

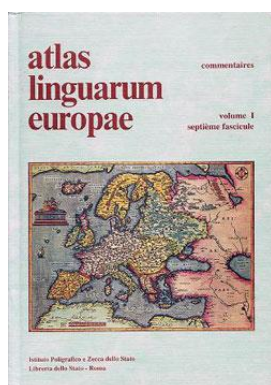
Projets de recherche

Sommaire

A. Atlas Linguarum Europae (ALE)	1
B. Atlas Linguistique Roman (ALiR)	2
C. Atlas Multimédia et Prosodique de l'Espace Roman (AMPER)	3
D. Atlas Linguistique Multimédia de la Région Rhône-Alpes et des régions limitrophes (ALMURA)	3
E. Documentation du Sarde	4
F. Projet ARC 5 2012 « Inventaire microtoponymique du massif de la Chartreuse »	5
G. Projet exploratoire Labex PERSYVAL	6
H. PEPS HuMain 2013	6

A. Atlas Linguarum Europae (ALE)

Ce projet a été fondé au début des années '70. Le Professeur émérite Michel Contini, dialectologue et phonéticien, collabore à ce projet : il est membre du Comité scientifique International. Il s'agit d'un atlas linguistique interprétatifs, de nouvelle génération. Son objectif est l'analyse comparée de la variation lexicale dans les dialectes des 21 domaines linguistiques que compte l'Europe regroupant des aires linguistiques génétiquement apparentées (domaines gallo-roman, domaine roman) ou supranationales et multilinguales.



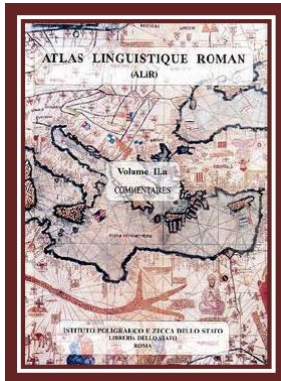
- Un réseau de 2631 points d'enquête.
- Chaque volume comporte un tome Atlas (les cartes), et un tome de commentaire (analyse linguistique de chaque carte).
- 9 volumes publiés ; 3 volumes en chantier.
- Projet parrainé par l'UNESCO.
- Membres du Comité national français : M. Contini et J.-E. Médélice (PR émérites, Univ. Grenoble Alpes & GIPSA-lab).
- Membre du Comité national italien : E. Carpitelli (PR, Univ. Grenoble Alpes & GIPSA-lab).

Cet atlas, ainsi que l'ALiR (ci-après), a permis le développement d'une approche théorique novatrice sur la variation lexicale prenant en compte la motivation à l'origine de la création des mots et aboutissant à un classement typologiques des variétés par référence à ce paramètre. Il existe plusieurs sortes de motivations :

- 1) onomatopéique : imitation de manifestations sonores à l'origine d'un grand nombre de désignations comme des zoonymes (grillon, coucou, caille...), des phénomènes naturels (tonnerre), ou des bruits du monde environnant ;
- 2) phonosymbolique : des unités phonétiques, isolées ou constituant des syllabes, simples ou redoublées, peuvent être porteuses d'information sémantique concernant, en particulier, la dimension, la distance, le mouvement, le degré d'une sensation, des sentiments, des impressions chromatiques.... ;

- 3) iconique : renvoie à l'aspect visible des référents ou à leurs qualités, à leurs relations avec les hommes, les animaux ou la nature, mais aussi aux traits socio-culturels, aux avancées techniques, aux idéologies dominantes, aux mythes, aux croyances, aux pratiques magico-religieuses ou à la religion en général - éléments qui caractérisent la vie des hommes depuis la nuit des temps.

B. Atlas Linguistique Roman (ALiR)



Directeur : M. Contini (PR émérite, Univ. Grenoble Alpes & GIPSA-lab).

Président : J. Veny (PR. Univ. de Barcelone).

Vice-présidents : P. García Mouton (PR. Univ. de Madrid/CSIC), N. Saramandu (PR. Univ. de Bucarest).

Directeurs adjoints : L. Massobrio (PR. Univ. de Turin), J. Saramago (Chercheur Univ. de Lisbonne/IJNIC).

Coordinatrice scientifique : E. Carpitelli (PR. Univ. de Nice)

Secrétaire scientifique : J.E. Médélice (PR, Univ. Grenoble Alpes & GIPSA-lab).

Participants équipe SLD : C. Chauvin-Payan (IE, Univ.

Grenoble Alpes & GIPSA-lab) et J.-P. Lai (IE, Univ. Grenoble Alpes & GIPSA-lab).

Le projet est né en 1986, à l'initiative du *Centre de Dialectologie de Grenoble*. Sa finalité est la réalisation d'un atlas linguistique interprétatif de deuxième génération, contenant la totalité des variétés dialectales de l'Europe romane sous les aspects lexicaux, phonétiques, phonologiques et morfo-syntaxiques.

- Un réseau de 1037 points d'enquête.
- Huit partenaires réunis par comités nationaux ou par domaines linguistiques : Portugal, Espagne (castillan, galicien, catalan), France, Belgique (romane), Suisse (romande, rhéto-romane, italo-romane), Italie, Roumanie, République de Moldavie.
- Comme pour l'ALE, chaque volume comporte un tome atlas, avec les cartes, et un tome avec l'analyse linguistique de chacune d'entre elles.
- 3 volumes imprimés ; 4^{ème} volume en cours d'impression.
- Une convention éditoriale est établie entre le Centre de Dialectologie et l'Istituto Poligrafico e Zecca dello Stato de Rome (Imprimerie Nationale de l'Italie).

C. Atlas Multimédia et Prosodique de l'Espace Roman (AMPER)



Michel Contini (PR émérite Univ. Grenoble Alpes & GIPSA-lab),

Antonio Romano (Univ. de Turin & Laboratoire de phonétique expérimentale *Arturo Genre*).

Secrétariat scientifique:

Jean-Pierre Lai (IE, Univ. Grenoble Alpes & GIPSA-lab)

Valentina de Iacovo (doctorante Univ. de Turin)

Silvia Gally (IE, Univ. Grenoble Alpes & GIPSA-lab),

Création et gestion de la base de données :

Albert Rilliard (LIMSI/ Orsay)

Coordination générale :

Ce projet naît en 2001 sous l'impulsion de Michel Contini, d'Antonio Romano, de Stefania Roulet et de Jean-Pierre Lai. Il s'agit d'un atlas linguistique qui s'intéresse à la description de la prosodie et, en particulier, de l'intonation des phrases interrogative et déclarative, dans les variétés romanes d'Europe et d'Amérique latine. Les données sont recueillies, en respectant un même protocole, au cours d'enquêtes de terrain réalisées par les spécialistes de chaque domaine linguistique. Elles sont ensuite analysées et interprétées, dans le respect d'une même approche méthodologique, en vue de la réalisation d'une typologie intonative des variétés étudiées et de l'évaluation des distances prosodiques relatives. [AMPER](#) fait partie d'une nouvelle génération d'Atlas, interprétatifs et multimédia, avec [une base de données interrogeable en ligne](#), sur internet.

- 110 enquêtes ont déjà été réalisées en Europe et en Amérique Latine.
- [Neufs équipes partenaires de recherche](#) : France, Italie, Portugal, Roumanie, Galice, Espagne, Catalogne, Asturies, Roumanie. De nouvelles équipes d'Amérique Latine participent au projet : Brésil, Chili, Venezuela, Cuba, Costa-Rica.
- 290 publications et 8 ouvrages, consacrés au projet.
- Publications ELLUG des numéros *Géolinguistique Hors-Série 3* et *Géolinguistique Hors-Série 4* (données figurant sur CD-ROM).

D. Atlas Linguistique Multimédia de la Région Rhône-Alpes et des régions limitrophes (ALMURA)



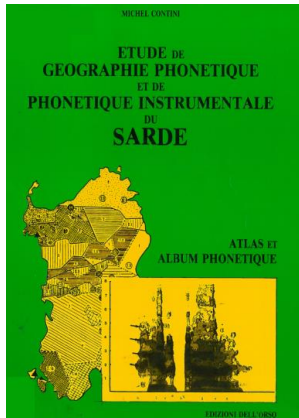
Image tirée de <http://www.atlas-almura.net/Atlas-Almura.html>

Direction: J.-E. Médélice (PR émérite, Univ. Grenoble Alpes & GIPSA-lab)

Nombreux participants associatifs et universitaires.

Il s'agit d'un projet qui explore en profondeur les aires linguistiques francoprovençales et occitanes de la région Rhône-Alpes (Isère, Drôme, Ardèche, Rhône, Loire, Ain, Haute-Savoie et Savoie). Si cet Atlas peut être considéré comme un atlas traditionnel dans la mesure où il présente des données «brutes», il appartient en même temps à la génération des «Atlas parlants» disposant d'une BD-audio consultable sur internet.

E. Documentation du Sarde



Coordination du projet : Giovanni Depau (MCF, Univ. Grenoble Alpes & GIPSA-lab).

Description du projet

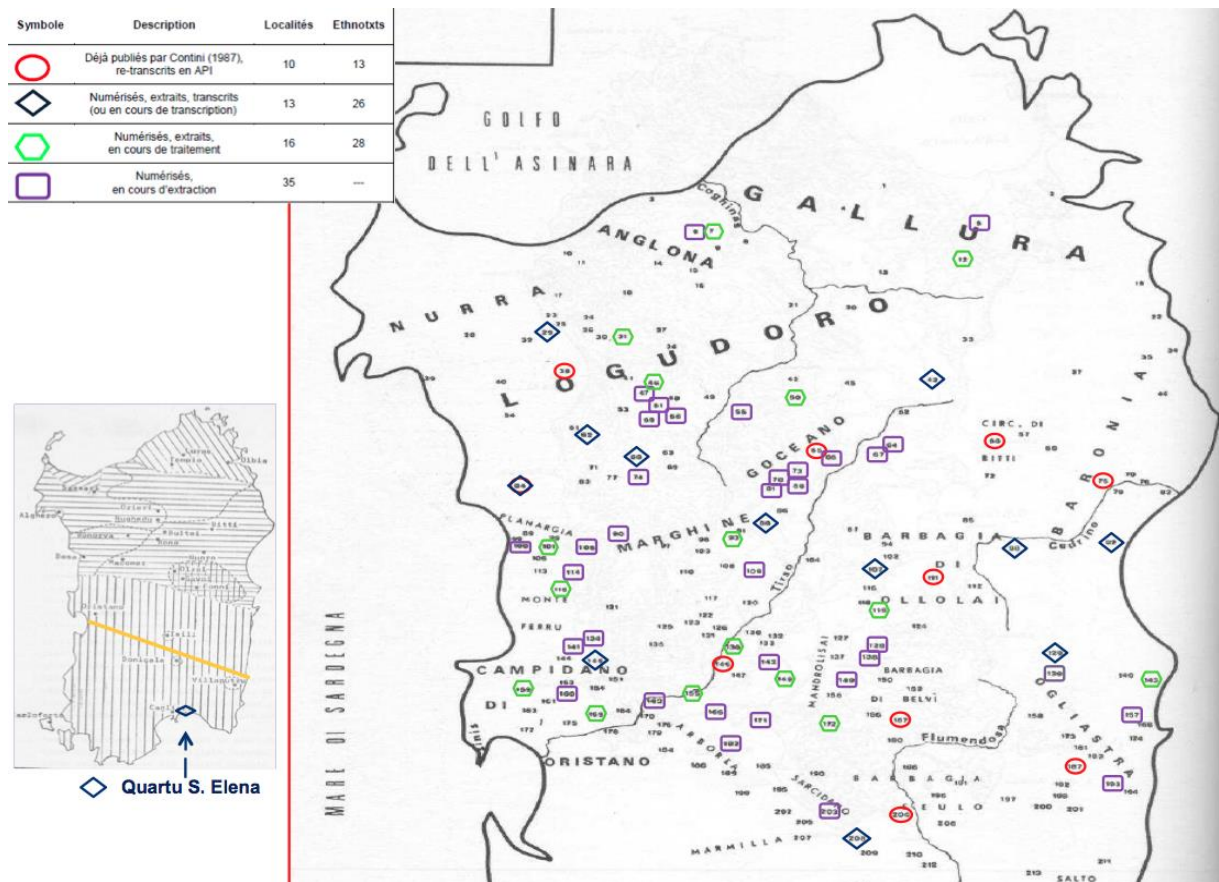
Il s'agit d'un projet de documentation ethnolinguistique concernant la Sardaigne. Le but étant de constituer et implémenter une base de données lexicale et d'éthnotextes concernant l'espace sarde centro-septentrional. Ce projet a bénéficié d'une bourse Explora Pro CMIRA, Rhône-Alpes 2010 pour une collaboration avec l'Institut de l'Atlas Linguistique Italien (ALI) de Turin (Italie).

Synthèse & Objectifs du projet

Les données faisant l'objet de ce projet sont constituées d'un corpus d'environ 250 heures de matériel audio recueilli par Michel Contini dans les années '60-'70 dans le but de fournir une documentation détaillée du phonétisme des variétés parlées dans le même domaine. À partir de ces recherches un atlas phonétique du sarde a déjà été réalisé dans les années 1980 (Contini, M. (1987), *Étude de géographie phonétique et de phonétique instrumentale du Sarde*, Alessandria : Dell'Orso).

L'objectif de ce projet est de sauvegarder le matériel (numérisation des bandes par J-P Lai – 2010), enregistré sur de vieilles bandes sonores destinées à l'usage, et d'en extraire un vaste ensemble d'éthnotextes et d'informations sur les pratiques discursives en sarde - celles se focalisant sur les activités de travail et sur les pratiques sociales. Cela nous permet de disposer de données du sarde pouvant être étudiées à travers l'adoption d'une approche à la fois linguistique et socioculturelle.

Répartition des ethnotextes dans l'espace géolinguistique sarde considéré.



F. Projet ARC 5 2012 « Inventaire microtoponymique du massif de la Chartreuse »

Après une phase préliminaire effectuée en 2011 (en partenariat avec le *Comité de recherches et de promotion de l'histoire et de l'art* à Voreppe / COREPHA) et destinée à mesurer l'ampleur du chantier ainsi que tester la motivation des éventuels partenaires, il a été décidé de mettre en place une série de conférences et d'ateliers afin de mobiliser le plus grand nombre. Il s'agit d'*Ateliers d'initiation aux techniques de recherche en toponymie* au cours desquels sont abordées les phases nécessaires à l'établissement d'un inventaire : découverte des sources indispensables (dictionnaires topographiques, cadastres ...). Ces ateliers incluent une initiation à la maîtrise (et peut-être la découverte) de l'outil Internet : « géo-portail / consultation en ligne des archives, des dictionnaires topographiques » ...

Le pourquoi d'une telle méthodologie réside dans le fait qu'un projet d'une telle ampleur ne peut en aucun cas se réaliser à la seule échelle universitaire pour plusieurs raisons, car les intervenants universitaires (enseignants / chercheurs) ne peuvent consacrer l'intégralité de leur temps de recherche à un seul projet et, surtout, parce qu'il est important d'associer les habitants du Massif de la Chartreuse. Les plus anciens d'entre eux sont de véritables mines d'informations : ils ont une connaissance sans égale du terrain et, pour la plupart, une bonne connaissance du francoprovençal, langue à l'origine d'une majorité de toponymes locaux.

Ce projet se veut d'une large ouverture vers le grand public d'où vient, paradoxalement, un grand nombre de questions touchant à la toponymie. En effet, les associations œuvrant pour la défense du patrimoine local et les Universités populaires sont en demande constante d'interventions sur ce sujet (et le nombre de participants est toujours très important).

G. Projet exploratoire Labex PERSYVAL

GeoDialect. Exploration des outils géomatiques pour le traitement et l'analyse des données géolinguistiques : application à la dialectologie.

Le projet exploratoire que nous proposons s'intéresse à l'amélioration de la gestion des documents cartographiques utilisés dans le traitement et l'analyse des données géolinguistiques, au moyen de méthodes géomatiques innovantes. Plus précisément, nous nous concentrons sur un outil de la dialectologie - les Atlas Linguistiques - et sur la manière dont nous pouvons transférer ces travaux de leur support initial papier, vers un support numérique et informatisé. Nous prenons en considération l'*Atlas Linguistique de la France* (Gilliéron et Edmont, 1902-1910), considéré comme le point de départ de la géolinguistique moderne (639 points d'enquêtes et autour de 1700 cartes). Il s'agit de récupérer les données qui le composent par le biais de moyens informatisés et plus particulièrement des outils de type Système d'Information Géographique.

Porteurs : Didier Demolin (PR, Univ. Grenoble Alpes [GIPSA-lab/UMR 5216]).

Participants LIG (UMR 5217), équipe Steamer : Paule-Annick Davoine (MCF), Marlène Villanova-Oliver (MCF).
Participants GIPSA-lab (UMR 5216), équipe SLD : Carole Chauvin Payan (IE), Silvia Gally (IE), collaborateurs: Michel Contini (PR émérite, Univ. Grenoble Alpes) et Giovanni Depau (MCF, Univ. Grenoble Alpes)

H. PEPS HuMain 2013

CartoDialect. Extraction d'informations sémantiques et géographiques à partir des données géolinguistiques : le cas de l'Atlas Linguistique de France.

Le projet CartoDialect s'inscrit dans l'axe « Analyse et usages, technologies de numérisation, reconnaissance d'images et de caractères, techniques de restauration numérique des manuscrits anciens » de l'appel à projet PEPS HuMain. Il s'intéresse à l'extraction d'informations sémantiques et géographiques contenues dans les documents cartographiques utilisés pour le traitement et l'analyse des données géolinguistiques de l'Atlas Linguistique de France, au moyen des méthodes de numérisation et de vectorisation et de géoréférencement/géocodage. Ce projet s'inscrit dans une démarche pluridisciplinaire regroupant des équipes en informatique (L3i), en géolinguistique (GIPSA-Lab, SLD) et en traitement de l'information géographique et géomatique (LIG-Steamer).

Porteur : Paule-Annick Davoine (MCF, Grenoble-INP [LIG-Steamer/UMR 5217]).

Participants LIG (UMR 5217), équipe Steamer : Marlène Villanova-Oliver (MCF, UPMF).

Participants GIPSA-lab (UMR 5216), équipe SLD : Carole Chauvin-Payan (IE, Univ. Grenoble Alpes), Didier Demolin (PR, Univ. Grenoble Alpes), Silvia Gally (IE, Univ. Grenoble Alpes)

Participants L3i (AE 2118), équipe ASPIC : Jean Christophe Burie (MCF, Univ. La Rochelle), Mikael Coustany (IR, Univ. La Rochelle), Jean-Marc Ogier (PR, Univ. La Rochelle).

**ANR 2015 : ECLATS : Extraction automatisée des Contenus géolinguistiques
d'Atlas et analyse Spatiale : application à la Dialectologie**

1 RESUME DU PROJET

The ECLATS project deals with the enhancement and the treatment of ancient maps, a historical and cultural heritage widely recognized as a very important source of information, but not easy to use. We focus on the Linguistic Atlas of France (ALF), built between 1902 and 1910. This cartographical heritage produces first-rate data for dialectological researches.

Dialectology addresses the study of the linguistic features of languages having a strong oral tradition such as local dialects. These linguistic features are diverse : phonetic, morphosyntactic, lexical, semantic or prosodic. They are also context-dependant: they evolve according to (geographical) space, time, and sociocultural environment. To study local dialects, corpuses of phonetic data have been transcribed into linguistic atlases. These atlases consist of a set of maps where, for a given notion, linguistic forms are saved and collected at several geographical points. Made up of 1900 maps, representing 639 points of inquiries, the ALF is a data corpus of 1.214.100 reliable lexical data, noted in a homogeneous way, collected from a single questionnaire, with details about places, dates and circumstances.

While theoretical approaches used for the construction of linguistic atlas are structured, reliable and homogeneous, the processing and the phonetic analysis of data, as well as the elaboration of interpretative maps, are still manually achieved, based on unstandardized approaches. Moreover, software solutions are missing in geolinguistics: the dematerialization of old atlases is not systematic; the extraction of data from old paper maps is made by hand and very time-consuming; the use of Geographic Information Systems and spatial analysis methods is undeveloped and lexical or interpretative maps are still hand-drawn. This situation delays both the processing and the diffusion of data and it also limits the cartographic production capacities, by impeding an efficient exploitation of geographical knowledge by researchers in dialectology.

The ECLATS project is a multidisciplinary project that will take 48 months. It gathers researchers in geomatics, computer science and geolinguistics. Its global purpose is the design and the development of innovative methods and tools for extracting and analysing the linguistic features and the geographical data included in collection of old cartographic documents used by researchers in dialectology. More precisely, four goals will be pursued:

- The design and definition of new models and standard format(s) for geolinguistic data fostering interoperability between geolinguistic software.
- The development of storage methods of digitalized maps in order to improve their use and dissemination; the development of new and innovative digitalization for the phonetic characters' recognition techniques for the automatic extraction of old map contents. Such techniques will be tested on the ALF but should be applicable on any old map document.
- The definition of a methodology implemented in a software suite dedicated to experts in geolinguistics, with advanced geovisualization and spatial analysis functionalities, for geolinguistic data processing. Such an (r)evolution is required and crucial in Geolinguistics for upgrading and optimizing both the production and the analysis of new interpretative atlases.
- To save a unique linguistic heritage and to promote a collaborative approach for encouraging the sharing and the dissemination of geolinguistic data.

Partners of the project are research groups in Computer Science specialized in geomatics (LIG, Grenoble), in digitalization of ancient documents (LIRIS, Lyon), in automatic content extraction (LIRIS & Li3) and in geolinguistics and dialectology (Gipsa-lab, Grenoble).

Porteur : Paule-Annick Davoine (MCF, Grenoble-INP [LIG-Steamer/UMR 5217]).

Participants LIG (UMR 5217), équipe Steamer : Marlène Villanova-Oliver (MCF, UPMF).

Participants GIPSA-lab (UMR 5216), équipe SLD : Silvia Gally (IE, Univ. Grenoble Alpes), Carole Chauvin-Payan (IE, Univ. Grenoble Alpes), E. Carpitelli (PR, Univ. Grenoble Alpes), G. Depau (MCF, Univ. Grenoble Alpes), M. Contini (PR émérite, Univ. Grenoble Alpes), JE. Médélice (PR émérite, Univ. Grenoble Alpes),

Participants L3i (AE 2118), équipe ASPIC : Jean Christophe Burie (MCF, Univ. La Rochelle), Mikael Coustany (IR, Univ. La Rochelle), Jean-Marc Ogier (PR, Univ. La Rochelle).

CONTEXTE ET PROBLEMATIQUE

La dialectologie s'intéresse à l'étude des traits linguistiques caractéristiques des langues à tradition orales comme les parlers locaux, appelés patois ou, encore, dialectes. Ces traits linguistiques peuvent être de nature très différente - phonétique, morphosyntaxique, lexicale, sémantique ou prosodique - et évoluent dans un espace géographique donné, dans le temps et au contact de la société : en linguistique, on parle de variations diatopique, diachronique et diastratique/diaphasique. Pour étudier les parlers locaux, la dialectologie s'est spécialisée dans la constitution de corpus de données descriptives, collectées via une méthodologie d'enquête qui repose sur des questionnaires, sur le choix de réseaux de points linguistiques et d'informateurs (Contini 2006). Le traitement et l'analyse des données de terrain se fait au moyen de supports cartographiques, sur lesquels sont portées sous forme de points les localités enquêtées et les formes linguistiques collectées en transcription phonétique. À chaque concept exprimé par une entrée lexicale sous la forme d'un titre de carte en français est associée une et une seule carte sur laquelle figurent toutes les formes dialectales transcrites phonétiquement désignant le concept en question. L'Atlas Linguistique de France fut le premier du genre : après une campagne de terrain menée entre 1897 et 1901 dans 639 localités, J. Gilliéron et E. Edmont publient les résultats de 1902 à 1910. L'ALF concerne le domaine des dialectes gallo-romans de France, ainsi que d'une partie de la Belgique, de la Suisse, de l'Italie, et débordé également sur le domaine catalan de France et, dans un volume spécifique, sur la Corse. La publication de l'œuvre, en format papier, comporte 35 fascicules, réunissant en 12 volumes, 1920 cartes géolinguistiques qui présentent une photo instantanée de la situation dialectale de la France à la fin du XIXe siècle. L'ALF peut être défini comme un atlas de première génération publiant des données brutes et constituant un corpus de plus d'un million de données lexicales fiables, transcrites de façon homogène à l'aide de l'alphabet phonétique Rousselot-Gilliéron. L'ensemble a été réalisé à partir d'un même questionnaire, précisant les lieux de référence, les dates, les témoins et les conditions des enquêtes. Il constitue une immense base de données sur support papier. C'est à partir de ces matériaux et des analyses qui en découlent que Gilliéron fonde la géographie linguistique du XXe siècle. L'ALF va ainsi servir de modèle à de nombreux travaux de dialectologues du monde entier¹ ; la France reste de son côté très engagée dans le domaine de la géolinguistique et l'atlantographie, avec la mise en place, dans les années 50 du XXe siècle, de la grande entreprise des Atlas Linguistiques de la France par Régions qui a pris en compte aussi des thèmes plus adaptés aux différentes réalités régionales. Les données des atlas linguistiques et notamment de l'ALF sont utilisées pour élaborer des atlas dits de deuxième génération c'est-à-dire interprétatifs où les données « brutes », de terrain, sont utilisées pour présenter de manière synthétique, en se fondant sur une analyse typologique, l'extension et la distribution des traits linguistiques dans une aire géographique prédéterminée. Par ailleurs, l'étude de la variation phonétique, morphologique et lexicale dans un espace dialectal met en évidence la distribution des différentes réalisations phonétiques (consonnes, voyelles, structures syllabiques, patrons accentuels), des éléments morphologiques, des aboutissants de différentes bases étymologiques. Elle peut se concentrer sur la dimension synchronique des phénomènes (analyse de la situation actuelle), aussi bien que sur la dimension diachronique (analyse de la variation des phénomènes dans le temps). Cette dernière perspective se fonde sur l'hypothèse que chaque localité conserve une phase du changement linguistique et contribue ainsi à la reconstruction d'un phénomène en diachronie : l'évolution dans le temps est donc projetée dans l'espace.

L'intérêt des recherches dialectales pour la linguistique n'est plus à démontrer et dépasse même le cadre de cette discipline. L'attention que les dialectologues portent sur la relation entre l'objet et le mot qui le désigne rejoint aussi les préoccupations d'autres domaines scientifiques. Les réflexions sur la variation dialectale dans les dimensions diatopique, diastratique et diachronique ouvrent à la collaboration avec les géographes, les sociologues, les historiens, les généticiens. Émergent alors plusieurs problématiques :

- 1) la sauvegarde des atlas : compte tenu de l'état de vieillissement de ce type d'ouvrages, il devient urgent d'engager un processus de numérisation systématique ;
- 2) la valorisation du contenu des atlas linguistiques et de l'ALF en particulier, aussi bien à des fins scientifiques que culturelles : au risque que ces atlas deviennent « des monuments reliquaires », mais aussi pour favoriser leur exploitation et contribuer à améliorer la connaissance de ce patrimoine culturel exceptionnel et unique, il est nécessaire de mettre en place des outils facilitant l'accès aux données et leur exploitation ;

¹ Les travaux de Gilliéron et Edmont ont inspiré aussi la création de l'*Atlas Linguistique d'Italie et de Suisse* (AIS) par Karl Jaberg et Jakob Jud (1928) et de l'*Atlante Linguistico Italiano* (ALI) par Matteo G. Bartoli (1924).

3) le traitement des données dialectales : l'outillage logiciel destiné à l'exploitation des données géolinguistiques est très en retard. Si l'approche théorique utilisée lors de la construction d'atlas linguistiques est structurée, fiable et homogène, le traitement et l'analyse phonétique, l'élaboration des cartes lexicales, ainsi que la réalisation des atlas interprétatifs, s'appuient encore sur des approches manuelles. Pour élaborer les cartes interprétatives, les géolinguistes doivent procéder à des regroupements de formes linguistiques visant à réduire la variation, tout en construisant manuellement des aires linguistiques définies par des isoglosses lexicales (lexique) ou des isophones (phonétique). Le recours aux systèmes d'information géographique (SIG), largement reconnus pour leur capacité cartographique et les outils d'analyse spatiale qu'ils proposent, est encore peu développé dans le domaine de la géolinguistique (Hoch & Hayes 2010 ; Sibling & al. 2012). La géovisualisation et l'analyse exploratoire des données spatiales ne sont pas utilisées.

Traditionnellement, les données d'un atlas linguistique apparaissent sous forme de cartes sur lesquelles les transcriptions phonétiques sont accolées à chaque point d'enquête. L'usage d'un SIG implique donc à minima de saisir les points d'enquête et les formes dialectales associées à chaque entrée linguistique, de les géocoder avant de les organiser sous forme de couches d'information géographique. Sans automatisation, l'ensemble des étapes décrites ici ne peut être réalisé à l'échelle d'un million de données à saisir pour l'ALF.

Par ailleurs, les données contenues dans les cartes de l'ALF, sont transcrites dans un alphabet phonétique spécifique, l'alphabet Rousselot-Gilliéron, et nécessitent donc une traduction en Alphabet Phonétique International afin de permettre leur diffusion auprès d'un plus large nombre de linguistes. Cette étape, jusqu'à présent effectuée manuellement par le dialectologue, augmente le temps consacré à la saisie des données, et nécessite aussi une certaine expertise. Ceci constitue un véritable frein à la diffusion et la réutilisation des données par les différentes communautés scientifiques concernées, et au développement de l'innovation dans le domaine aussi bien en termes de méthodes, d'outils que d'analyses.

OBJECTIFS

Le projet ECLATS a pour objectif de valoriser les atlas linguistiques anciens. Cette valorisation couvre deux aspects complémentaires :

- Permettre des recherches novatrices en géolinguistique et plus particulièrement en dialectologie en appliquant en particulier des techniques d'analyse spatiale et de géovisualisation innovantes aux données sémantiques et géographiques extraites de cartes anciennes,
- Faciliter l'accès tant aux documents originaux qu'aux données et résultats issus de leur exploitation. En nous appuyant sur l'emblématique Atlas Linguistique de France, il s'agit d'étudier à la fois les aspects méthodologiques et techniques pour intégrer au sein d'une plateforme dédiée à la géolinguistique un ensemble d'outils couvrant à la fois le processus d'extraction automatique de données à partir des cartes anciennes numérisées et leur interprétation à l'aide d'analyses spatiales et de traitements cartographiques appropriés. Ainsi, les principaux résultats visés par le projet sont :
 - Le développement de techniques d'extraction de contenus et de stockage des données et images numérisées visant à préserver et à faciliter l'exploitation et la diffusion des cartes de l'ALF
 - La définition de modèles combinant les dimensions linguistiques, spatiales et temporelles pour représenter les données géolinguistiques. La définition de méthodes d'analyse spatiale et de géovisualisation permettant d'optimiser la production d'atlas interprétatifs et plus généralement les recherches en géolinguistique. Ces méthodes s'appuieront bien entendu sur les modèles définis précédemment, mais devront également permettre de combiner les données issues de l'ALF avec d'autres types de données géographiques, géolinguistiques, géo-historiques (cartes ou données anciennes géo-référencées), socio-économiques, etc. permettant ainsi de mieux établir des relations entre processus linguistiques, environnementaux et sociétaux. L'intégration de ces données, modèles et outils au sein d'une plateforme destinée à la fois aux géolinguistes et dialectologues pour mener et partager leurs propres interprétations et aussi à la sauvegarde, la diffusion et la valorisation culturelle de ce patrimoine linguistique unique et du travail (méconnu) des géolinguistes auprès d'un public varié.

Pour le domaine de la géolinguistique, le projet ECLAT répond à un besoin de (r)évolution nécessaire et indispensable des processus de recherches. Au-delà de ces bénéfices immédiats pour la géolinguistique et la dialectologie, nous espérons plus généralement dégager un certain nombre de solutions pouvant être appliquées à la valorisation de toute carte ancienne dont on souhaite exploiter les symboles et leur localisation.

POSITIONNEMENT DU PROJET ET CARACTERE AMBITIEUX/NOVATEUR DU PROJET

Le projet ECLATS s'inscrit dans la continuité de travaux engagés en 2013 par le consortium dans le cadre de projets exploratoires financés par le Labex PERSYVAL² (GéoDialect) et la Mission interdisciplinaire du CNRS³ (CartoDialect), portant sur le traitement des données géolinguistiques via des approches géomatiques.

ETAT DE L'ART

Dès les années 90, plusieurs auteurs (Ormeling, 1992 ; Lee J. & Kretschmar, 1993), montraient l'intérêt d'utiliser les fonctions des SIG pour les études géolinguistiques, mais les recherches actuelles sur le sujet restent encore peu répandues (Luebbering & al. 2013). Les récentes avancées technologiques se sont surtout attachées à dématérialiser les atlas papiers⁴, afin de les mettre en ligne tel que l'Atlas Linguistique et Ethnographique de l'Italie et de la Suisse méridionale (<http://www3.pd.istc.cnr.it/navigais/>), et en associant parfois aux anciens atlas linguistiques les carnets d'enquête (cf. l'Atlas Linguistique de la Péninsule Ibérique⁵). D'autres projets sont allés plus loin en combinant les données brutes (sonores ou retranscrites) avec des données analysées et lemmatisées ainsi que des représentations de données sur carte, comme la Banque de données de la langue corse (<http://bdlc.univ-corse.fr/>) ou le Thesaurus Occitan⁶ (THESOC, Dalbera & al., 2012). Dans ces bases de données qui souvent disposent de modules phonétiques ou morphosyntaxiques sophistiqués l'intégration de fonctionnalités fondées sur le principe des SIG et de l'analyse spatiale pour l'exploitation et la représentation cartographique des données est très peu développée.

Or, Lee & Kretschmar (1993) évoquaient déjà l'intérêt des fonctionnalités de superposition et de relations spatiales proposées par les SIG, aussi bien pour identifier des compositions linguistiques en combinant différentes couches de données linguistiques, que pour mettre en relation ces dernières avec des informations géographiques décrivant les facteurs démographiques, politiques, environnementaux...

Du point de vue de l'analyse spatiale, différents auteurs se sont intéressés à l'analyse de la variation, en s'appuyant notamment sur des approches de modélisation telles que le *Gravity model* de Trudgill (1974) ou le *Wave model* de Bailey (1993) et Boberg (2000). Grieve (2011) analyse les variations linguistiques régionales en s'appuyant sur des approches statistiques telles l'analyse factorielle et l'analyse en cluster ou l'autocorrélation spatiale. Toutefois, la diffusion de ces approches est encore très limitée en géolinguistique.

Du point de vue de la cartographie des données linguistiques, certains auteurs ont travaillé sur différents types de représentations cartographiques et sur la sémiologie graphique à adopter en linguistique, sans toutefois définir des propositions concrètes et opérationnelles⁷ (Ormeling, 1992 ; Kanellos & al., 1999 ; Silber & al., 2012 ; Dell'Aquila & Iannàccaro, 2013). Les avancées les plus marquantes sur le plan de la visualisation cartographique se sont concentrées sur l'usage de cartes interactives intégrées dans des interfaces multimédia et multivues, comme par exemple le projet ALAVAL⁸ (Diémoz & Kristol, 2012). Face aux difficultés rencontrées pour réaliser des cartes interprétatives lisibles et efficaces Hoch et Hayes (2010) et Luebbering & al. (2013) évoquent la nécessité de mener une réflexion globale sur les types de représentations cartographiques et sur la sémiologie graphique adaptée aux données géolinguistiques et proposent quelques solutions méthodologiques et techniques concrètes.

Sur le plan de l'extraction de contenus, des travaux portant sur la dématérialisation de cartes anciennes (180 cartes de Cassini) en vue de la vectorisation de leurs frontières et de leurs symboles ont été réalisés dans le cadre de l'ANR GéoPeuple⁹. Sur le plan de l'analyse des images, un travail d'adaptation d'outils utilisés en vision par ordinateur pour la segmentation et la détection d'objets a été réalisé : il a conduit à la définition de caractéristiques de contenus plus spécifiques au document ancien (essentiellement fondées sur des histogrammes de gradients orientés) et a permis de souligner l'importance du processus de numérisation

² ANR--11-LABEX-0025, <https://persyval-lab.org/>

³ Projet Exploratoire PEPS « en réseau » HuMain 2013 & 2014 <http://www.cnrs.fr/mi/spip.php?article291>,

⁴ Nous ne tenons pas compte ici des atlas linguistiques qui se configurent en réalité comme des bases de données dialectales sonores sans prise en compte de la cartographie, comme c'est le cas de l'Atlas Multimédia de l'Espace Linguistique Roman (AMPER) :

<http://amper.limsi.fr/> et <http://dialecto.ugrenoble3.fr/AMPER/amper.htm> (Contini & al. 2002)

⁵ http://westernlinguistics.ca/alpi/?global_lang=fr

⁶ <http://thesaurus.unice.fr/index.html>

⁷ En particulier, Dell'Aquila & Iannàccaro (2013) proposent des types de cartographie de données issues d'enquêtes sociolinguistiques à grande échelle et non de données provenant d'atlas linguistiques.

⁸ Projet ALAVAL (Atlas Linguistique audiovisuel du francoprovençal valaisan)

<http://www5.unine.ch/dialectologie/Atlas001/001AtlasDemo.htm>

⁹ <http://geopeuple.coriolys.fr/pages/anr>

des cartes et la prise en compte du contexte comme cela est rendu nécessaire pour la compréhension du contenu (Divala & al. 2009 ; Heitz, 2008). La spécificité des cartes anciennes, la diversité des représentations par traits et par textures, la présence de nombreuses occultations et l'existence d'une grande variabilité de représentation des symboles sont restés des verrous difficiles à lever. L'automatisation complète d'un processus de détection d'informations sur des cartes anciennes se révèle être le fruit d'expertises pluridisciplinaires.