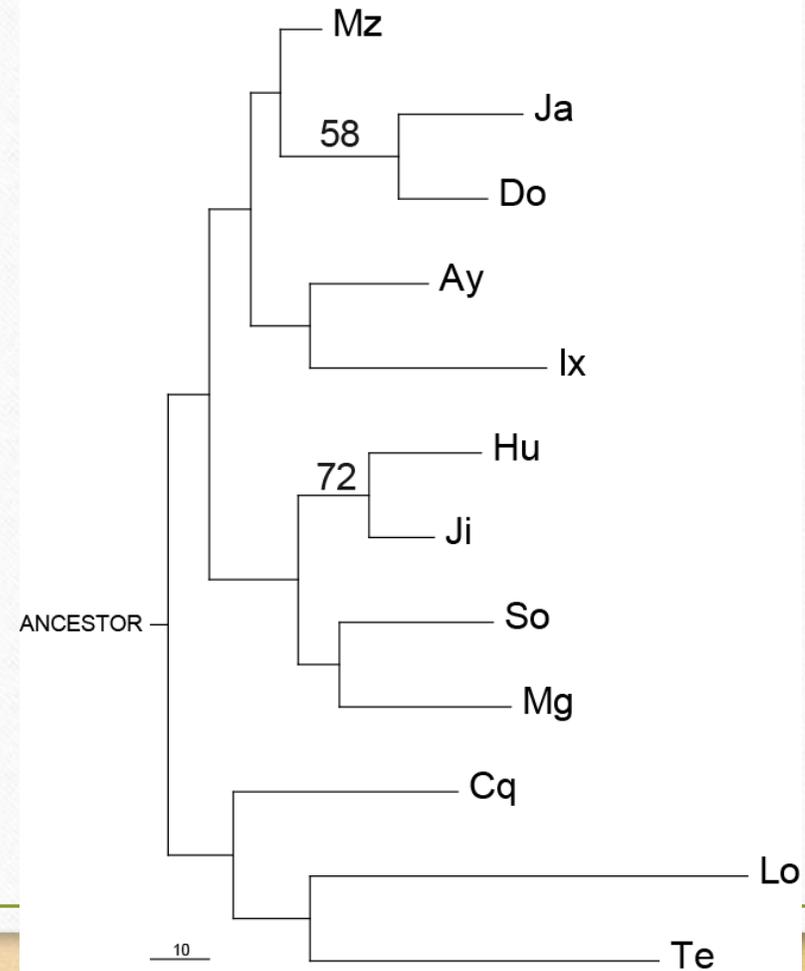
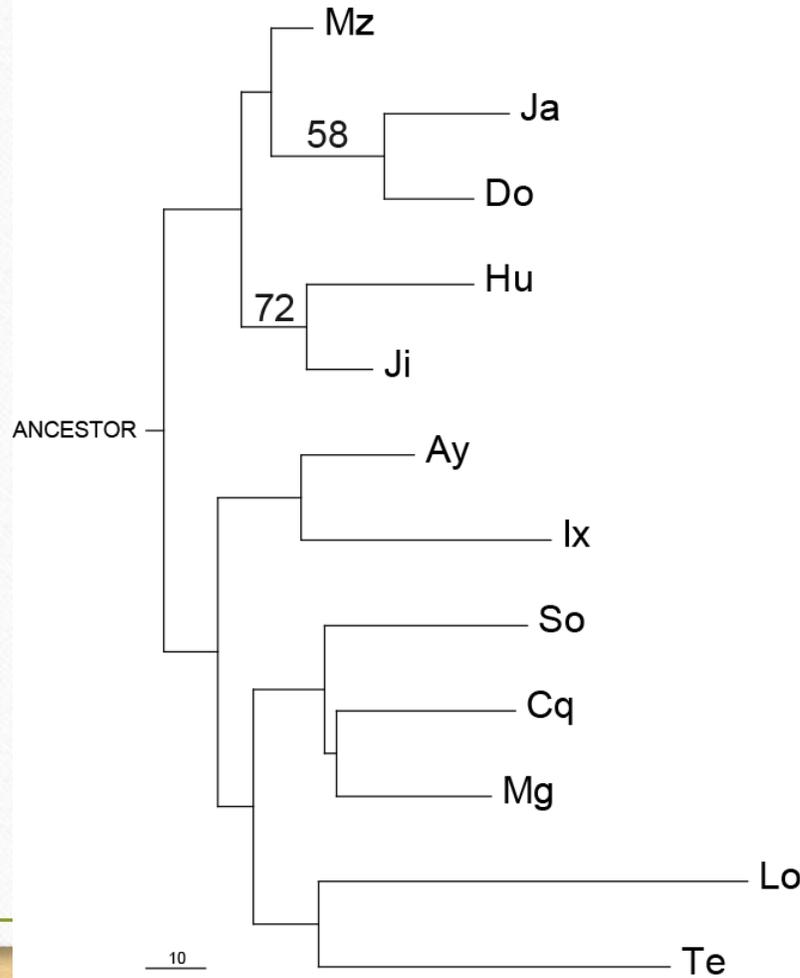
Finalement, la pondération ne change pas radicalement la représentation. Sans doute faut-il aller dans le détail de la distribution des états de caractères pour voir s'il y a des modifications de leur position dans l'arbre. Cela dit, l'arbre est mieux résolu (en particulier pour la position de Mz). Si le coupe Lo-Te change de groupe-frère, c'est à un niveau où, de toute façon, il n'y a aucune résolution.



2) Morphologie avec pondération

Branch and bound, Tree length = 469
Consistency index (CI) = 0.7036
Homoplasy index (HI) = 0.2964
CI excluding uninformative characters = 0.4654
HI excluding uninformative characters = 0.5346
Retention index (RI) = 0.5697
Rescaled consistency index (RC) = 0.4008
Two parsimonious trees

Deux arbres, mais assez différents. Seuls les deux groupes (Ja,Do) et (Hu, Ji) ont un sens, relatif



3) Phonologie avec pondération

Branch-and-bound search completed:

Score of best tree found = 477

Tree description:

Optimality criterion = parsimony

Character-status summary:

329 characters are excluded

Of the remaining 138 included characters:

All characters are of type 'irrev.up'

29 characters have weight 1

109 characters have weights other than 1

2 characters are constant

70 variable characters are parsimony-uninformative

Number of (included) parsimony-informative characters = 66

Character-state optimization: Delayed transformation
(DELTRAN)

Tree length = 477

Consistency index (CI) = 0.6541

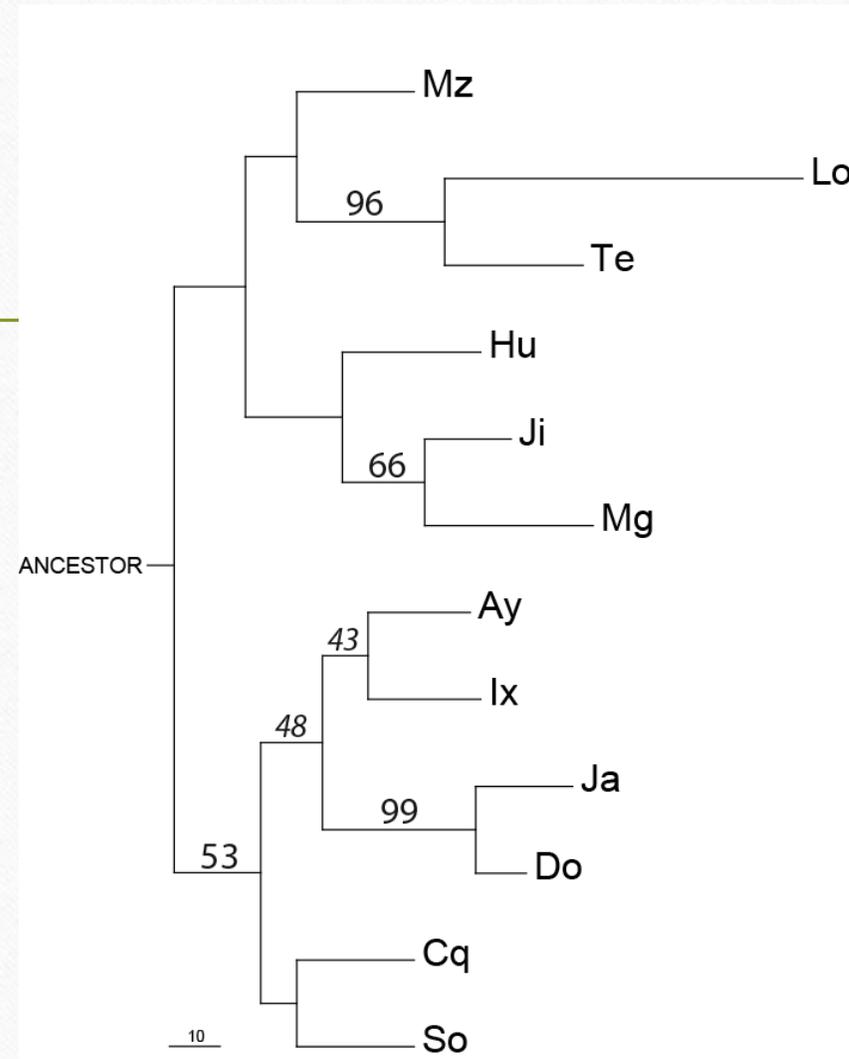
Homoplasy index (HI) = 0.3459

CI excluding uninformative characters = 0.4745

HI excluding uninformative characters = 0.5255

Retention index (RI) = 0.6504

Rescaled consistency index (RC) = 0.4254



3) Préfixe avec pondération

Tree description:

Optimality criterion = parsimony

Character-status summary:

294 characters are excluded

Of the remaining 173 included characters:

All characters are of type 'irrev.up'

55 characters have weight 1

118 characters have weights other than 1

108 characters are parsimony-uninformative

Number of (included) parsimony-informative characters = 65

Character-state optimization: Delayed transformation (DELTRAN)

AncStates = "standard"

Tree length = 581

Consistency index (CI) = 0.7745

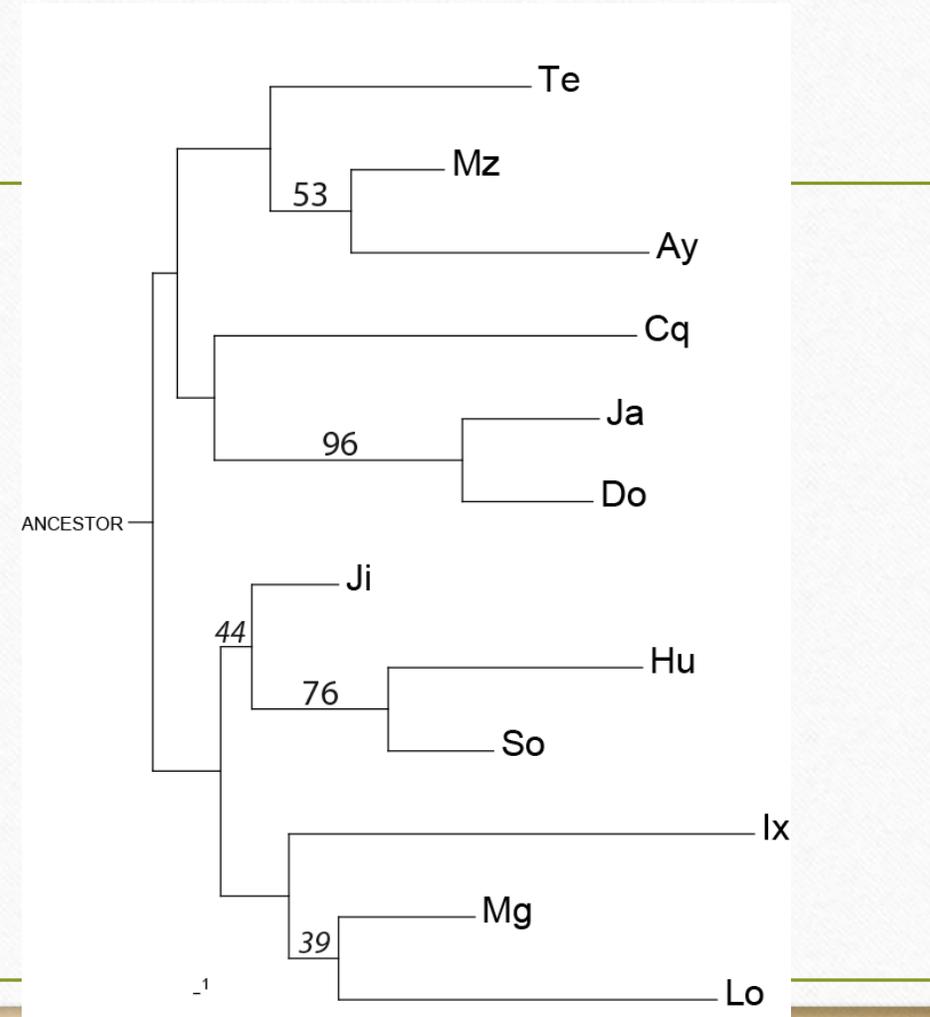
Homoplasy index (HI) = 0.2255

CI excluding uninformative characters = 0.5148

HI excluding uninformative characters = 0.4852

Retention index (RI) = 0.5747

Rescaled consistency index (RC) = 0.4451



J'ai également testé l'incongruence entre les trois types de données (Morphologie, Phonologie, Préfixe) avec le test ILD de Farris

.
L'hypothèse de congruence entre les arbres de ces trois types de données est rejetée avec une $p=0.001$

La congruence entre Morphologie et Phonologie et ou entre Morphologie et Préfixe n'est rejetée qu'avec une $p=0.07$

En revanche la congruence entre Phonologie et préfixe est rejetée avec $p=0.001$.

Conclusion : c'est entre Phonologie et Préfixe qu'il y a le plus de divergence en terme de phylogénie: les arbres ne sont pas comparables sérieusement !

*Dans le fichier MZ_apolist.xls
un CI de 1 signifie que l'état du caractère ne change qu'entre les deux nœuds indiqués
Un CI de 0.5 signifie que l'état du caractère change sur deux branches différentes
Un CI de 0.333 qu'il change sur trois branches différentes ...*

Même chose dans le fichier MZ_chglist.xls. Les résultats pour les caractères sont donnés cette fois par branches

Evidemment seuls les CI=1 fondent solidement un « clade ». Tous les autres changements sont des convergences (ou emprunts ?)