Dialectologia. Special issue, 12 *(2024), 5-50.*
ISSN: 2013-2247

Received 31 December 2023

Accepted 20 February 2024

Published June 2024

DOI: 10.1344/Dialectologia2024.2024.2

LANGUAGE DYNAMICS FOR DIALECT CLASSIFICATION: A SKETCH IN MINIATURE[1]

Jean Léo LEONARD [*]

Université Paul Valéry Montpellier 3 (UPV/M3)

leonardjeanleo@gmail.com

Abstract

The paradigm of *Language Dynamics* is still little known to dialectologists and language classification specialists. The aim of this article is to show how an approach in terms of the Theory of Complex Adaptive Dynamical Systems (CADS) can be of great benefit to the theories and methods of dialect classification. The principles, methods and tools of both qualitative dialectology (isoglosses, ethnolinguistic areas) and quantitative dialectology (cladistics and editing distance) are applied here to Occitan and Mazatec (Otomangue) in order to validate the *uniformity* (universality) of the complexionist method in terms of Critical Language Taxonomy (CTL).

Keywords: language dynamics, complexity, dialectometry, dialect classification, taxonomy

[1] This work is part of the research programme/project Dipralang (EA 739) *Sociolinguistique des contacts de langues et écologie linguistique*. This paper is dedicated to our colleague Carlos de Jesús Wagner Oviedo, Escuela Nacional de Antropología e Historia (Mx), whose recent contribution to Mazatec studies (and Mazatec dialect classification) is crucial. Great thanks too to Gotzon Aurrekoetxea, Universidad del Pais Vasco/Euskal Herriko Unibertsitatea (UPV/EHU), for his patient and efficient editorial help in reshaping the initial version of this paper, insisting on writing it all over again and to include the second case study, in order to make General System Theory central in this contribution –we would not have dared, otherwise, to do so.

[*] ITIC, bâtiment E, bureau 215, Route de Mende, 34090 Montpellier, France.

DINÀMICA DE LA LLENGUA PER A LA CLASSIFICACIÓ DIALECTAL: UN ESBÒS EN MINIATURA

Resum

El paradigma de la *Dinàmica Lingüística* encara és poc conegut pels dialectòlegs i especialistes en classificació lingüística. L'objectiu d'aquest article és mostrar com un enfocament segons la Teoria de Sistemes Dinàmics Adaptatius Complexos pot ser beneficiós per a les teories i mètodes de classificació dialectal. Els principis, mètodes i eines tant de la dialectologia qualitativa (isoglosses, àrees etnolingüístiques) com de la quantitativa (cladística i distància d'edició) s'apliquen aquí a l'occità i al mazatec (otomang) per validar la *uniformitat* (universalitat) del mètode de la complexitat en termes de Taxonomia del llenguatge crític.

Paraules clau: dinàmica lingüística, complexitat, dialectometria, classificació dialectal, taxonomia

LA DYNAMIQUE DES LANGUES POUR LA CLASSIFICATION DES DIALECTES : UNE ESQUISSE EN MINIATURE

Résumé

Le paradigme de la linguistique complexionniste (angl. *Language Dynamics*) est encore peu connu des dialectologues et des spécialistes de classification des langues. Les deux approches sont analogues, mais s'ignorent. Cet article entend montrer comment une approche en termes de théorie générale des systèmes et des systèmes dynamiques et adaptatifs complexes, peut rendre de grands services aux théories et méthodes de classification des langues. Des principes, méthodes et outils de la dialectologie qualitative (isoglosses, aires ethnolinguistiques) aussi bien que quantitative (cladistique et distance d'édition) sont tout à tout appliqués à deux langues de nature très différente, afin de valider l'uniformisme (universalité) de la méthode complexionniste en termes de *taxinomie critique des langues* : le mazatec (otomangue) et l'occitan (gallo-roman, indo-européen).

Mots-clés: linguistique complexionniste, complexité, dialectométrie, cladistique, classification, taxinomie

1. Introduction

It is well acknowledged that any classification of complex semiotic objects, such as languages in their typological, geolinguistic, and sociolectal diversity, is not inherently *given*, as a Thing of the World, but rather emerges as a *construct*. Phyla, language families, and diasystems are not natural entities, immediately observable as one would succeed to do so with plants or animal species. Furthermore, languages are in no way comparable to living organisms, contrary to the long-standing analogy inherited from the romantic period of comparativism and the era of grand reconstructions, leading to grand narratives. Nevertheless, a long tradition dating back

to the early days of modern linguistics, notably with the pioneering work of August Schleicher (1821-1868), tends to induce, for the Common Sense, a somewhat allegedly «entomological attitude» among linguists dedicated to the classification of languages or dialects from various linguistic domains. Though, we know this is not what *Language and Dialect Classification* (LDC) is about. Indeed, LDC is a reflexive field of General Linguistics, and shares a lot with crossdisciplinary fields and methods, such a *Complex Adaptive Dynamical Systems Theory* (CADS), which will provide us here with a methodological grid.

Whether in the spirit of a 'world language map' or in the delineation of internal divisions within well-known language domains like Gallo-Romance, Finnic, or Basque, classifications undergo a process of refinement and division, employing methods such as isoglosses; phonological, morphological, and lexical areas or dialectometry – cladistics and other quantitative and computational tools. Despite acknowledging a degree of 'variable geometry' within these subdivisions, the progression of classifications typically adheres to a linear logic, aiming for increased legitimacy among either specialists or the general public. However, the complexity of LDC necessitates consideration within a logic suitable for intricate phenomenology, characterized by multiple determinisms shaping structural order –such as hierarchizing the richness of functional inventories of functional units, either in «open systems» like lexicon and discourse or «closed systems» like phonology and morphology. This entails a *non-linear*, *vicarious logic* tolerating a good deal of *variable geometry*.

From the standpoint of theoretical and descriptive linguistics, the goal remains unveiling deep geolinguistic and diasystemic patterns (systemic, areal, dependency graphs, etc.) and correlating them with observable surface forms at taxonomic level, and this is quite a task, and at the same time, a very empirical and a critically theoretical one, too. Dealing with complex systems as linguistic stocks, families, languages, and collections of dialects making up intricate webs of polylectal grammars, LDC implies a high degree of theoretical tools and insights. LDC is *per se* a *complex*

field, handling *complex data*, which might benefit from concepts, methods and tools from Complex System Theory (CST).

The reductionist gesture employed here for the purpose of integrating LDC to *General System Theory* (GST) will duly follow unavoidable prerequisites: congruence of results with canonical knowledge – (ad)equation with acquired knowledge, visibility or discrete nature of the divisions obtained. Once this convergence checked, algorithmic playgame may start, enhancing emerging structures (dialects, subdialects, varieties, sub-varieties) that would not have been visible with usual methods, heading towards «invisible dialects», beyond «eponym dialects» (Nerbonne & Kretzschmar 2003). The former are dialects as we generally know them: corresponding to places we already know; the later are deeper geolinguistic cores or areas we cannot fathom with our own eyes — we need tools for that, as isoglosses or dialectometric choremes and taxonomies.

1.1 Complex Systems Theory for LDC: two case studies

However, over the past few decades, advancements in theoretical and quantitative dialectology have opened up alternative prospects. We will briefly outline two case studies that will serve as our experimental and critical observatory: first, Mazatec within the Popolocan languages (Eastern Otomanguean, Mexico), and second, Occitan (France) within the Gallo-Romance subdomain of the Romance *genus* or sub-family, within Indo-European) (see Brun-Trigaud 2023, in issue 1 of Diacleu project). We will infuse this approach with a slight historiographical hue, in line with the editorial stance of the present journal, while focusing on methodological points.

Our focus will be on the contribution of *Complex Adaptive Dynamical Systems* (CADS) (Gros 2015) to reshaping the classification of languages and dialects.[2] We will demonstrate how this approach, grounded in the vicariance of methods and

[2] Our approach here is part of the paradigm of Language Dynamics, in the rise since two decades : see Wichmann (2008), Heinsalu *& al.* (2020). Since 2004, Jeff Good and Simon Greenhill have been editing a journal specialized on this domain of research at Brill : *Language Dynamics and Change* (see https://brill.com/view/journals/ldc/ldc-overview.xml), of special relevance for LDC.

perspectives on one hand, and the non-linearity of methods for hierarchizing language components and sub-components, understood as *diasystems* (Weinreich 1954), on the other, proves to be highly heuristic. We shall begin by examining five dialect classifications of Mazatec (section 2) and two of Occitan (section 3), through the hermeneutics (subsections 2.2-6) and the methods of GST and CADS. Finally, we will conclude the article by considering potential avenues for future research (section 4).

1.2 A Glimpse at CADS premices and holistics

Firstly, we will delineate the foundations of CADS theory before establishing its connection with the objects and methods of dialectology and language classification. We will find that the two paradigms share much in common, primarily because they direct their attention towards complex sets of data, or, in more epistemological terms, complex ontologies. We will revisit the key principles that capture the architecture of Complexity, as articulated by one of the founders of CADS, Herbert Simon (1905-2001), an economist and sociologist, in a now-classic article (Simon, 1962). According to Simon, CADS, derived from General System Theory (GST) (Bertalanffy 1968), is grounded in two elementary notions inherent in any complex set of elements or agents in relation to space and time: *feedback* (superordinate of rank I) and *homeostasis* (superordinate of rank II, pertaining to the dialectics of entropy vs negentropy). The first accounts for the *interactions* among the components of an organized system, ensuring both the conditions for its maintenance or *self-organization* and its *reproduction*, based on *hierarchized (quasi)components* and *(quasi)sub-components*. The second elucidates the processes of *unification* and *aggregation* or, conversely, *division* and *disaggregation* among the variegated units constituting the internal structure of the system as a whole. In terms of GST, these units, interconnected, organize themselves like « matryoshkas dolls », from the infinitely small (ultimate constituents: elementary particles or atoms in physics; distinctive features in phonology or morphology) to the infinitely large

(macrostructures: planetary systems, galaxies, galaxy clusters, etc.), encompassing all intermediate states (cells and molecules in chemistry; lexemes and phrases, utterances, paragraphs and texts, discourse and narrative masses, etc.). These two *dynamic processes* of *feedback* and *homeostasis* are *self-regulated* through their *internal and external interactions*, linking mass and energy in a dialectic of *entropy* (intensification of exchanges, with increasing instability) vs *negentropy* (stabilization of states or decreasing instability).

Four dimensions (levels or patterns) must be considered for an ordered and evolving understanding, according to Herbert A. Simon, with all the spirit of one of the founders of CADS: (A) hierarchy and structuring of components, from simple to complex (*modularity*); (B) evolutionary lines of unification/fragmentation and (re)composition (*dynamism*); (C) internal structure and balance of constituent units (*componentiality*); (D) description and representational levels (*emergence*). These patterns or levels of understanding involve formalization processes, which can be qualitative (QUAL) or quantitative (QUANT) –level E– as well as realization or actualization in the form of knowledge objects –in other words, an implementation: level F.

The entirety of these coordinates in CADS-type research is summarized and configured in the diagram (Figure 1), which stands as an implication graph (Haspelmath 2001). It should be read from top to bottom and optionally by following paths connecting various moments of understanding, variable according to the complexity of the phenomena studied. Type I and II understandings are comprehensive operations that account for the entanglement and nature of relationships between components of all sizes in the studied system (interaction/feedback) and the intensity of these exchanges of matter, form, energy, or value (valence, marking, etc.), in terms of the regulation of information flows – more or less homeostatic states.

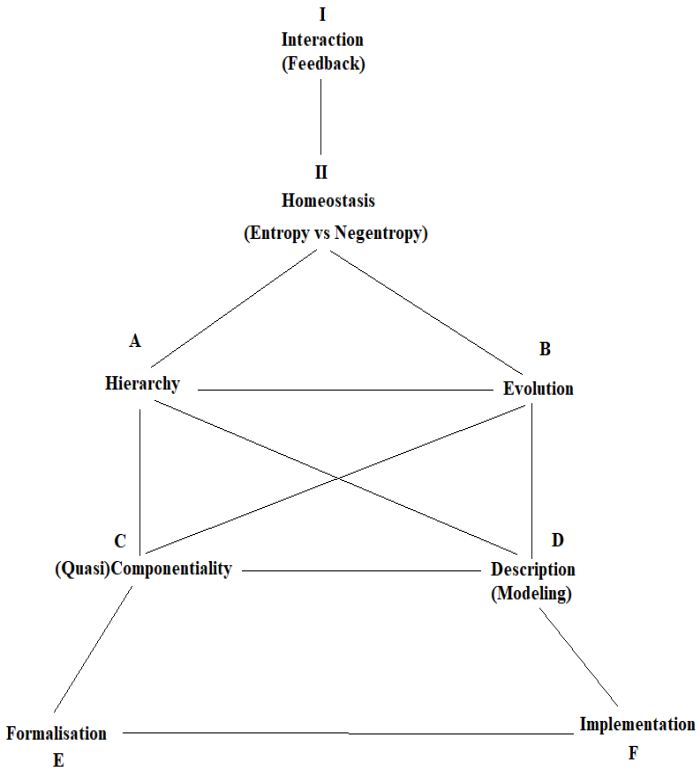


Figure 1. Herbert Simon's Sketch of CADS premises

The four vertices A-D make up a field of *interaction* and *homeostatic* adjustment, hierarchizing these *information flows* (A), generating *evolutionary trends* (B), arranging and rearranging the *(quasi)components* and (quasi)sub-components of the system (C), and bringing forth patterns of *variable geometry* (D). These patterns are made fully accessible or visible, along with their parts or (quasi)components, through *implementation* techniques such as computing or any form of instrumentation (QUANT). In the relationship between D and C, as no description or modeling of a complex system is inherently definitive due to the intrinsic complexity of the constituent parts of the system, the hierarchies available to the researcher are destined to be questioned during the discovery protocol, embedded in a historiographical and epistemological temporality (this parameter is referred to as "quasicomponentiality" in the graph; Simon speaks of «nearly decomposable systems», Simon (1962: 473-477): formalizations (vertex E of the graph) are therefore

subject to evolution (level B) both in their external form and in their relationships with other vertices of the graph. For example, a phylogenetic reconstruction, generating etyma, indeed constitutes a formalization, even if the only formal symbolic element reduces to indicating the protoforms reconstructed by an asterisk.

Similarly, the model of *cyclical concatenation* governing lexicon formation in Mazatec, according to Gudschinsky (1956), which will be discussed later, falls under a process of formalization (vertex E) –just like the theory of voice quality in phonology, which was discovered only late, with implications for the implementation (vertex F) of an internal classification of Mazatec dialects. Consequently, vertices E and F are connected in the graph and, therefore, they interact to varying degrees, by percolation, throughout the entire network of understanding levels represented by the graph: they percolate throughout the system, just as superordinates I and II at the top of the graph percolate at each level of this construction. Thus, the researcher in complex systems is somewhat akin to a genie released from Aladdin's lamp, an expert in problem-solving amidst the complexity of the world, capable of answering all questions (in the famous tale, of fulfilling all wishes), by pulling the strings of his magical power (the heuristic nature of his solutions) from multiple dimensions inaccessible to the layman, who moves at the surface level of the world.

As an application of this model to the field of language classification or dialectal domains, level of understanding (A) corresponds to the hierarchization into *phyla* or macro-groups of languages vs "linguistic families" vs sub-families, languages vs dialects, sub-dialects, and varieties; level (B) to the various possible or successive arrangements in describing the groupings and affiliations between these entities, as well as the different stages of discovering the constituent sub-families of the linguistic group; level (C) corresponds to the internal complexity of each of these groupings and the multiple phonetic laws and specific structural properties of each subgroup (for example, the hesitations regarding the relevance or truth of a subgroup like Italo-Celtic in the Indo-European classification provide a good example of the inherent «quasi-componentiality» of many sub-components of a classification, seen as a complex system); level (D) to the properties described and retained to establish these entities

and justify their groupings, as well as the levels of empirical description (for example, structural variation of the language vs sociocognitive perceptual patterns in terms of intelligibility, as we will see for Mazatec).

1.3 Introducing the Mazatec case study, as a "toy universe"

This analytical framework, applicable to all categories of constituent entities of varying sizes within the linguistic systems under consideration, will be exemplified, for instance, with Mazatec. Initially, an aporia inhibits level (A), resulting in the postponement of classifying this language within the classification system of Mexican languages. Once this obstacle is overcome, the proposed evolutionary lines to configure Mazatec within a family like the Otomanguean phylum remain opaque, prompting various reconfigurations and an ethnohistorical approach with an attempt at periodization (level B). The very nature of language structures and combinatory rules of both phonemes and morphemes nevertheless still pose serious typological problems, motivating a modelling by cycles of composition/fusion (analytical vs synthetic constructions) of functional units, at both the phonemic and lexical (level C). Finally, a reanalysis of observable natural groupings through the sociocognitive dimension, aided by interdialectal intelligibility tests, may lead to a partly vicariant (i.e. alternative) viewpoint, partly congruent with previous classification (description) attempts (level D).

We will now apply further this analytical framework to Mazatec as an exemplary case study, both in terms of the structural complexity of the language typologically and in terms of the vicariance of perspectives that the researcher can obtain by using various methods interchangeably and in combination. Only after this preliminary exploration of a language, which, although relatively unknown to the general public, is well-known to typologists, phonologists, and historians of linguistics, will we consider, for pedagogical purposes, applying the CADS approach to a domain much more familiar to most linguists interested in language classification: the Occitan domain, as a

component (itself complex, in terms of diasystemic structure) of the Gallo-Romance domain. At this stage, we can apply in a more sustained manner the higher order macroprocesses of *feedback* and *homeostasis*. The Mazatec case-study will provide a « toy universe » model for the application of GST/CADS hermeneutics (through dynamic, interactive hierarchies) and tools (algorithms, as Edit distance with *Gabmap* or cladistics with PAUP).

2. Mazatec: From classificatory apory to a pilot diasystem for theoretical dialectology

Mazatec (ISO 639-3), a language now well-known as belonging to the Popolocan subfamily of the Oto-Manguean phylum (eastern branch), is primarily spoken in the state of Oaxaca, Mexico, with approximately 240,000 speakers in 2020. Since the 1950s, Mazatec has experienced population movements to neighboring states, particularly Veracruz. The Mazatec diasystem (as per Weinreich 1954) is empirically and epistemologically intriguing due to its structural diversity, yet with dense geographical coherence, described as a compact vertical archipelago. Mazatec stands out for its vitality, accessibility, and richness in speakers and dialectal varieties (Ariano Cifuentes, Esteinou Dávila, Gómez Flores & Rodríguez Pérez-Abreu 2014; Léonard & Dell'Aquila 2014), surpassing other languages in its subfamily – the Popolocan languages, a peripheral branch of the Eastern Otomanguean stock. The most neutral endogenous term is *énnà* 'our language' (*én* 'language' =*nà* Poss1Pl.Inclusive).

The phonological and morphological complexity, tonal-grammar interaction (Léonard & Fulcrand 2015), diverse verbal inflection class patterns (Léonard & Kihm 2012, 2014), and geolectal lexical variability make Mazatec a challenge for general linguistics, linguistic typology, and theoretical and descriptive dialectology. Mazatec poses a challenge to traditional dialect classification by questioning subdivisions based on typological and geographical criteria. Interdialectal contact plays a fundamental role in classification due to successive population movements throughout history.

Various methods, such as comparison, isoglosses, dialectometry, and Complex Adaptive Dynamical Systems theory (CADS), have been applied to study Mazatec.

To what extent (and how) can an approach in terms of CADS assist the linguist in overcoming the aporias that may prevent the classification of a language within its potential affiliation group (level A)? The incompleteness of the grouping inevitably entails a questioning of the proposed classificatory complex, with a domino effect on the entire classification (levels B and C). In the case of Mazatec, the apparent complexity of linguistic structures may have led to difficulties at level (D), with inhibitory consequences, retroactively, at all preceding levels.

2.1 Mazatec dialect classifications

As any taxonomic endeavor concerning a linguistic domain results from a more or less complex and temporally extended set of successive problem-solving attempts (problem solving, cf. Simon 1962: 472-473), an overview of the seven successive classifications of Mazatec falls within level B of complexity understanding resulting levels in D & E modifications: any evolutionary line of technical solutions mobilizing discovery protocols constitutes, in itself, a process of natural selection, on the epistemological level. We will now address this dynamic by applying our four-part model as systematically as possible.

We will describe first the successive internal classifications of the Mazatec *dialect continuum* (or *diasystem*), indicating the methodological or theoretical framework of reference, as well as the hierarchies or taxonomies proposed by different authors (dialects, sub-dialects).

The first classification about Mazatec dialects was made within the Otomanguean Phylum.[3] The integration of Mazatec into the Popolocan genus within

[3] We use the term *phylum* here because the internal structural diversity of this group is analogous to that found in densely diversified language families worldwide, such as Niger-Congo or Afro-Asiatic. We use *language family* for linguistic stock (Indo-European, Uralic, Sinitic, etc.) and *genus* to refer to a

the Otomanguean complex will soon be suggested by the Mexican linguist, a pioneer in the history of classifying Mesoamerican languages, Francisco Pimentel (1875: 466). He formulated the hypothesis, based on some cognates gathered in one of his lists, of a relationship of Mazatec with a Mixteco-Zapotec macro-subgroup, applying his philological method, cross-referencing a wide variety of sources (word lists, translations of religious texts, grammars of evangelists, cf. Cifuentes 2002: 98-100). This hypothesis had ambivalent consequences in terms of taxonomic construction of Mazatec: on the one hand, it proved heuristic as it effectively integrated Mazatec into the Otomanguean complex, whose designation varied between "lenguas oaxaqueñas" and "macro-Mixteco" in Swadesh (1959), before a conflation was finally realized and accepted under the term "Otomanguean," bringing together the Oto-Pame subfamily comprising the Otomi-Mazahua languages and Pame to the north of the domain (cf. Soustelle 1937).

Mazatec constitutes a diversified diasystem, with varieties that are both mutually intelligible and relatively isomorphic in terms of cognates and typological structures, yet irreducible to the other three languages of the same *genus* or, in other words, of the same *subfamily* (see Heinsalu *& al.* 2020: 14, 20, 32, 38-40): Ixcatec, Chocho, Popoloca (see Adamou 2021 for successive hypotheses of affinities and contacts between these languages).

This branch, called *Popolocan*, internal to the Eastern Otomanguean, formerly called Zapotec-Mixtecan, confirmed by a study by González Casanova (1925), finds its consecration with María Teresa Fernandez de Miranda's article (1951), which concluded that Mazatec was more distant from Ixcatec, Popoloca, and Chocho than these three were from each other.

From this stage, the aporia that inhibited level A of understanding the taxonomic properties of Mazatec since Pimentel's encounter finally finds its resolution. In a cascade-like fashion, the other patterns of our model will soon be able to interact with each other (feedback), as we will see with Eric Hamp's vicariant model (1958), which

subfamily within a language family (Germanic, Celtic or Romance languages; Finnic or Volgaic or Permian, etc.).

proposes several stemmata (levels B & E of complexity understanding) within the Popolocan subfamily, but without yet delving into the internal structure of the Mazatec diasystem. Gudschinsky's "Popotecan" model (1959) still concerns a level B understanding but lays the groundwork for a subsequent level C understanding by doubly integrating Mazatec in terms of filiation within the phylum – a comparative step necessary before considering any internal taxonomy of the language –: on the one hand to the Mixtec ensemble (level A understanding), and on the other hand to the Popolocan subgroup. As shown below (Figure 3), while the author manages to place four Mazatec dialects in her "Popotecan" stemma, she does not find it necessary at this stage to configure them among themselves (level C). This latter task requires a discovery protocol at level D, which she will gradually establish using i) the model of cycles of morphological and lexical componentiality and ii) Bloomfieldian isogloss models (see below). E. Hamp (1958) adopts this hypothesis and systematically summarizes the phonological isoglosses of this author, distinguishing between phenomena of retention vs innovation, suggesting interferences between the two closest languages: Chocho and Popoloca. This approach allows him to outline a stemma in which Mazatec would be an outlier (or external member) of the Popolocan taxonomy, while Chocho and Popoloca would be two languages with closer affinities, flanked by Ixcatec as an external member of this internal cluster (Figure 2).

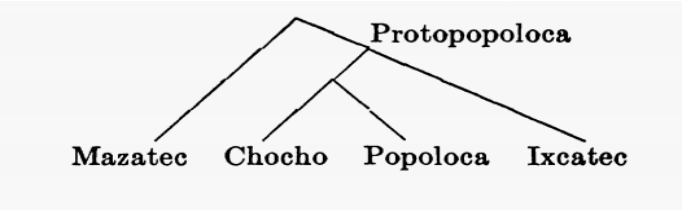


Figure 2. Eric Hamp's Popolocan *Stemma* (1958: 151)

A year later, Sarah Gudschinsky's (1959) published a second classification, which brings together the Popolocan languages within a macro-subfamily of Otomanguean that she names *Popotecan*, uniting Popolocan and Mixtecan in a broad sense, even

including Amuzgo in the second branch. The stemma (Figure 3) thus configures a group derived from Proto-Popotecan (PPtn), generating a descendant set from Protopopolocan (PPn) on the left vs a group descending from Protomixtecan (PMx) on the right. The first group splits into two: on the one hand, the three languages descending from Protopopoloca (PP): I (Ixcatec), P (Popoloca), and C (Chocho); on the other hand, the four dialectal varieties of Mazatec that the author managed to document in close partnership with the local team of the Summer Institute of Linguistics (S.I.L.) in the field (H: Huautla, SJ: Santa Maria Jiotes, Mz: Mazatlán de Flores, S: San Miguel Soyaltepec; see map 1 bellow). The second group concerns Mixtecan (MM: macro-mixtecan, PM: proto-mixtecan) and associates an external member (A: Amuzgo) with three unconfigured languages: M: Mixtec, C: Cuicatec, T: Triqui, as in Mechling 1912 and Lehman 1920.

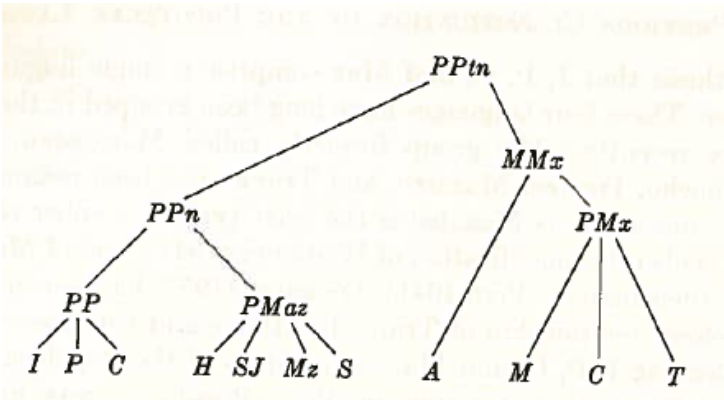


Figure 3. The Popotecan stemma, according to Sarah Gudschinsky (1959: 2)

Sarah Gudschinsky, relying on firsthand data collected with Kenneth and Eunice Pike in a Mazatec area undergoing forced modernization (Boege 1988). Linguists from the S.I.L. established themselves in Mazatec communities from the 1940s, providing Gudschinsky with a documentary foundation surpassing all previous compilations (e.g., Brinton 1892, Starr 1900, Belmar 1892). Relying on this first-hand documentation, Gudschinsky writes several comparative and dialectological studies on Mazatec, using various methods of reconstruction, glottochronology, and sequencing of cognate sets (Gudschinsky 1953, 1955, 1956, 1958). Her work lays the foundation for the

reconstruction of the Mazatec protolanguage and influences the early taxonomies describing the inner structure of its dialect continuum. These classifications are crucial for understanding the evolution of Mazatec within the Eastern Otomanguean phylum.

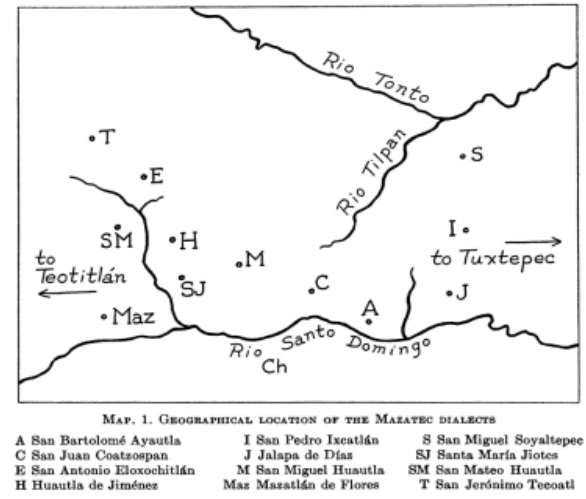
2.2 Gudschinsky's models

The question arises: how does the reconstruction and proper placement in a *Stammbaum* relate to the internal classification of a dialectal domain, as in this specific case? Once again, the consequences of the "natural selection" process, involving a sequence of attempts and problem-solving techniques, are obvious. Reconstruction, crucial for building and modelling a higher-order hierarchy of complexity (level A in Figure 1 above), holds significant importance for a language lacking written attestations of the "proto-language" or diachronic source. Moreover, the original typology of Mazatec, initially puzzling to Pimentel and early classifiers, becomes more predictable once attachment to a specific language group becomes feasible or gets some accomplishment for the specialists of the field.

From the perspective of *vicariance*, facilitated by the diversity of qualitative and quantitative methods applicable – a fundamental prerequisite of the CADS method – the gain is substantial. Methods such as cladistics, to be employed later, rely on deriving contemporary dialect forms from so-called "ancestral" characters (etymons). Another method, like edit distance, also used in this historical overview, requires a coherent encoding of dialectal forms beforehand, drawing on knowledge of evolutionary trends within the diasystem. These methodological motivations are tied to level D (description, modeling), with implications for constructing knowledge objects like maps and trees of levels B and C, as illustrated through the successive advancements of S. Gudschinsky below.

In 1958, she goes beyond the simple groupings of her previous Popolocan or Popotecan comparative works and achieves a first ordered and detailed classification of the internal geolectal components of the Mazatec continuum. The highlands (with

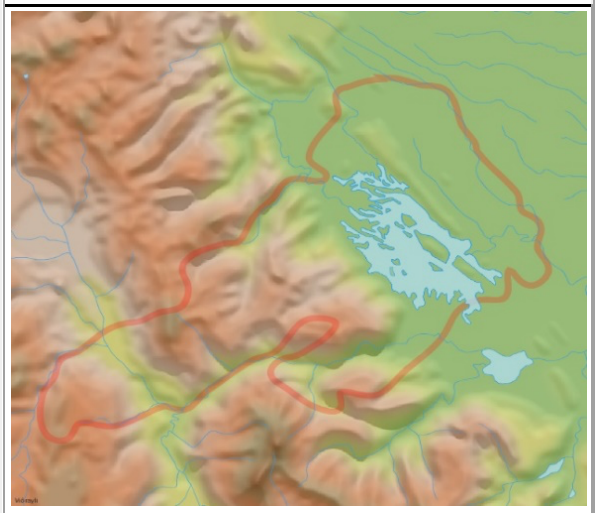
Huautla de Jiménez, i.e. HU, as the central point) are distinguished from the lowlands (two competing centers: San Pedro Ixcatlán (IX) to the west vs San Miguel Soyaltepec (SO) to the east, whose population was relocated starting in 1954 during the construction of the Miguel Alemán hydroelectric dam, causing the visible expansion to the east in the map (Map 2. In between lies a piedmont that includes San Bartolomé Ayautla (AY), San Felipe Jalapa de Diaz (JA), and Santo Domingo (DO). Most of the outer shape of the Mazateca ecological context and evolutive historical frame in the making up of agrarian communities, matches Charles Camproux's general principle in geolinguistics of « au fil de l'eau » basic pattern of settlement:[4] along main rivers, whatever the altitude. In short, physical geography runs anything else: human factors just cope with it — namely, with hydrography and orography of a settlement zone.



MAP. 1. GEOGRAPHICAL LOCATION OF THE MAZATEC DIALECTS

A San Bartolomé Ayautla	I San Pedro Ixcatlán	S San Miguel Soyaltepec
C San Juan Coatzospan	J Jalapa de Díaz	SJ Santa María Jiotes
E San Antonio Eloxochitlán	M San Miguel Huautla	SM San Mateo Huautla
H Huautla de Jiménez	Maz Mazatlán de Flores	T San Jerónimo Tecoatl

Map 1. Location of Dialectal Varieties Analyzed in (1958, map 1 p. 470)



Map 2. Orography of the Mazatec Dialectal Space (Vittorio dell'Aquila, CELE 2011)

The map of the left (Map 1) provides a list of locations where Gudschinsky (1958) had collected firsthand data at that time. It is evident that, due to the vicissitudes of recent history, the Mazatec area has expanded since 1954, primarily due to the

[4] See Camproux (1962: 764-766) about this kind of geolinguistic fractal based on orography and hydrographic basins, applied to the Gevaudan sub-dialect of Occitan Languedocian – a case study, in the Gallo-romance domain, but also for geolinguistic theory. See note 12 below, too.

displacement imposed on the populations of the lowlands (especially IX and SO) following the construction of the dam. The area of the Tropical Canyon or «Cañada» in the central-southwest of the Mazatec domain is highlighted (Map 2), in relation with map 1: the valley of the Santo Domingo River, leading eastward to the so-called «Papaloapan-Tuxtepec valley», and the river depression to the southwest, enclosed in the canyon, leading to Teotitlán del Camino.

From the perspective of a CADS approach, these details are far from insignificant. Based on the fundamental uniformity of Simon's four dimensions, the ecological and orographic tripartition between high, low, and middle lands along the vertical axis, with its climatic consequences, belongs to analysis level A, describing a hierarchical organization of the terrain with significant implications for the self-organization and condensation of human communities (level B). These communities, particularly agrarian ones, will develop evolutionary trends of unification/fragmentation of their practice communities, with a wide range of consequences on the formation of internal units and their equilibrium relationships, interactions, and feedback in terms of communication, and consequently linguistic variation, through the condensation of more or less extensive (or shared) dialectal varieties (level C), as we will see shortly with P. Kirk's intelligibility study (1970). It is also based on these categories and representations that not only linguists but also speakers describe or shape their perception of dialectal diversity (level D). Similarly, in what follows, we will trace S. Gudschinsky's path from this same level D of description and representational fields in her quest for methods and criteria enabling her to propose a classification and, consequently, a hierarchy and internal structure of the various components of Mazatec – in other words, a dialectal classification.

The groundbreaking article on the classification of internal components of the Mazatec dialect continuum by Sarah Gudschinsky (1958) was preceded by a brief study by the same author on glottochronology. In this short paper, she proposes a lexico-statistical method to study lexical relations between dialects within a diasystem (Gudschinsky 1955). She suggests basing the analysis on what she calls DIPS (*Degrees*

of Immediate Phonological Similarity) as an objective measure of lexical relations between dialects. DIPS are based on lexicostatistical techniques that provide a measure of structural relations between cognates, independently of any attempt to establish a chronology of phases of subdivision of the linguistic domain, unlike the classical approach in this field of studies, which aims to propose for each phylum, language family, or dialectal domain a periodization.

In doing so, she refocuses Swadesh's lexico-statistical method on the language itself and for itself. Her study (of the QUANT type, i.e. quantitative) examines six closely related Mazatec dialects: Huautla de Jiménez (HU), San Mateo Huautla or Huautepec (SM), San Miguel Huautla/Huautepec (MG), San Pedro Ixcatlán (IX), San Felipe Jalapa de Diaz (JA), and Soyaltepec (SO). She identifies a bias in lexicostatistical data that can be explained by recent borrowings between dialectal varieties within the same domain. In this case, she attributes this bias to phenomena of mutual borrowing, especially between the dialects of SM and HU, resulting in an apparent lexical proximity.

This approach, partly qualitative and quantitative, which stands in critical use of the then-dominant model in the classification of Native American languages – Morris Swadesh's lexicostatistics/glottochronology – is refined in Gudschinsky's unpublished master's dissertation[5] of 1956. In the glottochronological step of her research, which is quantitative in nature, as we have just mentioned, language facts *per se* were absent, or at least they only served as a basis for calculations aimed at describing geolinguistic phenomena at levels A, B, and C of our GST graph (Figure 1). In her taxonomic approach to Mazatec, S. Gudschinsky operated at level D of H. Simon's tripartition, but in terms of CADS methodology and in accordance with the uniformity of complexity capture levels, she will once again go through the process of discovery, but this time focusing on the constituents of the language itself and for itself. Initially, she will propose a heuristic model accounting for the internal structure and constituent units in terms of phonology and lexical morphology (level C), to confront, in terms of

[5] A summary can be found in print and revisited in Léonard (2020: 316-325).

evolutionary trends (level B), the products of morphemic concatenation cycles with varieties and dialect areas geolinguistically, in order to potentially holistically explain the major dynamics of hierarchization and structuring of areas (level A). To this end, she will propose two models: one we will refer to as “Concatenation Cycles”, and the other an isoglossic model “à la Bloomfield” In each case, the author implements (level F of the graph in Figure 1) a vicarious formalization (level E) toward problem resolution (achieving a classification of the internal structure of the diasystem).

The first approach provides a model of syllabication/resyllabication cycles and lexicon formation heuristic for grasping trends of diversification in the Mazatec diasystem. This morpho-lexicological approach crossed with morphonological rules accounts for phonological evolution within the dialect continuum through cycles of morphonological adjustments between polyvalent lexical roots of the Specifier+Polyvalent Root combined into fusional radicals. This approach of morpho-lexicology processed through morphonological rules allows describing and explaining phonetic changes in the dialect network. The central idea is that compositional templates, such as Classifier + Polyvalent Root for nouns, have merged into inseparable lexemes ([YCLAS[X]] > [YX]). This fusion results in radicals characterized by voice quality, such as modal, creaky, or breathy. The five main *Gudschinsky Cyclic Rules* (hereafter GCR:1-5) are defined as an internally deterministic model, based on phases of compression of morphemic sequences within lexemes and syntax. Each cycle can have cascading consequences on subsequent cycles, in structural and/or relative chronological relation. The cyclic rules proposed by Gudschinsky shed light on the linguistic evolution within the Mazatec dialectal continuum. They are a good example of how and why even linguistics units (lexemes, morphemes, phonemes) may be considered as « quasi-components », as their shape and concatenative constraints may strongly vary in space and time, but also within every single part of the whole (every

variety within the diasystem). Gudschinsky's « quasi-componential » cycles in the Mazatec dialect network shortly read as follows:[6]

GCR-1: Phase of Protomazatec, with consonant clusters and vowels following the Pike & Pike model (1947):[7] complex onsets and nuclei and no coda in autonomous specifying and lexical monosyllabic roots, eventually combined in couplets (disyllabic lexemes).

GCR-2: Introduction of a « metric reform » involving morphemes alternating between short and long forms, leading to lexical redundancy and optional CV | CVCV templates.

GCR-3: Identifiable evolution of templates based on internal consonant groups, implying a composition in the process of being absorbed into a fusional stem.

GCR-4: Prosodic compression of CVCV templates > CVT1-2 by Tonal contour (T_n), equivalent to prosodic absorption, suprasegmental, of previously segmental material.

GCR-5**:** Forms analyzable as compound words in some varieties vs indivisible lexemes in others (as eventually postulated in Jamieson, 1982).

These cycles result in *syllable deletion, reduction of internal vowels, nasalization effects, emergence of medial clusters*, and more. Dialect variation within the Mazatec diasystem involves phenomena such as juncture, reduction of vowel quality, cluster simplification, realignment of glottal features, and the use of prosodic tones. The second model delves a degree deeper into the hierarchy of structures or building blocks composing the complex system of language –from lexical morphology to phonology, through a seminal article published in 1958, which we believe has not received the attention it deserves from the community of linguists and dialectologists, despite the exemplarity of its method. In our view, what is lacking in this article by S. Gudschinsky is the combination of her qualitative (isoglossic) approach of 1958 with

[6] A detailed account and analysis of this QUAL (qualitative) modeling of concatenation constrains in Mazatec according to Gudschinsky (1956) is provided in Léonard (2020).

[7] The concept of correlation of voice quality (cf. Ladefoged & Maddieson 1996: 50-63) is not yet operational at this stage of language description.

the quantitative one she proposed in 1955, to constitute a « complexionist » approach, in terms of CADS, or even GST.

Gudschinsky's (1958) ground-breaking article is called by the author herself a "miniature", an elegant way of introducing a QUAL *model* for dialectological study encompassing both internal linguistic aspects and external factors. The author aims to apply the model of shared innovations, as proposed by the renowned neogrammarian Brugmann (1884). Gudschinsky's retrospective outlook is thus inherently directed towards neogrammarian methodology and classical diffusionist theories in comparative linguistics (Dyen, Dahl, and austronesianists).

Although she explicitly acknowledges that the differences among Mazatec dialects are minor and primarily involve phonological variation, with a bold, yet consistent set of cognates, Gudschinsky accomplishes this reductive classification of Mazatec dialectal areas (Figure 4).

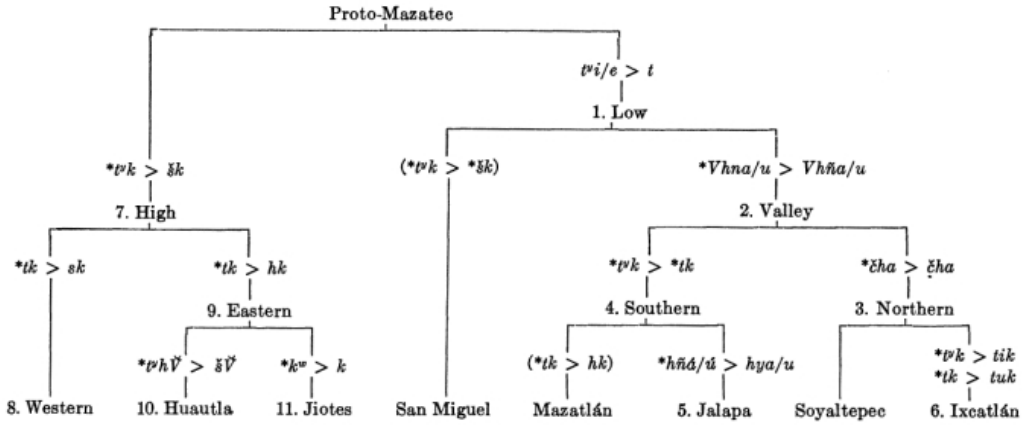


Figure 4. *Stemma* of Mazatec Dialects, According to Gudschinsky (1958: 471)

Gudschinsky's isogloss *stemma* therefore opposes lowlands to highlands from the perspective of external interpretants (i.e., determining geographical factors). Eleven isoglosses make up this system, with the first isogloss (**t^{j}i/e > t*) marking the lowlands and the seventh isogloss (**t^{j}k > ʃk*) characterizing the highlands. Successive subdivisions cascade from this model (Table 1).

Isogloss number	Phenomenon	Phonological process	Dialect areas
1	**tʲi/e > t*	Depalatalization T^{J}	Lowlands (Valley)
2	**Vhna/u > Vhɲa/u*	Palatalization of nasal sonorant HN^{J}	*idem*
3	**tʃha > tʂha*	Depalatalization of aspirated affricate TS^{H}	Northern Valley (IX & SO)
4	**tk > hk*	Debucalization of implosive coronal stop TK	Southern Valley (MZ)
5	**hnaɭ/uɭ > hja/u*	Approximantization of palatal sonorant HN	*idem* (JA)
6	**tʲk > tik, *tk > tuk*	Vocalization &/or epenthesis of V high nucleus in cluster $T^{(J)}K$	Northern Valley (IX)
7	**tʲk > ʃk*	Palatal Desaffrication of alveopalatal semi-affricate $T^{J}K$	Central Highlands / Cañada core (SM)
8	**tk > sk*	Preconsonantal assibilation TK	Western Cañada
9	**tk > hk*	Debucalization TK	Eastern Cañada
10	**tʲV > ʃV*	Hushing & desaffrication of alveopalatal semi-affricate T^{J}	*idem* (HU)
11	**k^{w} > k*	Delabialization K°	*idem* (JI)

Table 1. Gudschinsky's 'Bloofieldian' set of isoglosses for the Mazatec dialect network

In terms of CASD, Gudschinsky's stemma reveals the geolinguistic components, rooted in a landscape articulated between Upper and Lower Mazatec. We observe the emergence of a hierarchy at Level A, which subdivides into closely nested subcomponents (Level B), especially in the Lower Mazatec cluster: on the one hand, the foothills of the Sierra Mazateca through the innovation of San Miguel, near Huautla, as an external member of the Papaloapan Valley –achieved through a simple preconsonantal palatal desaffrication (**tʲk > *ʃk*) –; on the other hand, the internal block of this cluster, which contrasts the north and south of this territorial depression. The internal structure of these two branches is composite: to the south, the chain of Mazatlan and Jalapa dialects, to the north, the chain of Soyaltepec and Ixcatlan, which were two attractors with high diasystemic entropy (superordinate II in Figure 1 above) in Lower Mazatec. These four varieties are all "Town Dialects". However, as we will see in the study of dialectal intelligibility, they are more niches than radiating centers. It is also facing these «strong personalities» that one finds oneself inclined to a certain degree of indecision in defining whether they are autonomous dialects or sub-dialects –a categorial indecidability induced at Level C, in terms of «quasi-components».

Conversely, the highlands block (Upper Mazatec) presents a simple hierarchy of «bosses» or leading dialects (Huautla) surrounded by a satellite like Jiotes (only about 5 km away), flanked by a «secessionist» clade in terms of diasystemic division (northwest varieties). In terms of CADS, we see that each branch and sub-branch of the stemma unfolds a coherent yet diversified range of (sub)components, and they are definable in terms of the interaction regimes (from superordinate I in Figure 1) among them. These are the products of self-organization (from superordinates I & II in Figure 1) among human communities and their processes and acts of mutual adaptation throughout history. The importance of sensitivity to social, geographical, and historical context becomes clearer. Heterogeneity and flexibility are closely associated, as predicted by GST, but it can hardly predict such a diversity of ethno-linguistic situations, rooted in multiple temporalities and interaction regimes (Level B, evolutionary).

Mindful of these limitations, the author ventures further into methodological exploration. Beyond the model of successive innovations of consonantal patterns, Gudschinsky, in her «miniaturized model» (1958) of geolinguistic analysis, discerns multiple levels of variation —a matter of undeniable interest from the perspective of descriptive linguistic methodology. The author, faced with decisions on how to serialize, classify, and organize data, aims not only for a descriptive model of dialect diversity within a domain, but also for an explanatory understanding of the internal processes of diversification within the ethnohistorical region under scrutiny. In her reductionist approach —an assurance of methodological efficiency in descriptive linguistics, where cutting through the complexity of intertwined language facts is essential— Gudschinsky, having identified dialect areas and sub-areas through her Bloomfieldian consonantal variables that make up the basis of her model of successive innovations (see Figure 4 above), establishes the following premises for analysing observable series or layers of dialect diversification over time and space, from (i) to (x), in the order of her exposition:

i) Great divide (Neogrammarian phonetic laws condensed into major areas)

ii) Peripheral areas vs central innovations

iii) Lexical diffusion, instead of expanding Neogrammarian phonological innovations

iv) Polymorphism

v) Lexical lability

vi) Indecidability, making the tracing of isoglosses impossible

vii) Inference: ancient polymorphism, inherited, pre-partition of the sub-family

viii) Local semantic innovations (or motivational endemism)

ix) External factors

Last, but not least:

x) Buffer zones, as a result of local intricacies between all previous 9 factors.

This sequencing of variation parameters, acting both independently and in interactions with each other, enables the author to suggest an innovative narrative on the evolution of the Mazatec dialectal continuum throughout history, independently of the mechanical solution that a purely lexicostatistical approach or a Neogrammarian or Bloomfieldian approach might have provided. It allows her to suggest at a synopsis of ethnohistorical reconstruction spanning seven periods, from an initial period of homogeneity, with free variation of **a* and **u* in initial syllables across the entire continuum, to a last period in which recent changes unfold within the diasystem in individual varieties within the previously defined areas: this era intersects between Aztec hegemony and the Spanish "Conquest" (Gudschinsky 1958: 480-481).

In doing so, S. Gudschinsky grounds the synthesis of her results in temporality; she models the evolution of the Mazatec diasystem by relating it to ethnohistory as a dynamic system (Level B). Her formalization (Level E) is less trivial than it appears, as it goes beyond a *Stammbaum* (her *stemma*): she implements (Level F) a complex device consisting of 8 variational modules separately from her Bloomfieldian model in historical phonetics to explore the complexity of the diasystem's evolution in terms of quasi-componentiality (Level C). Her approach rivals CADS and GST: on the one hand, she initially uses reductionist methods to grasp the diasystem's architecture (Level A),

employing capture processes related to homeostasis (superordinate I): the large diffusion flows of Neogrammarian phonetic laws (criterion i), center-periphery dynamics, akin to Matteo Bartoli's model (criterion ii), partially random phenomena of lexical diffusion (criterion iii). On the other hand, from the quasi-componentiality level (Level C), she considers polymorphism (criterion iv), lexical lability (criterion v), indecisiveness of isogloss tracings for certain items (criterion vi), as well as residual local phonetic laws originating from the aggregation level represented by Popolocan (criterion vii, and cf. Hamp's model *supra*). Cognate relevance blurring through motivational endemism (criterion viii). Finally, criterion ix of external factors opens the perspective toward a higher order of complexity, involving ecology, material and cultural anthropology, history, and geography. This ascends the exploration of complexity to the levels of superordinates I and II respectively: interactions between human aggregates (village and rural communities in the verticality of settlement environments) and homeostasis of local condensation processes (formation of dialects, sub-dialects, and varieties) and unification. The third sociocognitive classification by Paul Livingston Kirk allows us to address this macrodimension, anchoring his empirical approach and discovery protocol in linguistically formatted and implemented data (Levels E and F) as tests of interdialectal intelligibility. We will explore the significant heuristic qualities of this approach, applying an analysis in terms of CASD and GST.

2.3 Dialect Intelligibility: a sociocognitive field of simulated interactions

In the third classification, Casad's framing of Kirk's (1970) results provides an unexpected portrayal of dialect groupings within a geolinguistic complex that is now contemplated more as a dialect network than as a dialect continuum. His representation takes into account the directionality of dialect intercomprehension, depicted by arrows between senders (i.e. speakers) and receivers (i.e. hearers), as explicitly shown in the diagram (Figure 5).

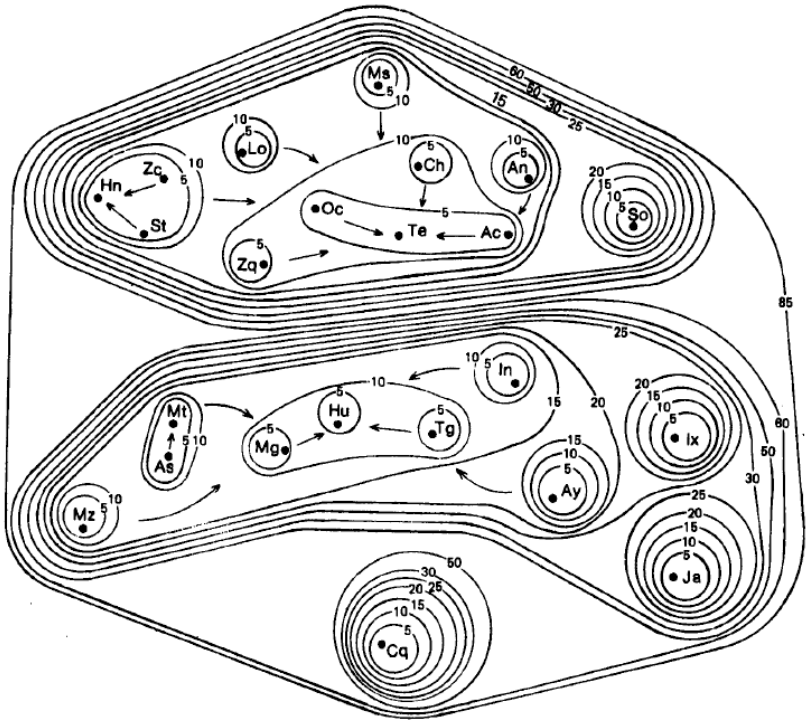


Figure 5. The Mazatec Interdialectal Intelligibility Network (Casad 1974: 47)

This network is consequently formed, in hindsight, by the following sets, now largely emancipated from any external ecological or orographic correlation:

1. A northern macro-group associating the dialects of the NW (including LO and TE) and a "town dialect" from the low plain (SO).
2. A central and southern macro-group encompassing both the central highlands (HU, MG) and the northern (IX) and southern (JA) Papaloapan-Tuxtepec valley, although IX joins the block of high and middle lands while JA remains a satellite of this larger ensemble.
3. A singleton or outlier not reducible to the two previous clusters: Chiquihuitlán (CQ), which turns out to be the variety, if not the most isolated, at least the one that communicates the least with the rest of the dialectal network.

These three remarks pertain to the trivial interpretation of this result formalized in a set-theoretical diagram by Casad. From the perspective we are interested in here,

this visual artefact is analyzed as follows: no longer merely from the standpoint of the speaker's utterance production – which provides the trivial data of cognate lists that form the basis of both comparative analysis and dialectometric processing of dialectal facts – but this time, from the receiver's perspective, the entire order of interaction (superordinate I) is inverted here in a way that simulates the conditions of homeostasis (superordinate II) in communication between speakers. The overall hierarchy (Level A) reveals three nearly decomposable subsystems (Level C), each with variable and heterogeneous degrees of entropy (in terms of the intricacy of interaction flows and their feedback). A first set (top of the diagram) seems relatively independent of the ecological and orographic factor (the tripartite model of high vs. middle and lowlands, Level D in the perception of external factors): it brings together varieties from the northwestern highlands (San Lorenzo, San Jerónimo Tecoatl, etc.) with those from the eastern lowlands (Soyaltepec: SO). These places have shared, over recent centuries, a microfundio economy oriented towards subsistence agriculture, explaining the configuration of these aggregates in the form of variety chains (Level D), whose interactions eventually converged toward the attractor of San Jerónimo Tecoatl (a satellite coffee center of Huautla, in the central highlands), while Soyaltepec, marginalized by the creation of the dam and water reservoir, eventually became peripheral.

The second major set corresponds to the varieties spoken in the heart of the Cañada (highlands), reflecting the socio-economic «attractor» that has been the coffee production system over the past two centuries, centered around Huautla de Jiménez, the main hub of the Mazatec region, including the middle lands. In turn, this subset is linked to "Town dialects" of the foothills and the Tuxtepec-Papaloapan valley, following a less intricate chain model: first Ixcatlán, then Jalapa to a lesser degree. Finally, the most «isolated» variety is Chiquihuitlán, nestled in the Tehuacán-Cuicatlán valley, with a tendency towards a corridor leading to mestizo towns like Tehuacán and Teotitlán, or in contact with Mixtec and Cuicatec varieties. In each of the two major groups, a homeostasis related to the evolution (Level B) of agrarian production

systems (Level D) favored groupings (Levels A and C), resulting in a relatively flat hierarchy (flat hierarchy) through negentropy (superordinate II in Figure 1 above), while three southeastern peripheral varieties (SO, JA, and CQ) found themselves further away from the rest of the dialectal network. We can thus observe the impact of external factors (Criterion ix of Gudschinsky's isogloss modeling) on the evolutionary dynamics (Level B) of hierarchical and grouping formation more or less clearly delimited (Level C) of dialect chains.

The fourth classification will further implement (Level F) this scenario, which differs significantly from Gudschinsky's ethnohistorical model. The method used will also be quantitative, and its formal framework (Level E) will come from the natural sciences, using the cladistic tool.

2.4 Phylogenetics (phonological component) Through Cladistics

Fourth classification: the phylogenetic groupings obtained through cladistic method are derived from «patristic distances», which take into account successive changes and the interdependence of evolutionary characters diachronically, through pairings configured in upstream stemmata or graphs indexed downstream into a matrix. This method goes beyond simple relationships of addition, deletion, or substitution on the surface – synchronically – as done by the editing distance or traditional similarity maps in dialectometry. The methodology of cladistics applied to dialectal continuums is detailed in Gaillard-Corvaglia et al. (2007, 2008); Gaillard-Corvaglia (2012). This method is well-established in population genetics and molecular biology since it was contrived by the German entomologist Willi Hennig (1913-1976, see Darlu & Tassy 1993), and has been recently applied to linguistic data under the supervision of Pierre Darlu (Inserm & CNRS, Fr) in partnership with linguists and dialectologists.

The groupings below align with Gudschinsky's classification and are supported by dialect intelligibility tests conducted by Kirk (1970) and Casad (1974). The detailed analysis of the 31 phonological variables processed here from cognate sets from Kirk

(1966) reveals distinctive traits within each cluster. Additionally, the classification considers less salient phenomena, such as processes of assimilation and dissimilation of vowel chains, diphthong reduction (*au* > *o*), mid-vowel lowering (*e* > *a*), nasal vowel fronting, prothetic vowels, depalatalization of palato-alveolar stops ($t^j i$ > *ti*), affrication of breathy sibilants, nasal onset deletion (*n* > –),voicing of prenasalized stops (*nt*, *nk* > *nd*, *ng*) and modalisation of creakiness, etc. In their classification of Mazatec dialects based on 31 phonological variables processed with PAUP (see https://paup.phylosolutions.com/), Léonard & al. (2012) emphasize a detailed analysis of phonological characteristics, while not excluding external factors in the final shape of the taxonomy, like agrarian economy, as mentioned above. In this respect, the Mazatec area is subdivided into three regions: Highlands (Alta), Midlands (Media), and Lowlands (Baja).

The clusters are labeled as follows:

Cluster A: Northern Mazatec

Alta North-West: San Jerónimo Tecoatl (TE), San Lorenzo Cuanecuiltitla (LO)

North-East Baja: San Miguel Soyaltepec (SO)

Cluster B: Southern Mazatec

South-West Cañada: Chiquihuitlán (CQ)

Central Western Baja: Santo Domingo (DO), San Felipe Jalapa de Díaz (JA)

Central East Baja: San Pedro Ixcatlán (IX), San Batolomeo Ayautla (AY)

Cluster C: Central Mazatec

South-Alta: San Miguel Huautepec (SM), Santa Maria Jiotes (JI)

Central Alta: Huautla de Jiménez (HU)

South-Western Alta: Mazatlán (MZ).

Léonard *& al.* (2016) showcases various phylograms derived from the cladistic analysis initially employed to achieve this synthesis. A significant advantage of the phylogram in Figure 6 is that it clusters varieties from the northwest subgroup LO and TE with Mazatlán (MZ) on one hand and with the clade of the central highlands

subgroup (HU, JI, MG) on the other hand – a result that the dialectometric approach using the editing distance has been unable to achieve, see Heinsalu & al. (2020: 52-55). The distance of a branch from the tree's root (ancestor: the proto-language, whose characters are established upstream in the diachronic evolution stemmata) indicates the degree of retention vs innovation of a variety: LO is by far the most innovative, notably due to its full-fledged Vowel Shift (**i* > *e*, **e* > *a*, **a* > *o*, **o* > *u*, **u* > *ɯ*). Index values such as 96, 66, 53, etc., in the phylogram's branches denote the robustness of the clades (Figure 6).

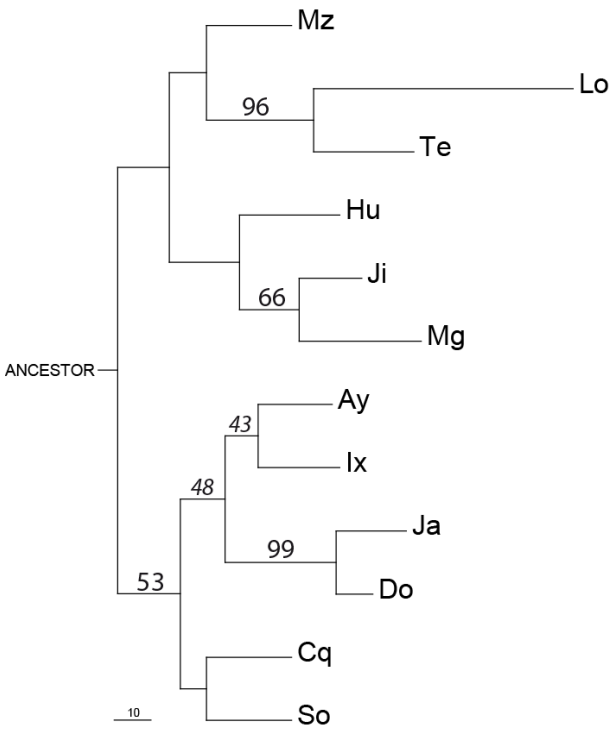
Cladistic Phylogram: phonology, pondered (Léonard *& al.* 2016, slide 43)	PAUP processing of the input data
	Branch-and-bound search completed: Score of best tree found = 477. Tree description: Optimality criterion = parsimony. Character-status summary: 329 characters are excluded. Of the remaining 138 included characters: All characters are of type 'irrev.up'. 29 characters have weight 1; 109 characters have weights other than; 2 characters are constant; 70 variable characters are parsimony uninformative. Number of (included) parsimony-informative characters = 66. Character-state optimization: Delayed transformation (DELTRAN) Tree length = 477. Consistency index (CI) = 0.6541. Homoplasy index (HI) = 0.3459. CI excluding uninformative characters = 0.4745. HI excluding uninformative characters = 0.5255. Retention index (RI) = 0.6504. Rescaled consistency index (RC) = 0.4254. We express our gratitude to Pierre Darlu for creating this phylogram and providing the technical documentation justifying its construction.

Figure 6. Cladistics of main isoglosses (Mazatec). Phonological Classification by Léonard & al. (2012) on Kirk's (1966) cognates, and cladistic phylogram for the 31 selected phonological variables (excerpt from Léonard *& al.* 2016)

This classification combines both external and internal factors and culminates in a synthesis of Gudschinsky's various classificatory series. However, similar to

Gudschinsky's stemma (1958), which aimed to achieve a single taxonomy, this taxonomy subsumes multiple intermediate classificatory stages that did not find a place in a brief article on dialectal typology.

2.5 The Mazatec Dialect Network revisited through typological phonology

Fifth classification: a team of young Mazatec linguists coordinated by the phonologist and dialectologist specializing in the Oto-Manguean domain, Mario Chávez Peón (*Proyecto nanginá*) recently conducted dialectological surveys on 16 varieties of Mazatec. This active and innovative group, both theoretically (Wagner Oviedo 2016) and methodologically, worked closely with INALI to work out a new classification of Mazatec varieties (Chávez Peón, Wagner Oviedo & Filio García 2018, 2023). This new contribution to the classification of Mazatec models (Level D) the geolinguistic variation of Mazatec in a qualitative manner, using around thirty isoglosses based on first-hand data. From the perspective of GST, the contribution of this proposal is exemplary of the impact that formalization (Level E) can have on any taxonomic task in general dialectology. Indeed, the team had to make choices on how to identify and qualify the isoglosses, conceived as typological variables, based on the language's characteristics, integrating accumulated knowledge (Level B) on the language structures as a diasystem (Level C). For example, while neither K. & E. Pike (1947) nor Gudschinsky (1956, 1958) were aware that complex syllabic constituents such as *ht, hk* («preaspirated»), *ʔt, ʔk* («preglottalized»), *nt, nk (*«prenasalized»), etc., could be interpreted (in other words, modeled –Level D) as respectively voiceless breathy, creaky, and voiced stops, or that «preaspirated» or «preglottalized» vowels were nothing but breathy and creaky vowels opposing modal vowels, symmetric to the series of nasal vs oral vowels; the Proyecto nanginá, on the other hand, integrates this typology into its description or modeling of the phonological constituents of the language (Level D), which amounts to innovatively formalizing (Level E) the diasystem. This approach leads to a vicariant description of the Mazatec diasystem –its

hierarchical classification into dialects, sub-dialects, and varieties (Level A) as well as the geolinguistic configurations of dialect chains (Level C).

The selected 31 phonological variables for delineating these zones are as follows: 1) vowel **i*, 2) vowel **æ*, 3) vowel **a*, 4) vowel **u*, 5) diphthong **aGi*, 6) Diphthong **aGu*, 7) opposition between vowels *o* and *u*, 8) opposition between vowels *e* and *æ*, 9) #V (vowels in absolute initial position), 10) laryngeal vowel (creaky), 11) whispered vowel (breathy), 12) aspirated stop, 13) glottal stop, 14) sonorant stop, 15) preaspirated stop, 16) aspirated fricative, 17) preaspirated nasal, 18) preglottalized nasal, 19) postaspirated nasal, 20) prenasalized preaspirated nasal, 21) prenasalized preglottalized nasal, 22) prenasalized series, 23) /sn/ sequence, 24) fricativization of breathy approximant **hw* into labial fricative/spirant, 25) retroflex consonants, 26) transformation of **n* into *n* (retention) or *l* (denasalization by lateralization) or *nd* (prenasalized stop), 27) transformation of **tVk*, 28) syllabic nasal, 29) tonal nasals *ñ/m*, 30) atonal nasals *ñ/m*, 31) opposition between *yu* and *ni*.

Traits 1 to 9 (T1-9) grasp Vowel Shift trends within the diasystem. Traits 10 to 24 account for the evolution of vowels and modal vs breathy or aspirated consonants and rephonologization processes (T23-24) as well as autosegmental nasal features (20-22), characteristic of the Mazatec diasystem and many Popolocan and Mixtecan languages. T25, crucial in Gudschinsky's taxonomy and salient in the speakers' minds, belongs to the family of palatalization and transphonologization phenomena of palatal contoids (alveopalatals vs. palato-alveolars). T26 is typically phonolexical, affecting a limited set of very specific but highly functional lexemes, already identified by Gudschinsky in her 1958 study: **ntihi* 'grass,' **ntiʔi* 'fire,' **ntjuhu* 'stone.' T27 corresponds to Gudschinsky's trait 6, specific to the IX area, as well as T28 of the syllabic nasal sonorant. Traits 29-30 concern nasals, similar to Gudschinsky's trait 5. T31 is morpholexical (nominal specification). This list thus synthesizes Kirk's (1966) synchronic (distributional analysis) and diachronic (comparative) approach and Gudschinsky's (1958) funnel-like isogloss taxonomy (criteria i-viii, p. 27-28) while incorporating advances in Otomanguean language typology within the framework of modern phonology.

The contribution of the taxonomy derived from this method, employing entirely new data, lies in the density of the first-hand explored network by investigators who are native Mazatec speakers, surveying the area familiar to them. This involved nearly 25 additional varieties compared to Kirk's comparative study (1966). The taxonomy[8] also takes into account recently established varieties resulting from the displacement caused by the flooding of the lowland basin in the eastern region, specifically in the areas of SO and IX.

To conclude this exercise applying GST and CASD premises to dialectal classification through Mazatec, we have observed how our senses and common sense struggle to grasp the complexity of a diasystem and its rootedness in spatial and temporal reality, in terms of information flows, self-organization, and the emergence of various types of varieties, endowed with ethnolectal –or typological, for the linguist properties. Moreover, from the standpoint of general systems theory (GST), Complexity begins with the assembly of phonemes into syllables and words (simple or complex lexemes), and scales through a range of linguistic, language faculty properties (as Universal Grammar vs local parameters), and sociocognitive components, embedded in temporality. We will now illustrate these points by applying the method to a more familiar domain, starting from a European language: Occitan (Gallo-Roman), using the THESOC database [9] and employing edit distance (a straightforward dialectometric procedure among others), on a complexity level much higher than in the case of Mazatec – a complex diasystem in miniature, as suggested by Gudschinsky (1958).

[8] The detail of dialect areas according to this description reads as follows: Zone 1: Huehuetlán, Ateixtlahuaca, Cuaunecultitla (LO), Tlacotepec, Tezonapa. Zone 2: Ocopetatillo, Acatepec, Eloxochitlán, Tecóatl, Yoloxochitlán, Zoquiapam (San Juan et Visaje), Los Cues. Zone 3: Mazatlán, Tecomavaca. Zone 4: Chilchotla (Centre, Municipalité et Barranca Seca), Huautla, Asunción, Huautepec, Tenango (Centre et Cerro Central). Zone 5: Independencia, Tilpam, Tezonapa, Pochota. Zone 6: Ayautla. Zone 7: Chiquihuitlán. Zone 8: Jalapa (majeur et jeune), Santo Domingo. Zone 9: Ixcatlán, Nuevo Ixcatlán. Zone 10: Temascal, Pescadito, Las Margaritas, Acatlán.

[9] Special thanks here to Guylaine Brun-Trigaud, from UMR 7320: Bases, Corpus, Langage, Nice. Guylaine Brun-Trigaud contrived the appropriate list of 71 cognates available in all atlas linguistics of the NALF and identified the corresponding phonological variables.

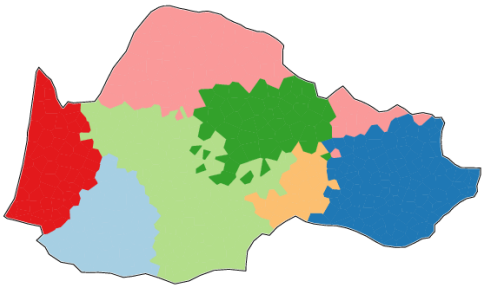
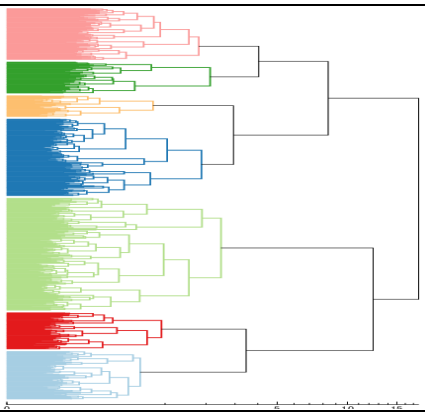
3. The Occitan Dialect Continuum

The paradigm of *Complex Adaptive Dynamical Systems* (CADS) encompasses a wide range of quantitative technical solutions for the multidimensional examination of emerging phenomena in both natural and social sciences. The works on Mazatec employing this approach fall within the emerging paradigm of *Language Dynamics* (Wichmann, 2008). We are now going to apply GST and CADS to a more familiar domain of application, well known to linguists and dialectologists: The Occitan dialect network (see Brun-Trigaud, in the 1st issue).

3.1 Joe Ward's Method (hierarchical cluster analysis)

Given that the diasystem we are about to approach from the perspective of CASD & GST is widely known in the fields of Romance Language Classification (RLC) and World Language Classification (WLC), we will only briefly mention the state of the art in RLC regarding the Occitan domain –a subset of the Romance subfamily of Gallo-Romance languages, alongside the domains of Oïl and Franco-Provençal.

We will immediately apply an algorithm integrated with edit distance, such as the Gabmap processing using the R software base: hierarchical agglomerative clustering, known for its congruence with canonical classifications in dialectology (Map 3; Figure 7).

Occitan, THESOC, Phonology, 71 items, WM, 8 classes: map. Database by Guylaine Brun-Trigaud.	Occitan, THESOC, Phonology, 71 items, WM, 8 classes: taxonomy. Database by Guylaine Brun-Trigaud.
	
Map 3. Occitan, THESOC, Phonology, 71 items, WM, 8 classes: map. Database by Guylaine Brun-Trigaud	Figure 7. Occitan, THESOC, Phonology, 71 items, WM, 8 classes: taxonomy. Database by Guylaine Brun-Trigaud. Gabmap (edit distance)

The implementation of a canonical taxonomy (level F), such as provided by the tool of the Joe Ward's Method or Hierarchical Ascendant Classification, consists of two flat hierarchies in equipollent relation: on the one hand, a north-Occitan macro-dialect with an oblique orientation (NW-SE), commonly referred to as North Occitan, but which nonetheless includes the Provençal subcomponent in all its complexity (Rhodanian, Maritime, Alpine); on the other hand, another macro-dialect, Center-Occitan, which integrates the Languedocian on one hand against the Gascon-Languedocian South-Occitan complex on the other. Here, we encounter Gudschinsky's « Great Division » at the initial nodes of the dendrogram, which is nothing more than a taxonomic tree based on three trivial operations (addition, deletion, modification of a segment), akin to a spell checker. However, this implementation is robust: it relies on 662 points of inquiry from linguistic atlases where Occitan was spoken at the time of the survey (1950-80), processed with Gabmap 71 items of phonological word forms through string-edit distance,[10] generating 44748 Instances or edit distance tests; on

[10] List of lemmas used for the treatment of phonological isoglosses in the THESOC corpus: *bee, magpie, lamb, to lamb, needle, tree, wheat, beef, quail, hat, chimney, shirt, horse, chestnut tree, goat, sky,*

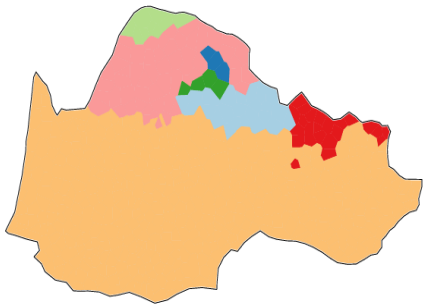
245029 characters, among which 99 unique characters; tokens 240562, among which 203 unique tokens.

3.2 Group Average: a reductionist, yet heuristic quantitative tool

The Occitan diasystem's configuration mirrors the division between highlands and lowlands, with the Massif Central playing a pivotal role. This volcanic complex contrasts with the Alps, influencing the partition between northern Occitan and the Languedoc. Limousin and Provençal mark its northern and southeastern boundaries, while Gascon sub-dialects flank Languedoc. This hierarchical structure, akin to Mazatec's vertical archipelago, suggests a self-organizing network. The Group Average algorithm, unlike Ward's, simplifies hierarchical groupings but obscures « invisible dialects » (Nerbonne & Kretzschmar 2003). Gabmap's simulations illuminate the Occitan dialectal network's complexity, showcasing the interplay of homeostasis (cf. superordinate II in Figure 1 above) and geographic nodes. Group Average acts as a centrifuge, revealing honey-like layers alongside granulated relief out of its spinning, reflecting accrued dynamics and contingencies in singular points and areas within the geolinguistic web or thread of diasystemic patterns. Group Average is particularly good at detecting these "invisible (sub)dialects" (Map 4; Figure 8).

---

scissors, key, neck, knife, thigh, butt, sheet, dice, water, staircase, fire, leaf, gall, thread, liver, hay, make, pitchfork, cold, knee, sheaf, acorn, wasp, wool, milk, lye, hare, moon, honey, fly, mule, ripe, blackberry, nest, walnut tree, eye, egg, goose, bird, stone, to rain, to fold, meadow, well, sun, supper, to sweat, soot, cow, calf, wind, viper, donkey, ladder, star. These lemmas cover the entire range of phonological isoglosses one can expect from this Gallo-Roman domain. Here is the list, just as previously done for several lists of phonological criteria relevant to the diasystem under scrutiny: Pretonic drop, tonic, -AL, -AL final, -arb/-amb, -ARE, -ARIU, -aticum, -ATUS, AU-, -B-, b/B, BL-, CA-, -CA-, -CE- (s/z), CL-, final -k, final -P, final -R, final -S, final -T, -CT > ch/jt, CU-, -D-, -DIA-, diphthong A + yod, diphthong E + L, diphthong gd, diphthong I + L, diphthong O, diphthong O + k, diphthong O + j, diphthong O + L, diphthong O + R, diphthong O + V, diphthong O + yod, diphthong U + L, -ELLUM, F-, -F-, feminine plural, form/frum-, -G-, g/ɣ, -GN-, group ch -dr-, -IC(U)L(US), -IC(U)LA, -js final, L-, -L-, -LI-, masculine plural, -MB-, metathesis, metathesis k-br/kr-b, -N-, nasal ending, nasalization ending, PL-, -QU-, -RBR, s- (s/ch), -SK, -sk-, -ST-, -T-, tg, -TR-, treatment of U, final -tz, U + L final, V-, final feminine vowel -e, final masculine vowel, tonic vowel, -z-. Both the selection of items and the description of variables for historical phonology have been provided by Guylaine Brun-Trigaud – great thanks to her again.



Map 4. Group Average, 7 diasystemic classes

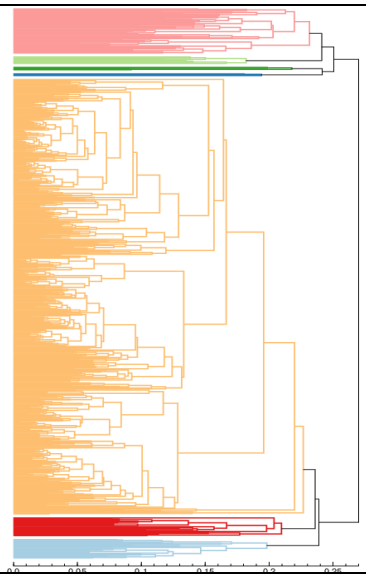


Figure 8. Group Average, *idem*, dendrogram

In this vicarious grasp of the Occitan geolinguistic and taxonomic space, an increasingly asymmetrical equipollence emerges: Northern Occitan narrows down to a highly composite mass, now consisting of a Northern Crescent, a Grand Limousin in the center, and, to the southeast, a densely concentrated double hub of Auvergnat varieties. Plot twist: the entire Central and Southern Auvergnat regions now join as external members to the grand central-southern clade. Meanwhile, the Vivaro-Alpine, situated north of Provençal in the northeast, interposes as an external member of the massive central bloc, confederating a chain of "heavy" dialects such as Provençal, Languedocian, and Gascon from East to West.

Considering the two typologies (Mazatec and Occitan), it seems to us that the heterogeneity of Northern Occitan mirrors that of high Mazatec, along with its highly innovative Northwestern subgroup – much like the Limousin dialect, more homeostatic in nature to the west compared to the more entropic Auvergnat to the east. However, the real revelation lies in the presence of three "invisible sub-dialects" of Auvergnat, distributed between two macro-areas, in accordance with the principle of quasi-componentiality (level C): an innovative twofoold hub, functioning like

beehives producing two highly concentrated varieties on one hand, countered by a buffer zone associated in a cascade of external aggregation. A Southern Auvergnat, aggregating to the Vivaro-Alpine, itself an external member (outlier) of the flat-hierarchy (level A) «Greater Southern Occitan» block.

The diagnosis stemming from the comparison of two entirely disparate diasystems, belonging to language families or phyla in no way related, is characteristic of an approach in terms of GST: all things considered (space and time), Mazatec and Occitan, on different scales, can be described (level D) in terms of geolinguistic hierarchy (level A) as complex systems, roughly following similar trends, akin to a compact "vertical archipelago" –if only for its unique typology in Romance languages, much like the Mazatec in its "Popolocan bath", not to mention "Popotecan". Moreover, considering the principle of quasi-componentiality of CADS, it remains uncertain whether Occitan is confined to what the maps above show or whether it should encompass a southern "leg" by incorporating Catalan.[11]

3.3 When Classification (also) Helps for Data Mining

Even better, following the principle of uniformity in GST, there is nothing preventing us from «testing» the heuristic properties, in terms of relevant variables for describing (level D) dialect variation, of the «Guschinsky Model», with its nine parameters (i to ix above). We hardly have the space to do so here, but let's mention a few avenues, heuristically: we have just seen two vicarious representations of variable (i) "major divisions" underlying the dialect hierarchy, from a classificatory standpoint. In line with what we proposed earlier, artificial intelligence could learn to rank the variables as well as the singular behaviors (occurrences) and the details of typological parameterization of the corpus used here (Table 2).

[11] Our personal standpoint is that this Catalan « leg » is neither proved empirically, nor needed theoretically for LDC. In the glottopolitical sphere, it has been the topic of much debates between activists since Frédéric Mistral, head of the Felibrige mouvement supported this idea at the beginning of the past century, but this does not concern LDC as a *technê* (τέχνη) or a science. See Rafanell (2006) about this issue, from a glottopolitical standpoint.

	Neogrammarian scheme (i)	Bartolian scheme (ii)	Lexical diffusion (iii)	Polymorphism (iv)	
	1	**2**	**3**	**4**	**5**
Group Average, 8 classes	'agassa' (magpie)	'foie' (liver)	'puits' (well)	'bœuf' (ox)	'feu' (fire)
Croissant Lim					
CLW_ALAL 66	*ʒˈasᵊ*	*fˈœdʒᵊ*	*pˈu*	*bˈø / bˈy*	*fˈø*
CLC_ALAL 35	*aʒˈas*	*fwˈɛ*	*pˈu*	*bˈø*	*fˈœ*
CLW_E_ALAL 23	*aʒˈas*	*fwˈɛ*	*pwˈi*	*bˈø*	*fˈø*
AUV					
AUV_S_ALAL 10	*dzˈasɔ*	*fødzˈø*		*bˈœᵊ*	*fjˈo / fju*
AUV_N_ALAL 11	*dzˈasɔ*	*fødzˈø*	*pwˈɔ*	*bjˈy*	*fjˈa*
Lim_C					
ALAL 48	*ðˈaʃɔ*	*fɛðˈɛ*	*pˈu*	*bjˈɔ*	*fjˈɔ*
ALAL 25	*dzˈasɔ*	*fœdʒˈi*	*pˈu*	*bʌˈɔº / bjˈɔº*	*fˈɛ*
Buf_Zone_N					
BZ_W_ALMC 20	*dzˈasa*	*fˈidʒə*	*pˈɔw*	*bjˈew*	*fjˈɔ*
BZ_E_ALP 106	*agˈasœ*	*fˈedʒe*	*pˈus*	*bˈyw*	*fwˈœ*
LGDC_W					
ALLOc 31.12 (Tls)	*aɣˈasɔ*	*fˈetʃe*	*pˈuts*	*bjˈɔw*	*fjˈɔk / fˈɔk*
ALLOc 11.02	*agˈasɔ*		*pˈutʃ*		

Table 2. Indexation « à la Gudschinsky » of some heuristic spots in the Northern vs Central dendremes and choremes, through Group Average detection of entropic *quasi-sub-components* (level C) of the Occitan Diasystem

For example, item 1 *agassa* (< Gothic "agasa") 'magpie' has several differential properties: this lemma is a complex component of the lexicon and, consequently, for the phonology of the language: a flat melody with a low vowel on the entire template, a voiced velar occlusive intervocalic, an accented vowel, an intervocalic fricative, and an unstressed final vowel. There is room here for no fewer than five « neogrammarian » laws of a high hierarchical level. Item 2 is much more intricate in terms of evolution (level B), as it contains an etymological vowel melody *i-a-u* (< Latin "FICATU" 'liver'), making the expected treatments more unpredictable (criterion iii). Items 3 to 5 are highly intertwined and seem to conceal inferential traits from underlying Gallo-Roman strata or adstratum (the "Oïl" type dominates in the Crescent for items 4 and 7), suggesting the application of Gudschinsky's criterion *vii*, but integrating the Oïl *superstratum* here. Algorithmic vicariance (Group Average instead

of Ward's Method) has brought forth "invisible dialects" instead of "eponymous dialects."

4. Discussion

In summary, in the foregoing, the sole implementation (level F) of an algorithmic solution (level E) based on a reductionist input mode such as Group Average has allowed the emergence of nodes of entropy in an ocean of negentropy (levels C and D of the graph in Figure 1) in the northwest of the Occitan domain, with two hotspots (two «hubs» or «individualization niches») in the center of what is usually considered as canonical Auvergnat. This description of the internal structure of the large-scale diasystem provided by this algorithm allows the extraction of initial data and the verification of typological parameter effects on the scale of these invisible entropic areas, beyond maps provided by heuristic tools. The geocomputational implementation thus urges the classifying linguist to a more attentive and targeted return to upstream data, making choices in terms of input, as indexed earlier in Table 2, where the detection of strongly innovative and diversified "hot zones" allows a properly linguistic analysis of the data. In this regard, GST and CADS are tasked with finding appropriate tools to explore complexity holistically, objectifying the hierarchies that manage the surface or underlie the diversification of dialects, sub-dialects, and varieties –all quasi-subcomponents of the complex system that any diasystem is.

Hence, we discern the potential contribution of GST to dialect classification: CADS emerges as a transdisciplinary metatheory, combining formalization (level E) and implementation (new level F) of descriptive models (level D) to assemble the ordered components (ranging from level C to A) of any complex system, whether in the realm of the living, the semiotic order, as much as in physics, chemistry, or sociology. This intersection between material sciences and human and social sciences harnesses and applies mathematical and computational methods (algorithms) for the quantitative treatment of data of all orders. Here, from the perspective of this approach to dialectal

classification, the method proves to meet the criterion of uniformitarianism and descriptive power. A significant lesson from the foregoing, however, is that GST and CADS are also useful as interpretive frameworks to re-examine past and present dialectologists' taxonomies, with the aim of contributing to the ultimate horizon, epistemologically speaking, which is the General Taxonomy of Semiotic Systems (GTSS), in terms of CADS. Such a transition, from the General Taxonomy of Languages, the object of the quest of language classification specialists, towards GST and CADS, by reconnecting with the fine grain of data and typological knowledge currently in the hands of linguists, would move towards a Critical Taxonomy of World Languages – another step towards General Dialectology.

In no way would the Occitan dialectal network "function" like that of Mazatec, however, due to differences in territorial scale and the number of locolects. Yet, as a compact vertical archipelago, in GST terms, it is as if the Mazatec almost simulates, a posteriori, a fractal opposing high-middle-low ecohuman zones –a kind of Otomanguean "Gévaudan".[12]

The systemic approach presented here is particularly well-suited to explore these spaces, structurally irreducible though they may be, in terms of the method for capturing the constitutive and evolutionary mechanisms of a diasystem. *General Systems Theory* and *Complex Adaptive Dynamical Systems Theory* then become both a paradigm relevant to *Language and Dialect Classification*, for mutual benefit, aiming at the heart of *General Linguistics*.

[12] In the conclusion of his essay on Gévaudan linguistic geography, Charles Camproux (1962: 759-775) describes Gévaudan as somewhat a kind of compact vertical archipelago, structured in a manner analogous to what we have seen in Mazatec, following a mountain/plains & valleys division, and above all, through settlement patterns "along the waterways," in river basins. For Camproux, this « au fil de l'eau » circularity of human aggregates through history represents a universal scheme of dialectal diversification, determined by physical geography. All further determinisms, among which human geography (feudalism, realms and governments) just adapt to these fundamental constrains of anchoring people in physical places on the maps. This material substrate of linguistic diversity in space and time makes it even more profitable for the methods of GST and CADS. From this standpoint, Camproux' remarks are fully welcome.

References

ADAMOU, Evangelia (2021) "The Popolocan languages", in Soeren Wichmann (ed.), *Languages and Linguistics of Mexico and Northern Central America: A Comprehensive Guide,* Berlin: Mouton de Gruyter. In press. halshs-03153566.

ARIANO CIFUENTES, Francisco Javier, Montserrat ESTEINOU DÁVILA, Jéssica América GÓMEZ FLORES, Kristie RODRÍGUEZ PÉREZ-ABREU (2014) "Bilingüismo: Situación actual del mazateco frente al español", *Fuentes Humanisticas, Lenguas Amerindias y Lingüistica*, 26-49, 35-45.

BELMAR, Francisco (1892) *Ligero estudio sobre la lengua mazateca,* Oaxaca: Imprenta del Comercio.

BERTALANFFY, Ludwig von (1968) *General System Theory*, New York: George Braziller.

BOEGE, Eckart (1988) *Los mazatecos ante la nación. Contradicciones de la identidad étnica en el México actual,* México: Fundo de Cultura Económica.

BRINTON, Daniel Garrison (1892) "On the Mazatec Language of Mexico and its affinities", *Proceedings of the American Philosophical Society,* 30-137, 31-39.

BRUGMANN, Karl (1884) Zur Frage nach den Verwandtschaftsverhältnissen der indogermanischen Sprachen, *Internationale Zeitschrift für allgemeine Sprachwissenschaft* 1.253.

BRUN-TRIGAUD, Guylaine (2023) "Gallo-Roman dialect classifications", in G. Aurrekoetxea, A. Ensunza, J. Škofic & H. Van de Velde (2022) "An introduction to dialect classifications in Europe", *Dialectologia*, Special issue, X, 127-154. <https://www.edicions.ub.edu/revistes/dialectologiasp2022/>

CASAD, Eugene (1974) *Dialect Intelligibility Testing,* Summer Institute of Linguistics of the University of Oklahoma.

CHÁVEZ PEÓN, Mario E., Carlos de Jesús WAGNER OVIEDO & Israel FILIO GARCÍA (2018), *Isoglosas fónicoléxicas para la dialectología de las lenguas mazatecas,* México: UNAM et INALI.

CHÁVEZ Peón, Mario, Eloy GARCÍA GARCÍA & Israel FILIO GARCÍA (2023) *Kjᵾgátúxuma xi tinchjáya kús'ín kᵾgᵾ'indú 'iénná ('ien Chíntjie), Norma de escritura del mazateco*, México: UNAM and INALI.

CIFUENTES, Barbara (2002) *Lenguas para un pasado, huellas de una nación*. *Los estudios sobre lenguas indígenas de México en el siglo XIX*, México: INAH, Plaza y Valdés.

CAMPROUX, Charles (1962) *Essai de géographie linguistique du Gévaudan,* Paris: Presses Universitaires de France.

DARLU, Pierrre; Pascal Tassy (1993) *Reconstruction phylogénétique. Concepts et Méthodes*, Paris: Masson. <http://sfs.snv.jussieu.fr/publications_sfs.shtml>

DYEN, Isidore (1953) "Review of 'Malgache et Maanjan: une comparaison linguistique' by Dahl Otto Chr. 1951", *Language*, 29, 577-590.

FERNÁNDEZ DE MIRANDA, Maria Teresa (1951) "Reconstrucción del potopopoloca", *Revista Mexicana de Estudios Antropológicos*, 12, 61-93.

GAILLARD-CORVAGLIA, Antonella, Jean-Léo LÉONARD & Pierre DARLU (2007) "Testing cladistics on dialect networks and phyla (gallo-romance and southern italo-romance)", in John Nerbonne, T. Mark Ellison, Grzegorz Kondrak (eds.), *Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*, Prague: Association for Computational Linguistics, 23-30. <https://aclanthology.org/W07-1304.pdf>

GAILLARD-CORVAGLIA, Antonella, Jean Léo Léonard & Pierre Darlu (2008) "Approche cladistique des vocalismes d'oïl à partir des données de l'A.L.F.", *Bollettino dell'Atlante Linguistico Italiano*, III-32, 36-78.

GAILLARD-CORVAGLIA, Antonella (2012) *De la cladistique à la linguistique. Une étude appliquée aux dialectes italo-romans méridionaux et salentins*, Alessandria: Edizioni dell'Orso. <file:///C:/Users/jleonar3/Downloads/2009PA030092.pdf>

GONZÁLEZ CASANOVA, Pablo (1925) "Los idiomas popolocas y su clasificación", *Anales del Museo de Arqueologia, Historia y Etnografia*, 4-3, 497-536.

GROS, Claudius (2015) *Complex and Adaptive Dynamical Systems. A Primer,* 4th ed., New York: Springer.

GUDSCHINSKY, Sarah Caroline (1953) "Proto-mazateco", *Memoria del Congreso Cientifico Mexicano*, 12, México: Universidad Nacional Autónoma de México, 171-174.

GUDSCHINSKY, Sarah Caroline (1955) "Lexico-statistical skewing from dialect borrowing", *International Journal of American Linguistics*, 21-2, 138-149.

GUDSCHINSKY, Sarah Caroline (1956) *Proto-Mazatec Structure,* Master of Art Thesis, manuscript, Pennsylvania University.

GUDSCHINSKY, Sarah Caroline (1958) "Mazatec dialect history", *Language*, 34, 469-481.

GUDSCHINSKY, Sarah Caroline (1959) “Proto-Popotecan. A Comparative Study of Popolocan and Mixtecan”, *International Journal of American Linguistics*, special issue**,** 25-2.

Hamp, Eric (1958) “Protopopoloca Internal Relationships”, *International Journal of American Linguistics***,** 24-2, *Franz Boas Centennial Volum*e, 150-153.

HASPELMATH, Martin (2001) *Indefinite Pronouns*, Oxford: Oxford University Press. Available on <https://academic.oup.com/book/27086/chapter/196428934>

HEINSALU Els, Jean Léo LÉONARD & Marco PATRIARCA (2020) *Languages in Space and Time. Models and Methods from Complex Systems Theory*, Cambridge: Cambridge University Press.

JAMIESON Carole (1982) “Conflated subsystems marking person and aspect in Chiquihuitlán Mazatec verb”, *International Journal of American Linguistics***,** 48, 139-167.

KIRK, Paul Livingston (1966) *Proto-Mazatec Phonology.* Ph. D. dissertation, University of Washington, UMI.

KIRK, Paul Livingston (1970) “Intelligibility Testing: The Mazatec Study”, *International Journal of American Linguistics*, 36-3, 205-211.

LADEFOGED Peter & Ian MADDIESON (1996) *The sounds of the world’s languages*, Oxford: Blackwell.

LEHMAN, Walter (1920) *Zentral-Amerika: Die Sprachen Zentral-Amerikas in ihren beziehungen zueinander sowie zu Sud-Amerika und Mexiko*, Berlín: Museum für Völkerkunde.

LÉONARD, Jean Léo (2020) “Flux de variation, auto-organisation, émergence: éléments d’écologie diasystémique”, in Razky Abdelhak, Inés Sfar, Olivier Soutet & Salah Mejri (eds.), *A variação nas línguas: universais compartilhados e idiomaticidade dinâmica / De la variation dans les langues : universaux partagés et idiomaticité dynamique*, Cagua: Letraria, 49-88.

LÉONARD, Jean Léo; Vittorio DELL’AQUILA (2014) “Mazatec (Popolocan, Eastern Otomanguean) as a Multiplex Sociolinguistic ‘Small World’”, in Urmas Bereczki (ed.), *The Languages of Smaller Populations: Risks and Possibilities. Lectures from the Tallinn Conference, 16–17 March, 2012*, Tallinn: Ungarian Institute’s Series, Miscellanea Hungarica <Hal-01137372>, 27-55.

LÉONARD, Jean Léo, Antonella GAILLARD-CORVAGLIA, Daniele DALMASSO & Vittorio DELL’AQUILA (2012) “The ALMaz (Atlas Lingüístico Mazateco): from geolinguistic data processing to typological traits”, *STUF - Language Typology and Universals*, 65-1, 78-94. <https://doi.org/10.1524/stuf.2012.0006>

LÉONARD, Jean Léo, Antonella GAILLARD-CORVAGLIA, Daniele DALMASSO & Vittorio DELL'AQUILA (2014) "Émergence forte versus émergence faible en écologie linguistique : le cas du mazatec", *Actes du colloque 2014 de l'Institut Universitaire de France*, Dijon, 27-28 mai, nature et culture, in Patrick Bonin & Thierry Pozzo (eds.), *Nature et Culture*, Saint-Étienne: Publications de l'Université de Saint-Étienne, 77-102.

LÉONARD, Jean Léo & Alain KIHM (2012) "Classes flexionnelles du mazatec et diasystème. Empirisme critique et formalisation", *Bulletin de la Société de Linguistique de Paris*, 56, 379-446. ⟨hal-01137254⟩.

LÉONARD, Jean Léo & Alain KIHM (2014) "Mazatec verb inflection: Revisiting Pike (1948) and comparing four dialects", in Jean Léo Léonard & Alain Kihm (eds.), *Patterns in Mesoamerican morphology*, Paris: Michel Houdiard Editeur, 26-76.

LÉONARD, Jean Léo & Julien FULCRAND (2015) "Tonal Inflection and dialectal variation in Mazatec", in Enrique Palancar & Jean Léo Léonard (eds.), *Tone & Inflection, Trends in Linguistics*. Studies and Monographs, 296, Berlin: Mouton de Gruyter, 165-195.

LÉONARD, Jean Léo, Anirban CHAKRABORTI, Antonella GAILLARD-CORVAGLIA, Els HEINSALU, Kiran SHARMA, Marco PATRIARCA & Pierre DARLU (2016) "Modular algorithmic complexity applied to the Mazatec diasystem", in Jean Léo Léonard, Didier Demolin & Karla Janiré Avilés González (eds.), *Proceedings of the International Workshop on Structural Complexity in Natural Language(s), Paris, 30-31 May 2016,* Paris: Paris 3 University - Labex EFL (PPC11), available on <http://axe7.labex-efl.org/node/353>.

MECHLING, William H. (1912) "The Indian linguistic stocks of Oaxaca, Mexico", *American Anthropologist*, 14-4, 643-682.

NERBONNE, John & William KRETZSCHMAR (2003) "Introducing Computational Techniques in Dialectometry". *Computer and the Humanities*, 37-3, 245-255.

PIKE, Kenneth, Eunice Pike (1947) "Immediate constituents of Mazatec Syllables", *International Journal of American Linguistics*, 13, 78-91.

PIKE, Kenneth (1948) *Tone Languages. A Technique for Determining the Number and Types of Pitch Contrasts in a Language, with Studies in Tonemic Substitution and Fusion*, Ann Arbor: University of Michigan Press.

PIMENTEL, Francisco (1875) *Cuadro descriptivo y comparativo de las lenguas indígenas de México*, vol. II, México: Imprenta de Andrade y Escalante.

RAFANELL, August (2006) *La il·lusió occitana: la llengua dels catalans, entre Espanya i França*, Barcelona: Quaderns crema.

SIMON, Herbert A. (1962) "The Architecture of Complexity", *Proceedings of the American Philosophical Society*, 106-6, 467-482. Available on <https://www2.econ.iastate.edu/tesfatsi/ArchitectureOfComplexity.HSimon1962.pdf>.

SOUSTELLE, Jacques (1937) *La famille Otomí-Pame du Mexique central*, Paris: Institut d'Ethnologie.

STARR, Frederick (1900) "Notes upon the ethnography of southern Mexico", in *Proceedings of the Daweport Academy of Sciences VIII*, Stanford: Center for the Study of Language and Information, 102-198.

SWADESH, Morris (1959) *Indian Linguistic Groups of Mexico*, Mexico: *Instituto Nacional de Antropología e Historia &* Escuela Nacional de Antropología e Historia.

WAGNER OVIEDO, Carlos de Jesús (2016) "Desarrollo diacrónico de la complejidad consonántica en las lenguas mazatecas, COLOV 7". <https://www.academia.edu/24238492>

WEINREICH, Uriel (1954) "Is a structural dialectology possible?", *Word*, 4, 388-400.

WICHMANN, Søren (2008) "The Emerging Field of Language Dynamics", *Language and Linguistics Compass* 2-3, 442–455. DOI: 10.111/j.1749-818x.2008.00062.x