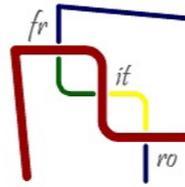




Université de Sofia « Saint Clément d'Ohrid »

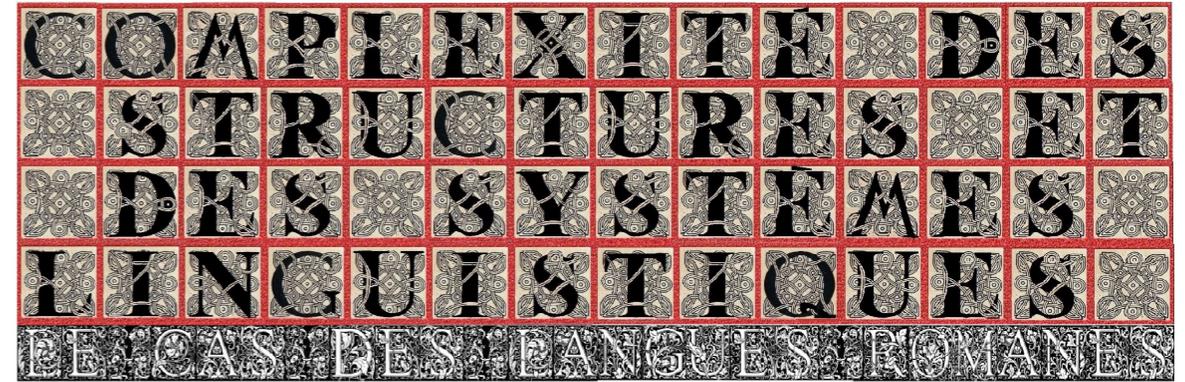
Département d'études romanes



Le présent volume réunit les textes revus, corrigés et augmentés de seize des vingt-trois communications présentées au colloque international *Complexité des structures et des systèmes linguistiques : le cas des langues romanes*, organisé les 24 et 25 novembre 2018 à l'Université de Sofia « Saint Clément d'Ohrid » à l'occasion du 80<sup>e</sup> anniversaire de la naissance du professeur de linguistique Krassimir Mantchev (1938-1997). Séduit par l'approche mentaliste de la psychomécanique du langage de Gustave Guillaume, K. Mantchev a contribué à la diffusion et au développement de la linguistique guillaumienne en Bulgarie grâce à plusieurs ouvrages qu'il a réalisés pendant sa carrière au Département d'études romanes de l'Université de Sofia et qui lui ont valu une reconnaissance internationale.

Les seize contributions sont organisées en trois parties thématiques, précédées d'un chapitre introducteur, présentant les enjeux théoriques et les diverses approches de la complexité, une des notions les plus controversées et débattues en linguistique ces vingt dernières années. Tirant ses sources des sciences exactes, la complexité se présente décidément comme un concept interdisciplinaire qui empiète largement sur les notions de difficulté, cardinalité, redondance, marque, intrication, sans s'identifier complètement avec aucune d'elles. Cette ambiguïté inhérente à sa signification n'en rend pas le terme moins opérationnel, comme le montrent les divers contributeurs au volume, qu'il s'agisse d'analyser la complexité de structures linguistiques isolées, celle des diverses composantes des systèmes linguistiques (phonologique, morphologique, syntaxique...), celle du langage comme phénomène, ou enfin celle du rapport entre la langue et son environnement non linguistique.

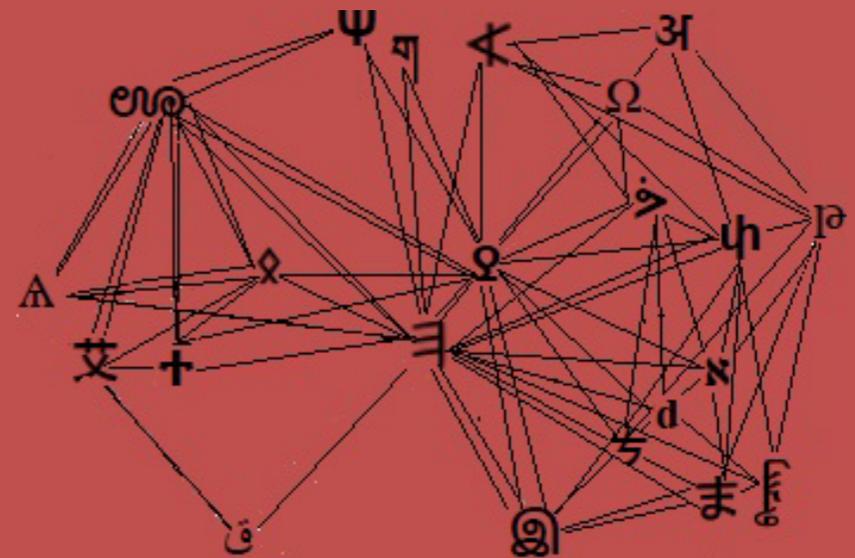
COMPLEXITÉ DES STRUCTURES ET DES SYSTÈMES LINGUISTIQUES : LE CAS DES LANGUES ROMANES



Département d'études romanes  
Université de Sofia « Saint Clément d'Ohrid »

## Complexité des structures et des systèmes linguistiques : le cas des langues romanes

Sous la direction de  
Ivaylo BUROV & Giuliana FIORENTINO



CU Romanistika

# 6

---

## Complexité et dialectologie quantitative et qualitative : éléments d'écologie diasystémique

Jean Léo LEONARD\*

A Lilianne Jagueneau, *in memoriam*

### § 1. Introduction

#### § 1. 1. Objectifs

Cet article<sup>1</sup> aura pour principal objectif, non pas de présenter un panorama général de l'*écologie diasystémique*<sup>2</sup> – on trouvera une « fresque » de ce genre

---

\* Sorbonne Université, STIH (EA 4509) [2014-2019] ; Université de Montpellier 3 (Paul-Valéry), Dipralang (EA 739) [2019-], [leonardjeanleo@gmail.com](mailto:leonardjeanleo@gmail.com).

<sup>1</sup> L'auteur de ces lignes est le principal rédacteur et endosse toute responsabilité pour d'éventuelles erreurs. Mais ce texte de synthèse qui se veut une contribution à la présentation de méthodes en linguistique variationnelle (ou dialectologie) adossées à des techniques et méthodes quantitatives issues de la Théorie de la Complexité n'aurait pas été possible sans l'apport considérable de collaborateurs, dont la liste (non exhaustive) est longue : Marco Patriarca, Els Heinsalu, Jürgen Rannap, Gilles Polian, Flore Picard, Kiran Sharma, Anirban Chakraborti, Vittorio dell'Aquila, Daniele Dalmassio. Ces co-auteurs implicites seront cités en notes à chaque fois qu'un fragment de leur travail sera mis à profit dans ce qui, sous le nom d'un seul auteur qui n'est en réalité que rapporteur, n'est autre qu'un travail collectif réalisé depuis 2010 sur un vaste chantier de recherches, commencé par l'obtention d'une chaire IUF (Institut Universitaire de France, projet MAMp) de 2009 à 2014, puis continué grâce à une série de financements : LabEx EFL (2011-), puis Parrot 2017-2018 (fondation Hubert Curien) et Idex Emergence (projet LaDyCa, 2017-2018). L'auteur du présent rapport de synthèse n'est jamais que l'humble et dévoué scribe de ce collectif de travail. Selon la formule consacrée, les recherches présentées ici ont bénéficié d'une aide de l'État gérée par l'Agence nationale de la recherche au titre du programme « Investissements d'Avenir » portant la référence ANR-10-LABX-0083 dans le cadre de deux opérations du LabEx EFL (PPC11 : Complexité et diffusion des systèmes phonologiques et EM2, axe 7 : élicitations croisées).

<sup>2</sup> J'entends par *écologie diasystémique* l'étude des conditions de diversification interne d'une langue ou d'une sous-famille de langues (au stade du « continuum dialectal »), tant du point de vue interne (variables lexicales, phonologiques et grammaticales), que du point de vue externe (dynamiques d'interactions entre communautés linguistiques parlant des variétés dialectales du continuum ; contraintes écologiques naturelles liées à la géographie et aux ressources alimentaires, aux terrains agricoles et cultures ; systèmes agraires et systèmes d'intercomplémentarité socioéconomique ou « distribution du travail » dans une région/aire écologique donnée). Comme dans toute approche écologique, les notions de *biotope*, de *niche*, de *bassin d'interactions*, sont ici fondamentales. Cf. Wendel (2005) qui résume ainsi les dynamiques d'écologie langagière : (i) *Diversité*, (ii) *Variation*, (iii) *Transmission*, (iv) *Sélection & adaptation* (ou *exaptation*), (v) *Niche*, (vi) *Rythmes et cycles de changement systémique, externe et interne*.

dans Léonard (2017b) –, mais plutôt d’illustrer, à l’aide de ce thème, ce qu’est le traitement de la *complexité* en linguistique, et plus particulièrement en *géolinguistique*. De là, il s’agit d’expliciter deux notions : d’une part, le concept de *systèmes complexes* (flux d’information > auto-organisation > émergence de structures) ; d’autre part, la notion de *simplicité*, permettant de décrire et d’expliquer comment une multitude de structures de surface peuvent procéder, en profondeur, de principes et d’unités constitutives simples (*cf.* Berthoz 2009). Non seulement tout le paradigme du générativisme en grammaire formelle est fondé sur cette dynamique élémentaire (simplicité > complexité), mais toutes les sciences modernes suivent également ce prérequis fondamental – également appelé *réductionnisme*, en opposition avec l’*atomisme*, sur le plan épistémologique. Le réductionnisme lui-même est appelé à être intégré dans une perspective *holiste*, dans le cadre de l’observation des systèmes complexes, qui subsume de multiples systèmes en interaction, faisant émerger des structures variables, quoique robustes et pertinentes.

Nous procéderons à l’aide d’exemples tirés de langues non indo-européennes, comme le tseltal (maya occidental), le mazatec (langue popolocane, otomangue oriental) et le numique (uto-aztécane). Du point de vue de la *Théorie de la Complexité*, les structures ne sont pas données en soi, ou au terme d’une analyse univoque : elles émergent en bouquets statistiques et stochastiques, en fonction de *phases* statistiques, ou de *seuils normalisés*<sup>3</sup>. Chaque objet de connaissance (graphe, carte, schéma, etc.) constitutif de ce bouquet n’en reste pas moins valide, et ne représente qu’un état parmi d’autres de la réalité observée. L’enjeu est donc de définir des *seuils de pertinence structurale*, qui varient au cas par cas, en fonction de la *cohérence* et de la *densité* des *variables* constitutives de la *trame* d’un *diasystème* (ou système de systèmes dialectaux, grammaire globale applicable à un réseau dialectal).

Quoi de plus complexe, en linguistique, qu’un réseau dialectal ? En effet, si une grammaire et un lexique sont en soi complexes, un diasystème forme une multiplicité de variantes d’un lexique et d’une grammaire – on démultiplie ainsi le nombre de plans d’analyse. Et pourtant, les méthodes quantitatives utilisées pour la théorie de la complexité permettent de rendre compte de manière *simplexe* (et non pas *simpliste*) et, partant, *heuristique*, de cette complexité.

Reste, pour le linguiste, une fois réalisée l’approche quantitative, la tâche la plus ardue, et qui incombe à lui seul : la synthèse et l’explication qualitative. Car, en dernier ressort, il ne doit pas se laisser subjugué par la seule magie des « mathématiques sévères », selon l’expression d’Isidore Lucien Ducasse, comte de Lautréamont. Cependant, afin de boucler la boucle, il sera aussi montré pourquoi et comment l’analyse *qualitative* doit primer d’un bout à l’autre de la recherche sur l’écologie diasystémique, et guider l’approche mathématique, car

---

<sup>3</sup> Voir diverses contributions allant dans ce sens sur la page <http://axe7.labex-efl.org/node/353>, notamment de Balev, Léonard & Duchamp et de Léonard, Patriarca & Heinsalu.

toute écologie, par sa vision holistique, globale, des équilibres systémiques, ne peut être appréhendée que par l'association de ces deux formes d'intelligence du réel : *qualitative* et *quantitative*.

### § 1. 2. Prémisses

Plutôt que d'opposer le structuralisme ou le fonctionnalisme à l'approche complexionniste (autrement dit, l'approche en termes de « systèmes complexes », ou selon les méthodes de la *Théorie de la Complexité*), je préférerai une approche en termes de *complémentarité*. Rien de plus nocif pour les sciences que ces dichotomies conçues dans une sorte de linéarité abstraite et idéaliste, alors que la bonne vieille approche cumulative, qui tient le plus grand compte des continuités, plutôt que des ruptures, reste la plus sûre, la plus féconde. Pas de *système complexe* sans *système*, pour commencer. Et il ne s'agit pas ici d'un simple truisme, puisque, actuellement, la notion même de *système* est remise en cause par nombre de linguistes, dans une perspective postmoderniste<sup>4</sup>. Une définition élémentaire du terme de *système*, en linguistique, consiste à poser qu'une langue est fondée comme système par un *lexique* (une collection d'entités lexicales, divisée ou non en parties du discours, à moins qu'elle ne se définisse par un ensemble de racines polyvalentes) et une *grammaire* (un ensemble de règles ou de contraintes d'ordre combinatoire). Dans le cas des langues à racines polyvalentes, dont les catégories lexicales ne sont pas « formatées » en amont par des indices formels, c'est, tout comme pour les lexèmes des langues isolantes, la composante combinatoire (grammaire au sens large : morphologie et/ou syntaxe), qui fonde le système.

La posture structuraliste – mais aussi générativiste – tendait à une certaine réification conceptuelle : les langues pouvaient se décrire en vase clos, hors de tout contexte d'usage, et sous forme de normes descriptives, fondant un modèle descriptif. Cela n'a cependant pas empêché les plus distingués parmi les structuralistes et fonctionnalistes d'apporter une contribution décisive à l'étude de la variation : la filiation qui va d'André Martinet à William Labov, en passant par Uriel Weinreich, est suffisamment connue pour éviter de s'étendre ici. Certes,

---

<sup>4</sup> Cf. Léonard (2017a) pour une critique radicale de cette même critique radicale de la langue comme système. On peut comprendre l'idée de la langue comme *continuum d'usage* plutôt que comme système clos, comme dans la linguistique de « contact, au-delà du contact » prônée par exemple par Matras (2009), dans une perspective tenant compte de l'écologie de la communication interlocuteurs. Mais ne s'agit-il pas, dans les faits, d'une forme de palinodie (propos autoréfutant), dans la mesure où le continuum d'usages est d'autant plus possible qu'il est fondé et délimité par un système (grammaire & lexique) ou par un assemblage de plusieurs systèmes en contact, au sein des répertoires, qui s'organisent eux-mêmes en un système complexe ? Ne confond-on pas non plus « système fermé » (celui de la description des langues *in abstracto*, comme collection d'entités lexicales et de règles de combinatoire) avec la notion de « système ouvert » (systèmes en interaction avec leur milieu externe) ? Je n'aborderai pas plus cette polémique ici, d'autant plus que je pense qu'il s'agit d'un faux problème et d'un faux débat : jusqu'à maintenant, aucun linguiste ni sociolinguiste n'a pu traiter de données linguistiques sans faire appel, qu'il le reconnaisse ou non sur le plan de la posture épistémologique, à la notion de système.

Uriel Weinreich pouvait encore poser comme une sorte de défi, au début des années 1950, la question de savoir si une dialectologie structurale était épistémologiquement envisageable (Weinreich 1954). Mais c'était pour mieux argumenter en faveur de la compatibilité des deux approches – variationniste et structuraliste –, afin de fonder la notion de *diasystème*, que j'utiliserai maintes fois ici. Le *diasystème* est le type même de construit structuraliste qui préfigure une intégration de la notion de système complexe dans l'analyse linguistique. C'est un *métabystème*, qui conçoit la langue comme *un système de systèmes* : une poupée russe, ou poupée-gigogne<sup>5</sup>, qui intègre de multiples « copies » emboîtées en son sein. Chacune de ces « copies » diffère cependant sur quelques « détails », ou variables, ou paramètres structuraux, mais pas seulement en termes de menues scories de formes. Comme les langues sont des systèmes non pas intelligents en eux-mêmes, mais portés par l'intelligence humaine, les « détails » qui font différer les dialectes entre eux sont des faits de structure, plus ou moins régulières, qui impliquent des réagencements ou des réinterprétations du système. Autrement dit, ce sont des variables (par exemple, en phonologie : diphtongaison, monophthongaison, (dé)palatalisation, (dé)vélarisation, (dé)labialisation, métathèse, etc.), ou des paramètres (en morphologie, par exemple, le nombre et la nature des classes flexionnelles, le choix des affixes ou procédés dérivationnels ou compositionnels ; en syntaxe, les règles de sous-catégorisation ou rection, la diathèse, etc.). Les systèmes complexes se définissent précisément comme des systèmes de systèmes. Les unités constitutives ne sont plus seulement, à ce niveau d'analyse, les seules parties du tout (unités fonctionnelles minimales, comme les affixes grammaticaux ou lexicaux, ou ultimes, comme les traits distinctifs en phonologie), mais les modules et sous-systèmes qui composent l'architecture générale d'un macrosystème.

Le Tableau 1 ci-dessous, repris à une contribution programmatique du sociolinguiste Albert Bastardas i Boada, qui a amplement contribué à développer le paradigme de la complexité en sociolinguistique et en contact de langues, confronte les principes de l'approche structuraliste (énoncés ou notions listées de S1 à S10, colonne de gauche du tableau) à ceux de ce qu'il appelle le paradigme de la « *complexics* », en anglais, qui n'est autre que celui de la Théorie de la Complexité (désormais, TC), de manière générale<sup>6</sup> (liste de C1 à C10 dans la colonne de droite du tableau).

---

<sup>5</sup> On trouvera une belle illustration de l'approche *diasystémique* à l'échelle d'une sous-famille de langues dans l'ouvrage de Colin Masica sur les langues indo-aryennes dans leur ensemble (Masica 1990).

<sup>6</sup> On ne répètera jamais assez combien l'ensemble des sciences sont redevables au sociologue Edgar Morin pour le développement de la *Théorie de la Complexité*. La liste de ses ouvrages théoriques et méthodologiques serait trop longue à mentionner ici, et ce serait oublier d'autres penseurs d'un grand impact, comme Ilya Prigogine, René Thom, Ludwig von Bertalanffy et tant d'autres, cf. le remarquable article d'Elvio Ceci dans le présent volume, ainsi que Massip-Bonet & Bastardas-Boada (2013) pour une application de la TC à la sociolinguistique.

**Tableau 1** : Comparaison entre l'approche structuraliste vs. complexionniste, selon Bastardas i Boada (2013 : 223), modifié et augmenté

Perspective « structuraliste »	Perspective « complexionniste »
S1 : Réification conceptuelle	C1 : Subjectivité et interprétation de la part de l'observateur (le point de vue crée l'objet)
S2 : Territoire	C2 : Cartographie, cartes mentales, projections multiples de plans ou de construits/structures
S3 : Vérité scientifique	C3 : Hypothèses de travail, théories successives et holographiques, vicariance
S4 : Eléments, unités discrètes	C4 : Eléments en contexte, unités interactives, imbrication, interdépendance, réseaux
S5 : Objets	C5 : Devenirs et processus
S6 : Statisme	C6 : Dynamiques fluides, changement, lignes évolutives
S7 : Logique « classique » (règles)	C7 : Logique diffuse, voie floue (contraintes)
S8 : Causalité linéaire	C8 : Causalité circulaire et rétroactive (cyclicité)
S9 : Dichotomisation et dualisme	C9 : Intégration et complémentarité
S10 : Planification	C10 : Auto-organisation (ou autopoïésis) et émergence

Un diasystème est un bon exemple de structure émergente de niveau complexe. Les trois notions fondamentales qui définissent ce type de structures sont :

- Flux d'information
- Auto-organisation (ou *autopoïésis*, *autopoïèse*)
- Emergence

A ce titre, toute carte géolinguistique offre au regard un flux d'information – même si je suis bien conscient que je parle là en termes d'artefact. Il en va de même de toute carte synthétique, présentant l'écheveau d'un faisceau d'isoglosses. Dans la réalité sociale et communicative de la langue, le flux d'information dialectale est constant. Dans un article sur la dialectologie perceptive, j'avais simulé statistiquement sur la base de deux idiolectes poitevins insulaires de l'île de Noirmoutier, la quantité d'information phonétique, phonologique, morphologique, lexicale et syntaxique susceptible d'être perçue par les locuteurs (Léonard 2002). Je mentionnerai rapidement ces résultats, qui testaient la perception d'un auditeur du hameau de La Bosse (environ 40 ans, marin pêcheur actif, de la commune de L'Epine) d'un idiolecte d'une locutrice âgée (environ 80 ans, couturière retraitée) de la commune voisine de La Guérinière, distante de 4,5 kms, sur un laps de temps d'une dizaine de minutes, en nombre absolu d'arrêts de magnétophone. Ce simple fragment minuscule de parole passée au filtre de la perception des différences inter-variétés (à moins de 6 kilomètres de distance) faisait apparaître 37 fléchages de la variation du grain fin des contrastes vocaliques

(macrovariable *Phonétique Voc*), 8 pour le gros grain (les oppositions phonologiques discrètes : macrovariable *Phonologie Voc*), 5 et 11 pour le consonantisme (phonétique et phonologie, respectivement), 5 pour la morphologie (surtout pronominale). Et ce, pour tout juste une dizaine de minutes d'écoute d'enregistrement...

Cet exemple rappelle combien, même lorsqu'on s'intéresse à un instantané perceptif sur le continuum de la parole et de l'usage, on ne peut rien saisir sans procéder à une analyse dans un cadre systémique (par exemple, en distinguant phonétique, au niveau *étique*, d'une part, de phonologique ou phonémique, au niveau *émique*, d'autre part). Surtout, il donne une idée, certes succincte mais concrète, alimentée selon le principe en C1, du flux d'information porté par la parole, au-delà des contenus (énoncés, interprétés en sémantique et en pragmatique), jusque dans la trame de la substance (phonétique) et de sa forme parlée (phonologie, morphologie).

Cet exemple montre aussi à quel point la territorialité (S2) n'est pas nécessaire pour porter de la variation, mettant en évidence que tout fragment de parole peut s'analyser de manière holographique, selon divers points de vue (phonétique vs. phonologique, point de vue du linguiste et point de vue du locuteur) et qu'il y a autant de faits réels que de structures émergentes selon le positionnement (du linguiste, du locuteur). Il montre aussi l'imbrication des niveaux d'analyse (S3 confronté à C3 dans le Tableau 1, autrement dit, la multiplicité des points de vue et la succession des hypothèses, plutôt que la vérité scientifique comme but ultime), car ce fragment de résultats quantitatifs rend certes compte du point de vue *émique* – celui de la perception du locuteur (cf. C1 dans le tableau) –, mais il intègre lui-même l'opposition entre niveau *phonétique* et niveau *phonémique*, si le linguiste prend le parti, comme je l'ai fait, de distinguer entre les deux sans son filtrage des réponses au test (énoncés en C4).

De manière générale, pour recentrer la lecture du Tableau 1, alors que l'approche structuraliste établit des unités discrètes et fonctionnelles bien délimitées (S4), la TC se nourrit d'objets complexes, de rang supérieur : elle traite des systèmes et sous-systèmes en interaction, les met en relation, en perspective (C4). Au lieu de se fixer sur des objets statiques (S5), elle s'intéresse à des processus et à des devenirs (C5), des états non seulement émergents, mais dynamiques – donc, fluides, changeants (C6 contre S6). Au lieu de fonder l'analyse des combinatoires sur des règles strictes (S7), elle la fonde sur une logique plus souple, en termes de contraintes (C7). Le lecteur phonologue aura reconnu, ici, l'esprit qui préside à la modélisation selon la Théorie de l'Optimalité, qui hiérarchise des contraintes en fonction de l'ancrage dans la langue (la fidélité lexicale) et dans le langage (la théorie des conditions de marquage, cf. la contribution dans ce volume d'Ivaylo Burov, ainsi que Burov 2017, 2018, 2019 pour une présentation plus détaillée de sa modélisation de la marque en phonologie). Surtout, la causalité, en TC, n'est pas conçue comme avant tout linéaire, selon une trajectoire inexorable (S8), mais comme non linéaire, ce qui induit la possibilité de retours, boucles

rétroactives, cyclicité (C8). La TC s'accommode très bien des modèles en cascade, mais ne contraint pas les niveaux d'instanciation des causes et des effets, qui peuvent entrer dans des trames et des treillis parallèles, dans des dimensions parallèles. Enfin, à la dichotomisation (noir contre blanc, jour et nuit, polarités (S9)), la TC préfère l'intégration et la complémentarité des états et des processus (C9) ; au déterminisme et à la planification, elle oppose l'auto-organisation, et la co-émergence d'états multiples ou de potentialités (C10).

### § 1. 3. Etat de l'art et plan de l'argumentaire

A titre d'exemples, concernant la notion de *complexité structurale*, on consultera utilement Miestamo, Sinnemäki & Karlsson (2008), Léonard, Demolin & Avilés González (2016) ; la jonction entre dialectologie, géolinguistique et systèmes complexes s'est développée grâce notamment à Kretzschmar (2009, 2010, 2014, 2015) dans une perspective postkurathienne<sup>7</sup>. On trouvera d'autres exemples d'approches holistiques envisageant la variation diasystémique dans son contexte géohistorique et social dans Léonard & dell'Aquila (2014), Léonard *et al.* (2014, 2015a, b, 2016). Une approche diasystémique complexionniste d'ordre qualitatif est développée dans Léonard & Kihm (2015) et Léonard & Fulcrand (2018). Une synthèse de divers modèles de diffusion et d'organisation géospatiale des diasystèmes est donnée dans Léonard (2017b). Je renvoie à Gribbin (2004) pour l'une des meilleures synthèses de vulgarisation sur les principes et les méthodes de la théorie de la complexité, ainsi qu'à Benkirane (2013), qui présente une série d'entretiens avec une vingtaine de chercheurs issus de tous les domaines de la connaissance, utilisant la notion de complexité ou de systèmes complexes.

En § 2. 1 je montrerai à partir d'une étude de cas maya (réseau dialectal tseltal, maya occidental, Mexique sud-oriental) en quoi la perspective complexionniste permet de faire jouer la vicariance des points de vue<sup>8</sup> – à savoir, le fait qu'un seul résultat sur une seule composante du système de la langue n'est pas suffisant, et que c'est à travers une gamme de résultats répartis entre diverses composantes du système (ici, du *diasystème*), qu'on obtient une vision étoffée d'une situation d'écologie diasystémique. Dans la section § 2. 2, je confronterai

<sup>7</sup> Hans Kurath (1891-1992), auteur et coordinateur du *Linguistic Atlas of New England*, 3 vols., New York 1939-1943. Sa vision de la géolinguistique se voulait holistique, car elle voyait dans la variation dialectale et sa diffusion un monde d'indices sur les voies de peuplement, les modes d'urbanisation, les échanges marchands, etc. dans une approche qui préfigure l'holisme du complexionnisme. Sans verser dans l'anachronisme, un fondateur européen de la géolinguistique tel que Jules Gilliéron, tout en manifestant implicitement une vision des réseaux dialectaux en termes de systèmes ouverts, restait très orienté vers des questions de théorie générale (réfutation de l'inexorabilité des lois phonétiques, procédés de collusion homonymiques et de « réparation »), comparé à un fondateur nord-américain comme Hans Kurath – qui, notons-le, était d'origine autrichienne, et donc transposait à sa façon une tradition de pensée européenne, transplantée et adaptée au continent américain.

<sup>8</sup> Cf. le magnifique essai d'Alain Berthoz sur cette notion centrale en TC (Berthoz 2013).

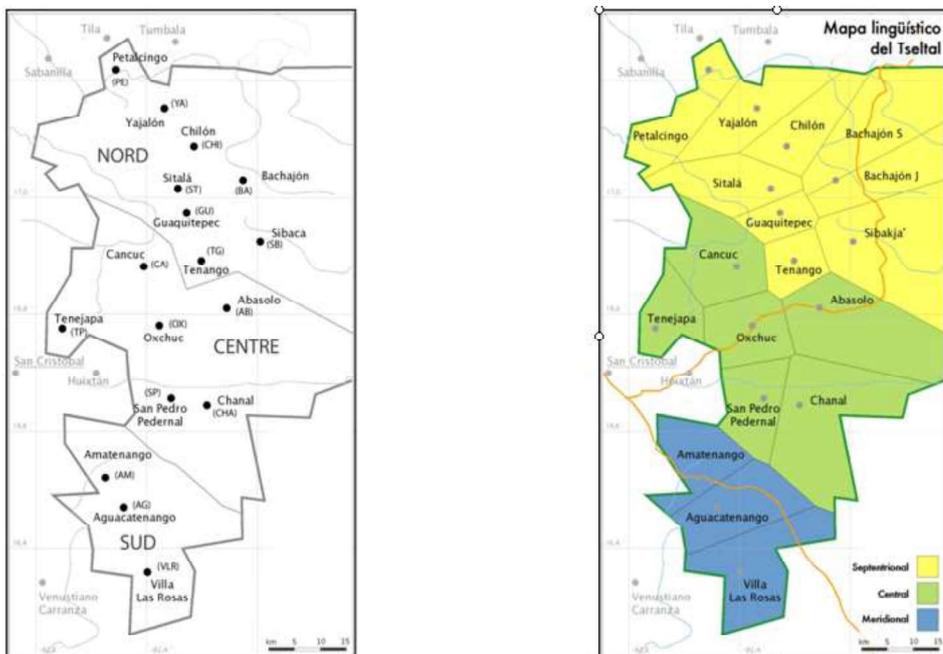
deux modèles d'écologie diasystémique : en § 2. 2. 1 le bassin de traits compact et verticalisé entre hautes, moyennes et basses terres du mazatec, langue otomangue orientale du Mexique ; en § 2. 2. 2 le modèle de diffusion radiale des dialectes ou langues numiques (uto-aztécane), à l'est de la Californie. Ces deux études de cas se fonderont sur des données linguistiques, et feront jouer diverses techniques quantitatives. Cette fois, la vicariance tiendra à la complexité algorithmique, sur la base de l'application de l'algorithme de Levenshtein, ou distance d'édition. Ce parcours complexionniste de langues non indo-européennes sera complété en § 3 par des faits gallo-romans. Nous pourrons alors conclure, dans la section § 4, sur les horizons et perspectives de recherche qu'offrent ces ressources méthodologiques, et leurs conséquences épistémologiques.

## § 2. Etudes de cas

### § 2. 1. Vicariance des composantes : tseltal

Dans ce monde diasystémique, les versants (la verticalité, ou la déclivité écologique, entre hautes, moyennes et basses terres) ne sont pas nécessairement un facteur-clé. La métaphore du *bassin* est préférable à celle de l'*archipel*, même si par certains aspects, ce monde peut être peuplé de singletons – ou d'îlots, comme c'est le cas pour le sud du réseau dialectal tseltal (maya occidental) – cf. Léonard *et al.* (2015a).

Figure 1 : Localités enquêtées dans le cadre du projet ALTO (et données Alisto)  
Cartes de Vittorio dell'Aquila (CELE, Milan/Vaasa)



L'ensemble de la zone peut se décrire comme un haut plateau très encaissé au nord et au centre, qui se fond dans des plaines au sud. L'histoire des mouvements de populations de cette région est assez complexe : après une période pré-colombienne durant laquelle cette zone faisait partie de l'aire centrale du monde maya classique et postclassique, composé de seigneuries relativement autonomes, la période coloniale (1525-1821) s'est caractérisée par des *reducciones* sous l'égide du fraile Pedro Lorenzo de la Nada, qui fit venir des populations des basses terres à l'est du domaine tseltal, transplantées surtout au nord et aux confins septentrionaux du centre du tseltal occidental. Puis l'époque libérale, au XIX<sup>e</sup> siècle, se caractérisa par la domination latifundiaire, qui brassa également les populations. On trouve ainsi des traits nettement méridionaux dans des parlers septentrionaux, par transplantation.

La région centrale, mais aussi le nord, a vu sa démographie bouleversée par les révoltes indigènes contre les abus des propriétaires terriens et des administrateurs de cette province reculée du Nouveau monde hispanique – que même les réformes agraires de l'époque révolutionnaire, au début du XX<sup>e</sup> siècle, n'atteignirent pas vraiment, tant la classe dirigeante locale faisait régner une terreur conservatrice (les romans de B. Traven, alias Traven Torsvan, ou Berick Torsvan sont éloquents sur les luttes paysannes et indigènes durant cette période). Le sud de l'aire tseltale occidentale a connu davantage de stabilité, sous la protection de puissants couvents, mais a fait l'objet d'une intense castillanisation qui fait que le parler le plus méridional, Villa las Rosas (VLR), est une variété actuellement « en danger ». Il est impossible d'entrer davantage dans les détails de cette histoire tourmentée. Nous allons plutôt voir ce que l'étude quantitative du diasystème fait apparaître – l'indice utilisé ici ne sera pas l'algorithme de Levenshtein, qui calcule les différences, mais l'indice de similarité, qui calcule les similitudes, à partir des données du projet ALTO – *Atlas Linguistique du Tseltal Occidental*, en préparation au Ciesas Sureste, Chiapas, en partenariat avec l'auteur de ces lignes (cf. la version électronique en cours de développement, Alisto (Atlas Linguistique Sonore du Tseltal Oriental : <http://alisto.org/inicio>). Il serait vain et redondant de déployer ici le détail des variables phonologiques, morphologiques et lexicales, qui ont été abondamment décrites ailleurs (cf. surtout Léonard & Polian 2014, Polian & Léonard 2017, Léonard *et al.* 2015a).

Afin de comparer les valeurs d'une même variable dans deux dialectes différents, on identifie d'abord la coïncidence entre les deux dialectes, c'est-à-dire les formes partagées<sup>9</sup>. Par exemple, si D1 présente {*a*, *b*} et D2 présente {*a*,

---

<sup>9</sup> Les paragraphes qui suivent, sur le calcul de l'indice de similarité, ont été rédigés par Gilles Polian, et sont extraits de Polian & Léonard (2009). Tous les calculs dialectométriques sont de Gilles Polian (Ciesas Sureste, San Cristobal de las Casas, Chiapas, Mexique), en collaboration avec Alain Polian (DR émérite, Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, Paris). Les schémas des Figures 2, 3 et 4 ont été réalisés par Marco Patriarca et Els Heinsalu (Institut de Physique et de Biochimie, Tallinn), sur la base de la moyenne normalisée, observée sur trois seuils de similarité (cf. Léonard *et al.* 2015a). Qu'ils en soient également remerciés.

$c\}$ , la forme partagée est seulement  $a$ . Si les valeurs sont  $\{a, (b)\}$  d'un côté et  $\{a, b, c\}$  de l'autre, les formes partagées sont  $a$  et  $b$ .

Formellement, on peut décrire un dialecte  $K$  par la formule :  $D_K = \{a_{\alpha,i}, P_K(a_{\alpha,i})\}$  où  $a_{\alpha,i}$  représente l'ensemble des variables,  $\alpha$  représente les différentes possibilités d'une variable donnée  $i$ , cet indice variant de 1 à 19. Dans ce cas,  $P_K(a_{\alpha,i})$  est la proportion de la variable  $a_{\alpha,i}$ .

On identifie alors d'abord l'intersection entre les deux groupes de valeurs, c'est-à-dire la proportion minimum partagée. Par exemple, si l'on a  $\{a, b, c\}$  d'un côté et  $\{a, d\}$  de l'autre, la forme partagée est  $a$ . Dans le premier cas,  $a$  présente une proportion de 0, 33, et dans le deuxième cas de 0, 5. L'intersection est donc de 0, 33. Ceci peut s'écrire à l'aide de la formule suivante, où *inf* signifie qu'on prend la plus petite des deux valeurs :

$$(1) \quad R_{K,L}(i) = \sum_{\alpha} \inf [P_K(a_{\alpha,i}), P_L(a_{\alpha,i})]$$

La deuxième partie du calcul de l'IS (ou Indice de Similarité) consiste à faire la moyenne entre les proportions respectives des formes partagées. Dans l'exemple précédent, les proportions sont de 0, 33 et 0, 5, ce qui nous donne une moyenne de 0, 42. Soit :

$$(2) \quad \langle R_{K,L}(i) \rangle = \frac{\sum_{\alpha} \sum_{I=K,L} P_I(a_{\alpha,i})}{2}$$

Finalement, l'indice de 0 à 1 s'obtient en faisant la moyenne entre l'intersection et la moyenne des proportions, c'est-à-dire entre les deux valeurs obtenues au moyen des formules précédentes. Dans notre cas, la moyenne entre 0, 33 et 0, 42 donne 0, 38. Il est facile de vérifier que s'il n'y a aucune forme partagée, le résultat sera de 0, et que si toutes les formes sont partagées, le résultat sera de 1. Soit :

$$(3) \quad I_{K,L}(i) = \frac{R_{K,L}(i) + \langle R_{K,L}(i) \rangle}{2}$$

A titre d'exemple, une matrice d'indices de similarité, à partir de laquelle sont obtenus les résultats d'un des tableaux les plus congruents avec la classification traditionnelle des dialectes tseltal (à savoir la Tableau 2), sera présentée après la Figure 4, à laquelle elle correspond.

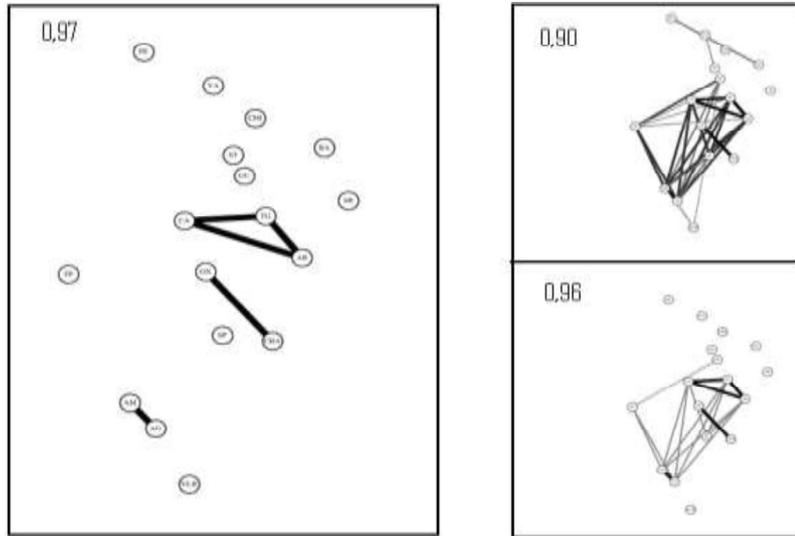
Pour l'instant, voyons l'intéressante asymétrie qui émerge de l'indice  $w$ , entre les seuils de similarité de la phonologie (Fig. 2) et de la morphosyntaxe (Fig. 3), et qui n'était aucunement prévisible en suivant les prémisses S2 (territorialité stricte), S3 (vérité scientifique à un seul choix), S4 (Eléments, unités discrètes), et S6 (Statisme) du Tableau 1. Les deux premières séries de résultats

feront pencher la balance pour les options épistémologiques C2 (Projections multiples), C3 (Modèles vicariants, vision holographique), C4 (Unités interactives, imbrication, interdépendance, réseaux) et C6 (Dynamiques fluides, changement, lignes évolutives).

Le principal apport du calcul de similarité entre les trois composantes que sont la phonologie, la morphosyntaxe et le lexique tient dans la spectaculaire asymétrie des configurations, lorsqu'on fait appel à l'indice  $w$  (moyenne normalisée de l'indice de similarité interponctuelle) : la première composante regroupe le sud et le centre contre le nord, tandis que la seconde composante, au contraire, regroupe le nord et le centre contre un sud qui s'avère dépourvu d'unité de ce point de vue – seulement des singletons. Enfin, le traitement du lexique fait apparaître la tripartition classique entre nord, centre et sud, avec un sous-dialecte au centre-ouest (Tenejapa).

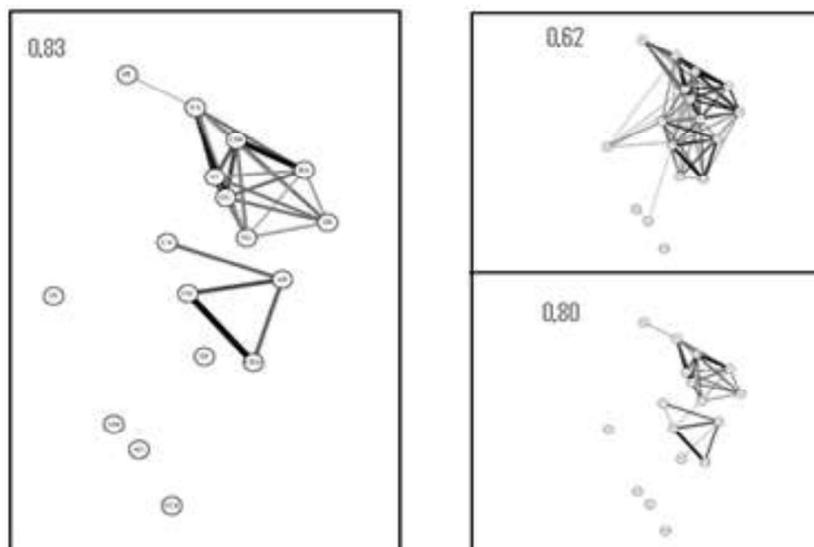
Les échelles diffèrent notablement en fonction de la nature plus ou moins binaire des catégories : le lexique a nécessairement des indices plus tranchés, avec des écarts plus contrastés (un item lexical est attesté ou ne l'est pas, ou bien on observe du polymorphisme, mais la saisie des données reste catégorique), tandis que la phonologie se caractérise par des variables plus intriquées, sur un continuum (ainsi, le vocalisme du tseltal varie peu dans les timbres, mais beaucoup dans la structure des noyaux, qui peuvent être simples ou complexes, articulés ou réarticulés, etc., ce qui forme des *nexti catégoriels*, ou des catégories intriquées davantage que immédiatement discrètes). Dans la Figure 2, le plus haut seuil de similarité phonologique (0, 97, donc une valeur très proche de la similitude totale, qui serait de 1, 00) fait émerger trois agrégats : un triangle au centre, et deux barres au centre-sud et au sud. Ce sont là des *chorèmes*, ou des *cœurs*, les noyaux aréologiques (relatifs aux dynamiques d'aires géolinguistiques). Toutes les autres variétés (ou sommets du graphe) ne peuvent afficher un tel degré de similitude entre elles, et apparaissent comme des singletons, pour le moment – ce sont autant d'îlots *par défaut*, à ce stade de l'analyse. Plus on abaisse le niveau de similitude, et plus on voit émerger des structures complexes et connectées : le degré 0, 96 montre bien l'union du sud et du centre (hormis un singleton tout au sud, VLR), avec des effets de trame du plus haut intérêt, car Tenejapa, au centre-ouest, se connecte avec une variété de la partie méridionale du nord et seulement avec elle. De même, le chorème central reste très sélectif dans ses liens avec les autres points du réseau. La quasi-totalité des variétés septentrionales reste à l'état d'archipel de singletons – une chaîne non liée, là encore, par défaut, à ce stade de l'analyse. Enfin, quand on abaisse à 0, 90 le seuil de similarité – un niveau relativement trivial pour un réseau dialectal –, alors la trame des relations entre les sous-composantes du diasystème devient plus explicite : le centre et le sud forment bel et bien un conglomérat dense, tandis que le nord constitue un ensemble bien moins compact, et que la trame de ses connexions reste élémentaire (un simple vecteur transversal, avec juste une connexion vers le centre).

**Figure 2** : Regroupements en fonction de l'indice  $w$  pour la phonologie dans ALTO



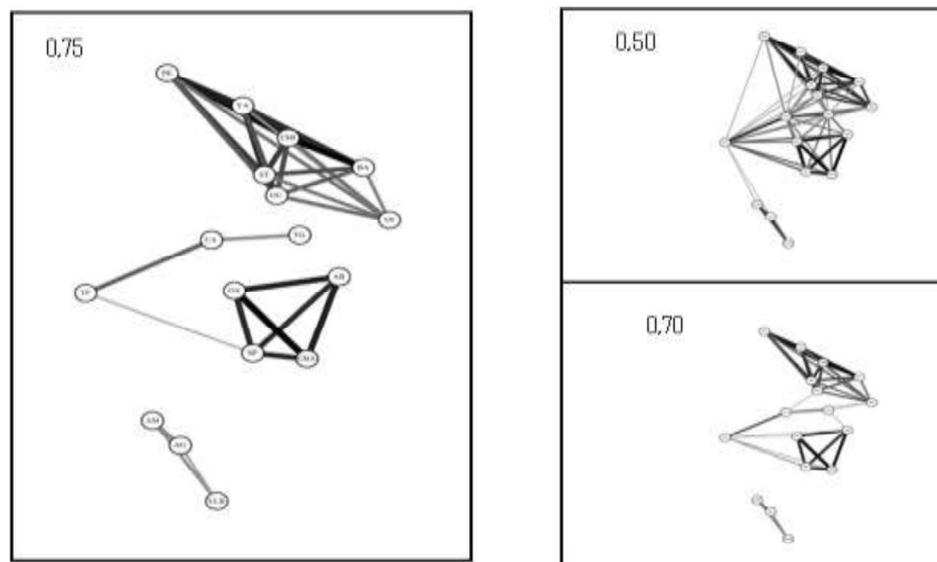
La morphosyntaxe, dans la Figure 3, présente l'image-miroir, quasiment, de la précédente configuration : au seuil 0, 83, deux sous-réseaux dialectaux à trame dense émergent au centre et au sud (*cf.* Polian & Léonard 2009, pour les matrices de similarité). Ces deux grappes (ou *cliques*) se rejoignent par une arête seulement au seuil 0, 80, et elles forment une pelote dense au seuil 0, 62. Cette fois-ci, c'est clairement le nord et le centre qui font alliance contre le sud, et qui ne s'allient que de manière clairsemée au centre-ouest (Tenejapa).

**Figure 3** : Regroupements en fonction de l'indice  $w$  pour la morphosyntaxe dans ALTO



Enfin, cette méthode, qui consiste à évaluer la distance des sommets d'un graphe par rapport à la moyenne des similarités sur la base d'un indice 1,00 de moyenne normalisée, apporte des résultats saisissants, lorsqu'on l'applique sur une base de données lexicales. Cette composante permet de détecter dès le seuil le plus élevé de similitude (0,75) trois chorèmes parfaitement congruents avec la tripartition nord vs. centre vs. sud généralement acceptée, en incluant également Tenejapa au centre-ouest comme un sous-dialecte du dialecte central (Hopkins 1970, Campbell 1987). La trame est dense d'emblée, dès ce niveau d'analyse, y compris pour le nord, dont les liens, pour la phonologie, semblent *a posteriori* exsangues, quand on les compare, tels qu'ils apparaissaient dans la Figure 2 à la pelote solidement tressée qui émerge du seuil minimal de différenciation, dans la Figure 4.

Figure 4 : Regroupements en fonction de l'indice  $w$  pour le lexique dans ALTO



La catégorialité du lexique fait qu'aucun secteur du diasystème n'apparaît comme singleton, et que les trames y sont clairement hiérarchisées – la relation de Tenejapa aux autres variétés centrales est, à ce titre, riche d'enseignements. Les seuils 0,70 et 0,50 ne font que confirmer ces effets de mise en relief de la trame diasystémique, au point que cette fois, au niveau 0,50, même le sous-réseau méridional est interconnecté – et le fait que cette transitivity diasystémique passe par Tenejapa n'a rien d'anodin.

Tableau 2 : Indices de similarité pour 36 items lexicaux issus des enquêtes ALTO

	PE	YA	CHI	BA	ST	GU	SB	TG	CA	AB	OX	TP	SP	CHA	AM	AG	VLR
PE		93.0	96.9	95.6	91.4	87.5	84.0	56.5	62.8	32.6	26.0	53.0	29.0	25.3	46.5	30.9	38.0
YA	93.0		93.9	92.6	93.4	88.5	79.0	55.5	55.8	31.6	25.0	46.0	28.0	24.3	45.5	33.9	35.0
CHI	96.9	93.9		98.5	91.3	88.9	83.9	56.4	62.6	32.5	25.9	52.9	27.9	25.1	45.4	29.8	36.9
BA	95.6	92.6	98.5		89.0	87.1	81.6	56.1	62.4	33.3	26.6	52.6	29.6	25.9	46.1	30.9	37.6
ST	91.4	93.4	91.3	89.0		91.4	83.4	59.4	65.6	37.8	32.4	55.9	33.4	31.6	47.9	34.3	38.4
GU	87.5	88.5	88.9	87.1	91.4		85.5	68.5	70.3	43.4	39.5	57.5	40.5	38.8	49.0	36.4	34.5
SB	84.0	79.0	83.9	81.6	83.4	85.5		71.0	70.8	47.9	42.0	58.0	43.0	41.3	45.5	35.9	40.0
TG	56.5	55.5	56.4	56.1	59.4	68.5	71.0		81.8	70.4	66.5	67.5	67.5	65.8	44.0	39.4	38.5
CA	62.8	55.8	62.6	62.4	65.6	70.3	70.8	81.8		67.9	66.8	87.3	67.8	66.0	44.3	40.6	36.8
AB	32.6	31.6	32.5	33.3	37.8	43.4	47.9	70.4	67.9		95.6	72.1	92.6	94.9	39.1	41.0	37.6
OX	26.0	25.0	25.9	26.6	32.4	39.5	42.0	66.5	66.8	95.6		71.0	93.0	99.3	36.5	38.9	33.0
TP	53.0	46.0	52.9	52.6	55.9	57.5	58.0	67.5	87.3	72.1	71.0		76.0	72.3	51.5	50.1	44.0
SP	29.0	28.0	27.9	29.6	33.4	40.5	43.0	67.5	67.8	92.6	93.0	76.0		94.3	43.5	41.9	40.0
CHA	25.3	24.3	25.1	25.9	31.6	38.8	41.3	65.8	66.0	94.9	99.3	72.3	94.3		37.8	40.1	34.3
AM	46.5	45.5	45.4	46.1	47.9	49.0	45.5	44.0	44.3	39.1	36.5	51.5	43.5	37.8		81.4	80.0
AG	30.9	33.9	29.8	30.9	34.3	36.4	35.9	39.4	40.6	41.0	38.9	50.1	41.9	40.1	81.4		76.9
VLR	38.0	35.0	36.9	37.6	38.4	34.5	40.0	38.5	36.8	37.6	33.0	44.0	40.0	34.3	80.0	76.9	

En somme, cette méthode des seuils de moyenne normalisée de l'index de similarité (désormais, nous y ferons référence sous le terme « indice  $w$  ») permet de visualiser la structure profonde du réseau dialectal avec une remarquable finesse de grain. Et il se trouve que ce réseau dialectal se caractérise à la fois par une robuste tripartition, comme le montre le lexique (Figure 4), avec en outre une nette asymétrie entre le nord et le sud dans leur relation avec le centre. Le jeu des trois phases de visualisation des relations de similarité permet de voir de manière quasiment holographique différentes étapes de constitution de ce diasystème. Ce type de réseau dialectal, au fond très hétérogène, bien que fondé sur un noyau structural très robuste, est plus que jamais récurrent dans un monde globalisé et en voie d'urbanisation exponentielle. Cette méthodologie pourrait être utilisée avec succès en dialectologie urbaine, mais aussi dans les situations de créolisation et de pidginisation.

## § 2. 2. Contrastes entre structures émergentes : mazatec vs. numique

### § 2. 2. 1. Mazatec : un bassin de traits compact verticalisé

Le mazatec est une langue otomangue du sud-est du Mexique, parlée par plus de 220 000 locuteurs, dans un biotope tropical fortement diversifié, sur trois paliers géographiques auxquels correspondent trois systèmes de cultures post-coloniales : basses terres (Ixcatlán et Soyaltepec : système canne à sucre), terres moyennes (Jalapa de Díaz : élevage), et hautes terres (Huautla : système café, en crise depuis les années 1980). A cette tripartition s'ajoute le Cañon tropical de Cuicatlán (avec le bourg mazatec de Chiquihuitlán), que caractérisent sa densité de microclimats ainsi que le contact de langues et de sociétés méso-américaines

(mazatec, cuicatec, mixtec, chinantec, etc.). Classée réserve de la biosphère (*Reserva de la Biosfera Tehuacán-Cuicatlán*), la région s'étend sur une surface de 490 817 hectares).

Depuis soixante ans, les populations autochtones de la région mazatèque affrontent une crise profonde et multiforme, à tous les étages du biotope complexe que nous venons de décrire. La structure agraire de *microfundios* (petite propriété, polyculture) et les réseaux familiaux et de solidarité (les « loyautés primordiales » selon les anthropologues), mais aussi l'accès à la terre et les fondements mêmes de la territorialité des populations autochtones ont été défaits dans les basses terres par la construction d'un barrage hydroélectrique<sup>10</sup> et l'essor de structures latifundiaires liées à la monoculture de la canne à sucre (grandes propriétés, agro-industrie). Dans les terres moyennes, c'est le complexe agro-industriel lié à l'élevage intensif qui a remis en cause les liens sociaux et territoriaux. Enfin dans les hautes terres, l'effondrement du *système café* (plantations de café et circuits commerciaux liés à cette production) a recentré l'économie des centres urbains vers le commerce et les services, dans une région très désavantagée par le relief. Le *système milpa* (l'agriculture traditionnelle, fondée sur la polyculture familiale de subsistance) et le marché du travail local ne suffisent plus à retenir la population sur place, qui tend à émigrer vers les centres urbains ou l'Amérique du nord.

Le Tableau 3 énumère un ensemble de variables relevant de la phonologie diachronique. Il permet de donner une idée des conditions de complexité structurale caractéristiques de la phonologie du mazatec et de son diasystème. On consultera également Gudschinsky (1958a, b), pour une contextualisation de ces variables au sein du popolocan, ou une première modélisation des dynamiques isoglottiques du mazatec. Les variables retenues ici le sont à titre de simple échantillon. Il va de soi que la quantité de phénomènes que prend en compte le calcul de distance d'édition (algorithme de Levenshtein) est bien plus étendue. Ces variables ont le mérite de faire ressortir l'incidence de propriétés typologiques de la langue, typiquement otomangues, comme la corrélation de qualité de voix (soufflée vs. craquée) et ses réalignements syllabiques ou sa neutralisation, selon les dialectes, ou encore, les occlusives prénasalisées (qui ne sont autres que

---

<sup>10</sup> La Presa Miguel Alemán, mise en service en 1959, s'étend sur 47 800 hectares de terres englouties, avec une capacité de 8 millions de m<sup>3</sup>, 725 million kilowatts/an de la centrale hydroélectrique de Temascal : 22 000 paysans indigènes furent alors réinstallés dans l'Etat voisin de Veracruz. Si l'on tient compte que le barrage en question est relié par un canal à une autre structure du même type, la Presa Cerro de Oro, les proportions s'amplifient pour atteindre un volume de 13 380 millions de m<sup>3</sup>. Ces détails comptent, car cette phase de « développement » économique et technologique de la région a accru la complexité du diasystème mazatec, avec les nouvelles variétés dialectales issues de ce déplacement, et par le contact intensifié avec l'espagnol, assimilateur, dans les basses terres. En outre, le deuxième barrage, de la Presa Cerro de Oro, a été facteur de contact entre chinantec et mazatec, deux langues otomangues d'une grande complexité structurale, mais très différentes, quoique génétiquement reliées. Ces dimensions du changement écolinguistique restent à étudier.

des sonantes nasales postoralisées au contact de noyaux syllabiques oraux, *cf.* Do Bui 2018).

Le protosystème du mazatec reconstruit par Gudschinsky (1958a) et repris par Kirk (1966) est le suivant pour les consonnes : six occlusives et mi-occlusives (ou affriquées) *\*t, \*tʃ, \*k, \*kʷ, \*ts, \*tʃ*, deux fricatives sibilantes *\*s, \*ʃ*, trois sonantes nasales *\*m, \*n, \*ɲ*, deux approximantes ou glides *\*j, \*w*, et deux glottales *\*h, \*ʔ*. Une liste de consonnes comme celle-ci est représentative des pivots contoïdes. Il faut compter également avec la qualité de voix (craquée, notée à l'aide d'un tilde souscrit *vs.* soufflée, notée par deux points souscrits :  $\underset{\cdot}{C}$  et  $\underset{\cdot\cdot}{V}$ ) ainsi qu'avec la nasalité (occlusives prénasalisées, voyelles nasales). Concernant les voyelles, le système est pentavocalique (*\*i, \*e, \*a, \*u, \*o*), avec des jeux d'alternances distributionnelles pour les deux voyelles haute et moyenne postérieures /u/ et /o/ ; il faut compter également avec une classe de huit diphtongues (*\*ia, \*iu, \*ei, \*ai, \*ui, \*ue, \*ua, \*au*), qui partagent toutes un vocoïde haut *i, u* comme pivot.

Le statut des voyelles dites « réarticulées » (VʔV) ou, selon nos propres termes, « réaspirées » (VhV) relève, à notre avis, d'une des propriétés prosodiques et syllabiques de la qualité de voix – à notre sens, ces noyaux complexes sont des voyelles respectivement craquées et soufflées en situation de plénoglossie, autrement dit, en position forte. Nous les noterons donc VʔV et VhV plutôt que  $\underset{\cdot}{V}V$  et  $\underset{\cdot\cdot}{V}V$  – le calcul de distance d'édition ne s'accommode guère de diacritiques, et préfère des objets simples et discrets à analyser, en termes de présence *vs.* absence. Cependant, en tant que linguiste, je souscris pleinement à la théorie de Golston & Kehrein (1998, 2004), selon laquelle la glottalisation se manifeste en mazatec par des traits autosegmentaux qui suivent le *principe de l'alignement senestre* (reporter la réalisation du trait phonologique sur la gauche du segment :  $\underset{\cdot}{C} \rightarrow ʔC$ ,  $\underset{\cdot}{C} \rightarrow hC$  ;  $\underset{\cdot}{V} \rightarrow ʔV$ ,  $\underset{\cdot}{V} \rightarrow hV$ , réalisations qui expriment la corrélation de qualité de voix (modale *vs.* craquée *vs.* soufflée)). En ce qui concerne l'inventaire tonal, l'état initial reconstitué par Gudschinsky et Kirk consiste en quatre tons de niveau (*level tones*) : Haut (H), mi-haut (h), moyen (M) et bas (B), et trois tons montants : Moyen-Haut (MH), Bas-Haut (BH) et Bas-Moyen (BM), soit les classes *\*H, \*h, \*M, \*B* et *\*MH, \*BH* et *\*BM* (Kirk 1966 : 8). Mais les listes de cognats de Paul Kirk ne prennent en compte le facteur tonal que pour les trois principaux dialectes urbains que sont Huautla (HU), Jalapa (JA) et San Miguel Soyaltepec (SO)<sup>11</sup>, si bien que le calcul de distance d'édition n'a pu en tenir compte.

<sup>11</sup> Abréviations des localités listées dans cette section et dans le Tableau 3 : Ay = Ayautla, Cq = Chiquihuitlán, Do = Santo Domingo, Ix = San Pedro Ixcatlán, Ji = Jiotes, Hu = Huautla, Ja = Jalapa, Lo = San Lorenzo, Mg = San Miguel Huautla, So = San Miguel Soyaltepec, Te = San Jerónimo Tecoaatl (conventions reprises de Kirk 1966).

Tableau 3 : Echantillon de variables diasystémiques mazatèques

Protoforme (Kirk 1966)	Exemple	Glose	Ex. Loc.	Processus
*ntuhu >	ntu	large	Hu	Apocope (réduction VhV)
	ntuhuu		Lo	Rétraction de V haute vélaire
	nduhu		Do	Voisement d'occlusive prénasalisée (nC)
*ntiʔju >	nĩʔju	fourmi	Mz	Simplification sonantique de nC
	nĩju		Te	Modalisation d'une attaque <i>creaky</i>
*tʃintu >	tʃintu	ver	Ay	Dissimilation séquence V hautes
	tʃintiu		Te	Propagation de V haute ou épenthèse
*nʔiʔi >	nʔiʔi	feu	Ay	Dépalatalisation
	liʔi, lʔi /li/		Hu	Latéralisation de nC palatal
*nkahau >	nkaho	grotte	Ji	Monophthongaison AU
> *nkaihau >	nkiho		Te	Dissimilation palatale intersyllabique
	nkeiho		Cq	Diphthongaison décroissante palatale EI
*nkahau >	nkohwu		Lo	Rehaussement/labialisation de V basse
*ntʃit̃ >	ntʃit̃	épis de maïs	Ix	Réduction palatale de V haute vélaire
> *ntʃit̃ >	ntʃit̃ē		Lo	Abaissement de V haute palatale
*sinē >	sinĩ	jaune	Mz	Assimilation progressive de V haute palatale
* <sup>h</sup> nka /ŋka/ >	nka	aile	Te	Modalisation d'attaque initiale nC <i>breathy</i>
* <sup>h</sup> a /ʃa/ >	sa	amer	Cq	Modalisation d'attaque initiale (fricative) <i>breathy</i>
* <sup>h</sup> awa >	tʃuwa	blanc	Te	Affrication cacuminale & labialisation de V basse
* <sup>h</sup> eʔe /teʔe/ >	t <sup>h</sup> aʔa /taʔa/	sorcellerie	Lo	Abaissement de V moyenne palatale
* <sup>h</sup> ue /sue/ >	asu	chaud	So	Prothèse V basse dans monosyllabe
	su		Ja	Réduction séquence UE
*nãũ >	ʔãũ	cinq	Hu	Amuïssement d'attaque nasale initiale
*nãũ > ãũ >	ʔõ /õ/		Ay	<i>Idem</i> & Monophthongaison AU
*nãũ > ãũ > ʔõ >	ʔũ /ũ/		Ix	<i>Idem</i> & rehaussement de V moyenne vélaire
... > ʔõ > ʔũ >	ũ		Lo	<i>Idem</i> & rétraction de V haute vélaire & modalisation V
*t <sup>h</sup> au /tʃau/ >	* tʃo	œuf	Mz	Monophthongaison AU
*t <sup>h</sup> au /tʃau/ >	t <sup>h</sup> u /tʃũ/		Te	<i>Idem</i> & rehaussement V moyenne vélaire
*t <sup>h</sup> au > t <sup>h</sup> u /tʃũ/ >	t <sup>h</sup> i /tʃi/		Lo	Antériorisation de V haute secondaire
*jaha >	jiha	lombric	Ay	Dissimilation VhV homorganique
*jaha > jiha >	ʔiha /iha/		Te	<i>Idem</i> & Glottalisation de yod attaque initiale
*nʔki >	ʃki	potion	Hu	Assibilation de nC palatal préconsonantique
*nʔki > ʃki >	hki /ki/		Mz	<i>Idem</i> & débuccalisation <i>breathy</i> chuintante implosive

La base de données des listes de cognats de Paul Livingston Kirk est considérée comme les réponses à une série de « tests » de phonologie diachronique menés dans douze localités du domaine mazatec. Les « tests » sont symbolisés par l'index  $k$ , allant de  $k = 1$  au nombre total de tests représentés par l'ensemble des données  $N_T$ . Les localités, quant à elles, sont représentées par les index  $i$  (ou  $j$ ), allant de  $i = 1$  ( $j = 1$ ) au total des localités représentées :  $i = N_L$  ( $j = N_L$ ). La réponse par exemple à un test  $k$  mené dans une localité  $i$  est représentée par la séquence  $a_i^k$  où  $k$  vaut pour le test et  $i$  pour la localité. Toutes les données recueillies à partir des tests se combinent dans une matrice de séquences  $A$ , comme illustré ci-dessous :

$$(4) \quad A = \{a_i^k\}, k = 1, \dots, N_T; i = 1, \dots, N_L.$$

La distance de Levenshtein  $L(a, b)$  revient donc à une mesure triviale du niveau de différenciation entre deux chaînes  $a$  et  $b$ . Etant donné deux séquences  $a$  et  $b$ , par exemple  $a = nt\text{fu}\tilde{i}$  et  $b = nt\text{fi}\tilde{i}$  'épis de maïs', la distance de Levenshtein  $L(a, b)$  sera définie comme le nombre minimum d'opérations impliquées dans ce couple, en termes d'insertions, d'effacements, ou de toute autre manipulation nécessaire pour changer  $a$  en  $b$  et inversement. Ce mode de calcul présente l'avantage d'obéir à une définition simple, et d'être facile à manier. En effet, l'algorithme de Levenshtein peut se calculer indépendamment du type de commutation ou de correspondance en cause (insertion ou effacement, modification), du type de caractères ( $a \rightarrow e$  ou  $a \rightarrow i$ ,  $nd \rightarrow l$  ou  $i'i \rightarrow i$ ), et de l'ordre d'application de ces règles. Mais il permet également de pondérer et d'attribuer des poids différents à ces opérations, en fonction d'une modélisation en amont. Voici, à titre d'exemple, les calculs qui peuvent se faire de façon triviale sur les séquences  $nt\text{fu}\tilde{i}$ ,  $nt\text{fi}\tilde{i}$ ,  $nt\text{fi}\tilde{e}$  :

$$(5) \quad \begin{array}{lll} nt\text{fu}\tilde{i} \rightarrow nt\text{fi}\tilde{i} & \text{'u'} \rightarrow \text{'i'} \text{ éditer} & L(nt\text{fu}\tilde{i}, nt\text{fi}\tilde{i}) = 1 \\ hnka \rightarrow nka & \text{'h'} \rightarrow \langle \text{zero} \rangle \text{ effacer} & L(hnka \rightarrow nka) = 1 \\ nt\text{fu}\tilde{i} \rightarrow nt\text{fi}\tilde{e} & \text{'u'} \rightarrow \text{'i'} \text{ éditer} \ \& \ \text{'r'} \rightarrow \text{'e'} & L(nt\text{fu}\tilde{i}, nt\text{fi}\tilde{e}) = 2 \end{array}$$

L'analyse aréologique quantitative, basée sur la liste d'étymons fournie en annexes, d'après la base de données issue de Kirk (1966), se fonde sur la distance de Levenshtein<sup>12</sup>. La matrice de distances obtenues est la suivante, sur une échelle de 0 à 1, pour 117 items représentatifs, toutes composantes de la langue comprises (matrice sans données négatives) :

<sup>12</sup> Cf. Bolognesi & Heeringa (2002), Heeringa (2004), Beijering, Gooskens & Heeringa (2008), Brown *et al.* (2008). L'algorithme de Levenshtein est disponible en ligne à l'adresse : <https://github.com/coltekin/Gabmap/blob/master/doc/flowcharts/levenshtein-tokenized.eps>. Cf. également Léonard *et al.* (2015b), pour des propositions de pondération, ainsi qu'une confrontation avec d'autres techniques, comme les distances patristiques de la cladistique, en collaboration avec Pierre Darlu.

**Tableau 4** : Matrice de distance de Levenshtein, base de données de Kirk (1966).  
 Traitement des données : Vittorio dell'Aquila.

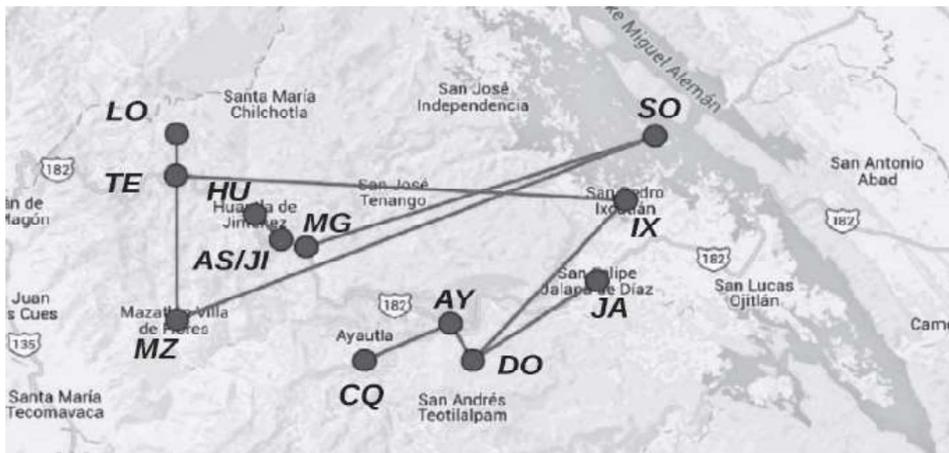
	AY	CQ	DO	HU	IX	JA	JI	LO	MG	MZ	SO	TE
AY		0.28	0.20	0.32	0.21	0.24	0.30	0.52	0.29	0.27	0.24	0.29
CQ	0.28		0.30	0.38	0.30	0.33	0.37	0.54	0.34	0.35	0.30	0.34
DO	0.20	0.30		0.33	0.19	0.11	0.33	0.54	0.27	0.26	0.24	0.28
HU	0.32	0.38	0.33		0.32	0.30	0.21	0.53	0.25	0.30	0.24	0.33
IX	0.21	0.30	0.19	0.32		0.22	0.31	0.53	0.29	0.27	0.24	0.25
JA	0.24	0.33	0.11	0.30	0.22		0.32	0.55	0.28	0.28	0.25	0.28
JI	0.30	0.37	0.33	0.21	0.31	0.32		0.55	0.33	0.28	0.24	0.28
LO	0.52	0.54	0.54	0.53	0.53	0.55	0.55		0.55	0.33	0.50	0.50
MG	0.29	0.34	0.27	0.25	0.29	0.28	0.33	0.55		0.25	0.24	0.31
MZ	0.27	0.35	0.26	0.24	0.27	0.28	0.28	0.33	0.25		0.22	0.29
SO	0.24	0.30	0.24	0.24	0.24	0.25	0.24	0.50	0.24	0.22		0.26
TE	0.29	0.34	0.28	0.33	0.25	0.28	0.28	0.50	0.31	0.29	0.26	

En principe, la méthode du plus proche voisin (*Nearest Neighbor Network*) est censée regrouper des localités en fonction de regroupements en classes sur la base de traits majoritairement partagés. Cette méthode simule le parcours le plus économique d'un représentant de commerce pour visiter de nombreux lieux dans un espace circonscrit en une seule journée. On sait que la réalité joue des tours, dès qu'on quitte l'ordre des principes abstraits et que l'histoire ou la géographie s'en mêlent, ou malgré l'un ou l'autre, comme c'est le cas ici – rien de nouveau sous le soleil, de ce point de vue, et aucun statisticien n'a lieu d'être surpris. Le résultat produit par l'algorithme sur la base des 117 cognats de Kirk est en effet confondant et riche en informations sur la structuration du réseau dialectal et, partant, sur la géohistoire de cette sous-région du bassin du Papaloapán : les regroupements sont surtout transversaux dans l'espace mazatec et, hormis la grappe constituée dans les hautes terres par Huautla ([Hu], centre cafetier) et ses plus proches bourgs (Santa Maria Jiotas [Ji] et San Miguel Huautepéc [Mg], au sud), aucun ne suit le tracé routier qui va de Teotitlán de Flores Magón, à l'ouest, à San Lucas Ojitlán (de langue chinantèque), à l'est, qui est pourtant la colonne vertébrale de la région mazatèque, depuis la construction de la route panaméricaine. Mais ici, hormis le fait que le lien avec Soyaltepec passe d'abord par Jiotas et San Miguel (So-Ji/Mi) pour atteindre Huautla, dans une logique de « voisin le plus proche », ce sont les anciens chemins de traverse, sur la longue durée, qui témoignent de l'histoire des interactions et du peuplement de l'aire mazatèque, plutôt que les voies de communication modernes. Ainsi, un premier vecteur relie la grappe de Huautla (Hu) et ses deux sous-dialectes (Ji & Mg) d'une part, et le bourg cafetier et *milpero* (agriculture de subsistance) du Canyon de Teotitlán,

Mazatlán de Flores Magón (Mz), à l'ouest, d'autre part. Ce dernier est relié à cet antique bourg des basses terres qu'est San Miguel Soyaltepec (So), à l'est. L'arête du graphe est ancrée dans l'île aujourd'hui appelée Isla del Viejo Soyaltepec, située à 200 m d'altitude, ce qui lui a permis d'échapper à l'engloutissement sous les eaux de la retenue du barrage Miguel Alemán. Le village n'a plus que 1300 habitants environ, et n'est accessible que par *lancha*, mais ce fut un centre important avant la construction du barrage. Sa population est résiliente – elle a résisté avec ténacité et courage à la relocalisation que l'Etat mexicain a d'abord organisée de manière civile, à l'aide de l'Institut National Indigéniste, puis imposée par la force, en faisant intervenir l'armée.

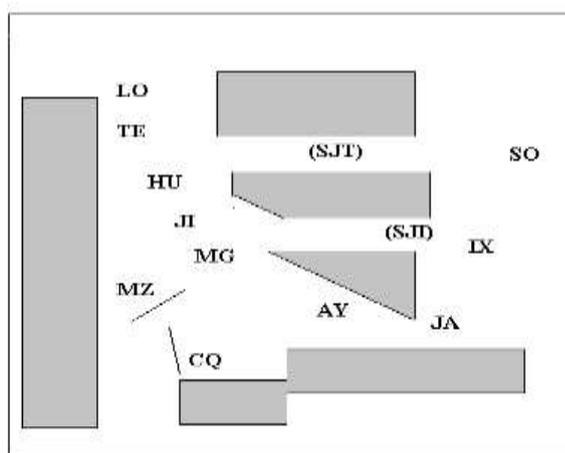
Autrement dit, voilà réunis par l'algorithme du plus proche voisin trois dialectes considérés comme différents : Hu (hautes terres centrales), Mz (Canyon) et So (basses terres orientales), qui ne sont pas aisément intercompréhensibles, même s'ils partagent de nombreux traits structuraux, notamment rétentifs. Ce triangle forme un premier chorème transversal. Un deuxième chorème transversal est constitué, d'ouest en est, de San Pedro Ixcatlán (Ix), important bourg du centre-ouest des basses terres, en lien avec le village de Santo Domingo (Do) au pied du piémont mazatec, d'une part, et de San Jeronimo Tecoatl (Te), dans les hautes terres nord-occidentales, qui a pour satellite le petit bourg de San Lorenzo (Lo), dont la variété dialectale est, par ailleurs, très innovante (*cf.* dendogramme, ci-dessous). Un troisième chorème est formé de la chaîne Chiquihuitlán (Chq), Ayautla (Ay), Jalapa de Diaz (Ja), en lien avec le précédent, en passant par Santo Domingo (Do), qui fait office de pivot. De même, Mazatlán dans le Canyon, à l'ouest, fait office de chaînon qui relie les deux premiers chorèmes. Toutes ces structures émergentes sont transversales et font apparaître, à travers les trois chorèmes, trois « itinéraires » d'interactions ou voies de peuplement de l'aire mazatèque.

Figure 5 : Réseau dialectal mazatec en fonction de l'appariement par « voisin proche ». Données de Kirk (1966), diagramme de Marco Patriarca.



Les trois chorèmes peuvent être appelés Vecteur central (So-Hu-Mz), Vecteur périphérique (Ix-Te-Do) et Vecteur méridional/piémont (Chq-Do-Ja). Ce complexe de trois chaînes dialectales fait exploser les classifications traditionnelles et le cadre géographique avec ses divisions entre hautes terres, basses terres, canyon et piémont, et implique une situation comme celle schématisée dans la Figure 6. Les variétés documentées par Paul Kirk y sont disposées dans un bassin, interrompu ou confiné par les masses grises, avec des localités qui ont pu fonctionner en relais, même si leur développement urbain n'est que récent, telles notamment les localités entre parenthèses abrégées en SJT (San José Tenango) et San José Independencia (SJI, fondé en 1925). La première est un bourg cafetier encastré au centre de la Sierra mazateca, dans une cuvette assurant la transition entre hautes et basses terres (sous-dialecte de type HU) ; la deuxième est un centre agricole (sous-dialecte mixte entre HU et IX, à forte teneur HU). Une barre oblique fine à l'endroit de Chiquihuitlán (CQ) et de Mazatlán (Mz) indique leur cloisonnement orographique – bien que Mazatlan soit plus aisément accessible, par Huautla, que Chiquihuitlán, autrefois reliée à Huautla par un pont suspendu, aujourd'hui abandonné. On retiendra, comme le montre explicitement le schéma de la Figure 6, que la montagne mazatèque est poreuse : loin de constituer une barrière infranchissable entre hautes et basses terres, elle est un vecteur d'échanges. Les premiers peuplements l'ont sans doute contournée par le sud en passant par Do et Ay pour remonter vers Hu et Mz, depuis So, mais les interactions entre Ix et Te se sont sans doute développées et intensifiées ultérieurement, et le couloir de San José Tenango a dû constituer une sorte d'escalier pour la montée et la descente depuis les basses terres centrales, à date plus ou moins récente<sup>13</sup>.

Figure 6 : Schématisation du bassin de traits dialectaux mazatec



<sup>13</sup> L'auteur de ces lignes est, de ce point de vue, d'une autre opinion que celle exposée dans Patriarca, Heinsalu & Léonard (2018 : 121-124), qui suppose un hermétisme de la dorsale montagneuse centrale.

Le modèle à trois, voire quatre étages écologiques (hautes terres, piémont, basses terres, canyon) n'est pas invalidé pour autant, car comme nous le verrons avec d'autres résultats et modes de calcul, il rend bien compte de la trame interne des chorèmes transversaux. Il reste probant pour les paquets, ou sous-ensembles de variétés dialectales. Mais il est transcendé à ce niveau d'analyse, conformément au principe de vicariance des points de vue, caractéristique de l'approche complexionniste.

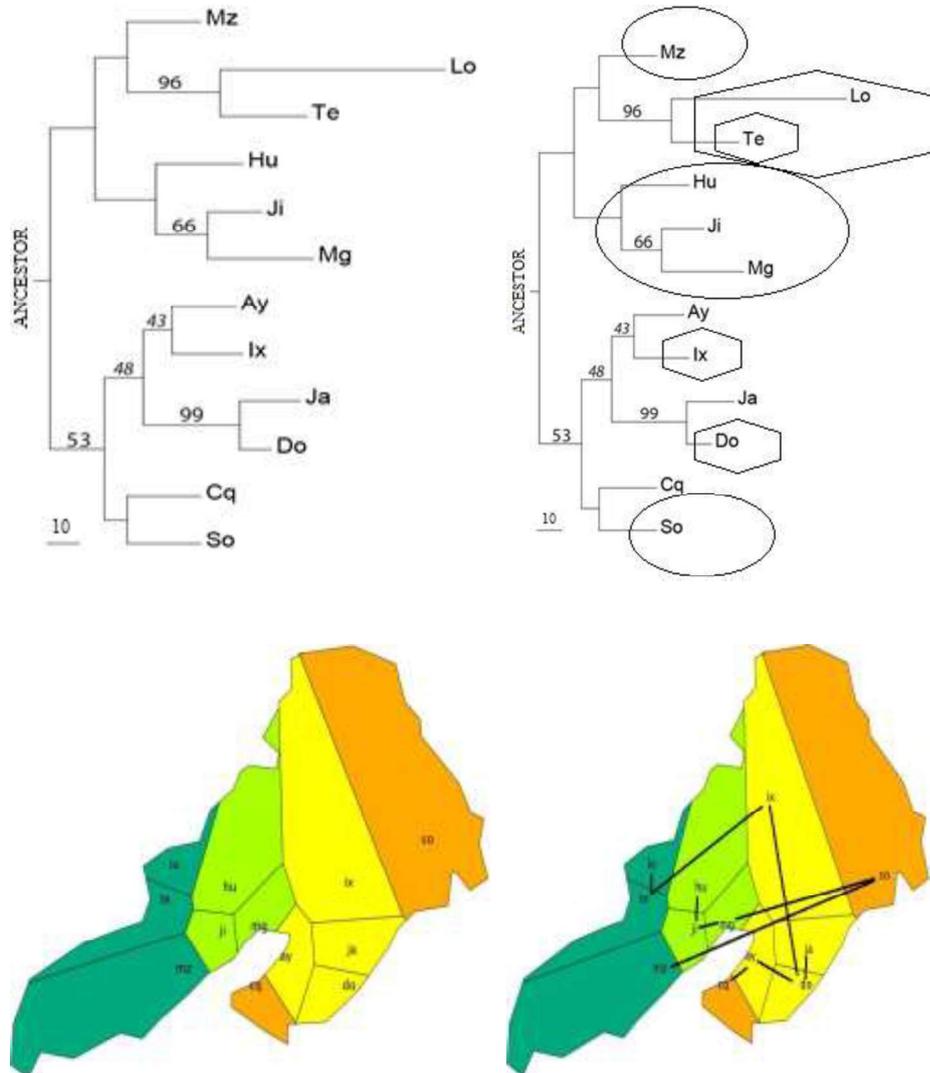
A titre comparatif, voici ce que donne une analyse de type cladistique, sur un échantillon de données phonologiques, réalisée à l'aide du logiciel PAUP<sup>14</sup> : à gauche, le cladogramme, sur un échantillon de données pondérées (*cf.* Figure 7 *infra*), à droite, les variétés dialectales du chorème central sont entourées par des cercles, tandis que celles constituant le chorème périphérique sont entourées par des losanges. Le reste des localités (Ay, Ja, Cq) se répartit dans le chorème du piémont ou chorème méridional. Les biais statistiques sont suggérés par des indices à la racine des clades (index bas, de 42 = clade incertain, index élevé, par exemple de 96 = clade hautement probable). Une telle représentation permet d'envisager la diversité dialectale moins en termes de clades ou de groupes cloisonnés à proprement parler qu'en termes de bassin de traits (Mufwene 2001), comme le suggère également le schéma dans la partie droite de la Figure 10 (*cf.* Patriarca, Heinsalu & Léonard 2018 pour une simulation de la diffusion du mazatec dans son « bassin dialectal »).

Les mêmes relations sont mises en regard dans deux cartes schématiques, présentées en bas de la mosaïque de schémas de la Figure 7 à la page suivante. Il importe de souligner que, dans ces phylogrammes, plus une branche d'un clade est éloignée de la racine de l'arborescence, plus la variété dialectale en question est innovante (ou distante de l'ancêtre) par rapport à la proto-langue (cette propriété distingue d'ailleurs les phylogrammes à proprement parler des simples dendrogrammes). Tous les graphes de dérivation diachronique ayant servi à l'élaboration des matrices cladistiques sur PAUP étaient ancrés sur des étymons reconstruits par Kirk, si bien que les écarts de branches rendent compte des rétentions ou, au contraire, des innovations cumulées. Les variétés plus « conservatrices », ou plutôt, *récentrices*, figurent donc plus proches de la racine de l'arbre.

---

<sup>14</sup> Cladogramme réalisé par Pierre Darlu (Institut national de la santé et de la recherche médicale) en 2012, sur un total de 477 arbres possibles. Données techniques : il s'agit d'un arbre de parcimonie, excluant 329 caractères et n'en gardant que 138 ; 66 caractères sont probants pour le calcul de parcimonie. Optimisation DELTRAN. Autres détails techniques : 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.

Figure 7 : Cladogrammes (analyse cladistique, à gauche) croisés avec le réseau de proximité projeté sur les clades (à droite), cladogramme par Pierre Darlu



Cependant, une autre vision est encore possible. Une autre série de résultats a été produite à partir d'une liste restreinte de cognats, cette fois uniquement issus du paradigme des substantifs, ce qui permet une approche plus unitaire, sur la base d'une phénoménologie moins hétérogène que la liste de 117 cognats. Cette fois, c'est sur un échantillon de 311 substantifs que porte le calcul de distance d'édition (*cf.* matrice Tableau 5).

**Tableau 5** : Matrice de distance d'édition : 311 substantifs de la base de données issue de Kirk (1966)

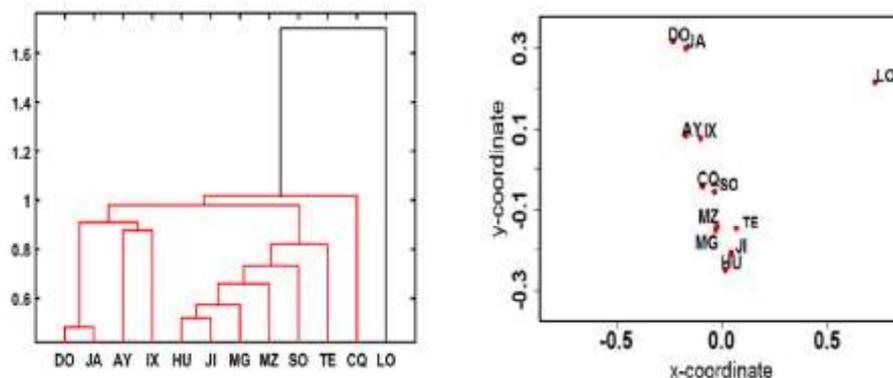
#	AY	CQ	DO	HU	IX	JA	JI	LO	MG	MZ	SO	TE
AY	0,000	0,632	0,629	0,668	0,606	0,607	0,636	0,981	0,562	0,573	0,582	0,708
CQ	0,632	0,000	0,717	0,703	0,666	0,704	0,589	0,978	0,627	0,645	0,636	0,688
DO	0,629	0,717	0,000	0,689	0,585	0,334	0,643	1,000	0,608	0,639	0,620	0,703
HU	0,668	0,703	0,689	0,000	0,593	0,655	0,346	0,897	0,402	0,481	0,519	0,550
IX	0,606	0,666	0,585	0,593	0,000	0,599	0,616	0,937	0,574	0,639	0,519	0,586
JA	0,607	0,704	0,334	0,655	0,599	0,000	0,617	0,945	0,594	0,604	0,585	0,675
JI	0,636	0,589	0,643	0,346	0,616	0,617	0,000	0,841	0,377	0,426	0,462	0,502
LO	0,981	0,978	1,000	0,897	0,937	0,945	0,841	0,000	0,883	0,892	0,884	0,870
MG	0,562	0,627	0,608	0,402	0,574	0,594	0,377	0,883	0,000	0,446	0,490	0,539
MZ	0,573	0,645	0,639	0,481	0,639	0,604	0,426	0,892	0,446	0,000	0,511	0,567
SO	0,582	0,636	0,620	0,519	0,519	0,585	0,462	0,884	0,490	0,511	0,000	0,574
TE	0,708	0,688	0,703	0,550	0,586	0,675	0,502	0,870	0,539	0,567	0,574	0,000

Le dendrogramme à gauche de la Figure 8 rend compte de la configuration des composantes géolinguistiques du réseau dialectal mazatec ; le nuage de localités à droite de la figure rend compte du positionnement multidimensionnel de ces mêmes entités. Cette seconde technique est utilisée en astrophysique pour représenter des constellations stellaires, afin de déceler des groupements d'étoiles, qui peuvent être assez éloignées, alors qu'elles apparaissent alignées à la vue. Cependant, la grille de lecture est simple : plus les objets sont rapprochés, plus ils sont semblables, plus ils sont éloignés, plus ils sont dissemblables, ce qui vaut aussi pour les nuages et sous-nuages (ici, Do et Ja se distinguent de Ay et Ix, tandis que le reste du réseau, sauf Lo qui est éloigné de tous dans sa propre dimension, sont plus semblables entre eux que les cinq variétés mentionnées). Du point de vue complexionniste, ces deux représentations sont intéressantes aussi bien pour ce qu'elles attestent que pour leurs biais. Ce qui est attesté ici est, pour le dendrogramme, une répartition très semblable à celle que nous venons de voir : variétés du piémont et des basses terres centrales à gauche de l'arbre (chaîne Do-Ja, Ay-Ix) d'une part, variétés des hautes terres (Hu-Ji-Mg, Te) et du Canyon (Mz) d'autre part ; isolement relatif de Chiquihuitlan (Cq) et isolement patent de San Lorenzo (Lo), en ce qui concerne les deux branches dextres.

L'intrication de Cq avec tous les parlers sauf Lo est intéressante, mais elle reste discutable, car cette variété devrait venir rejoindre le premier sous-ensemble, des basses terres et piémont, comme dans le graphe du voisinage proche. La position externe de Lo est exagérée et demande à être corrigée, car quiconque se penche sur ce parler et fait du terrain dans son aire (qui ne comprend pas moins de quatre variétés : San Pedro Ocopetatillo, Santa Ana Ateixlahuaca, San Lorenzo, et Mazatzongo de Guerrero dans l'Etat de Puebla), sait qu'il est issu du sous-dialecte des hautes terres nord-occidentales, dont le seul représentant ici est Te. Une fois de plus, on voit que les conditions de vérité sont toutes relatives. La cladistique avait bien mieux saisi la relation entre Te et Lo que ne le fait ici le

calcul de distance redimensionné par la distance de Ward (le dendogramme) ou le positionnement multidimensionnel. Ceci dit, les deux objets de connaissance disent à quelques détails près la même chose (y compris l'écart entre Do-Ja et Ay-Ix apparaît dans le dendogramme dans l'écart des clades par rapport à la racine du sous-clade senestre). A chaque fois, c'est une composante de la vérité qui nous apporte une information nouvelle, à laquelle nous n'aurions pas eu accès sur la base d'un seul mode de traitement des données, ou d'une seule technique (ou d'un unique algorithme), mais ce n'est pas toute la vérité ou la réalité. La réalité n'est pas accessible à travers une image unique, ou un produit final, mais à travers un prisme ou un puzzle, constitué de multiples parties, ou fragments, ou pour ainsi dire, de points de vue.

Figure 8 : Distance de Ward et positionnement multidimensionnel, cognats substantivaux, données de Kirk (1966) (extrait de Léonard *et al.* 2016)



Une interprétation que l'on peut donner *a posteriori* des configurations géolinguistiques mazatèques est qu'il s'agit d'un bassin de traits en partie verticalisé sur un versant écohumain de hautes terres aux basses terres, et flanqué de deux canyons (à l'endroit de Mz et de Cq), et que ce bassin est compact. De ce point de vue, la grande leçon de la trame des nuages issu du positionnement scalaire, à droite de la Figure 8, est que ce réseau dialectal fonctionne par paquets proches, d'une grande densité : la chaîne Hu-Ji-Mg-Mz-Te d'une part, la paire Cq-So s'autre part, puis les paires Ay-Ix et Do-Ja du piémont, plus éloignées entre elles que ne le sont les parlers des hautes terres. Quant à Lo en cavalier seul, loin de tous les autres parlers, c'est une erreur ou un biais de calcul, car sa place devrait être à la même distance de Te que Cq-So l'est de Ay-Ix. Mais on apprend de tout, même (et surtout), des erreurs.

Nous avons donc là un bel exemple de réseau dialectal compact, avec dispersion progressive et liens transversaux d'une grande densité. Un bel exemple de diversité dans l'unité et d'unité dans la diversité.

Voyons maintenant le panorama que nous offre, à l'aide des outils utilisés en TC, un réseau dialectal d'une toute autre nature, situé bien plus au nord-ouest,

sur le même continent, dans une famille de langues très différente : le numique (uto-aztécane). Cette étude de cas nous fera voir l'effet d'un tout autre modèle de diffusion-réaction d'un flux d'information dialectale et de ses phases ou modalités d'auto-organisation pour aboutir à des structures émergentes. Avec le numique, nous avons à faire à un modèle de diffusion radiale, en éventail, à partir de trois souches ou cellules-matrices, avec expansion rapide et extensive. Un tout autre modèle que celui de diffusion-réaction fractale<sup>15</sup> par chorèmes transversaux et par *chunks* ou petits paquets dialectaux du mazatec. Tout autre également que le modèle asymétrique du tseltal, qui distribuait ses configurations sous forme d'aires différenciées fédérant des faits de discours (la morphosyntaxe), d'interlocution et d'intelligibilité immédiate (phonologie), ou de structures conceptuelles (lexique) dans un espace travaillé par une histoire conflictuelle, tiraillé entre contraintes congrégatives (métiers, distribution du travail) et transactionnelles (commerce, transport), au sein d'un biotope relativement unitaire dans un dédale de canyons enchevêtrés, et caractérisé par un polycentrisme urbain, au sud-est du Mexique.

### § 2. 2. 2. Numique : un modèle de diffusion-réaction radial

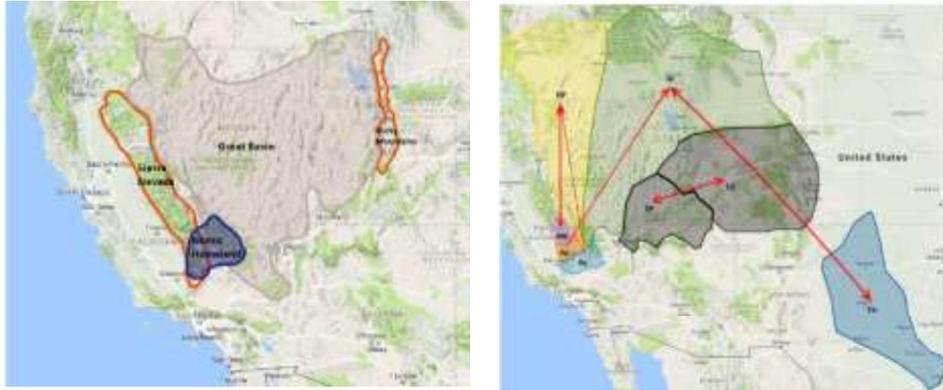
À la différence des deux études de cas précédentes, les données numiques sont de seconde, voire de troisième main (Iannucci 1973). Aucun membre de l'équipe mentionnée dans la note 1 du présent chapitre n'a eu l'occasion de réaliser des enquêtes de terrain dans l'aire où ces langues sont parlées. En revanche, nous avons pu recueillir et systématiser deux listes de cognats (299 entrées lexicales), qui ont été analysées à l'aide de méthodes quantitatives par Rannap (2017), qui a rédigé un rapport de recherches dans lequel il utilise toute une gamme de techniques à partir de matrices de dissimilarité produites à l'aide de l'algorithme de Levenshtein.

Les langues numiques (dont on trouvera la liste permettant d'identifier les variétés et leurs abréviations dans la Figure 10, *infra*) présentent un grand intérêt pour l'étude des phénomènes de diffusion dans l'espace (ou *dynamiques aréologiques*) en raison de leurs conditions historiques d'expansion, caractérisées par une diffusion radiale à partir de deux ou trois noyaux-souches. La carte à droite de la Figure 9 présente le contexte écologique : un vaste plateau accidenté, entrecoupé de plaines steppiques (le Grand Bassin), au centre, flanqué de deux chaînes montagneuses : à l'ouest, la Sierra Nevada, à l'est, les Montagnes Rocheuses. C'est au sud de la Sierra Nevada qu'aurait émergé la niche formative des langues numiques, qui ont ensuite connu une vaste expansion vers le nord, faisant émerger le NP (Northern Paiute), vers le centre (le shoshone, puis, bifurquant vers le sud-est, le comanche), le SP (Southern Paiute) et l'Ute.

---

<sup>15</sup> Un modèle de diffusion-réaction simule ou modélise la géométrie de la diffusion de motifs géométriques dans l'espace ou sur des surfaces (taches sur la fourrure du léopard, anneaux colorés sur les serpents, etc.). Les fractales sont des structures minimales itératives, dont l'auto-organisation, en phase d'expansion, fait émerger des structures complexes (*cf.* Sapoval 1997).

Figure 9 : Aire numique : contexte géographique (à gauche) et réseau de similarité par plus proche voisin (à droite). Extrait de Rannap (2017 : 5, 25)



La division classique des langues numiques est reproduite à gauche de la Figure 10 (en gras, les langues du corpus retenu ici), tandis qu'à droite on trouvera un dendrogramme tenant compte des distances interdialectales, calculées par Rannap à l'aide de la distance d'édition (*cf.* matrice : gradient de gris en fonction de la distance interponctuelle, ex. couleur claire = différenciation minimale).

Figure 10 : *Stammbaum* classique des langues numiques (Rannap 2017 : 6, 26)

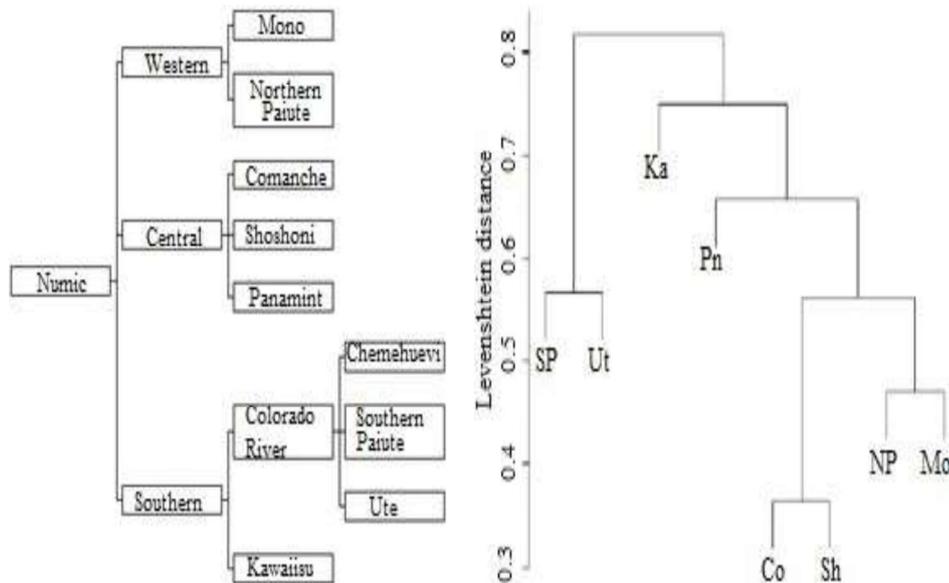


Figure 10 (suite) : *Stammbaum* classique des langues numiques  
(Rannap 2017 : 20)

Langues	Co	Sh	Pn	NP	Mo	SP	Ka	Ut
Co	0	0.364	0.610	0.518	0.611	0.724	0.762	0.871
Sh	0.364	0	0.602	0.506	0.617	0.693	0.673	0.759
Pn	0.610	0.602	0	0.751	0.667	0.868	0.928	0.983
NP	0.518	0.506	0.751	0	0.470	0.725	0.666	0.820
Mo	0.611	0.617	0.667	0.470	0	0.783	0.723	1
SP	0.724	0.693	0.868	0.725	0.783	0	0.700	0.568
Ka	0.762	0.673	0.928	0.666	0.723	0.700	0	0.875
Ut	0.871	0.759	0.983	0.820	1	0.568	0.875	0

Les Tableaux 6 et 7 aux pages suivantes présentent un échantillon de données et décrivent en petites majuscules un ensemble de *paramètres* (davantage que de *processus*) phonologiques à l'œuvre dans ce réseau dialectal<sup>16</sup>. Trois grandes familles de paramètres dominent, et sont notées en petites majuscules dans ces deux tableaux (à la différence des simples processus phonologiques triviaux énumérés dans le Tableau 3 pour le mazatec, qui correspondaient à autant de faits calculables par distance d'édition) : (i) deux tendances liées à la corrélation de force consonantique – occlusives et sonantes *fortis* (gémées : /k/, n:/, ou bien préaspirées /hk, hn/), vs. *lenis* (simples : /k, n/<sup>17</sup>, ou subissant une lénition active par voisement ( $p, k > b, g$ ) ou spirantisation consécutive ( $b, g > \beta, \gamma$ ), non prise en compte); (ii) dénasalisation et assimilation régressive ( $\eta k > kk$ ) ou au contraire, maintien de la nasalité et assimilation progressive ( $\eta k > \eta \eta$ ) du cluster vélaire  $*\eta k$ , phénomènes qui ont pour conséquence d'augmenter le contingent de gémées dans le lexique, et qui débouchent éventuellement sur des cycles ultérieurs de lénition (par exemple,  $\eta k > k > k > g$ ); (iii) des alternances vocaliques endémiques. Une colonne suit les étymons pour proposer une approche phonémique, au-delà des aléas diachroniques et de la reconstruction proposée par Iannucci, qui fait dériver des consonnes gémées de préaspirées reconstruites, alors qu'il vaut mieux, à mon sens, faire dériver celles-ci de gémées étymologiques. Là encore, une *vicariance* des points de vue est souhaitable.

<sup>16</sup> Selon Freeze & Iannucci (1979), le consonantisme du protonumérique est, dans les grandes lignes, le suivant : occlusives  $*p, *t, *ts, *k, *kw, *ʔ$ , fricatives  $*s, *m, *n, *\eta$  (en position médiane ou prévélaire), approximantes  $*w, *j, *h$ ; le vocalisme :  $*i, *i, *e, *a, *u, *o, *ə$ . Cf. aussi Davis (1966).

<sup>17</sup> On notera que l'occlusive sourde simple vélaire peut s'uvulariser ( $k > q$ ). L'occlusive coronale peut aussi s'affriquer en position prépalatale.

Tableau 6 : Echantillon de cognats de la base de données tirée de Iannucci (1973)

	Repr. phonémique	Co	Sh	SP	Glose	Processus
*kihma	/kim:a/	kima	kim:a		venir	SON FORTIS   LENIS
*ahna(=pi)	/an:a(=pi)/	ahna	ahna	aŋapu	aisselle, aile	& PREASP   VELAR
*ahpə	/ap:V/	ahpəʔ	ap:ə	nanahpə	père	OCCL FORTIS   PREASP
*a(h)ta(=pi)	/at:a(=pi)/	ahrabə	ahtahpə	ahtakəpə	joue	& RHOT COR
*ama(h)(taN)	/amaʔ=tan=pi/	amat:ampə		amahtaN	reins	OCCL FORTIS (SANDHI)   PREASP
*kəmaa/(ʔa)		kəmaʔ	kəma(a)		tranchant, lame	MAINTIEN SON V_V
*awah, *awaN	/awaʔ/N/	a:ʔ	a:n	a:h	corne	(REDUC) APPROX W V_V
*aŋka(h), eŋka(h)	/eŋkaʔ/	eka(pi)	eŋka	aŋka(h)	rouge	DENAS & ASS REGR ŋk ALT V # GLOTT #
*əhpəʔi, *e(ʔ)wi	/Vp:Vʔi/	əhəpi	əp:əi	ahpəi	dormir	OCCL FORTIS   PREASP   EPENTH
*hipi	/hipi/		hipi	ipi	boire	MAINTIEN OCCL SIMPL
*kope	/kope/	kobe	kope	kopa, kupa	visage	OCCL SIMPL   LENIS   ALT V / #C
*etə	/etV/	etə	eti, etə	etsə	arc, fusil	OCCL SIMPL   AFFR / V
*hota	/hota/	hora	hota	ota	creuser	OCCL SIMPL   RHOT COR
*katə, kahtə (dur.)	/kat:V/	karə	katə, kat:ə	katə	s'asseoir	OCCL SIMPL   RHOT COR
*hake	/haka/	hakarə	hakatən		qui, quoi	MAINTIEN OCCL SIMPL
*kuk <sup>w</sup> i	/kuk <sup>w</sup> i/		puip:ə (?)	uk <sup>w</sup> i	herbe	OCCL LABVEL   LAB (?)
*naŋka/i, *neŋki	/naŋka/	naki, naka 'ouïr'	neŋki, naŋka 'ouïr'	naŋkapa, naŋka 'ouïr'	oreille, ouïr	DENAS & ASS REGR ŋk ALT V

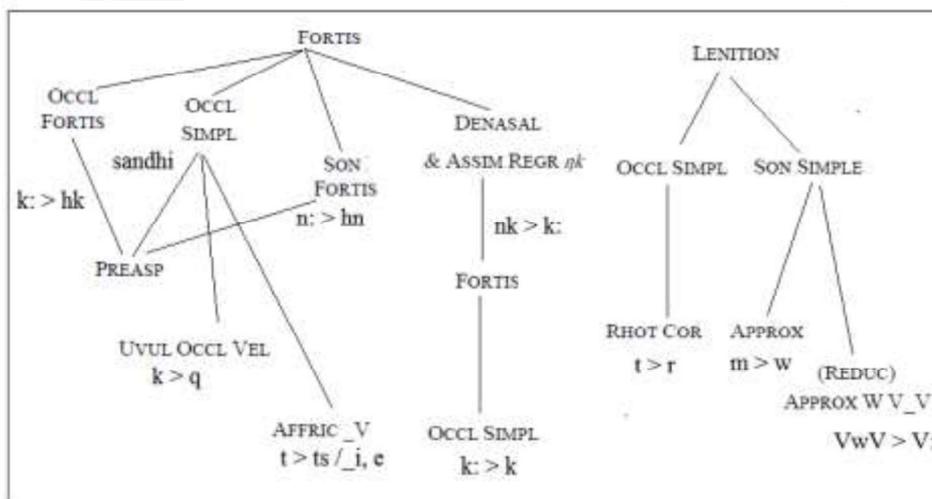
Dans l'ordre, les *paramètres diasystémiques* retenus dans le Tableau 6 sont les suivants, afin d'en expliciter la teneur et de rendre lisibles les abréviations : sonantes fortis vs. lenis, préaspiration ou vélarisation, occlusives fortis vs. préaspirées ; rhotacisme d'une occlusive coronale, occlusive fortis en position de sandhi ou préaspiration, maintien de sonante intervocalique, dénasalisation et assimilation régressive d'un nexus *ŋk* ; occlusive simple vs. lenis, alternance de voyelle initiale de racine après attaque radicale ; occlusive simple vs. affrication prévocale ; occlusive labiovélaire vs. coalescence labiale (*\*k<sup>o</sup> > p*) ; dénasalisation (et) assimilation régressive d'un nexus *ŋk* ; alternance vocalique. A cela s'ajoutent d'autres paramètres déclarés dans le Tableau 7 comme l'approximantisation de sonantes simples (*m > w*), la réarticulation de noyaux glottalisés (*a:; awa > aʔa*), l'uvularisation d'occlusives vélaire, etc.

Tableau 7 : Echantillon de cognats de la base de données tirée de Iannucci (1973), suite

	Pn	Mo	NP	Glose	Processus
*kihma		kihma	kihma	venir	SON FORTIS → PREASP
*ahna(=pi)	(non documenté dans la source)			aisselle, aile	
*ahpə	un:ap:ə			père	OCCL FORTIS
*a(h)ta(=pi)		atapə	atapə	joue	SIMPLIFICATION
*ama(h)(taN)		awawono	amatahpə	reins	SON SIMPLE → APPROX (LENITION V_V)
*kəmaa/(ʔa)		kəwa:	kəma(ʔa)	tranchant, lame	
*awah, *awaN	un:awa	awa(h)	a(a), aʔa:	corne	(REDUC) APPROX W / V_V   REARTICUL GLOTT
*e/anjka(h)	anjka:is:əh	ahqa		rouge	DENASAL & ASSIM REGR $\eta k$ & UVUL OCCL VEL
*əhpəʔi, *e(?)wi	əp:əʔi	əwi	eʔwi	dormir	OCCL LAB FORTIS → APPROX V V
*hipi	hipituʔi	hipi	hipi	boire	MAINTIEN OCCL SIMPL
*kope	ojkope	qope	kopa	visage	UVUL OCCL VEL
*etə		etəh	atə	arc, fusil	MAINTIEN OCCL SIMPL
*hota			təho(h)na	creuser	RADICAL MODIFIÉ
*katə, kahtə (dur.)	kat:ə	qatə, qahtə	ka(t)tə	s'asseoir	OCCL SIMPL   & RHOT COR
*hake		haqeh	haka	qui, quoi	UVUL OCCL VEL
*kuk <sup>w</sup> i	huk <sup>w</sup> ip:ə			herbe	MAINTIEN OCCL LABVEL
*nanjka/i, *neŋki	*ənanjki (POSS 2SG)	nahqa	nahka	oreille, ouïr	DENASAL & ASSIM REGR $\eta k$ UVUL OCCL VEL & PREASP

L'intrication des trois familles de processus phonologiques est représentée dans la Figure 11, qui reprend les principaux paramètres des tableaux ci-dessus.

Figure 11 : Intrication des mécanismes de fortition et lenition en numique

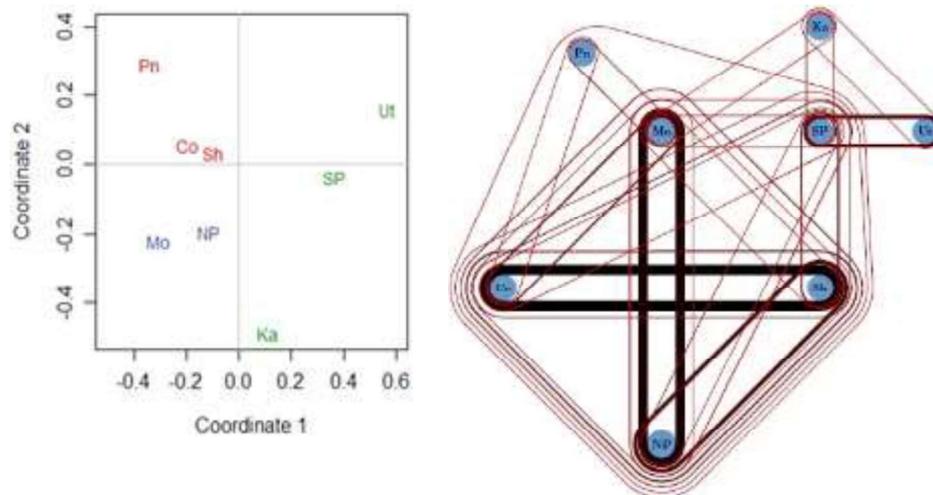


La Figure 12 présente, à gauche, un diagramme de positionnement multi-dimensionnel, dont la typologie diffère notablement de celui examiné précédem-

ment pour le mazatec : la dispersion des objets est bien plus grande, et elle est moins liée à la géographie (Co et Sho y sont collés en une paire, alors que géographiquement très éloignés, comme le montrait la Figure 10) que dans le cas du mazatec ; en outre, les trois « niches » sud-occidentales, matrices des variétés en expansion (Mo, Pn & Ka<sup>18</sup>), sont associées dans des zones respectives du diagramme à leurs extensions secondaires (Mo avec NP, Pn avec Sho & Co, tandis que Ka hésite entre se rallier à la paire Mo-NP ou à SP). Toutes ces informations revêtent une grande importance pour l'ethnohistoire de la diversification de cette sous-famille uto-aztécane et pour l'histoire de l'Amérique du nord.

Enfin, le graphe à droite de la Figure 12 a été calculé selon la méthode glottométrique, qui met l'accent sur les innovations successives, en écartant les rétentions (*cf.* François 2014), dans la logique du graphe décisionnel de la Figure 11. Les indices ethnohistoriques apportés par ce dernier objet de connaissance sont très précieux : on constate que les paires Sho-Co et Mo-NP, qui forment une sorte de grande « croix » centrale ou nexus stratégique, constituent le moteur des innovations au sein du diasystème, tandis que SP & Ut en forment un autre nexus, latéral cette fois, ayant Mo pour relais dans la « niche » initiale, flanqué de Ka. La relation de cette seconde « croix » avec Sho est de toute évidence due au contact. Aucun diagramme seul ne détient toutes les clés du problème, entre la part de rétentions et d'innovations, de matrice originelle (la « niche » sud-occidentale du piémont californien) et d'expansions secondaires, et de relations liées au contact plutôt qu'à l'héritage, mais la lumière se fait peu à peu à partir de la confrontation de chacun d'eux avec tous les autres.

Figure 12 : Positionnement multidimensionnel et graphe glottométrique  
(Rannap 2017 : 29, 32)



<sup>18</sup> Ka = Kawaisu.

Ce caractère kaléidoscopique de l'éventail des objets de connaissance produits par les méthodes quantitatives est l'une des plus grandes vertus de l'approche complexionniste, et le cas numique nous permet<sup>19</sup> de voir que l'analyse linguistique, loin de jouer le rôle de parent pauvre de l'analyse, joue un rôle d'autant plus grand qu'il n'est de bon produit quantitatif, dans ce domaine, qui ne prenne sa justification dans les principes et méthodes de l'analyse structurale en linguistique. La boucle est de nouveau bouclée, réconciliant le point de vue structuraliste et le point de vue complexionniste, comme dans les deux colonnes en regard de la Tableau 1, par laquelle tout ce parcours a commencé.

### § 3. Retour sur la matrice gallo-romane

#### § 3. 1. L'incidence du gascon landais

Ce parcours complexionniste à travers des langues non indo-européennes demande à être complété par des faits romans. Mais j'irai plus loin encore : je vais évoquer à quel point des recherches qu'on pourrait qualifier de « marginales », au regard des critères actuels d'évaluation de la recherche, ont pu préfigurer ou avoir une incidence sur l'avancement des connaissances, à un très haut niveau épistémologique. Je pense à la monographie bien connue des dialectométriciens, mais largement inconnue des linguistes, et même de nombre de romanistes, d'un des enquêteurs de l'Atlas linguistique et ethnographique de Gascogne, coordonné et édité par le bien regretté Jean Séguy (1973) : Théobald Lalanne, dit « l'abbé Lalanne ». Un récent article de Jean-Louis Fossat (2016) rappelle la portée épistémologique de ce document qui s'apparente à un rapport de recherches, de portée quasiment confidentielle, consultable dans de rares bibliothèques :

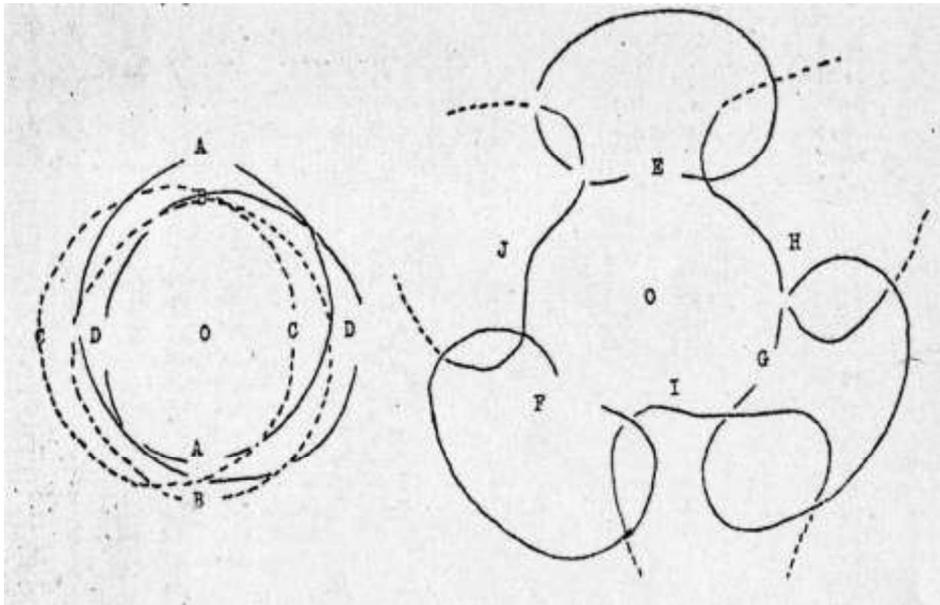
« Dès 1950 Th. Lalanne, dit l'abbé Lalanne, à partir d'un travail précis de cartographie analytique et synthétique, était un des premiers à dénoncer, à partir de leur tracé, le leurre des isoglosses et des frontières dialectales géométriques droites, fussent-elles calculées en distances euclidiennes. Mais personne à cette date, en sciences humaines, ne disposait des outils d'analyse exploratoire permettant des tracés de classes hiérarchiques, fermées ou ouvertes. Toutefois, dans sa thèse sur l'indépendance des aires en Gascogne maritime (Lalanne 1950), il mettait en évidence l'existence de ce qu'on appellerait aujourd'hui *sous-ensembles flous*, autrement dit *graduels* dans des *petits mondes* fractals, que tous les géolinguistes appellent *parlers dialectaux*. » (Fossat 2016 : 113)

---

<sup>19</sup> J'ai le plus souvent utilisé le pronom « je » dans mon argumentaire, jusqu'à présent. Maintenant que *nous*, *le lecteur et moi*, avons avancé dans la lecture des données et l'exploration de la Complexité, j'alternerai avec un « nous » participatif. Il ne s'agit donc pas d'une incohérence énonciative, mais au contraire, d'une stratégie d'inclusion de l'allocutaire dans un processus d'essence maïeutique.

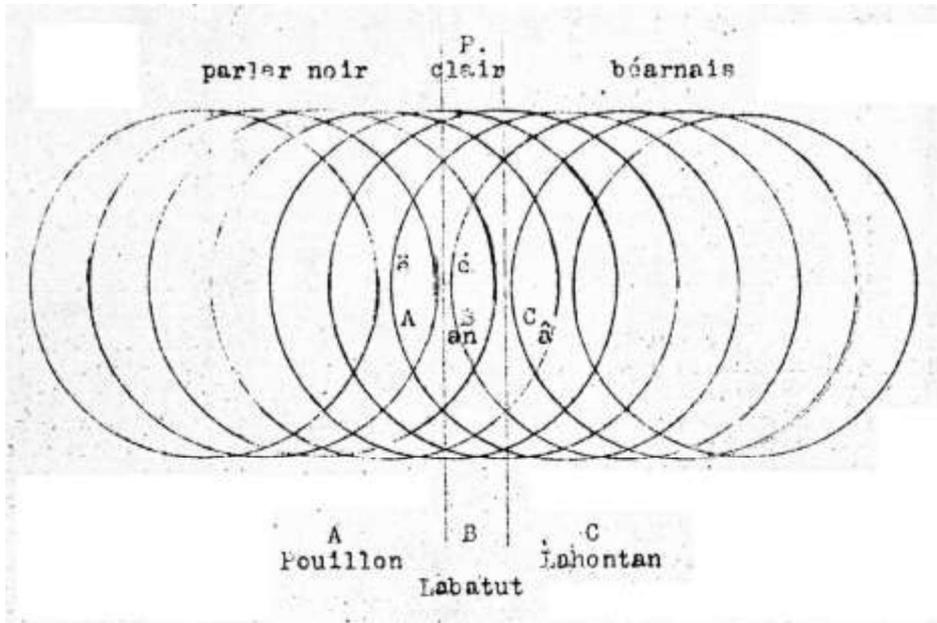
L'auteur reproduit le diagramme suivant, extrait de l'argumentaire de Lalanne (Figure 13), qui remplirait de joie un mathématicien comme Lotfi Aliasker Zadeh, concepteur des ensembles flous en mathématiques. Ces deux schémas représentent deux modalités d'ensembles intersécants : soit par fluctuations et dynamique vibratoire, à gauche, soit par maillage ou tuilage (Séguy), à droite.

Figure 13 : Représentation par Lalanne du partitionnement territorial organisé en sous-ensembles flous (Lalanne 1949 (II) : 93), repris de Fossat (2016).



Fossat reprend également à Lalanne le diagramme des sous-ensembles sécants, ou ensembles sériels en forme de chenille (Cf. la Figure 14 à la page suivante), opposant trois blocs dialectaux (indiqués par A, B et C, respectivement) : à gauche et à l'ouest, le « parler noir » (où \**e* mi-fermé tonique du latin tardif est passé à une voyelle labiopalatale moyenne de type /ø/, notée ici <ë>), au centre le « parler clair », où la voyelle moyenne antérieure étirée s'est maintenue (notée <é>), et dans la marge droite, le bloc du béarnais, comme variété sud-orientale de gascon, qui partage le maintien de /e/, et où la dénasalisation opère par exemple sur la voyelle basse : <an> vs. <â>. Les localités de Pouillon, Labatut et Lahontan peuvent être considérées comme trois échantillons des types A, B, C au sein du diasystème (même si ce terme est étranger à Lalanne).

Figure 14 : Sous-ensembles sécants (Lalanne 1949 : 2, fasc. 1, conclusions II), repris de Fossat (2016 : 115)



Lalanne quantifie et justifie sa vision qui ressortit *a posteriori*, à la logique floue et aux formes fractales, par un argumentaire aussi énergique que stylé. Tout comme la prose de Jules Gilliéron et de Georges Millardet, celle de Lalanne est un délice, comme en témoigne ce passage, qui fonde la notion même d'*aires indépendantes*, déjà clairement énoncée par Millardet (1923 : 202-235) par le passé, mais qu'il développe à l'aide d'une cartographie quantitative :

« Il faut le reconnaître, l'expérience [consistant à cumuler les isoglosses lexicales] est décevante, et il semble inutile de la poursuivre. Les bourrelets attendus n'apparaissent nulle part. Chaque mot a son aire propre et indépendante. Il ne connaît que l'adversaire qu'il a devant lui, et n'a cure de s'aligner avec ses frères dialectaux, qui de leur côté reculent ou avancent comme ils le peuvent, sans s'informer s'ils bousculent les poteaux-frontières de vénérables dialectes ou d'illustres vicomtés. Je ne m'étonnerai donc pas si, en fin d'étude, je n'ai pas rencontré, dans mon secteur, deux aires superposables sur toute leur étendue. Rien n'est plus vraisemblable, si l'on veut bien distinguer – et il semble naïf d'y insister – les *aires libres* ou *naturelles* et les *aires imposées* ou *artificielles*. Lorsque l'occupant, pour se procurer de l'huile, décrète que tous les cultivateurs de France devront ensemercer 10 ares de tournesol, l'aire de cette plante est forcée ou artificielle ; mais l'occupant disparu, le tournesol n'est plus cultivé sur les terrains, sous les climats qui lui conviennent, et il revient à son aire naturelle. Dans les langues officielles, toutes les aires sont forcées ; elles

s'étendent, au nom de la loi, jusqu'aux limites de la souveraineté de l'État ; mais dans les dialectes, le lexique, la phonétique, la grammaire ne connaissent que des aires naturelles. Chaque mot, chaque forme lutte *séparément* pour sa vie ; son aire se contracte, se dilate, suivant les qualités d'expression, d'euphonie, suivant ses harmoniques plus ou moins accordées au génie local. La concurrence et l'adaptation jouent à plein, et seules ; tant pis pour les mots lourds, sourds, boiteux, inexpressifs ; leurs aires seront contaminées, grignotées, envahies ; elles cesseront d'être superposables ... Si elles l'ont jamais été. Mais cette lutte pour la vie n'est pas le combat en rangs serrés d'un hypothétique dialecte contre un autre, c'est la lutte en ordre dispersé des francs-tireurs isolés qui s'ignorent ». (Lalanne 1949 (I) : 15)

Il faut renouer avec la lecture de ces textes (de Lalanne, et des dialectologues du début du siècle passé, comme Millardet, Bloch, Terracher, etc.) pour déceler, sous la verve stylistique, une vision riche en perspectives, pour l'étude des phénomènes complexes de la vie du langage dans le temps et dans l'espace, avec la tension épistémologique qui oppose, par exemple, le phonéticien empiriste Georges Millardet au sémanticien lexicaliste qu'est Jules Gilliéron.

Mais le grand apport de ce que Jean-Louis Fossat appelle le « *parcours de Lalanne* » consiste à démontrer l'universalité des structures non linéaires et fractales en dialectologie, que l'approche microdialectologique est particulièrement à même de faire apparaître.

« La classification hiérarchique qui ne permet pas de visualiser ces faits d'empiètement est par conséquent prise ici en défaut. Depuis il a été établi que ce type d'organisation ne représentait en aucun cas une caractéristique spécifique du gascon mais une organisation universelle, qui rend compte aussi bien des gradients d'occitanité que de catalanité, de lusitanité, de basquité et de francité en domaine d'oïl. » (Fossat 2016 : 117)

Dans les études de cas précédentes, la tseltalité, la mazatéquité ou la numicité s'avérait variable, mais de manière davantage endogène. Ceci dit, l'assertion de Fossat s'applique également dans ces contextes non indo-européens, en fonction des hiérarchies et des interactions internes aux réseaux polycentriques de ces domaines, car le grand universal, en matière de diasystème et de réseaux dialectaux, est bel et bien la polycentricité.

L'étude de cas romane suivante – le poitevin-saintongeais, en domaine gallo-roman d'oïl occidental – est également exemplaire de la dynamique polycentrique dans un réseau dialectal.

### § 3. 2. Epilogue poitevin-saintongeais

Suivant l'hypothèse de Lalanne, et sous la férule du même Fossat, Lilianne Jagueneau (1987), dans sa thèse d'État sur la structuration de l'espace linguistique entre Loire et Gironde, fait clairement apparaître ces *micro*, *méso* et *macro*-

structures, aussi universellement floues que localement fractales, à partir de moments statistiques dans l'arbre des similarités interponctuelles qu'elle tire de l'examen exhaustif des données de l'Atlas Linguistique et ethnographique de l'Ouest (Massignon & Horiot 1971-1983), sur la base de caractérisations typologiques (type d'oïl exogène, type poitevin-saintongeais endogène, type d'oc). La Figure 15 par exemple, fait émerger les chorèmes minimaux du domaine poitevin-saintongeais.

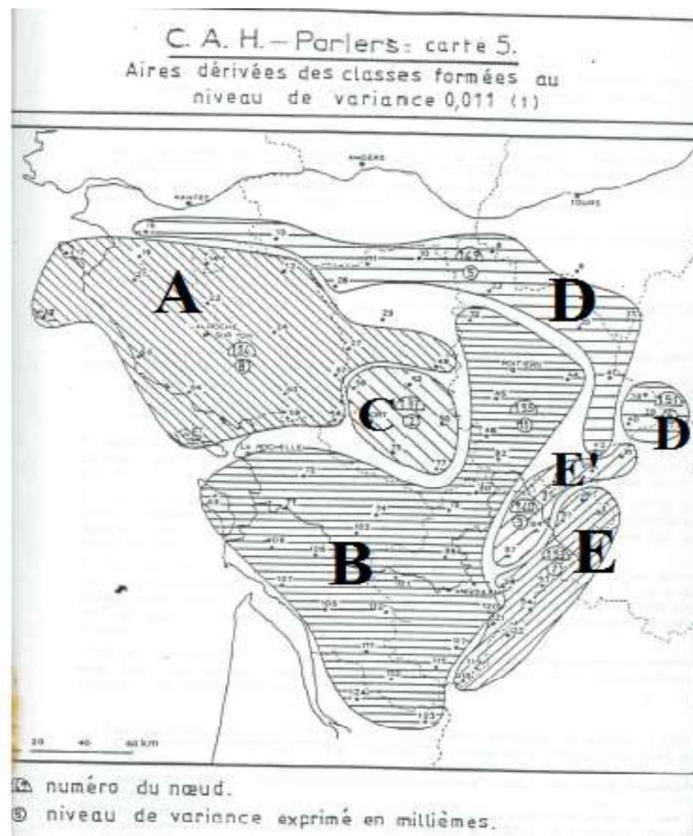
Figure 15 : Sous-ensembles d'incidence minimale du fait dialectal : niveau de variance 0, d'après Jagueneau (1987 : 187)



Pour quiconque connaît ce terrain dialectologique, ces regroupements sont éloquents : on y reconnaît la constellation sous forme de mosaïque des micro-parlers à l'échelle des « terroirs ». Il y a là autant de chorèmes qu'il y a d'aires et tous suivent cette logique « aveugle » de proximité immédiate, que décrit la citation de Lalanne ci-dessus. L'écologie évoquée ici est celle de la parcelle territoriale, faisant émerger des agrégats caractérisés par leur densité typologique. Dans ces poches, ou ces niches, ces couvées, la similarité est maximale, et on peut s'attendre à ce que l'intercompréhension le soit également, du moins *a priori*. La forme des aires stylisée sur la carte évoque des brins, ou des molécules plus

ou moins étirées ou joufflues. Certaines se tordent en contournant leurs voisines, comme au sud, d'autres sont intersécantes, comme au centre-sud et au sud-est ; d'autres sont discontinues, à de grandes distances, comme au nord-ouest. Ces tracés, bien qu'inévitablement grossiers, en fonction des contraintes du dessin bidimensionnel en tant que schématisation, font pourtant émerger de grandes tendances structurales : aires diffuses au nord-ouest, aires isolées au centre et au nord-est, mais intense condensation et interaction au centre-sud et surtout, au sud, en contact avec l'occitan. Ces sortes d'amibes flottent dans une sorte de plasma géolinguistique relativement inerte, non corrélé à ces nodules (les points de l'atlas non reliés entre eux par des aires). Les dynamiques territoriales de peuplement et repeuplement (au nord-ouest), d'intense campanilisme interne (au centre-sud) et de contacts fragmentés avec le monde d'oc sous au moins deux de ses dialectes majeurs (l'arverno-limousin, ou occitan septentrional, et le gascon, ou occitan sud-occidental), résument les grandes lignes de l'écologie dynamique de ce bassin de confluence du domaine d'oïl et du domaine d'oc au sein du gallo-roman.

Figure 16 : Sous-ensembles d'incidence maximale du fait dialectal : niveau de variance 0,011, d'après Jagueneau (1987 : 201)



La carte schématique de la Figure 16 résorbe l'atomisme de l'image précédente, et montre clairement la structuration de cet espace géolinguistique et, partant, de l'histoire régionale : cinq blocs s'opposent, à ce niveau majeur de convergence dialectale (niveau de variance de 0, 011) : un bloc nord-occidental compact en forme de masse, constitué essentiellement de la Vendée, que l'on pourrait étiqueter Bloc A ; un bloc centre-méridional en forme de louche (Bloc B), qui rassemble la Saintonge et la région de Poitiers. Ces deux aires ne sont plus des chorèmes au sens minimaliste, mais des aires confédératives de sous-dialectes agrégés ou associés. En plein centre, l'œil de Niort constitue, lui, un chorème, de rang C, sous forme de nodule, qui rompt la continuité entre les blocs A et B. On peut y voir la concurrence entre deux centres urbains majeurs, dont celui du bloc C niortais serait le plus résilient, en termes d'endogénéité poitevine-saintongeaise. Enfin, deux ensembles complexes flanquent ce complexe central : au nord, un bloc D, en continuité avec le Bassin parisien, qui peut se décrire comme le front du français central (ou variété d'oïl centrale), au sud-est, un bloc E dont E est le front occitan, et dont E' est le liseré intermédiaire (ou amphizone). De même, D est flanqué d'un nodule D', au centre-est, relativement autonome et davantage endogène dans son domaine, par rapport à la variété « parisienne ».

On voit que non seulement le terme composé « poitevin-saintongeais » est empiriquement confirmé, mais on distingue également son tropisme, vers la Vienne et Poitiers, qui a fait office de couloir de « francisation » ou de pénétration du français, entendu comme cette langue « artificielle » exogène, qui vient se superposer sur la trame typologique poitevine. Dès lors, on comprend mieux que les Saintongeais renâclent aujourd'hui à accepter la notion de « poitevin-saintongeais », tant la relation génétique entre les deux composantes de ce domaine est floutée par les interférences verticales (normées), et tant la composante poitevine est elle-même divisée entre le nodule enkysté du niortais, au centre, comme variété urbaine (bloc C), et l'aire expansive nord-occidentale (bloc A), davantage rurale, et diversifiée, on l'a vu, de manière discontinue (ou densément fractale).

La tension entre types endogènes (A et C) et types à incidence exogène (B et E'), sous la pression externe de deux types puissamment exogènes (D et E, respectivement français central et occitan), peut se décrire comme un champ de force, propre à générer des tensions identitaires (c'est le cas depuis deux décennies, où les mouvements associatifs défenseurs du « poitevin-saintongeais » vs. du « saintongeais » sans « fâcheux trait d'union » s'affrontent, dans le contexte d'aménagement glottopolitique pour la revitalisation de ces variétés (cf. Julien-Labruyère 1996, Léonard 2019 : 579). La société n'est jamais inerte, et le champ social est davantage mû par le binarisme, afin de mobiliser des intérêts, des idées et des actions, que par une logique floue ou graduée, si bien que la science atteint là son « seuil de Peter ». Du moins, a-t-elle accès à des niveaux de complexité et de nuance difficile à atteindre d'une autre manière. On retrouve là le volet droit du Tableau 1, et notre boucle réflexive est ainsi bouclée, laissant entrevoir la

dimension psycho-sociale qui vit et agit, derrière le rideau épais des données et des faits observables.

#### § 4. Conclusion

Nous n'avons en effet cessé de parcourir un itinéraire, à partir de la confrontation entre l'approche structuraliste et l'approche complexionniste, dans une perspective cumulative, initialement d'ordre générique, où les données linguistiques étaient provisoirement absentes. Nous avons ensuite commencé à observer des résultats ancrés dans des faits de langue avec la remarquable asymétrie en tseltal, entre les différents paliers du diasystème que sont les composantes phonologique, morphosyntaxique et lexicale. Puis, nous avons palpé des données sonnantes et trébuchantes avec les études de cas mazatèque et numique, ce qui nous a permis de jongler avec différentes matrices et différents résultats algorithmiques. Au fond, pour résumer, je dirais que l'approche complexionniste nous a permis de payer un tribut à ce propos du regretté Jean Séguy, fondateur de la dialectométrie, qui qualifiait l'activité de tracer des isoglosses sur des cartes de « jeu sinistre et vain ». A travers la *Théorie de la Complexité*, les dynamiques géolinguistiques prennent corps, se chargent de vie et de vitalité, et se racontent à travers de multiples facettes, par le biais d'un éventail d'objets de connaissance. On retrouve alors le plaisir de la science, tout comme Roland Barthes parlait du « plaisir du texte ». Un plaisir du texte qui emporte d'autant plus lorsqu'on lit la prose de Jean Séguy, Jean-Louis Fossat ou Théobald Lalanne, chercheurs de terrain, et trameurs de concepts.

Certes, ce parcours en écologie diasystémique aurait pu commencer par l'examen de faits romans, comme ceux du gascon landais et du poitevin-saintongeais. Mais ce faisant, il aurait été difficile de sortir de nos schémas ordinaires (par exemple, opposition d'une variété centrale prestigieuse et expansive *vs.* les variétés rurales ou les langues régionales récessives) et de notre zone de confort. L'examen de langues non indo-européennes, comme le tseltal, le mazatec et les langues numiques nous a permis d'observer une gamme de phénomènes, d'interactions entre groupes humains situés dans le temps et dans l'espace, et de contraintes adaptatives au milieu, au-delà de nos routines. C'est dans ce sens que les méthodes d'analyse de données en termes de systèmes complexes prennent toute leur valeur : par le déploiement de la diversité du réel et des sociétés humaines, ainsi que par la montée épistémologique. De même que la totalité d'un phénomène est plus que la somme de ses parties, le monde des langues, observé à travers sa variation, est bien plus que la somme de sa diversité, et peut se décrire comme le produit non linéaire et fractal de la relation entre l'homme et son milieu ambiant, naturel et social.

### *Bibliographie*

- BASTARDAS I BOADA Albert, 2013. « Complexity and language contact: A socio-cognitive framework », in S. Mufwene, F. Pellegrino & C. Coupé (eds.), *Complexity in language: Developmental and evolutionary perspectives*. Cambridge University Press, pp 218-244.
- BEIJERING Karin, Charlotte GOOSKENS & Wilbert HEERINGA, 2008. « Predicting intelligibility and perceived linguistic distance by means of the Levenshtein algorithm », in M. van Koppen & B. Botma (eds.), *Linguistics in the Netherlands 2008*. Amsterdam, John Benjamins, pp. 13-24.
- BENKIRANE Réda, 2013. *La complexité, vertiges et promesses. 18 histoires de sciences*. Paris, Le Pommier.
- BERTHOZ Alain, 2009. *La simplexité*. Paris, Odile Jacob.
- , 2013. *La vicariance. Le cerveau créateur de mondes*. Paris, Odile Jacob.
- BOLOGNESI Roberto & Wilbert HEERINGA, 2002. « De invloed van dominante talen op het lexicon en de fonologie van Sardische dialecten », in D. Bakker *et al.* (eds.), *Gramma/TTT: tijdschrift voor taalwetenschap*, 9 (1). Nijmegen University Press, pp. 45-84.
- BROWN Cecil, Eric HOLMAN, Søren WICHMANN & Viveka VELLUPILAI, 2008. « Automated classification of the world's languages: a description of the method and preliminary results », *STUF Language Typology and Universals*, 61, pp. 285-308.
- BUROV Ivaýlo, 2017. *Фонетични и фонологични универсали при асимилационните процеси. Теория за относителната бинарност и йерархия на признаците*. Sofia, CU Romanistika.
- , 2018. « Combien les traits distinctifs sont-ils binaires ? Vers une théorie de la binarité relative », *Bulletin de la Société de linguistique de Paris*, CXIII (1). Paris/Louvain, Peeters, pp. 27-78.
- , 2019. « L'harmonisation vocalique du français dans le contexte des harmonies de hauteur : de l'universel au particulier », *Bulletin de la Société de linguistique de Paris*, CXIV (1). Paris/Louvain, Peeters, 55-93.
- , ce volume. « Complexité des phonèmes et des systèmes phonémiques : une alternative pour formaliser la notion de marque », in I. Burov & G. Fiorentino (dir.), *Complexité des structures et des systèmes linguistiques : le cas des langues romanes*. Sofia, CU Romanistika, pp. 77-104.
- CAMPBELL Lyle, 1987. « Tzeltal dialects: New and old », *Anthropological Linguistics*, 29 (4), pp. 549-570.
- CECI Elvio, ce volume. « Structure conservativa del Linguaggio », in I. Burov & G. Fiorentino (dir.), *Complexité des structures et des systèmes linguistiques : le cas des langues romanes*. Sofia, CU Romanistika, pp. 105-122.
- DAVIS Irvine, 1966. « Numic consonantal correspondences », *International Journal of American Linguistics*, 32 (2), pp. 124-140.
- DO BUI Bien Nhat, 2018. *Grammaire descriptive de l'amuzgo de Xochistlahuaca : documentation d'une variété amuzgoane (otomangue orientale) de 'langue en danger'*. Thèse de doctorat, Sorbonne Université.
- FOSSAT Jean-Louis, 2016. « Le parcours de Th. Lalanne : du recueil de données (1947-1952) aux cartes synthétiques de l'Atlas linguistique de la Gascogne maritime », *Géolinguistique*, 16, pp. 108-129.
- FRANÇOIS Alexandre, 2014. « Trees, waves and linkages: Models of language diversification », in C. Bowerman & B. Evans (eds.), *The Routledge Handbook of Historical Linguistics*. London, Routledge, pp. 161-189.
- FREEZE Ray & David IANNUCCI, 1979. « Internal classification of the Numic languages of Uto-Aztecan », *Amerindia*, 4, pp. 77-92.
- GOLSTON Chris & Wolfgang KEHREIN, 1998. « Mazatec onsets and nuclei », *International Journal of American Linguistics*, 64 (4), pp. 311-337.

- , 2004. « A prosodic Theory of laryngeal contrasts », *Phonology* 21, pp. 1-33.
- GRIBBIN John, 2004. *Deep simplicity, chaos, complexity and the emergence of life*. London, Penguin
- GUDSCHINSKY Sarah, 1958a. « Proto-Popotecan. A comparative study of Popolocan and Mixtecan », *International Journal of American Linguistics*, 25 (2), pp. 1-118.
- , 1958b. « Mazatec dialect history », *Language*, 34, pp. 469-481.
- HEERINGA Wilbert, 2004. *Measuring dialect pronunciation differences using Levenshtein distance*. PhD thesis, Rijksuniversiteit Groningen.
- HOPKINS Nicolas, 1970. « Estudio preliminar de los dialectos del tzeltal y del tzotzil », in N. McQuownn & J. Pitt-Rivers (eds.), *Ensayos de antropología en la zona central de Chiapas*. México, Instituto nacional indigenista, pp. 185-214.
- IANNUCCI David, 1973. *Numic historical phonology*. PhD dissertation, Cornell University.
- JAGUENEAU Lilianne, 1987. *Structuration de l'espace linguistique entre Loire et Gironde, Analyse dialectométrique des données de l'Atlas Linguistique et Ethnographique de l'Ouest*. Thèse de Doctorat d'État, Toulouse.
- JULIEN-LABRUYERE François, 1996. « Le fâcheux trait d'union », *Aguaine, Revue de la Société d'Ethnologie et de Folklore du Centre-Ouest (SEFCO)*, pp. 77-80.
- KIRK Paul, 1966. *Proto-Mazatec phonology*. PhD dissertation, University of Washington.
- KRETZSCHMAR William, 2009. *The linguistics of speech*. Cambridge University Press.
- , 2010. « Language variation and complex systems », *American Speech*, 85 (3), pp. 263-286.
- , 2014. « Emergence of 'new varieties' in speech as a complex system », in S. Buschfeld *et al.* (eds.), *The evolution of Englishes: The dynamic model and beyond*. Amsterdam, John Benjamins, pp. 142-159.
- , 2015. *Language and complex systems*. Cambridge University Press.
- LALANNE Théobald, 1949. *L'indépendance des aires linguistiques en Gascogne maritime*, 2 fascicules ronéotypés, à compte d'auteur, Toulouse.
- LÉONARD Jean Léo, 2002. « Microcosmic perceptual dialectology and the consequences of extended linguistic awareness. A case study of Noirmoutier Island (France) », in D. Preston & D. Long (éds.), *Handbook of Perceptual Dialectology*, 2. Amsterdam/Philadelphia, John Benjamins, pp. 219-248.
- , 2017a. « Reconstruire au-delà de la déconstruction. Requalifier les langues et réaménager la pensée critique », in S. M. Berkaïne *et al.* (éds.), *Construction/déconstruction des identités linguistiques, Actes du colloque 'Jeunes Chercheurs'*. Montpellier, Connaissances et Savoirs, Collection Langues et Société, pp. 53-78.
- , 2017b. « Modéliser les diasystèmes et les aires dialectales : exploration de huit mondes géolinguistiques », in A. Iglesias *et al.* (éds.), *Gotzon Aurrekoetxea lagunarterik hara*. Bilbao, Universidad del País Vasco, pp. 57-86.
- , 2019. « L'éventail de l'aménagement linguistique 'de par en bas' dans le domaine d'Oïl : sociolinguistique contrastive des domaines Poitevin-Saintongeais, Gallo et Bourguignon-Morvandiau », in F. Fouchard *et al.* (coord.), *La recherche en études françaises : un éventail de possibilités / La investigación en Estudios Franceses : un abanico de posibilidades*. Editorial Universidad de Sevilla, pp. 571-580.
- LEONARD Jean Léo & Gilles POLIAN, 2014. « Le Cañon des isoglosses qui bifurquent : ALTO (*Atlas Lingüístico del Tzeltal Occidental*), les temps et les lieux du diasystème », in G. Forlot & F. Martin (éds.), *Regards sociolinguistiques contemporains. Terrains, espaces et complexités de la recherche. Carnets de l'Atelier de Sociolinguistique du CERCLL-LESCLAP*. Paris, L'Harmattan, pp. 185-214.
- LÉONARD Jean Léo & Vittorio DELL'AQUILA, 2014. « Mazatec (Popolocan, Eastern Otomanguan) as a Multiplex Sociolinguistic 'Small World' », in U. Berezcki (ed.), *The Languages of Smaller Populations: Risks and Possibilities. Lectures from the Tallinn Conference, 16-17 March, 2012*. Tallinn, Hungarian Institute's Series, Miscellanea Hungarica, pp. 27-55.
- LÉONARD Jean Léo, Daniele DALMASSIO, Vittorio DELL'AQUILA & Antonella GAILLARD-CORVAGLIA, 2014. « Emergence forte versus émergence faible en écologie linguistique : le cas du

- mazatec », in P. Bonin & T. Pozzo (éds.), *Nature et Culture*. Publications de l'Université de Saint-Étienne, pp.77-102.
- LEONARD Jean Léo & Alain KIHM, 2015. « Mazatec verb inflection: A revisiting of Pike (1948) and a comparison of six dialects », in J. L. Léonard & A. Kihm (eds.), *Patterns in Mesoamerican morphology*. Paris, Michel Houdiard Editeur, pp. 26-76.
- LEONARD Jean Léo, Gilles POLIAN, Els HEINSALU, Marco PATRIARCA, 2015a. « Variación dialectal de la morfología tseltal (Maya occidental) en los ámbitos morfológico, fonológico y léxico: un enfoque holístico del diasistema », in J. L. Léonard & A. Kihm (eds.), *Patterns in Mesoamerican morphology*. Paris, Michel Houdiard Editeur, pp. 280-303.
- LEONARD Jean Léo, Els HEINSALU, Marco PATRIARCA & Pierre DARLU, 2015b. « Modeling regional variation from EAS: Complexity and social aggregates », in G. Aurrekoetxea, A. Romero, A. Etxebarria (eds.), *Workshop EUDIA-4: Linguistic variation in the Basque language & Education*, I. Bilbao, Servicio editorial de la Universidad del País Vasco, pp. 145-172.
- LÉONARD Jean Léo, Didier DEMOLIN & Karla Janiré AVILEZ GONZALEZ (eds.), 2016. *Proceedings of the International Workshop on Structural Complexity in Natural Language(s) (SCNL)*, mai 2016. Paris, Université Paris 3 – Labex EFL (PPC11), accessible sur <http://axe7.labex-eefl.org/node/353>.
- LEONARD Jean Léo, Els HEINSALU, Marco PATRIARCA, Kiran SHARMA & Anirban CHAKRABORTI, 2016. « Patterns of linguistic diffusion in space and time: The case of Mazatec », in F. Abergel *et al.* (eds.), *Econophysics and sociophysics: Recent progress and future directions*. Springer, pp. 227-255.
- LÉONARD Jean Léo & Julien FULCRAND, 2018. « Inflectional class shifts in the Mazatec diasystem: Innovation, contact and metatypy », *STUF* 71-3. Mouton de Gruyter, pp. 429-473.
- MASICA Colin, 1990. *Indo-Aryan languages*. Cambridge University Press.
- MASSIGNON Geneviève & Brigitte HORIOT, 1971-1983. *Atlas linguistique et ethnographique de l'Ouest : Poitou, Aunis, Saintonge, Angoumois*, Paris, Éd. du CNRS, 3 vol.
- MASSIP-BONET Angels & Albert BASTARDAS I BOADA (eds.), 2013. *Complexity perspectives on language, communication and society*. New York, Springer.
- MATRAS Yaron, 2009. *Language contact*. Manchester, Cambridge University Press.
- MIESTAMO Matti, Kaius SINNEMÄKI & Fred KARLSSON (eds.), 2008. *Language complexity: Typology, contact, change*. Amsterdam, John Benjamins.
- MILLARDET Georges, 1923. *Linguistique et dialectologie romanes. Problèmes et méthodes*. Montpellier, Société des Langues Romanes.
- MUFWENE Salikoko, 2001. *The ecology of language evolution*. Cambridge University Press.
- PATRIARCA Marco, Els HEINSALU & Jean Léo LEONARD, 2018. « Applicazioni alla linguistica dei metodi e modelli della teoria dei sistemi complessi », in L. Costamagna *et al.* (ed.), *Mutamento linguistico e biodiversità*. Roma, Il Calamo, pp. 103-143.
- POLIAN Gilles & LEONARD Jean Léo, 2009. « La morphologie dans ALTO (Atlas Linguistique du Tseltal Occidental). Réseau dialectal et systèmes à décideurs multiples », *Géolinguistique*, 11, pp. 149-201.
- , 2017. « Herramientas cuantitativas en el proyecto ALTO (*Atlas Lingüístico del Tseltal Occidental*) y evaluación de la zona dialectal sur », in G. Guerrero & L. Orozco (eds.) *Estudios de Variación geolingüística*. México, Secretaría de Cultura/Insti-tuto Nacional de Antropología e Historia, pp. 445-484.
- RANNAP Jürgen, 2017. *Mathematical analysis of Numic languages*. Bachelor's Thesis, University of Tartu, Faculty of Science and Technology, Institute of Mathematics and Statistics.
- SAPOVAL Bernard, 1997. *Universalités et fractales*. Paris, Flammarion.
- SÉGUY Jean, 1973. *Atlas linguistique de la Gascogne (ALG)*, 6 vol. Paris, Éditions du CNRS.
- WEINREICH Uriel. 1954. « Is a structural dialectology possible? », *Word*, 10 (2-3), pp. 388-400.
- WENDEL John, 2005. « Notes on the Ecology of Language », *Bunkyo Gakuin University Academic Journal*, 5, pp. 51-76.