

# კვლევითი ანგარიში RESEARCH REPORT

Jean Léo Léonard  
Sorbonne Université

## Le projet LaDyCa (Language Dynamics in the Caucasus, IDEX Emergence 2017-18): avancées et résultats

Mots-clés: *Caucase, Géorgie, Kartvélien, Géorgien, Abkhaze, Morphologie, Phonologie, Modélisation, Conjugaison, Complexité, Ecologie linguistique, Dialectologie, Dialectométrie, Algorithme (de Levenshtein)*

### Introduction

Nous présenterons ici les grandes lignes et quelques avancées et résultats d'un projet international en linguistique théorique, descriptive et appliquée, réalisé en étroite collaboration avec l'Université d'Etat Ilia à Tbilissi, pour la période 2017-18. Dans la mesure où cette collaboration est appelée à se pérenniser, étant donné les enjeux d'un très grand intérêt bilatéral pour les échanges franco-géorgiens, mais aussi en raison de la contribution remarquablement efficace et productive de l'université géorgienne partenaire, nous avons accepté avec plaisir de synthétiser les objectifs et les résultats de cette initiative de recherche par un article qui réalise un compromis entre les exigences d'un rapport scientifique pour spécialistes (notamment linguistes et spécialistes de systèmes complexes ou de théorie de la complexité), et un compte-rendu des activités et des résultats lisible par un public plus ample.

Cette double contrainte s'impose d'autant plus que le projet LaDyCa a commencé avec un objectif somme toute plutôt théorique et méthodologique, pour intégrer de plus en plus, sur le terrain, la dimension d'un travail social et d'un partenariat pour le transfert de compétences et de données entre les milieux universitaires français et géorgiens. Il nous est aussi apparu qu'une

telle orientation mixte – recherche fondamentale et appliquée – s'avérait la plus fructueuse, la plus viable et la plus éthique, dans une perspective à moyen et long terme. Nous avons aussi été frappés par la convergence à tous les niveaux de ces deux centres d'intérêt dans une perspective de continuité avec un important précédent, pourtant réalisé dans une aire géographique très éloignée (la Méso-Amérique). Preuve, si besoin est, que les sciences humaines s'inscrivent dans une dynamique universelle, globale, de production de connaissances, de méthodes et de solutions pour l'avenir et le devenir de l'humanité et de sa relation équilibrée à l'environnement, comme nous allons le voir. Des méthodes et des savoir-faire acquis en contexte mésoaméricain, dans le „Nouveau Monde“ s'avèrent probants et fructueux lorsqu'on les transpose dans „l'Ancien monde“, tout en se régénérant aux avancées technologiques globales, comme les nouvelles méthodes de dialectologie quantitative et de traitement automatique de la parole, ou la Théorie de la Complexité, ajoutant à la transversalité géographique les atouts de l'interdisciplinarité scientifique. Ce qui va suivre peut donc se lire comme le récit de la construction d'un pont intercontinental (de la Méso-Amérique au Caucase), intercommunautaire (de l'UE à la Géorgie et la Transcaucasie), et interdisciplinaire (de la Théorie de la Complexité aux sciences appliquées). En cela, le projet LaDyCa est héritier de la tradition d'Alexander von Humboldt, fondée sur les *analogies* et les *contrastes* à grande échelle entre observables (Andes, Altaï, Oural; steppes d'Asie centrale et pampa sud-américaine, etc.), dans une visée d'écologie des systèmes<sup>1</sup>.

### 1.1. Description du projet

Le projet LaDyCa (*Language Dynamics in the Caucasus*) est un projet de l'IDEX Emergence<sup>2</sup> de nature interdisciplinaire (linguistique, mathématiques, informatique, histoire et géographie), financé dans le cadre de la fusion entre les universités de Paris 6 (Pierre et Marie Curie, faculté de sciences et de médecine) et Paris 4 (Sorbonne Université, faculté de lettres). Dirigé par l'auteur de ces lignes pour la partie française, il a été rendu possible par le partenariat avec l'Université d'Etat Ilia à Tbilissi, avec Mme le professeur Tamar Makharoblidze comme partenaire géorgienne. Le présent article a été rédigé à l'aimable demande de la revue *Kadmos*, et tient lieu de rapport scientifique de cette initiative internationale, à un mois de la clôture du projet (février 2017-juin 2018).

1 Cf. Wulf Andrea, *The Invention of Nature. The Adventures of Alexander von Humboldt, The Lost Hero of Science* (Londres: John Murray, 2015).

2 Les projets IDEX (Initiative d'Excellence) financent, selon les termes du ministère „des projets de recherche de haut niveau présentant une prise de risque importante avec une approche créative et originale“.

Le thème central de LaDyCa est l'écologie linguistique dans le Caucase, avec la Transcaucasie comme zone centrale, et le Caucase septentrional comme deuxième cercle d'étude. Dans la première région, les recherches ont porté sur les langues kartvéliennes, avec une approche qui optimise certains outils de traitement de la diversité linguistique et dialectale (comme l'algorithme de Levenshtein, les modèles de Markov ou *Hidden Markov Models*, l'analyse PFM<sup>3</sup> de la flexion verbale), tandis que dans la deuxième région, elles se sont limitées à des avancées exploratoires (principalement des enquêtes phonologiques, sur la base de divers protocoles). La coopération avec la Géorgie, grâce à l'Université d'Etat Ilia de Tbilissi, mais aussi grâce à la collaboration de collègues de nombreuses institutions universitaires géorgiennes, a permis d'aboutir à des résultats concrets et encourageants, dont nous allons rendre compte dans le présent article. La contribution de Gilles Authier, directeur de recherches à l'EPHE (Ecole Pratique des Hautes Etudes, Paris), a également été d'un grand secours, à partir de mars 2018, permettant d'accéder à une langue nakh-daghestanienne – le dargi. Le projet se situe, du point de vue de sa problématique générale d'écologie linguistique (et, plus précisément, d'*écologie diasystémique*)<sup>4</sup>, dans la continuité d'une précédente initiative analogue, menée dans le cadre de l'IUF (*Institut Universitaire de France*), en 2009-14, sur les langues de Méso-Amérique: le MAmP (*MesoAmerican morphoPhonology*)<sup>5</sup>, qui utilisait également des méthodes quantitatives (dialectométrie) pour l'étude des diasystèmes d'une part, et le modèle PFM pour l'analyse de langues à morphologie complexe, d'autre part, mais n'avait pas eu recours aux chaînes de Markov pour la segmentation automatique à partir de corpus oraux<sup>6</sup>.

3 *Paradigm Function Morphology*: théorie de la flexion dans les langues du monde développée par Gregory Stump, Olivier Bonami, Raphael Finkel, etc. Voir notamment Stump, *Inflectional Morphology. A Theory of Paradigm Structure*, 2001 et Bonami and Stump, "Paradigm Function Morphology", 2016.

4 On doit à Uriel Weinreich ("Is a structural dialectology possible?", 1954), ce concept heuristique de *diasystème*, qui se réfère à un *métasystème* incluant les différentes variétés dialectales constitutives d'un *continuum dialectal*: loin d'être un système homogène, la langue saussurienne intègre des variétés géo – ou sociolectales, qui entretiennent un rapport de corrélations mutuelles sur les différentes fonctions structurales qui définissent l'inventaire des unités et leurs régimes combinatoires. Un diasystème peut se décrire également, en termes non plus structuralistes, mais générativistes, comme une grammaire polylectale (cf. Le Guern *et al.*, *Principes de grammaire polylectale*, 1983). L'écologie diasystémique étudie comment cet hologramme structural polylectal qu'est, *a posteriori*, le diasystème, est en partie déterminé par des facteurs externes d'écologie des échanges et des flux de populations dans un espace habité.

5 Ce projet est décrit en détail sur la page Internet <http://jll.smallcodes.com/home.page>, où l'on trouvera également le rapport d'activité (en anglais) rendant compte des résultats.

6 Au sujet de cette technologie, voir en anglais [https://en.wikipedia.org/wiki/Hidden\\_Markov\\_model](https://en.wikipedia.org/wiki/Hidden_Markov_model) et, en français [https://fr.wikipedia.org/wiki/Mod%C3%A8le\\_de\\_Markov\\_cach%C3%A9](https://fr.wikipedia.org/wiki/Mod%C3%A8le_de_Markov_cach%C3%A9).

Dans une logique cyclique d'interfécondation, ce projet mésoaméricain s'était largement inspiré lui-même d'hypothèses et de résultats obtenus dans le cadre de recherches d'écologie linguistique ou d'ethnohistoire caucasienne, notamment l'article de Johanna Nichols<sup>7</sup>, qui a suggéré le terme générique de „linguistique des milieux alpins“ (*Alpine Linguistics*) et celui de „verticalité“ qui, inspiré de la notion de John Victor Murra d'*archipel vertical*<sup>8</sup>, se donne pour objectif d'étudier dans une perspective d'écologie historique les mouvements d'implantation, de transplantation et de diversification des langues du monde sur plusieurs étages environnementaux: les hautes terres échangent continuellement des ressources et des populations avec les terres de moyenne altitude (*piémonts*) et les basses terres ou les littoraux. A ce titre, les langues kartvéliennes sont indéniablement un observatoire privilégié d'une telle dynamique, jusque dans la répartition actuelle des langues qui composent cette famille, selon une distribution verticale, du svan et de ses dialectes en milieu „alpin“ (ou montagnard), aux langues zanes des piémonts (mingrélien au nord-ouest, laze au sud-ouest), en passant par le géorgien, qui occupe un bassin central déployé sur toute une gamme de régions écologiques et d'altitudes, depuis les hautes terres de Khévi, Khévsuréti, Pskhavi, jusqu'aux plaines de Kakhétie et au piémont d'Iméréti.

L'idée élémentaire est que si la géographie et le milieu ambiant déterminent la diversité des langues, et notamment leur fragmentation dialectale, ce déterminisme opère moins par l'effet des obstacles naturels (montagnes, cours d'eau), que par le contrôle des sociétés humaines sur les ressources naturelles, et leurs systèmes d'échanges et de réciprocité (commerce ou troc selon les époques et les conditions économiques, mais aussi intermariages, alliances politiques, organisation communale, accès à la terre, aménagement du territoire, modes de gouvernance, rayonnement culturel, coalitions résistantes ou résilientes face à des hégémonies invasives, etc.). Comme le décrit élégamment le géographe O'Sullivan<sup>9</sup>, la complexité des relations en géographie humaine se laisse analyser à travers trois ordres de structures:

- (a) *Phylogénétique* (complexité des *déterminismes* par caractères hérités)
- (b) *Ontogénétique* (complexité des *types* et des *schèmes* ou des *structures*)
- (c) *Épigénétique* (complexité environnementale des *relations* et des *représentations* secondaires).

7 Nichols J., "The Origin of the Chechen and Ingush: A Study in Alpine Linguistic and Ethnic Geography", *Anthropological Linguistics* 46-2, (2004): 129-155.

8 Murra J.V., *The Economic Organization of the Inca State*, Ph.D. diss., Department of Anthropology (University of Chicago, 1956).

9 O'Sullivan David, "Complexity Science and Human Geography", *Transactions of the Institute of British Geography*, (2004): 282-295.

Ainsi, le projet LaDyCa a tenté de se donner les moyens de saisir la complexité phylogénétique interne des dialectes géorgiens (cf. section 2.1 *infra*), tout en tenant compte des *ontologies* (notamment, les facteurs externes: les facteurs géohistoriques, qui façonnent l'histoire et définissent temporairement du moins, les individus et les groupes sociaux en tant que tels) et des facteurs épigénétiques (l'incidence de l'environnement et des phénomènes d'adaptation et d'exaptation au cours des âges, dans la formation des *ontologies*). Une source d'inspiration notable a également été la géographie culturelle multidimensionnelle telle que la développe le géographe américain Paul Adams<sup>10</sup> (2012), qui cherche à faire apparaître, au-delà de la surface des construits tenus pour acquis, des structures transversales profondes, comme par exemple, en termes de relations intergouvernementales, la résilience forte, en Scandinavie, d'un réseau dense de relations entre pays du nord, hérité d'avant l'intégration de chacun des pays scandinaves à l'UE, et davantage relié à la Suisse (notamment par les institutions liées à la neutralité géostratégique) qu'avec les pays Baltes (alors que les liens historiques entre Suède et l'Estonie ou la Lettonie sont pourtant indéniables), qui sont quant à eux davantage engagés dans des liens avec d'autres pays de taille comparable, et géographiquement éloignés, comme l'Islande et la Slovaquie. De même, la méthode de réticulation des liens entre 179 organisations intergouvernementales européennes fait apparaître des interactions géopolitiques *relativement* plus fortes entre l'Autriche et les pays du triangle de Visegrád qu'entre l'Autriche et le reste de l'Europe<sup>11</sup>.

Nous verrons bientôt comment des méthodes quantitatives aboutissant à des représentations sous forme de graphes rendant compte de réseaux font émerger, en dialectologie, des structures à la fois congruentes avec les connaissances intuitives, et des configurations moins attendues, sans pour autant être insolites ou aberrantes. Leur grand intérêt est qu'elles représentent des patrons d'interaction plutôt que des entités cloisonnées ou fermées. Ce sont des *construits dynamiques* plutôt que des *entités statiques*. On peut révéler ces structures à l'aide de *seuils quantitatifs*. Elles sont toutes hautement indicelles de tendances profondes qui président à l'ordonnement de la complexité *phylogénétique* (la classification des dialectes d'une langue comme le géorgien, par exemple), et elles incitent à approfondir et affiner les recherches sur les facteurs déterminants (la complexité *ontologique*) ou cumulés et associés (la complexité *épigénétique*).

Rappelons brièvement les prémisses de toute approche écologique: (i) Di-

10 Adams Paul, "Multilayered regionalization in Northern Europe", *GeoJournal*, 77-3 (2012): 293-313.

11 Adams, "Multilayered regionalization in Northern Europe", 304.

versité, (ii) *Variation*, (iii) *Transmission*, (iv) *Sélection & adaptation (ou exaptation)*, (v) *Niche*, (vi) *Rythme du changement (ou des mutations)*<sup>12</sup>. De ce point de vue, *diversité* et *variation* sont au cœur de toute approche dialectologique. La *transmission*, la *sélection* et *l'adaptation* forment la trame de la diversité, dans les espaces dialectaux (une aire géographique traversée par des isoglosses au sein d'une même langue, voire d'une famille de *langues proches* comme les langues kartvéliennes)<sup>13</sup>, dans la mesure où les *systèmes* linguistiques (au niveau saussurien de la *langue*) génèrent des *normes* sociolinguistiques (les dialectes et sous-dialectes), qui constituent des ensembles plus ou moins stables d'*usages* (les parlers locaux, relevant de la *parole*)<sup>14</sup>, qui entrent en compétition pour le prestige et la stabilisation sociale, de génération en génération. Les *normes* successives s'adaptent à des conditions sociales – une écologie socioculturelle – et locales ou régionales d'*habitus* linguistique – ou manière de parler, usages

12 Repris de John Wendel (source: [http://www.u-bunkyo.ac.jp/center/library/image/fsell2005\\_51-76.pdf](http://www.u-bunkyo.ac.jp/center/library/image/fsell2005_51-76.pdf)) – notre traduction. On trouvera une problématique semblable dans la communication accessible en ligne sur le lien <http://axe7.labex-efl.org/node/295>, qui analyse la „résonance“ géolinguistique d'une enclave mixtèque en domaine mazatec, dans l'Etat mexicain de Oaxaca. Voir aussi <http://axe7.labex-efl.org/node/250> au sujet de l'écologie diasystémique du mazatec, qui a fait office de „banc d'essai“ pour la méthode.

13 Il importe de préciser qu'il est clair, pour nous, que les langues kartvéliennes, à savoir les langues zanes (mingrélien et laze), le géorgien et le svan, forment trois ensembles de langues phylogénétiquement liées, sans être des „dialectes“ subordonnés les uns aux autres pour autant. Par ailleurs, nous avons pu constater que le réseau dialectal du géorgien à proprement parler est, somme toute, assez unitaire et peu différencié, comparativement au réseau dialectal svane, où les différences interdialectales sont plus marquées – en outre, sur des distances géographiques courtes. Le Zan forme un continuum dialectal discontinu, avec le mingrélien au nord-ouest et le laze au sud-ouest et, bien qu'il s'agisse de deux langues avec une gamme de variétés internes, la continuité structurale est assez remarquable entre les deux sous-ensembles. Par „dialectes“ nous entendons ici le plus souvent des variétés du géorgien comme l'imérétien, le pshavien, le gourien ou l'adjarien, etc. qui, elles, s'ordonnent indéniablement selon une logique de continuum dialectal au sein d'une même langue – tendance qui a été renforcée par la longue tradition écrite du géorgien comme „langue de culture“, dans le sens de langue du culte et langue administrative, mais aussi de *lingua franca proche* (à la différence du turc, du russe ou du persan, langues hégémoniques de superposition à différents moments de l'histoire et dans différentes régions de la Géorgie actuelle, qui ont peu ou prou imposé une diglossie de type fishmanien) transversale à l'ensemble kartvélien. Malgré ces différences structurelles fortes entre zan, svan et géorgien, on observe actuellement, de manière de plus en plus prégnante depuis la restauration de l'indépendance de la Géorgie en 1991, les effets d'une diglossie de type plutôt fergusonien, avec la superposition du géorgien standard ou littéraire sur les langues kartvéliennes proches, ainsi que sur les dialectes du géorgien à proprement parler.

14 Ici, nous reprenons à Eugenio Coseriu la tripartition entre *système* et *parole* ou *l'usage* (la langue *versus* la parole saussuriennes), auxquelles s'ajoute la *norme* comme notion intermédiaire sur ce continuum, allant des *formes* du système aux *réalisations* (cf. Coseriu, 1982: 11-113), et aussi, pour une implémentation taxinomique en sociolinguistique (Pineda, „Diasistema lingüístico y cohesión interna. Un estudio lingüístico desde la pan-cronía“, 1985).

– qui vont favoriser telle solution structurale de la langue plutôt que telle autre. Ces choix peuvent résulter du jeu du hasard et de la nécessité, et surtout des représentations sociales que projettent les locuteurs, mais elles peuvent aussi dans certains cas, résulter de réinterprétations, ou de réanalyses – en ce cas, il ne s’agit plus seulement d’adaptation, mais plutôt d’*exaptation*: en typologie dialectale, de *métatypies*, comme par exemple les formations analogiques dans la flexion verbale ou nominale. Ces choix structuraux opèrent dans des séries de *niches*, ou lieux de vie, de liens et d’échanges matériels et culturels (notamment, à travers le prestige, local ou plus étendu, dont peuvent bénéficier les différentes options structurales déployées par les normes et les usages à partir d’un même système transdialectal, ou *diasystème*).

## 1.2. Questions posées, objectifs

Les questions que posait initialement le projet LaDyCa étaient les suivantes:

1) Peut-on appliquer aux langues caucasiennes les techniques qui ont porté leurs fruits dans un autre domaine très éloigné géographiquement et linguistiquement (la Méso-Amérique), et comment doit-on adapter la méthode? Cette question humboldtienne (cf. *supra*) est liée, sur le plan épistémologique, au caractère plus ou moins irréductible ou incommensurable de toute situation géohistorique et sociohistorique, du point de vue des observables en sciences humaines. C’est donc une question de fond, sur le plan méthodologique, du moins.

2) Comment affiner les techniques de traitement des données dialectologiques dans une perspective de dialectologie socio-historique? Ne peut-on, à l’aide des technologies nouvelles, comme la reconnaissance automatique de parole, aller au-delà de la méthode classique en dialectométrie, consistant à fonder le traitement des données sur des transcriptions (et donc, des construits de linguiste, en tous cas de transcripteur)? En quoi les résultats diffèrent-ils? Quel est le gain empirique et théorique? Quelles sont les limites du traitement des données brutes en aval, sans passer par le filtre des constructions théoriques en amont?

3) Si l’écologie (socio)linguistique apporte des méthodes et des connaissances utiles sur le plan documentaire, pour la recherche universitaire, mais aussi du point de vue de l’aide à la décision, en aménagement du territoire, que peut apporter un projet de dialectologie à cheval sur trois familles de langues si différentes qu’elles ne forment pas nécessairement un phylum (en tous cas, pas de manière définitivement probante), dans des situations géopolitiques si différentes? Mieux encore: quelle est la part de travail social et d’impact sociétal

que l'on est en mesure (et pas seulement en droit) d'attendre d'un tel projet?

En somme, la première question interroge la méthode et les conditions de reproductibilité des modes de traitement des *données* (les faits de langue) et d'interprétation des *construits* externes (les ontologies que sont les facteurs externes, comme la verticalité, liée aux milieux alpins ou de versants montagneux en articulation avec des paliers écologiques). La deuxième question interroge la catégorisation des données, et la portée des construits théoriques qu'utilise le linguiste (par conséquent, cette fois, les construits internes à la linguistique). Enfin, la troisième question interroge, au-delà des deux premières, qui concernaient davantage la recherche fondamentale, le champ des possibles, sur le plan des applications, en recherchant un impact sociétal positif. Elle relève certes de la recherche appliquée, mais elle ne peut se comprendre ni trouver de réponse sans interagir, en boucle, avec les deux précédentes, et elle s'inscrit dans la logique de la *linguistique du développement social*, actuellement en pleine émergence<sup>15</sup>.

Dans quelle mesure y a-t-il compatibilité des observatoires et des observables? La réponse à cette question est globalement positive, pourvu que l'on se donne les moyens de „cadrer“ les différentes configurations géolinguistiques. De ce point de vue, le profil du réseau dialectal géorgien est clairement celui d'un espace géolinguistique à la fois scindé en deux grands domaines (occidental et oriental), certes, mais surtout celui d'un *bassin de traits*, plutôt que d'un espace polycentrique, où plusieurs dialectes seraient en concurrence avec un potentiel de rayonnement. Le seuil de 80% de similarité (non pondérée) par rapport à la moyenne normalisée de distance interponctuelle, à partir d'une base de données de 244 entrées de cognats retenus pour l'analyse phonologique, compilées à partir de diverses sources de dialectologie géorgienne<sup>16</sup>, fait clairement apparaître cette propriété générale du diasystème géorgien: la trame qui relie les variétés parlées dans le territoire national est dense<sup>17</sup>.

Plus on concentre le regard sur les zones centrales, de piémont et de plaines, plus le réseau est compact, tandis que les parlers montagnards septentrionaux, comme le toushe ou le mtioulétien, sont soit isolés, soit moins densément et

15 Agresti Giovanni, *Diversità linguistica e sviluppo sociale* (Milano: Franco Angeli editore, 2018).

16 Mentionnons, entre autres sources Besarion Jorbenadze, *kartuli dialekt'ologia*, I 1989 & II 1998 ainsi que Gigineishvili, Ivane, Topuria Varlam, Kavtaradze Ivane, *kartuli dialekt'ologia* (Tbilissi, 1961), et de nombreuses monographies locales des divers dialectes et variétés prises en compte dans la base de données de 244 cognats.

17 La base de données utilisée à cette fin a été compilée par Hélène Gérardin, postdoctorante du projet LaDyCa, entre septembre 2017 et 2018 ; le traitement algorithmique a été réalisé par Flore Picard, doctorante contractuelle en sciences du langage à Sorbonne Université, à l'aide du logiciel Gephi (cf. <https://gephi.org/>).



uniformément reliés aux complexe central. Les dialectes ingiloin (Ingilouri, ინგილოური) d'Azerbaïdjan et le fereydanien (Pereidnuli, ფერეიდნული) de Fereydan en Iran sont également soit isolés, soit reliés de manière diffuse à une partie du réseau, mais les liens pointent vers la Kartlie et la Kakhétie (par le relais du kiziquien), donc vers le centre-est et l'est du réseau dialectal.

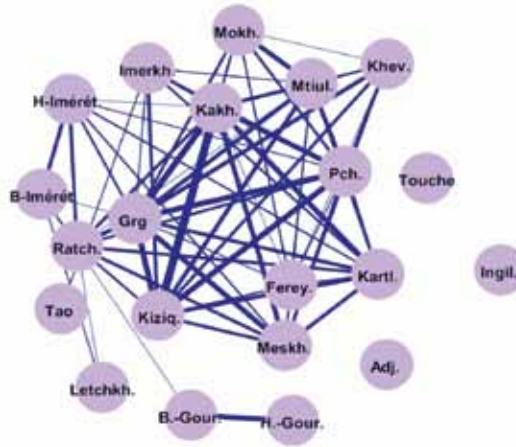


Figure 1. Seuil de 80% de similarité (non pondérée) par rapport à la moyenne normalisée de distance interponctuelle entre variétés dialectales de géorgien<sup>18</sup>

Ce qui est frappant dans cette configuration, c'est le caractère *compact* du réseau dialectal. Le fait qu'une continuité émerge entre deux dialectes aussi distants que le ratcha, au nord-ouest et le haut et bas gourien, au sud-ouest, conforte l'idée qu'on se trouve face à un bassin de traits (*feature pool*, selon la terminologie de Salikoko Mufwene<sup>19</sup>, dans la mesure où les normes et les usages sont relativement fluides, non compartimentés dans des frontières régionales étanches. Par ailleurs, l'hypothèse du déterminisme des paliers écologiques est dans l'ensemble confirmée, comme en Mésio-Amérique, car les normes des hautes terres tendent à former des sous-groupes, face aux normes des piémonts et des plaines, mais toujours selon une division longitudinale (ouest contre est).

La seconde question visait à affiner les techniques de traitement des don-

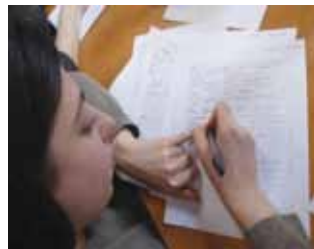
18 Picard Flore *et al.*, "Levenshtein algorithm applied to Kartvelian phonological data", The Second Language Dynamic in the Caucasus International Workshop, 26<sup>th</sup> of February, Ilia State University, Tbilisi, Georgia (2018).

19 Mufwene Salikoko, *The Ecology of Language Evolution*, Cambridge Approaches to Language Contact (Cambridge: Cambridge University Press, 2001) ; Mufwene Salikoko, *Complexity perspectives on language, communication, and society*, in Massip-Bonet Á. & Bastardas-Boada A., 197-2018 (Springer Verlag, 2012).

nées dialectologiques. C'est ce que Claude Montacié, professeur de linguistique et informatique à Sorbonne Université a entrepris de réaliser à l'aide de modèles de Markov appliqués aux abondantes données recueillies par Hélène Gérardin dans deux régions du sud-ouest de la Géorgie (Adjarie et Gourie). La collecte s'est effectuée en août et septembre 2017 auprès de 86 informateurs et informatrices né(e)s entre 1927 et 2006, à l'aide d'un questionnaire de 117 entrées lexicales choisies en fonction de leur pertinence phonologique, présenté sous forme de stimuli visuels, afin d'éviter les interférences avec le géorgien standard. Les enquêtes de terrain, enregistrées et filmées à l'aide d'un caméscope haute définition, ont été rendues possibles par le très efficace travail de coordination et d'accompagnement de deux jeunes chercheurs géorgiens, docteurs en linguistique de l'Université de Batumi: Natia Abashidze et Shota Rodinadze, que nous ne saurions assez remercier et féliciter pour leur aide en Adjarie. Le traitement actuellement en cours sur les données sonores consiste, à l'aide du logiciel PRAAT et d'un algorithme pour la transcription automatisée de corpus oraux, à obtenir des regroupements entre idiolectes directement à partir des propriétés du signal acoustique, selon une modélisation qui tient compte des propriétés des 34 phonèmes du dialecte gourien<sup>20</sup>. Non seulement cette technique permettra d'affiner la classification des normes et des usages observables pour le gourien d'abord, puis pour l'adjarien, mais elle permettra aussi de doter le géorgien, standard et dialectal, d'outils de transcription automatique – une retombée de grande utilité, en tant que ressource technologique. Nous comptons l'an prochain étendre cette élaboration de ressources au dargi, langue daghestanaise de la Fédération de Russie, à laquelle nous avons eu accès grâce à notre informateur Gasangusen Sulaibanov, auprès de qui nous avons pu enregistrer à Paris avec Anastacia Loskot, étudiante de master en Sorbonne, une dizaine d'heures de données phonologiques de bonne qualité. Dans la mesure où ces systèmes de reconnaissance et transcription automatique de la parole nécessitent de s'aligner sur des schèmes phonémiques de langues de comparaison, et que nous avons pu en laboratoire utiliser le russe et le turc comme parangons afin de guider les chaînes de Markov de détection des phonèmes, des syllabes et des mots, pour le géorgien sud-occidental mais aussi pour le tchéchène et l'ingouche (travail réalisé dans le cadre du projet LaDyCa par Françoise Guérin, du Lacito, Maître de conférences HDR à Sorbonne Université), ce sont pas moins de quatre langues caucasiennes qui seront bientôt outillées à l'aide de ces outils technologiques.

20 Montacié Claude et Hélène Gérardin, "Markov Models applied to Gurian data", *The Second Language Dynamic in the Caucasus International Workshop*, 26<sup>th</sup> of February, Ilia State University, Tbilisi, Georgia (2018).

La troisième question portait sur les retombées, et la dimension de travail social du projet LaDyCa, dans la mesure où le projet qui a servi de cadre méthodologique initial, le MAmP, avait une très forte dimension sociétale, notamment de formation de communautés de pratiques diverses à la linguistique ou à l'approche dialectique de leurs langues maternelles –un volet d'activités abondamment représenté sur le site Internet du LabEx EFL<sup>21</sup>. A ce titre, nous avons pu expérimenter à ILIAUNI la formation à l'enquête de terrain par méthode des élicitations croisées (confrontation de plusieurs solutions de transcription d'une même élicitation par un groupe de transcrip-teurs issus de différents dialectes), qui a donné des résultats très encourageants (cf. clichés ci-dessous)<sup>22</sup>.



Atelier de grammaire mingrélienne, 28 février 2018, ILIAUNI

21 Voir les nombreux liens sur des produits de ce travail de formation et de production de matériaux didactiques ou d'élicitations de données sur la page <http://axe7.labex-efl.org/taxonomy/term/12>.

22 Par ailleurs, le projet a commencé avec un premier atelier méthodologique, en mai 2017, suivi en février 2018 d'un atelier de présentation des résultats, tous deux tenus à ILIAUNI. Les conférences présentées lors des deux ateliers *Language Dynamic in the Caucasus* International Workshop des 23 mai 2017 et 28 février 2018 à l'Université d'Etat Ilia, Tbilisi, sont accessibles en ligne sur les liens <https://www.youtube.com/watch?v=hi8vcHBnmc8> et [https://www.youtube.com/watch?v=0SGy\\_CqIwGE](https://www.youtube.com/watch?v=0SGy_CqIwGE). On accède par ce dernier lien à trois des conférences citées dans les références à la fin du présent article.

## 2. Résultats

### 2.1.1. Kartvélien: complexité phylogénétique des dialectes géorgiens

Nous avons d'ores et déjà évoqué les résultats de la dialectométrie appliquée à un corpus de données compilées à partir de sources dialectologiques géorgiennes. Il est temps d'entrer dans le détail de ces résultats, et d'en fournir davantage d'illustrations. La distance de Levenshtein appliquée à notre base de données compilées figure dans la matrice ci-dessous<sup>23</sup>. Ces distances sont faibles: la fourchette entre le géorgien standard (Grg) et les dix-neuf variétés ici prises en compte varie entre 0,093 (avec le pshavien, connu pour ses rétentions consonantiques, qui le rapprochent du géorgien standard) et 0,518 (ingilien, dialecte géorgien du nord-ouest de l'Azerbaïdjan). Peu de variétés atteignent un seuil équivalent à seulement 0,305 sur l'index 1.00 (ex. le toushe), et seules quelques variétés, principalement situées au sud-ouest, dépassent le seuil de 0,220 (adjar, tao, gourien). Ce sont des valeurs très faibles, comparativement à celles auxquelles nous étions habitués pour des langues méso-américaines comme le mazatec, qui oscillaient toutes sur une fourchette de 0,24 à 0,54 sur 1.00<sup>24</sup>. C'est le signe à la fois d'une tendance cyclique au brassage des populations, surtout dans les régions centrales et orientales, cruellement touchées par les guerres et les déportations au cours des derniers siècles<sup>25</sup>, mais aussi et surtout, du puissant rayonnement du géorgien standard, comme langue de culture et comme langue nationale.

	Grg	Khev	Psh	Mokh	Mtiul	Touche	Kart	Kakh	Kiziq	Ferey	Ingil	Meskh	Ratch	Imérét	Letchk	Adj	Tao	Imerkh	H-Gour	B-Gour	
Grg	1.000																				
Khev	0.122	1.000																			
Psh	0.093	0.138	1.000																		
Mokh	0.115	0.187	0.147	1.000																	
Mtiul	0.122	0.144	0.138	0.209	1.000																
Touche	0.186	0.24	0.181	0.31	0.31	1.000															
Kart	0.127	0.188	0.144	0.198	0.178	0.238	1.000														
Kakh	0.076	0.189	0.189	0.174	0.128	0.136	0.097	1.000													
Kiziq	0.093	0.142	0.148	0.184	0.128	0.165	0.118	0.098	1.000												
Ferey	0.133	0.192	0.178	0.172	0.183	0.217	0.172	0.13	0.173	1.000											
Ingil	0.518	0.428	0.428	0.442	0.414	0.484	0.424	0.396	0.414	0.414	1.000										
Meskh	0.154	0.183	0.137	0.198	0.189	0.244	0.134	0.114	0.14	0.183	0.169	1.000									
Ratch	0.119	0.201	0.151	0.223	0.188	0.24	0.143	0.142	0.18	0.202	0.211	0.137	1.000								
Imérét	0.182	0.204	0.177	0.263	0.211	0.337	0.172	0.181	0.208	0.237	0.212	0.173	0.184	1.000							
Letchk	0.18	0.247	0.212	0.25	0.222	0.323	0.186	0.203	0.213	0.234	0.204	0.202	0.188	0.175	1.000						
Adj	0.26	0.282	0.261	0.313	0.286	0.388	0.239	0.252	0.276	0.306	0.289	0.279	0.19	0.219	0.178	1.000					
Tao	0.225	0.274	0.241	0.318	0.273	0.372	0.242	0.258	0.278	0.291	0.313	0.281	0.202	0.217	0.184	0.249	1.000				
Gour	0.238	0.274	0.251	0.282	0.268	0.368	0.238	0.254	0.267	0.289	0.314	0.288	0.178	0.176	0.188	0.287	0.23	1.000			
H-Gour	0.178	0.268	0.248	0.276	0.251	0.32	0.184	0.183	0.143	0.204	0.228	0.223	0.217	0.208	0.229	0.288	0.182	0.181	1.000		
B-Gour	0.284	0.304	0.282	0.303	0.303	0.404	0.245	0.277	0.291	0.301	0.349	0.279	0.238	0.229	0.2	0.245	0.286	0.301	0.278	1.000	
B-Gour	0.288	0.308	0.284	0.243	0.308	0.381	0.248	0.271	0.281	0.337	0.381	0.308	0.241	0.238	0.182	0.28	0.284	0.323	0.27	0.301	1.000

Figure 2. Matrice de distances de Levenshtein non pondérée pour 244 cognats représentatifs du continuum dialectal du géorgien<sup>26</sup>

23 Les abréviations se réfèrent aux variétés suivantes: Grg: géorgien, Vx grg: vieux géorgien, pchave, khevsour, tianétien (ce sous-dialecte peut être considéré comme un mélange de pshavien, de khévsour et de touche), mokhève, mtioulétien, touche, kartlien, kakhétien, kiziquien (sous-dialecte du kakhétien), fereydan, ingilo, djavakh, samtskhétien (djavakh et samtskhétien sont deux variantes ce qu'on peut appeler dialecte meskhète), ratchien, imérétien, lechkoumien, adjar, tao, imerkhève, gourien ; H: haut / B: bas.

24 V. Léonard Jean Léo et al., *Patterns of Linguistic Diffusion in Space and Time: The Case of Mazatec*, in F. Abergel et al. (eds.), *Econophysics and Sociophysics: Recent Progress and Future Directions* (New Economic Windows, Springer, 2017), 227-255.

25 Cf. Assatiani Nodar et Alexandre Bendianashvili, *Histoire de la Géorgie* (Paris: L'Harmattan, 1997).

26 Picard Flore et al., "Levenshtein algorithm applied to Kartvelian phonological data".



Figure 3. Carte des localités/dialectes retenus dans la base de données dialectales compilée par Hélène Gérardin<sup>27</sup>. Cartographie: Hélène Gérardin.

Les deux dendrogrammes ci-dessous rendent compte des configurations géolinguistiques à l'aide de la méthode de classification ascendante hiérarchique, ou *Hierarchical Cluster Analysis* (Ward's Method). Le premier présente les résultats non pondérés, le deuxième tient compte d'une échelle de pondération, qui donne davantage de poids aux faits de langue relevant du *système* et de la *norme* qu'aux faits de *parole*. Le tableau 1 énumère à titre d'exemple une vingtaine d'entrées de la base de données, en illustrant par une variante d'un dialecte retenu *ad hoc* comme exemple un ensemble de variables phonologiques, auxquelles est attribué un coefficient de pondération à trois degrés (faible, moyen et fort). Parmi les variables à poids „fort“ figurent, de (1) à (4) le changement de qualité vocalique  $i > ü$ , la métathèse de  $-r-$  formant une *muta cum liquidae* secondaire, l'antériorisation d'une vélaire, la prénasalisation d'une affriquée médiane. Parmi les variables d'incidence structurelle moyenne, de (5) à (15), on peut mentionner la chute d'une consonne médiane (6), d'une fricative postérieure initiale (7), la latéralisation d'une rhotique (8), la contraction ou monophthongaison d'une diphtongue décroissante (9), l'apocope dans des formes à haut rendement dans l'usage (10), la simplification de groupe consonantique initial sifflante-occlusive (11), l'alternance d'une occlusive vélaire/uvulaire dans des radicaux (12), ou l'alternance  $g/v$  (13), la nasalisation conditionnée d'une latérale (14), l'alternance  $p/f$  en contexte intervocalique (15). Tous ces phénomènes constituent des isoglosses relativement fortes, qui

<sup>27</sup> Picard Flore *et al.*, “Levenshtein algorithm applied to Kartvelian phonological data”.

permettent de dessiner des aires dialectales sur la base d'aires phonologiques plus ou moins endémiques ou spécifiques.

	<i>Entrée</i>	<i>Variété (ex)</i>	<i>Variante</i>	<i>Variable phonol.</i>	<i>Coefficient</i>	<i>Glose</i>
1.	<i>sircxvili</i>	<i>Ingil.</i>	<i>sürcxülü</i>	<i>i-ü</i>	<i>fort</i>	'honte'
2.	<i>bevri</i>	<i>H-Imérét.</i>	<i>brevi</i>	<i>metathèse</i>	<i>fort</i>	'beaucoup'
3.	<i>rogorc</i>	<i>Adj.</i>	<i>raperac</i>	<i>g-p</i>	<i>fort</i>	'comment'
4.	<i>viçro</i>	<i>B.-Gour.</i>	<i>vinçro</i>	<i>+n</i>	<i>fort</i>	'étroit'
5.	<i>cecxli</i>	<i>Adj.</i>	<i>cecxli</i>	<i>-c</i>	<i>moyen</i>	'feu'
6.	<i>didi</i>	<i>Ratch.</i>	<i>dii</i>	<i>chute de consonne médiane</i>	<i>moyen</i>	'grand'
7.	<i>ambavi</i>	<i>Kartl.</i>	<i>hambavi</i>	<i>+h initial</i>	<i>moyen</i>	'histoire, nouvelle'
8.	<i>bevri</i>	<i>Letchkh.</i>	<i>brevli</i>	<i>r-l</i>	<i>moyen</i>	'beaucoup'
9.	<i>dainaxa</i>	<i>Imerkh.</i>	<i>dīnaxa</i>	<i>ai-i</i>	<i>moyen</i>	<i>voir, AOR3sg</i>
10.	<i>erti</i>	<i>Ingil.</i>	<i>er</i>	<i>apocope</i>	<i>moyen</i>	'un'
11.	<i>ştumari</i>	<i>Kiziq.</i>	<i>ştumari</i>	<i>Simplification St-</i>	<i>moyen</i>	'invité'
12.	<i>rogorc</i>	<i>Khev.</i>	<i>raqelac</i>	<i>g-q</i>	<i>moyen</i>	'comment'
13.	<i>rogorc</i>	<i>Ratch.</i>	<i>ravarc</i>	<i>g-v</i>	<i>moyen</i>	'comment'
14.	<i>romelic</i>	<i>Khev.</i>	<i>romenic</i>	<i>l-n</i>	<i>moyen</i>	'lequel'
15.	<i>keipi</i>	<i>Ferey.</i>	<i>keifi</i>	<i>p-f</i>	<i>moyen</i>	'festin'
16.	<i>cxeni</i>	<i>Mokh.</i>	<i>sxeni</i>	<i>c-s</i>	<i>faible</i>	'chaud'
17.	<i>carieli</i>	<i>Imerkh.</i>	<i>carjeli</i>	<i>i-j</i>	<i>faible</i>	'vide'
18.	<i>ķargad</i>	<i>B-Imérét.</i>	<i>ķaat</i>	<i>-consonne médiane</i>	<i>faible</i>	'bon'
19.	<i>rom</i>	<i>H.-Gour.</i>	<i>rume</i>	<i>o-u</i>	<i>faible</i>	'que'
20.	<i>bevri</i>	<i>B.-Gour.</i>	<i>breuli</i>	<i>v-u</i>	<i>faible</i>	'beaucoup'

Tableau 1. Echantillon de données et de variables phonologiques pondérées, de la base de données dialectales compilée par Hélène Gérardin<sup>28</sup>.

Parmi les variables „faibles“, on trouve des variables relevant de la parole, soit parce que déterminées par des tendances phonétiques de surface, surtout co-articulatoires (16-17, 19), soit par l'ancrage lexical à haut rendement dans l'usage (18), davantage que par une modification en profondeur de l'inventaire phonémique ou des règles de combinatoire phonotactique – à la différence des précédentes variables, „fortes“ et „moyennes“.

28 Picard Flore *et al.*, "Levenshtein algorithm applied to Kartvelian phonological data".

Les résultats non pondérés donnent l'image de la figure 4 ci-dessous: on retrouve la grande bipartition longitudinale entre géorgien occidental et géorgien oriental. La première aire se subdivise entre parlers du nord-est (khev, mokh, mtiul) et fereydanien *versus* deux autres chorèmes<sup>29</sup>, constituées par les variétés du sud-est (kakh, kiziq) et des variétés „centrales“, incluant le géorgien standard aux côtés du pshavien, sous-dialecte oriental montagnard. A ce groupe central étagé des hautes terres au piémont central, s'agrège dans le même sous-ensemble le kartlien, le ratchien mais aussi le meshkien, selon la même dynamique verticale. La seconde aire comprend les dialectes occidentaux, regroupés en entités plus compactes, puisque les deux variétés haute et basse gourienne d'une part, et haute et basse imérétienne d'autre part, forment deux sous-groupes discrets – le dernier, en relation plus étroite avec le letchkhumien, là encore dans une dynamique de verticalité proche. Le toushe, dialecte pourtant nord-oriental, vient se ranger par défaut à la périphérie de ce groupe occidental.

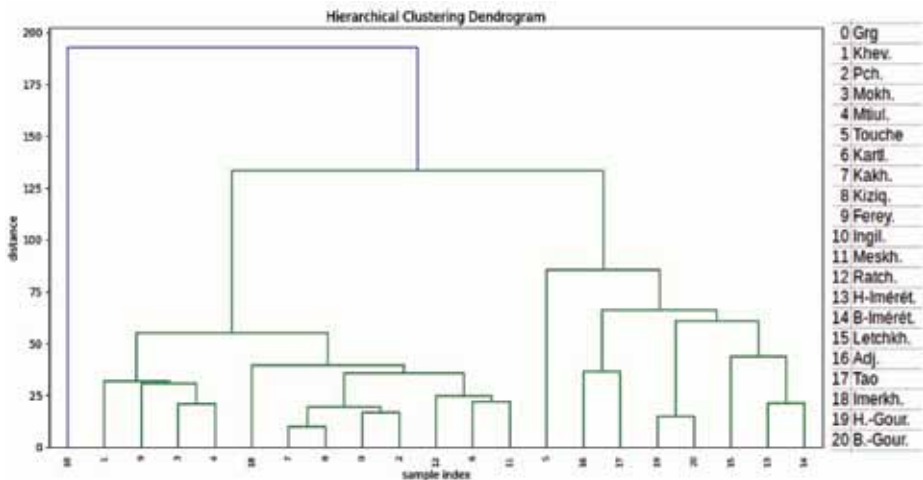


Figure 4. Résultats *non pondérés*, continuum dialectal géorgien: classification ascendante hiérarchique<sup>30</sup>

29 Un *chorème* est un noyau géographique dense: une région dont les propriétés de géographie humaine (ici, géolinguistiques) sont compactes et font que cette aire forme un cœur, une aire discrète ou, en termes figurés, une *contrée*. Le chorème en géolinguistique ne constitue pas nécessairement pour autant un dialecte à part entière ou en soi, mais il peut aussi tenir lieu de noyau dur, ou centre dense d'une aire dialectale. V. Léonard, *Éléments de dialectologie générale* (2012), 132-161.

30 Picard Flore *et al.*, "Levenshtein algorithm applied to Kartvelian phonological data".

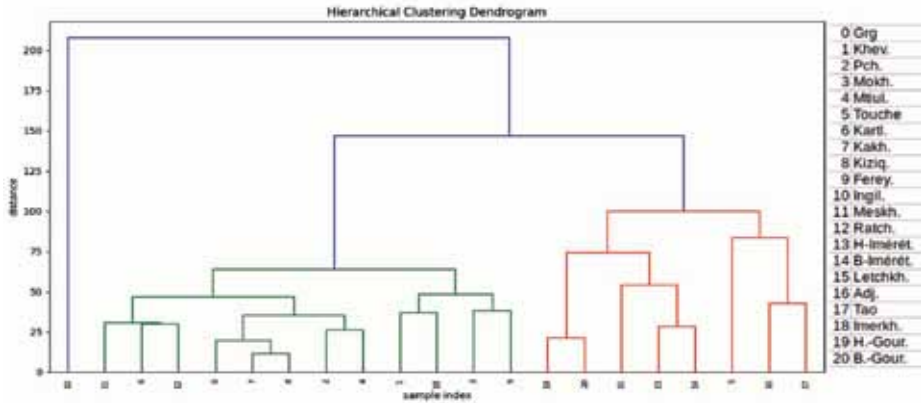


Figure 4. Résultats *pondérés*, continuum dialectal géorgien:  
classification ascendante hiérarchique<sup>31</sup>

Les résultats pondérés ont moins vocation à donner une image exacte de la complexité phylogénétique (les „familles“ de dialectes regroupés en fonction des traits hérités, de rétention aussi bien que d’innovation), que des dynamiques d’innovation et de différenciation forte entre normes. Cette fois-ci, la structure interne du dendrogramme regroupe un ensemble complexe de variétés orientales, parmi lesquelles le géorgien standard, le kartlien, le rachien et les parlers montagnards du nord-est d’une part: d’autre part, l’imérétien associé au gourien et au letchkhien, qui contrastent avec l’adjarien et le tao du nord-est anatolien, de la région de Trebizond. On retrouve encore une incidence, quoique périphérique, d’une variété orientale montagnarde – le toushe.

Le graphe du seuil de 75% de similarité (graphe de gauche, en haut) donne une image plus nette et cohérente des relations entre les variétés: les variétés de diaspora comme le toushe et l’ingilien ou le fereydonien gravitent autour d’un réseau composé de deux chorèmes et d’une arête transversale. Le chorème central forme un losange pshavien-meshkien-kartlien-kakhetien-kiziqien et mtulien lié au géorgien standard, liée au mokhevien et au ratchien ; cette arête forme un trapèze en se rattachant au flanc droit de ce treillis notamment avec le géorgien, et les parlers gouriens, de l’ouest, viennent s’y agréger. Le seuil de 70% de similarité (graphe en haut à droite) vient conforter cette configuration en renforçant la cohérence, puisque cette fois, les deux variétés d’imérétien sont connectées aux parlers gouriens. Mais on voit que l’adjar est une variété occidentale très innovante, qui ne vient se rattacher au réseau imérétien et gourien qu’au seuil de 60% de similarité (graphe en bas au centre).

31 Picard Flore *et al.*, “Levenshtein algorithm applied to Kartvelian phonological data”.



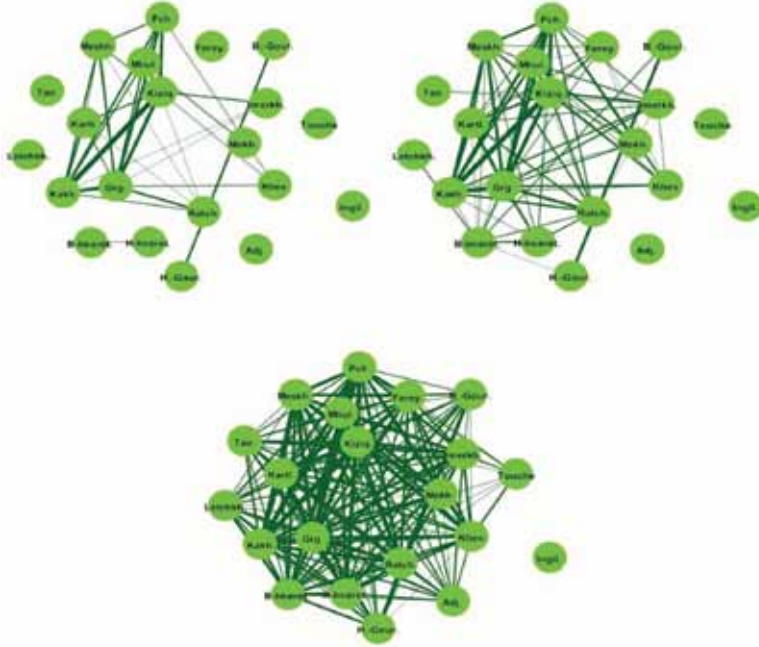


Figure 5. Seuils de similarité pondérés normalisés: 75%, 70% et 60%<sup>32</sup>

### 2.1.2. Complexité ontogénétique: le verbe kartvélien

Un deuxième chantier interne au projet LaDyCa est constitué d’une tentative d’application de la théorie *Paradigm Function Morphology*<sup>33</sup> aux systèmes flexionnels des langues kartvéliennes, en commençant par un article qui revisite le système de la flexion verbale du géorgien standard<sup>34</sup>. Selon ce modèle, on divise en trois blocs les représentations (et, partant, les techniques de segmentation des morphèmes lexicaux et grammaticaux). On distingue les RCR (Règles de Choix de Radicaux) des Règles d’Exponences (suffixation, clitisation) et des Règles Morphologiques (RMP). Plutôt que d’aborder le système flexionnel d’une langue par les verbes réguliers, on s’intéresse en premier lieu aux paradigmes irréguliers, qui sont le plus indicatifs de la structuration profonde du système. De ce point de vue, le verbe irrégulier *tkma* ‘dire’ révèle de manière exemplaire les principaux mé-

32 Picard Flore *et al.*, “Levenshtein algorithm applied to Kartvelian phonological data”.

33 Cf. Stump, *Inflectional Morphology. A Theory of Paradigm Structure*, 2001, et Bonami and Stump, “Paradigm Function Morphology”, 2016.

34 Nous reprenons ici, sans les formalisations, quelques éléments d’un article en cours de rédaction, dont les grandes lignes ont été présentées dans Makharoblidze et Léonard, “Paradigm Function Morphology applied to Georgian verb inflection”, 2018.

canismes de structuration de la flexion verbale en géorgien standard: à la manière d'un verbe irrégulier comme „aller“, en français, qui oppose, à partir de trois lexèmes étymologiques hérités du latin, les formes supplétives 'je vais, tu vas, il va' au singulier du présent de l'indicatif (< lat. VADERE) aux formes de pluriel 'nous allons, vous allez' (< lat. \*ALLARE) du même temps et mode, généralisées à l'imparfait ('j'allais, tu allais, il allait, nous allions, vous alliez, ils allaient'), lesquelles s'opposent en bloc aux formes du futur indicatif et du conditionnel à radical issu de lat. IRE ('j'irai(s), tu iras/-ais', etc.), le verbe *tkma* du géorgien oppose un radical en *ambob* '(ra)conter' à un jeu de radicaux formés sur une racine TKV dont le consonantisme peut se renforcer, avec éjectivation et uvularisation de l'occlusive vélaire (-*tq'v*-), ou rester sous sa forme par défaut - *tkv*-, associée à diverses voyelles versives (ou indices actanciels). Le tableau ci-dessous rend compte de cette diversité des thèmes flexionnels, par supplétion (*ambob*/TKV), RMP de fortition (*tq'v* contre *tkv*) et enfin, préfixation de voyelle versive, avec une nette dichotomie en fonction des Participants de l'Acte de Parole (PAP) dans le paradigme du parfait indicatif (avec voyelle versive *u* – pour les accords sujet de 3 sg et pl, contre voyelle versive *i* – pour les sujets 1 et 2 sg et pl; pour les deux autres tiroirs du parfait, la voyelle versive est *e*-).

	<i>ats'mq'o</i> Présent, indicatif	<i>uts'q'vet'eli</i> Imparfait, indicatif	<i>ats'mq'os k'avshirebiti</i> Présent, subjonctif
Série I/a	<b>-ambob-</b>	<b>-ambob-</b>	<b>-ambob-</b>
	<i>mq'opadi</i> Futur, indicatif	<i>kholmeobiti</i> Conditionnel	<i>mq'opadis</i> <i>k'avshirebiti</i> Future, subjonctif
Série I/b	<b>-itq'v-</b>	<b>-itq'v-</b>	<b>-itq'v-</b>
		<i>ts'q'vet'ili</i> Aoriste, indicatif	<i>k'avshirebiti</i> Optatif
Série II		<b>-tkv-</b>	<b>-tkv-</b>
	<i>turmeobiti</i> Parfait, indicatif	<i>turmeobiti</i> Plus-que-parfait	<i>k'avshirebiti</i> Parfait, subjonctif
Série III	<b>-i/utkvam-</b>	<b>-etkv-</b>	<b>-etkv-</b>

Tableau 2. Formes supplétives relevant des RCR en géorgien standard: *tkma* 'dire'

TAM Séries	Tiroirs		
I/a	Type A PRESENT v-, -s, -s v-_t, -t, -en	Type B IMPARFAIT v-_di, -di, -da v-_it, -it, -nen	Type B' SUBJONCTIF PRES v-_de, -de, -des v-_et, -et, -nen
I/b	Type C FUTUR v-_i, -i, -is v-_it, -it, -ian	Type B CONDITIONNEL v-_di, -di, -da v-_it, -it, -nen	Type B' SUBJONCTIF FUTURE v-_de, -de, -da v-_det, -det, -dnen
II	Type C' AORISTE v-_i, -i, -a v-_it, -it, -es	Type D OPTATIF v-_a, -a, -as v-_at, -at, -an	
III	Type A' PARFAIT m-_s, g-_s, -s gv-_s, g-_t, -t	Type D' PLUS-QUE-PARFAIT m-_a, g-_a, -a gv-_a, g-_at, -at	Type D/A' PARFAIT SUBJONCTIF m-_as, g-_as, -as gv-_as, g-_at, -at

Tableau 3. Affixes relevant des RE en géorgien standard: tkma 'dire'

En laissant de côté le marquage des personnes sujet 1 sg et pl, régulièrement préfixé, on peut identifier quatre procédés de marquage suffixal de l'accord sujet (et TAMV<sup>35</sup>): un type A, modèle de la flexion de personne sujet, qui se réalise par des suffixes sigmatiques (en – s au singulier) ou consonantiques (-t et – en), analysables comme formes primaires de l'accord subjectal; un type B, qui associe présence ou absence d'une occlusive coronale voisée – d-, et un vocalisme en – i/-a, qui neutralise au singulier tout marquage primaire du sujet tandis que le marquage pluriel fait contraster – t pour les PAP (1 et 2) contre – nen pour 3pl. Le type C est principalement à marquage vocalique simple (-i versus – a), avec des contrastes sigmatiques à la 3 sg (futur) ou pl (aoriste). Enfin, le type D généralise un marquage en – a. Toute la série III du parfait a recours à une gamme de préfixes de personne phonologiquement „forts“, issus de la série objectale, en raison de la stratégie de marquage actanciel oblique de ces tiroirs de parfait, qui jouent sur un effet de diathèse sémantique et pragmatique – comme en français, la tournure oblique «qu'il *me/lui* soit/fût dit que...“

Un article en anglais est en cours de finition pour appliquer PFM à la flexion non seulement du géorgien standard, mais des langues zanes et du svan. Un atelier de grammaire mingrélien a été réalisé le 28 février à l'Université

35 Temps, Aspect, Mode, Voix.

d'Etat Ilia avec Mme Rusudan Gersamia, enseignante et chercheuse de mingrélien, originaire de Zugdidi, consistant à éliciter des paradigmes à partir de ceux du géorgien standard, que transcrivent des étudiants originaires de diverses régions dialectales. Nous avons également eu la chance de pouvoir enregistrer de nombreux paradigmes de dialectes svan en élicitant quatre locutrices svanes travaillant à l'institut de linguistique Chikobava. Une étudiante de master 2 de Sorbonne Universités, Anaïs Tran Ngoc, a rédigé et soutenu (un mémoire qui applique PFM aux données mingréliennes et svanes enregistrées à cette occasion.) Cette initiative est également un des produits secondaires du projet méso-américain précédent, qui avait utilisé le modèle PFM pour décrire la complexité du diasystème mazatec<sup>36</sup>, réputé, tout comme celui des langues kartvéliennes, parmi les plus complexes parmi les langues du monde. Cette approche avait en effet permis non seulement de modéliser la diversité interne de la flexion verbale du mazatec, mais également de recueillir des données dialectales de zapotec et d'autres langues méso-américaines<sup>37</sup> (cf. Léonard 2017). Des propositions didactiques en termes de travail social, pour l'enseignement de la grammaire de ces langues, avaient également été élaborées par des collectifs d'enseignants. On pourrait envisager de reproduire le même genre d'activités en Svanétie ou en Samogrelo.

## 2.2. Abkhaze

Enfin, parallèlement au domaine kartvélien, le projet LaDyCa a été en mesure d'explorer quelques aspects de la phonologie de l'abkhaze, réputée parmi les plus complexes dans les langues du monde. Une tentative de modélisation du système phonologique de cette langue a été proposée pour le colloque international du réseau français de phonologie de juillet 2018, avec pour titre „*dynamique allophonique du vocalisme „vertical“ en abkhaze (ap<sup>h</sup>swa biz<sup>w</sup>a: caucasien nord-occidental)*“. Par ailleurs, une étudiante de master 2 de la Sorbonne, Ghalia Deraï, a rédigé et soutenu un mémoire de morphologie comparative, en synchronie, du complexe verbal dans trois langues CNO (oubykh, kabarde et abkhaze).

L'abkhaze (აჭყა ბызაჲა, ap<sup>h</sup>swa biz<sup>w</sup>a) est une langue caucasienne de la famille nord-occidentale (désormais CNO = Caucasique Nord-Ouest, branche oubykh-abkhaze), parlée par un peu plus de 100 000 locuteurs, dans la République autonome d'Abkhazie (Géorgie) ainsi qu'en situation de diaspora en Turquie, Syrie, Amérique du nord, etc. suite à diverses vagues historiques de

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

37 V. Léonard, *Éléments de dialectologie générale*, 2012.

„déplacements“ et de migrations<sup>38</sup> depuis le XIX<sup>ème</sup> siècle. La langue a successivement été codifiée en alphabet arabe et cyrillique ; elle compte six dialectes (division de type latitudinale: nord/centre/sud). Les données utilisées ici seront issues, à titre préliminaire, de la variété standard, en usage à Sukhumi, sur la base d'un questionnaire phonologique conçu par John Colarusso<sup>39</sup> (2018) à partir des données de Yanagisawa & Tsvinaria-Abramishvili<sup>40</sup>.

Si l'abkhaze n'est pas à proprement parler une langue peu décrite<sup>41</sup>, en revanche, nombre de dilemmes empiriques et théoriques restent à élucider, comme pour les autres langues de la famille CNO (kabarde, adyghé, etc.), notamment dans la perspective d'une modélisation phonologique à finalité réductionniste, afin de mieux comprendre la nature asymétrique de son inventaire segmental (entre 57 et 71 consonnes, selon le dialecte contre deux à trois voyelles) sur le plan typologique, et de désentrelacer la complexité des réalisations de surface, notamment sur le plan de l'allophonie vocalique, y compris sur le plan morphophonologique<sup>42</sup>. En outre, nombre d'études ont pour base empirique des textes oraux glosés, issus de la tradition orale, d'une indéniable qualité pour la morphosyntaxe et l'approche culturelle, mais on manque encore de données de première main, ciblées avec précision sur le plan des variables phonologiques (classes consonantiques de cavité orale – CO –, de résonances périphériques, labiale et pharyngale, de palatalité, de phonation – glottalité –, etc.), prises en compte dans le questionnaire de Colarusso.

Par exemple, pour le kabarde, Aert Kuipers propose la gamme allophonique suivante de vocoïdes (hors voyelles longues, issues de diphtongues décroissantes comme *əi, əw > [i:, u:]*, etc.), selon une taxinomie (critères classificatoires) qui tient compte des premières observations structuralistes, qu'on doit à N. Trubeckoj et à Jakovlev (tableau 1):

	<i>Avant</i>		<i>Central</i>		<i>Arrière</i>	
	<i>(plein)</i>	<i>plein</i>	<i>semi-arrondi</i>	<i>plein</i>	<i>semi-arrondi</i>	<i>arrondi</i>
<i>Elevé (higher)</i>	[i]	/ə/	[y, ʊ]	[ɨ]	[ʊ]	[u]
<i>Abaissé (Lower)</i>	[e]	[æ]	[ø]	/a/	[ɔ]	[o]

Tableau 4. Champ allophonique vocalique pour les phonèmes /ə/, /a/ du kabarde selon Aert H. Kuipers<sup>43</sup>, remanié.

38 Chiribka Viacheslev, *Abkhaz* (Munich: Lincom, 2003): 6.

39 Colarusso John, "Abkhazian Elicitation List" (manuscrit, inédit, 9 pages), 2018.

40 Yanagisawa Tamio & Tsvinaria-Abramishvili Ana, *Analytic Dictionary of Abkhaz* (Tokyo: Hituzi Syobo Publishing House, 2010).

41 Chiribka, *Abkhaz*, 17.

42 Cf. Colarusso John, *A Grammar of the Kabardian Language* (Calgary: The University of Calgary Press, 1992) pour une avancée dans ce domaine.

43 Kuipers Aert, *Phoneme and Morpheme in Kabardian (Eastern Adyghé)* (Janua Lin-

D'une part, nous proposerons une taxinomie plus parcimonieuse, basée sur une modélisation en termes de Géométrie des Traits (désormais, GT) et sur la base de relevés de phonétique expérimentale, à l'aide du logiciel PRAAT<sup>44</sup>, et en comparant avec une étude pilote sur le kabarde<sup>45</sup>, et des données comparatives de même ordre<sup>46</sup>.

	<i>+Avant</i>		<i>-Avant, - Arrière</i>	<i>+Arrière</i>
	<i>-Labial</i>	<i>+Labial</i>	<i>+RTR</i>	<i>+Labial</i>
<i>+ATR</i>	[i]	[y, ʏ]	-	[u]
<i>-ATR</i>	[ɪ]	[ə]	[ə, ɤ] /a/	[ɔ]

Tableau 5. Champ allophonique vocalique pour les phonèmes /ə/, /a/ de l'abkhaze standard.

D'autre part, nous tirerons les conséquences de cette grille réductionniste du tableau 2, ainsi que de la théorie des *traits linguaux de racine* (ATR: *Advanced Tongue Root* et RTR: *Retracted Tongue Root*) prônée par Vaux<sup>47</sup> dans le cadre de la GT, le consonantisme des langues CNO pouvant se décrire comme relevant d'un type déployant – outre les traits catégoriels de la représentation, pour la sonantinité ou la consonantinité, un éventail de traits de classe majeure répartis entre la COS (Cavité Orale Supérieure) et la COI (Cavité Orale Inférieure), avec pour sous-classes en COS les traits de lieu {labial, coronal +/-antérieur, dorsal +/-antérieur, guttural} et de manière {+/-continu}, associés à des *traits de résonance périphérique* {+/-arrondi} ainsi que des *traits linguaux* tels que {latéral} d'une part, {+/-ATR, +/-RTR} d'autre part, et des traits de COI laryngaux {+/-voisé, glotte fermée}. L'interaction de ces traits contoidaux qui enrichissent le domaine des attaques, toutes susceptibles de constituer des racines lexicales ou morphémiques (grammaticales) monoconsonantiques, avec un noyau léger assurant une syllabation par défaut (vocoïdes relevant du schwa), ou associés à une valeur marquée (la voyelle basse /a/, brève, ou longue, si associée à une fricative laryngale latente: /ah, ha/ = [a:]), se réalise

guarum, 8, Mouton & CO's-Gravenhage, 1960): 23.

44 Boersma Paul & David Weenink, *Praat: doing phonetics by computer* [logiciel libre], version 6.0.37, 2018, <http://www.praat.org/>.

45 Gordon Matthew & Applebaum Ayla, 2006. "Phonetic Structures of Turkish Kabardian", *JIPA* 36-2, (2006), 159-186.

46 Colarusso John 1994, "How to Describe the Sounds of the Northwest Caucasian Languages". In Howard Aronson (ed.), *Non-Slavic Languages of the USSR, Papers from the Fourth Conference* (Columbus, Ohio: Slavica Publishers, 1994), 61-113.

47 Vaux Bert, "The status of ATR in Feature Geometry". *Linguistic Inquiry* 26.1, 1996, 175-182 ; Vaux Bert 1997. "ATR Harmony in the Altaic Languages," *Proceedings of the 8<sup>th</sup> Biennial Non-Slavic Languages Conference*, Chicago, May 1993 (Chicago: Chicago Linguistic Society, 1997).

sur une gamme de traits. Cette gamme fait se contraster, hors des instances réalisées par *défaut* ou de manière non marquée, les valeurs chromatiques marquées PALATAL, LABIAL, PHARYNGAL, avec les corrélats de tonalité respectifs DIESE, BEMOLISE, GRAVE. Une contrainte mécanique de type PCO (Principe du Contour Obligatoire) conditionne l’harmonisation de ces traits sur les domaines nucléaires attribués aux attaques, comme dans les exemples suivants:

- (1) /i b z<sup>j</sup> ə j a ħ ə j t ?/<sup>48</sup> → [i b z<sup>j</sup> l j a ħ ə j t ?] ‘j’ai entendu ta voix’  
 (2) /j ə ħ w: ə j t ?/ → [j ə ħ w: ə j t ?] ‘nous l’avons dit’  
 (3) /ħ ə ħ w ə j t ?/ → [ħ ə ħ φ y i t ?] ‘nous avons crié’

En (1), le trait pharyngal (+RTR) du noyau de syllabation /ə/ est dés-associé, entre les valeurs palatales (diésées) de la fricative palatoalvéolaire à sa gauche et de l’approximante palatale suivante, aboutissant à une réalisation haute spécifiée [-ATR], avec allophone [i]. En (2), le trait bémolisé d’arrondissement associé à la fricative pharyngale à gauche n’a aucun effet sur le noyau bas phonémique subséquent, qui en revanche se rehausse (trait +ATR, ou „atérisation“) au contact du glide palatal subséquent. Enfin, en (3), le trait de bémolisation (ou trait labial) se déploie sous deux expressions allophoniques, contoïdale par fricativisation d’une part et approximante labiopalatale d’autre part, tandis que le vocoïde par défaut s’assimile avec le glide, qu’il syllabifie. Les contraintes de syllabification suivent des contraintes métriques élémentaires, dont on rendra compte par les positions squelettales dédiées dans les représentations en GT.

Ce sont là quelques effets de bascule de traits de tonalité et de syllabité, qu’une modélisation de type GT tenant compte des traits de racine linguale et des effets de PCO interdomaines syllabiques permet de saisir en vue d’une approche plus économique des conditions d’allophonie dans ce système réputé complexe, mais qui, comme toute langue naturelle, répond à des contraintes structurelles simples. Selon cette conception des interactions attaque-noyau et des contraintes métriques de syllabation, nous verrons que, finalement, ce type de système résout de manière tranchée et originale des conflits d’ordre général de concaténation de traits et de syllabation qui semblent opérer de manière plus simple dans des langues qui ne présentent pas une asymétrie aussi tranchée entre la densité de contoïdes et de vocoïdes. A ce titre, les langues de la famille CNO ont beaucoup à apporter à une théorie générale des traits phonologiques et de la syllabe, par le biais d’une complexité apparente, mise au profit de la résolution de conflits universaux entre traits de sonorité, traits de tonalité, et grilles métriques.

48 Les segments soulignés portent l’accent d’intensité – qui est distinctif, en abkhaz.

### 3. Perspectives

Ce bref survol des méthodes et des résultats ou premières avancées du projet LaDyCa permet de suggérer un certain nombre de pistes de recherche fondamentale, mais aussi appliquée, dans le domaine interdisciplinaire qui a été le nôtre, à la fois dans le cadre strict de ce projet, mais également dans ceux qui l'ont précédé, comme le projet méso-américain MAMP:

- 1) Le transfert de connaissances et de méthodes entre domaines géographiques et historiques ou civilisationnels très différents reste possible, voire s'avère être très fécond, pourvu que l'on centre l'attention sur des concepts généraux, comme la *verticalité* (notion d'archipel vertical), ou la *complexité des dynamiques géolinguistiques* ou des *structures* des langues naturelles (en morphologie flexionnelle, en phonologie).
- 2) Les retombées didactiques, sous forme d'ateliers thématiques, sont également possibles, en suivant les mêmes principes que ceux mis en application en Méso-Amérique: des collectifs de travail, notamment d'étudiants avancés en linguistique, ou d'instituteurs et de promoteurs culturels, travaillant à documenter (enregistrer et filmer), noter et systématiser des paradigmes de variétés dialectales des langues zanes ou du svan, peuvent apporter beaucoup à la revalorisation, à la conservation et la pérennisation ou la revitalisation de ces langues. De même, le questionnaire abkhaze pourrait, sous une forme remaniée, permettre d'envisager une approche sociophonétique de variétés d'abkhaze d'Abkhazie, mais aussi d'Adjarie et de la diaspora.

Enfin, tout se résume à une idée qui peut sembler déconcertante *a priori*, mais que nous avons expérimentée à diverses reprises, toujours avec succès: il n'est de science formelle qui ne trouve son application jusque dans les écoles primaires, pourvu qu'on sache enseigner les savoirs avec art. C'est cette „*touche artistique*“, ou „*art du pédagogue*“, qui permet de passer du complexe au simple, de l'abstrait au concret, de la science au jeu, de l'idée et de la pensée à la vie quotidienne. C'est là toute la raison d'être de ce que Jean-Louis Fossat, dialectologue toulousain émule de Jean Séguy, appelait la *dialectologie sociale*, et qui s'avère être l'un des horizons les plus marquants du projet LaDyCa pour son utilité auprès des sociétés du Caucase – l'un des foyers de complexité linguistique les plus dynamiques au monde, depuis des millénaires.



## References

- Adams, Paul. "Multilayered regionalization in Northern Europe". *GeoJournal*, 77-3 (2012), 293-313.
- Agresti, Giovanni. *Diversità linguistica e sviluppo sociale*. Milano: Franco Angeli editore, 2018.
- Assatiani, Nodar & Alexandre Bendianashvili 1997. *Histoire de la Géorgie*. Paris: L'Harmattan, 1997.
- Bonami Olivier & Gregory T. Stump. "Paradigm Function Morphology". Hippisley, Andrew & Stump, Gregory T. *Cambridge Handbook of Morphology*, 449-481. Cambridge University Press, 2016,
- Boersma, Paul & David Weenink. *Praat: doing phonetics by computer* [logiciel libre], version 6.0.37, 2018. <http://www.praat.org/>.
- Chiribka, Viacheslev. *Abkhaz*. Munich: Lincom, 2003.
- Colarusso, John. *A Grammar of the Kabardian Language*. Calgary: The University of Calgary Press, 1992.
- Colarusso, John. "How to Describe the Sounds of the Northwest Caucasian Languages". H. Aronson (ed.), *Non-Slavic Languages of the USSR, Papers from the Fourth Conference*, 61-113. Columbus, Ohio: Slavica Publishers, 1994.
- Colarusso, John. "Abkhazian Elicitation List" (manuscrit, inédit, 9 pages), 2018.
- Gordon, Matthew & Applebaum, Ayla. "Phonetic Structures of Turkish Kabardian", *JIPA*, 36-2, (2006), 159-186.
- Kuipers, Aert. *Phoneme and Morpheme in Kabardian (Eastern Adyghe)*. *Janua Linguarum*, 8. Mouton & CO's-Gravenhage, 1960.
- Le Guern Michel, Alain Berrendonner & Georges Puech. *Principes de grammaire polylectale*. Lyon, Presses Universitaires de Lyon, 1983.
- Léonard, Jean Léo. *Eléments de dialectologie générale*, Paris, Michel Houdiard Editeur, 2012.
- Léonard, Jean Léo. „Classes flexionnelles et diasystème: notes d'un terrain totonaco-tepehua“. In Iglesias, Aitor *et al.* (éds.), *Iñaki Gaminde lagunarterik hara* (Mélanges en l'honneur de Iñaki Gaminde). Bilbao: Universidad del País Vasco, 2017 (sous presse).
- Léonard Jean Léo & Alain Kihm. "Mazatec verb inflection: Revisiting Pike (1948) and comparing four dialects". In Léonard, J.L. & Kihm, A. (eds), *Patterns in Mesoamerican morphology*, 26-76. Paris: Michel Houdiard Editeur, 2014.
- Léonard Jean Léo, Els Heinsalu, Marco Patriarca, Kiran Sharma and Anirban Chakraborti. *Patterns of Linguistic Diffusion in Space and Time: The Case*

- of *Mazatec*. In F. Abergel *et al.* (eds.), *Econophysics and Sociophysics: Recent Progress and Future Directions*, 227-255. New Economic Windows, Springer, 2017.
- Makharoblidze Tamar & Jean Léo Léonard. "Paradigm Function Morphology applied to Kartvelian verb inflection", *The Second Language Dynamic in the Caucasus International Workshop*, 26<sup>th</sup> of February, Ilia State University, Tbilisi, Georgia, 2018.
- Montacié Claude & Hélène Gérardin. "Markov Models applied to Gurian data". *The Second Language Dynamic in the Caucasus International Workshop*, 26<sup>th</sup> of February, Ilia State University, Tbilisi, Georgia? 2018.
- Murra, John Victor, *The Economic Organization of the Inca State*, Ph.D. diss., Department of Anthropology, University of Chicago, 1956.
- Mufwene, Salikoko *The Ecology of Language Evolution*, Cambridge Approaches to Language Contact. Cambridge: Cambridge University Press, 2001.
- Mufwene, Salikoko *Complexity perspectives on language, communication, and society*. In Massip-Bonet Ángels. & Bastardas-Boada Albert. Springer Verlag, 2012, 197-218.
- Nichols Johanna, 2004, "The Origin of the Chechen and Ingush: A Study in Alpine Linguistic and Ethnic Geography", *Anthropological Linguistics*, 46-2, 2004, 129-155.
- O'Sullivan, David. „Complexity Science and Human Geography“, *Transactions of the Institute of British Geography*, 2004, 282-295.
- Picard, Flore, Hélène Gérardin & Jean Léo Léonard. "Levenshtein algorithm applied to Kartvelian phonological data", *The Second Language Dynamic in the Caucasus International Workshop*, 26<sup>th</sup> of February, Ilia State University, Tbilisi, Georgia, 2018.
- Pineda, Miguel Angel. "Diasistema lingüístico y cohesión interna. Un estudio lingüístico desde la panconía", *Sociolingüística andaluza, Metodología y estudios*. Publicaciones de la Universidad de Sevilla, 1985, 57-80.
- Stump, Gregory T. *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge: Cambridge University Press, 2001.
- Weinreich, Uriel. „Is a structural dialectology possible?“, *Word*, 4, 1954, 388-400.
- Stump, Gregory T. & Raphael Finkel. "Principal parts and morphological analysis. Presented at Workshop 'Morphological Complexity: Implications for the Theory of Language'", Harvard University, 22 January, 2010.
- Vaux, Bert. "The status of ATR in Feature Geometry". *Linguistic Inquiry* 26.1, 1996, 175-182.
- Vaux, Bert. "ATR Harmony in the Altaic Languages," *Proceedings of the 8<sup>th</sup> Bi-*

- ennial Non-Slavic Languages Conference*, Chicago, May 1993. Chicago: Chicago Linguistic Society, 1997.
- Wulf, Andrea. *The Invention of Nature. The Adventures of Alexander von Humboldt, The Lost Hero of Science*, Londres, John Murray, 2015.
- Yanagisawa Tamio & Tsvinaria-Abramishvili, Ana. *Analytic Dictionary of Abkhaz*. Tokyo: Hituzi Syobo Publishing House, 2010.