

ANNEXES

Statuts et règlement du laboratoire

UMR de linguistique : tableau récapitulatif

La recherche à la BnF

Proposition de convention avec la BnF : CODEX

CAT COD : Appel à colloque 2008

Revue de Sémantique et Pragmatique (2007-2010)

Revue CORELA

CERLICO : Présentation

CERLICO : 24^e colloque (Tours 2009)

Collaboration avec le LATTICE

Rapport de la thématique : « Construction du sens »

Journée d'étude sur la temporalité : programme

ESLO : VARILING (présentation)

GUYANE : Encyclopédies

GORDF : Séminaire

PROHEMIO : dépouillement des ouvrages pour un dictionnaire des figements

PROHEMIO : Sommaires de la revue

CRITERR (Orléans) : recherches en cours

RFP : Appel à communication et programme

MSH : Projet collaboratif

Statuts et règlement du LLL contrat 2008 – 2011
(adoptés : assemblée générale du 17 janvier 2008)

Laboratoire Ligérien de Linguistique, EA 3850 / ILF – CNRS

Directeur Gabriel Bergounioux (Université d'Orléans)

Directeur-adjoint, responsable de l'équipe de Tours, Jean-Michel Fournier

1. Objectifs

Recherche linguistique, formation des doctorants, support de formation de master, convergence des recherches sur le plan régional (Fédération FELINCO – Master international FELINCO), partenariats multiples selon les objectifs scientifiques.

2. Organigramme thématique

A Tours (3 axes et 1 atelier thématique)

- morphophonologie
- analyse du discours
- diachronie de l'anglais
- construction du sens

A Orléans (5 équipes)

- ESLO
- Langues d'Afrique et créoles
- Langues de Guyane
- GORDF
- PROHEMIO

3. Organismes de rattachement

Université François-Rabelais, Ecole Doctorale SHS, UFR Lettres & Langues)

Université d'Orléans, Ecole Doctorale SHS, UFR Lettres, Langues et Sciences Humaines

4. Membres : critères de définition des membres permanents et associés
Membres permanents : rattachement principal au laboratoire

Membres associés : collaborateurs sur programmes : professeurs émérites, doctorants, enseignants-chercheurs en rattachement secondaire, PRAG, enseignants de l'IUFM, chercheurs étrangers...

5. Droits et obligations des membres

- Droits : accès aux locaux et au matériel, support financier de leur activité de recherche, voix délibératives (ou consultatives pour les associés) à toutes les décisions du groupe de recherche (administratives, organisationnelles, financières)
- Obligations
 - ➔ Participation aux séminaires et implication dans des projets de recherche au nom du laboratoire
 - ➔ Encadrement doctoral : respect de la charte des thèses, des règles de l'Ecole Doctorale
 - ➔ Signature des publications sous l'appellation Laboratoire Ligérien de Linguistique

6. Institution et fonctionnement

- Assemblée Générale

sont membres de droit de l'assemblée générale tous les membres permanents du laboratoire. Les doctorants du laboratoire et les membres associés participent avec voix consultative. L'assemblée générale se réunit au moins une fois par an, élit le Conseil de laboratoire, adopte le bilan de l'année écoulée et les perspectives pour les années à venir.

- Conseil de laboratoire

Le Conseil de laboratoire est composé de 10 membres, en règle générale les membres des bureaux de site, sous réserve de validation par l'assemblée générale annuelle

- Bureaux du laboratoire

Afin d'aboutir à un fonctionnement efficace, il a été décidé d'instituer un bureau par site :

- un directeur (Orléans) et un directeur adjoint (Tours) – cette répartition valant pour ce contrat, sans engagement pour les suivants.
- un responsable des finances et des moyens par site ayant pour mission de suivre les entrées et les dépenses sur chaque site
- un représentant des filières de formation (parcours recherche des Masters et doctorats) par site
- un chargé des activités scientifiques (séminaire, colloques...) par site,
- un représentant des doctorants par site,
(en cas d'affectation de personnel., un représentant des IATOS sera ajouté au Conseil)

- Directeur

Le Directeur, et le Directeur-adjoint qui lui correspond sur l'autre site, ont la responsabilité de l'animation du laboratoire, de son activité et du respect de ses règles. Dans chacun des sites, ils représentent le laboratoire auprès des instances universitaires et des partenaires. Seul le Directeur est habilité à représenter l'ensemble du laboratoire. Ils sont désignés par élection de l'assemblée générale pour la durée du contrat.

- Responsables d'équipe

Les responsables d'équipes sont élus par les membres permanents du laboratoire engagés dans l'équipe concernée

7. Budget

- Sources : outre le budget récurrent alloué par le Conseil Scientifique, toute source contractuelle accessible
- Répartition : selon les besoins et les disponibilités, dans le respect des engagements contractuels

Gestion : les fonds sont alloués par le bureau de site après demande motivée, puis laissés à la libre utilisation de l'équipe ou du chercheur concerné sous réserve de justification. La répartition du budget est soumise à l'assemblée.

LISTE DES LABORATOIRES CNRS

Section 34 / INSHS

Recension établie au 1^{er} janvier 2008
 SHS département principal / section 34 principal ou secondaire
 En gras, les UMR centrées sur les sciences du langage ; rayées, celles qui ont été dés-UMRisées.

ILE-DE-FRANCE

26 laboratoires dont 13 UMR centrées sur les SDL (1 supprimée) = 12 UMR

Laboratoire de Phonétique et Phonologie	Jacqueline VAISSIERE	LPP	PARIS
Structure formelle du langage : typologie et acquisition, métrique et poétique	Clive PERDUE	SFL	ST DENIS
Langues et civilisations à tradition orale	Zlatka GUENTCHEVA- DESCLES	LACITO	VILLEJUIF
Laboratoire de linguistique formelle	Alain KIHM	LLF	PARIS
Modèles, Dynamiques, Corpus	Bernard LAKS	MODYCO	NANTERRE
Laboratoire d'ethnologie et de sociologie comparative	Anne-Marie PEATRIK		NANTERRE
Lexiques, dictionnaires, informatique	Salah MEJRI	LDI	VILLETANEUSE
Histoire des théories linguistiques	Sylvie ARCHAIMBAULT	HTL	PARIS
Centre de recherche en épistémologie appliquée	Paul BOURGINE	CREA	PARIS
Langues, textes, traitements informatiques, cognition	Laurence DANLOS	LATTICE	MONTROUGE
Langues—Musiques Sociétés	Frank ALVAREZ-PEREYRE	LMS	VILLEJUIF
Institut Jean-Nicod	Pierre JACOB	IJN	PARIS
Institut des Textes et Manuscrits Modernes	Pierre-Marc DE BIASI	ITEM	PARIS

Centre d'études des langues indigènes d'Amérique	Francisco QUEIXALOS	CELIA	VILLEJUIF
Langage, langues et cultures d'Afrique noire	Martine VANHOVE	LLACAN	VILLEJUIF
Centre de recherches linguistiques sur l'Asie orientale	Redouane DJAMOURI	CRLAO	PARIS
Institut d'Histoire et de Philosophie des Sciences et des Techniques	Jacques DUBUCS	SHS	PARIS
Cultures, Langues, Textes	Isabelle PASTOR-SOROKINE		VILLEJUIF
TGE-ADONIS	Gilbert PUECH		PARIS
Relais d'information sur les sciences de la cognition (RISC)	Jean LORENCEAU		PARIS
Institut de linguistique française (ILF)	Christiane MARCHELLO-NIZIA	FR	PARIS
Typologie et universaux linguistiques : données et modèles	Stéphane ROBERT	FR	PARIS
Communication et politique	Isabelle VEYRAT-MASSON		PARIS
Description et modélisation en morphologie	Bernard FRADIN	GDR	PARIS
Sémantique et modélisation (SEM)	Francis CORBLIN	GDR	PARIS
Linguistique latine	Michèle FRUYT	GDR	PARIS

PROVINCE ET OUTRE-MER

16 laboratoires dont 12 UMR centrées sur les SDL (3 supprimées) = 9 UMR

Interactions, Corpus, Apprentissage, Représentations	Lorenza MONDADA	ICAR	LYON
Cognition, Langues, Langages, Ergonomie	Jacques DURAND	CLLE-ERS	TOULOUSE
Laboratoire PRAXILING	Jacques BRES	PRAXILING	MONTPELLIER
Centre de recherches sur la langue et les textes basques	Bernard OYHARCABAL	IKER	BAYONNE
Dynamique du langage	François PELLEGRINO	DDL	LYON
Bases corpus et langage	Tobias SCHEER	BCL	NICE
Laboratoire Parole et Langage	Philippe BLACHE	LPL	AIX EN PROVENCE
Centre d'Etudes Supérieures de la Renaissance	Marie-Luce DEMONET	CESR	TOURS
Moyen-Âge	Patrick CORBET	SHS	NANCY
Analyse et traitement informatique de la langue française	Jean-Marie PIERREL	ATILF	NANCY
Langues, textes et communications dans les espaces créolophones et francophones	Lambert-Félix PRUDENT	SHS	ST DENIS
Savoirs, Textes, Langage	Fabienne BLAISE	STL	VILLENEUVE D ASCQ
UMS de la Maison de la recherche de l'université de Toulouse le Mirail	Bertrand JOUVE	UMS	TOULOUSE
Dynamiques sociolinguistiques	Régine DELAMOTTE-LEGRAND	DYALANG	MONT ST AIGNAN
Centre de Recherches inter-langues sur la signification en Contexte	Jacques FRANCOIS	CRISCO	CAEN
Approche pluridisciplinaire de la production verbale écrite	Denis ALAMARGOT	GDR	POITIERS

LABORATOIRES EN RATTACHEMENT SHS SECONDAIRE

ILE-DE-FRANCE

6 laboratoires

SHS département secondaire / section 34 secondaire

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI)	Patrick LE QUERE	ST2I	ORSAY
Institut Jean le rond d'Alembert	Gérard MAUGIN	ST2I	PARIS
Vulnérabilité, adaptation et psychopathologie	Roland JOUVENT	SDV	PARIS
Centre d'analyse et de mathématique sociale (CAMS)	Henri BERESTYCKI	ST2I	PARIS
Laboratoire des Usages en Technologies d'Information Numériques (LUTIN)	Dominique BOULLIER	ST2I	PARIS
Laboratoire psychologie de la perception	John Kevin O'REGAN	SDV	PARIS

PROVINCE ET OUTRE-MER

9 laboratoires

SHS département secondaire / section 34 secondaire

GIPSA	Jean-Marc CHASSERY	ST2I	ST MARTIN D'HERES
Laboratoire d'Informatique de Grenoble (LIG)	Brigitte PLATEAU	ST2I	ST MARTIN D'HERES
Institut de recherche en Informatique de Toulouse (IRIT)	Luis FARINAS DEL CERRO	ST2I	TOULOUSE
Laboratoire de psychologie sociale et cognitive (LAPSCO)	Michel FAYOL	SDV	CLERMONT FERRAND
Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen (GREYC)	Régis CARIN	ST2I	CAEN
Laboratoire d'informatique Fondamentale de Marseille (LIF)	Paul SABATIER	ST2I	MARSEILLE
Groupe d'imagerie neuro-fonctionnelle	Bernard MAZOYER	SDV	CAEN
Laboratoire Langage, Mémoire et Développement Cognitif (LMDC)	Jean-Francois ROUET	SDV	POITIERS
Laboratoire lorrain de recherche en informatique et ses applications (LORIA)	Karl TOMBRE	ST2I	VANDOEUVRE LES NANCY

LA RECHERCHE A LA BNF

(version du 25 juin 2010 - extraits)

site Internet : <http://www.bnf.fr/fr/>

La recherche subventionnée

Principalement cofinancée par la BnF et par le département de l'Enseignement supérieur, de la recherche et de la technologie du ministère de la Culture et de la Communication, la recherche subventionnée constitue une part très importante de l'activité de recherche au sein de la Bibliothèque. Elle répond à des exigences scientifiques précises en termes d'obligation de résultats, de travail en partenariat et de programmation rigoureuse dans le temps.

Elle est organisée dans le cadre de plans triennaux, dont celui en cours porte sur les années 2010-2012. Cette organisation remonte à 1995 et le nouveau plan lancé est donc le sixième de ce type. Au cours de cette période, l'accent est mis sur l'évaluation des programmes au sein de comités scientifiques particuliers et du Comité de la Recherche de la BnF.

La recherche non subventionnée

À côté des programmes de recherche subventionnés, l'activité de recherche non subventionnée est également essentielle et responsable en grande partie du rayonnement des différents départements de la Bibliothèque. Elle relève davantage de l'activité scientifique courante de ces derniers et est financée sur leur budget.

Informations sur les départements

Les travaux scientifiques menés répondent à l'une des missions essentielles de la Bibliothèque nationale de France :

- * leurs champs d'investigation sont, par essence, aussi vastes et divers que les collections de la bibliothèque elle-même

- * leurs niveaux d'ambition sont également divers (du dénombrement ou signalement d'objets au catalogue raisonné...) et exigent des travaux de quelques années à plusieurs générations...

- * leur publication s'effectue sous des modalités très différentes (catalogues de collections ou d'expositions, répertoires, bibliographies, publications savantes, articles de revue, communications dans des manifestations universitaires souvent internationales...)

- * enfin, dans le domaine de compétence et d'expertise qui est le leur, certains conservateurs de l'établissement ont acquis aujourd'hui une envergure comparable à celle de prédécesseurs comme Léopold Delisle, Ernest Babelon ou Jean Adhémar, qui ont marqué l'histoire de la recherche et de l'érudition. Ces travaux contribuent de façon essentielle au rayonnement de la bibliothèque.

Axes de recherche

Qu'ils soient subventionnés ou non, les programmes de recherche se développent selon les mêmes axes :

- * Bibliographie : production d'outils pour la recherche : inventaires, catalogues raisonnés, répertoires... ;

- * Conservation : recherche de nouveaux procédés de conservation aussi bien des supports « papier », numériques, microformes... que de leur contenant, à savoir boîtes d'archivage, pochettes...

- * Histoire du livre : étude et production d'outils concernant l'histoire du livre ;

- * Iconographie : étude méthodologique des représentations plastiques que la Bibliothèque conserve (peintures y compris enluminures, gravures, photographies quel que soit leur support, affiches, dessins, etc.) avec leurs sources, significations et classements ;

- * Patrimoine de la musique et du spectacle : création de catalogues thématiques et de répertoires, recensement de sources musicales et d'arts du spectacle ;

- * Numismatique : études sur les collections de monnaies du département des Monnaies, médailles et antiques en relation avec les autres cabinets numismatiques français et étrangers ;

- * Supports numériques et nouvelles technologies : mise au point d'outils de structuration des documents électroniques ;

- * Supports, usages et circulation de l'écrit : étude des supports d'information autres que le livre imprimé sur le plan matériel et intellectuel.

Evolution

La recherche s'inscrit aujourd'hui dans ces perspectives d'évolution :

- * un effort de resserrement des programmes,

- * un rééquilibrage des travaux de recherche entre les départements s'appuyant sur une longue tradition et des départements dont les domaines sont plus récents,

- * la mise en place d'une programmation, l'amélioration du suivi, le développement du faire savoir et de la valorisation,

- * le développement de partenariats et une plus grande ouverture vers le monde de la recherche, en particulier par le biais de conventions avec le Centre National de la Recherche Scientifique et les unités mixtes de recherche (UMR), de l'appel à chercheurs lancé chaque année auprès des jeunes doctorants français et étrangers.

Les programmes de recherche font l'objet de publications scientifiques.

Bibliothèque nationale de France
Plan triennal de la recherche 2010 – 2012

Bibliographie

In Principio 2	2012	Manuscrits
Catalogue des manuscrits enluminés d'origine germanique	2012	Manuscrits
Inventaire des incunables scientifiques de la bibliothèque de l'Arsenal	2012	Bibliothèque de l'Arsenal
Un cabinet savant à l'époque des Lumières : Jean-Baptiste d'Anville	2013	Cartes et plans
Les portulans conservés en France	2012	Cartes et plans

Conservation

Conservation des documents audiovisuels 2012		Audiovisuel
--	--	-------------

Histoire du livre

Renouard. Imprimerie et librairie parisienne du XVI ^e siècle	s.d.	Réserve des livres rares
--	------	--------------------------

Iconographie

Mise en valeur des fonds d'architectes	2012	Estampes et photographie
Fonds des architectes aux XIX-XX ^e s.	2012	Mission pour la gestion de la production documentaire
Manuscrits japonais à l'époque d'Edo	2012	Manuscrits

Numismatique

Catalogue des monnaies de l'Empire romain	2012	Monnaies, médailles et antiques
Trésors Monétaires	s.d.	Monnaies, médailles et antiques
Catalogue des monnaies celtiques	2012	Monnaies, médailles et antiques

Patrimoine de la musique et du spectacle

Catalogue thématique de l'œuvre
de Jean-Philippe Rameau s.d. Musique

Répertoire international des sources
musicales (R.I.S.M.) s.d. Musique

Supports numériques et nouvelles technologies

Caractérisation des Documents
Numérisés et prédiction de performances
des techniques d'OCR et d'Indexation 2013 Conservation

Supports, usages et circulation de l'écrit

Reliures françaises des collections
de la BnF, du Moyen Âge au XIX^e siècle 2012 Réserve des livres rares

Convention CNRS – BnF

La convention signée le 2 mars 2005 par le président de la BnF, Jean-Noël Jeanneney, et Bernard Larrouturou, directeur général du CNRS, reconduit pour une nouvelle période de cinq ans la coopération engagée en 1997 par les deux institutions autour de la production, de l'usage et de la conservation des supports d'information.

L'objectif est de poursuivre ensemble des programmes de recherche variés : fabrication, histoire et conservation du papier, histoire du livre, numismatique, patrimoine musical français, bibliographies spécialisées, traitement automatisé du langage, production d'outils d'accès aux sources documentaires, nouvelles technologies de l'information et de la communication (NTIC)...

La coopération entre les deux institutions consiste en un échange de compétences et de savoir-faire de leur équipes respectives. Elle passe aussi par la mise à disposition de locaux et de matériels ainsi que par la gestion et la valorisation concertée des résultats issus de ces recherches.

Cette nouvelle convention-cadre marque la volonté commune du CNRS et de la BnF de renforcer encore leur partenariat, engagé depuis huit ans.

Les programmes communs de recherche CNRS-BnF
Sept unités mixtes de recherche du CNRS sont particulièrement concernées par cette coopération scientifique avec la BnF :

- le Centre de recherches sur la conservation des documents graphiques (CRCDG – Ministère de la culture / MNHN / CNRS),
- l'Institut de recherche et d'histoire des textes (IRHT),
- l'Institut d'histoire moderne et contemporaine (IHMC – CNRS / ENS)
- le Centre d'Études Préhistoire, Antiquité, Moyen-Age (CEPAM – CNRS / Université de Nice),
- l'Institut de recherche sur le patrimoine musical en France (IRPMF – CNRS / MCC / BnF)
- le Centre de Recherche et de Restauration des Musées de France (C2RMF – CNRS / MCC)
- l'Institut de Recherche sur les Archéomatériaux (IRAMAT – Université Bordeaux 3 / Université d'Orléans / CNRS)).

Le CRCDG et la BnF coopèrent sur deux programmes scientifiques : l'étude de l'influence des traitements de restauration sur la conservation des documents papiers comportant des encres ferrogalliques et la recherche d'un cuir de tannage végétal pour la conservation des livres et reliures anciennes dégradées.

L'IRHT et la BnF vont poursuivre leur collaboration autour de la description, l'étude et le catalogage des manuscrits médiévaux et des premiers imprimés qui prendra fin en 2007. Le partenariat commun développe des projets de co-édition, la formation des personnels ainsi que la diffusion et la valorisation du patrimoine culturel et de la recherche scientifique (production et diffusion de revues communes par des liens ou des portails informatiques...).

L'IHMC et la BnF continuent leur partenariat pour la réalisation en commun de la *Bibliographie annuelle de l'histoire de France*.

Le CEPAM et la BnF coopèrent jusqu'en 2006 sur le Catalogue des monnaies impériales. L'objectif est d'étudier, de recenser et de reclasser les monnaies impériales des collections du Cabinet des médailles de la BnF ainsi que de constituer une base de données et de publier un catalogue.

L'IRPMF et la BnF poursuivent leur collaboration sur le programme RISM « Répertoire et inventaire des sources musicales » qui consiste à inventorier et décrire des manuscrits musicaux appartenant à la période 1570-1800, conservés dans les bibliothèques françaises. Il s'agit de continuer le travail engagé de catalogage des manuscrits musicaux et de procéder à sa publication.

Le C2RMF mène des recherches conjointes avec le CRCDDG et les ateliers de restauration de la BnF.

Enfin, le Centre Ernest-Babelon (de l'IRAMAT) et la BnF collaborent depuis de nombreuses années dans deux domaines, la numismatique et la codicologie. En numismatique, plusieurs axes de recherches sont conduits par les deux organismes : la réalisation de catalogues de monnaies gauloises et la caractérisation des métaux monnayés, en vue d'études dans le domaine de l'histoire économique et monétaire. En codicologie, le partenariat porte sur l'étude des manuscrits latins et arabes, en se fondant sur la caractérisation des encres et des pigments et colorants des enluminures.

Pour la constitution d'une UMR BnF / LLL / CNRS
- CODEX -
(Centre d'Opérabilité des Données pour l'Exploitation des Corpus Sonores)

Un objet scientifique : le corpus sonore

Si les corpus sonores ont émergé il y a une centaine d'années comme de possibles moyens de recherche à la disposition des sciences humaines et sociales, il a fallu attendre la diffusion du magnétophone pour qu'il soit possible de recueillir aisément de la parole sur le terrain. Il en est résulté un accroissement des enregistrements disponibles tel que seule l'informatique peut en rationaliser l'exploitation. Ainsi, en une vingtaine d'années, la référence à des sources contrôlables est devenue inhérente à l'argumentaire scientifique des linguistes, des anthropologues, des sociologues... C'est à cette pratique, sinon nouvelle du moins récente, qui conditionne largement le travail de laboratoire, que répond le champ d'investigation circonscrit par cette UMR avec une double spécificité :

- d'*objet*, puisqu'elle centre les investigations sur une classe particulière de corpus : ceux qui réunissent des matériaux langagiers pérennisés, conservés avant tout dans leur forme sonore, après qu'ils ont été établis ou collectés à des fins d'étude et de transmission, avec l'objectif de rendre compte de leurs propriétés sans obérer leur conservation ;
- de *méthode*, en unifiant des appropriations fragmentaires par une maîtrise de l'ensemble du processus qui constitue un corpus, de sa conception à sa consultation, avec une attention particulière concernant les procédures d'archivage.

Bien que l'étude soit volontairement restreinte aux documents sonores, trois types d'éléments seront associés dans l'inventaire :

- les enregistrements sous la forme où ils ont été réalisés et/ou conservés, ainsi que les duplications effectuées successivement,
- les annotations, qu'elles soient intégrées dans l'enregistrement (sous forme de plages ou de commentaires insérés) ou qu'elles soient adjointes sur des supports graphiques au moyen de transcriptions (orthographiques, phonétiques...), d'indications situationnelles et interactionnelles, etc.
- les outils et les systèmes de description qui assurent la contextualisation par l'assignation des locuteurs ou de la situation, par l'indexation des contenus, mais aussi qui préservent la maîtrise des techniques de saisie, de gravure et de lecture de l'enregistrement, un corpus sonore étant indissociable des dispositifs qui en assurent la restitution.

Les corpus ainsi entendus correspondent à des produits conçus à l'intérieur d'une pratique scientifique pour être éprouvés par des analyses et pour se prêter à des applications. Ils peuvent devoir leur existence à un programme qui les a constitués ou bien résulter d'une sélection dans des fonds préexistants. Quelles que soient les sollicitations, les communautés participant à l'exploitation du document sonore ont pour premier souci la préservation du matériel recueilli, ce qui signifie non seulement une exploitation non destructive des ressources mais aussi l'actualisation continue de l'accès aux contenus, pour surmonter les conséquences de transformations technologiques qui renouvellent constamment les formats et les appareils de lecture, et pour compenser la dégradation du contenant et du contenu des données. La pérennisation des documents suppose la définition d'une méthodologie dont les principes ont été définis par les acteurs eux-mêmes dans *Corpus oraux, Guide des bonnes pratiques*, sous la direction d'O. Baude (2006).

Deux approches complémentaires

L'accroissement des réserves de documents sonores et leur prise en charge patrimoniale ont mis en évidence la fragilité des supports et l'obsolescence des systèmes de restitution. Les fonctions de collecte et de conservation, de stockage et d'analyse, de balisage et d'indexation, de mise à disposition et d'application, effectuées séparément dans les laboratoires et les centres d'archivage, ne correspondent plus aux exigences induites par l'évolution des technologies et des pratiques scientifiques. La création d'une UMR, une *unité mixte de recherche* au sens plein du terme, prend acte qu'une division du travail, qui a prévalu jusqu'à maintenant, ne se justifie plus et qu'il faut établir de nouvelles formes d'expertise et de connaissance par l'addition des expériences et des compétences.

Alors que le document sonore a été appréhendé jusqu'à présent comme un objet, transférable de l'équipe qui le recueille à l'institution qui le conserve, le programme de l'UMR conçoit les corpus dans leur dynamique, c'est-à-dire comme un processus, un enchaînement d'opérations solidaires, depuis les hypothèses théoriques sous-jacentes à une décision de collecte jusqu'à la communication au public. Dans l'intervalle, le document sonore demande à être constamment retravaillé afin que l'archive demeure comme un témoignage vivant.

D'abord, son contenu peut être rattaché à des séries indéfiniment extensibles. Si même il se trouve catalogué dans une collection pour laquelle l'enrichissement est tenu pour achevé, ses modalités de conservation (codage et formatage) continuent d'accompagner sa transmission. Surtout, les annotations qui en assurent le traitement ne peuvent s'interrompre dans une évolution qui, en contribuant à l'enrichissement continu du corpus, en modifie l'accès et les lectures.

Dans cette perspective, la collaboration entre les parties contractantes se pense dans une solidarité des gestes scientifiques, une complémentarité des objectifs et des modes de fonctionnement, avec des recoupements et des relais qui confortent la pertinence du projet, sa cohérence et plus encore sa nécessité. C'est ce qui ressort du circuit du document dont on reconstituera le parcours comme indiqué dans le tableau :

Légendes et sigles

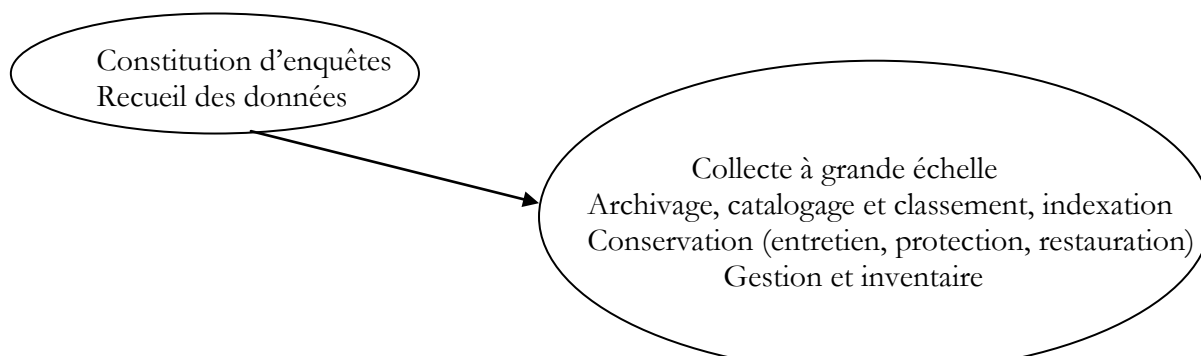
- UMR : Unité Mixte de Recherche (Université / Grand Etablissement / CNRS)
- EA : Unité Propre de Recherche à l'Enseignement Supérieur / Equipe d'Accueil
- CORAL : Centre Orléanais de Recherche en Anthropologie et Linguistique
- LLL : Laboratoire Ligérien de Linguistique
- BnF : Bibliothèque Nationale de France (Ministère de la Culture)

*Circuit du document et phases de traitement
selon les compétences des parties contractantes dans l'UMR*

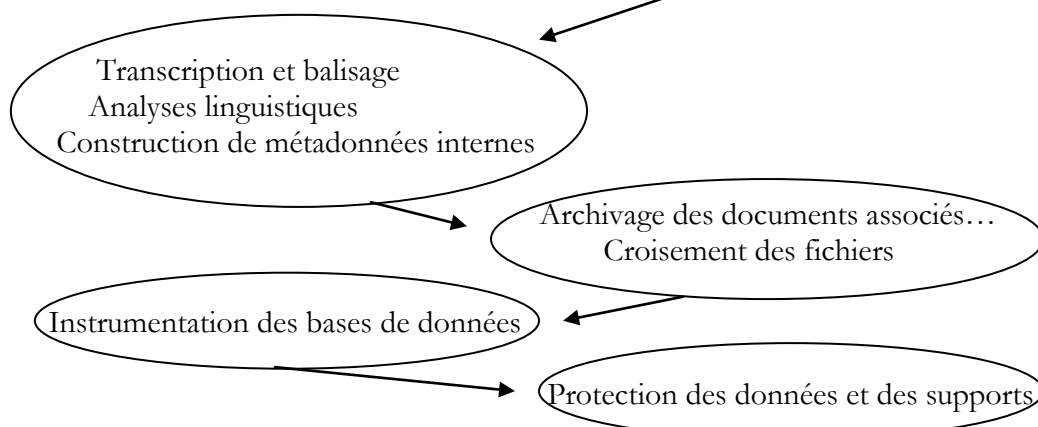
EA LLL / U. d'Orléans et Tours

BnF / Département de l'Audiovisuel
Service des Documents Sonores

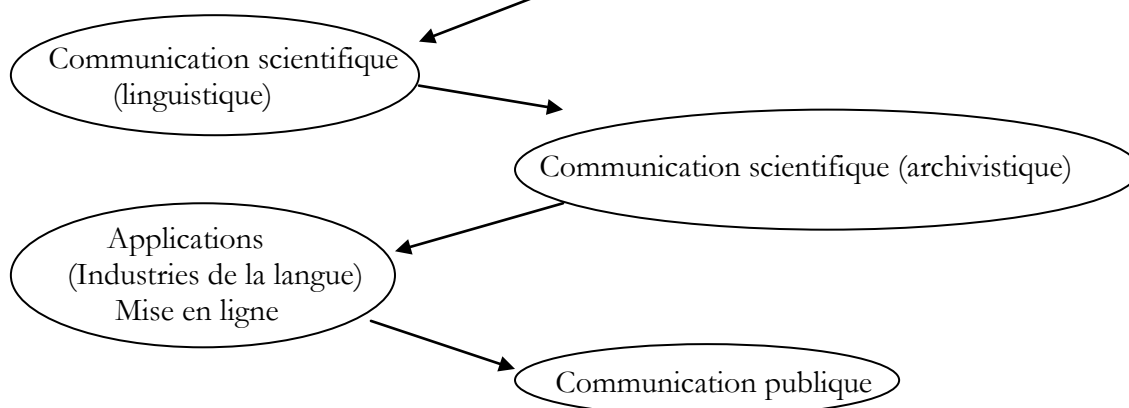
1. Création de corpus



2. Exploitation des corpus



3. Diffusion des corpus



A partir de cette circulation entre deux institutions, qui représentent aussi deux types de compétences qu'il s'agit de réunir, l'une des perspectives de l'UMR est de parvenir à produire des usages différenciés et multiples des corpus. En lieu et place d'un document collecté par un laboratoire pour une seule exploitation finalisée, l'objectif serait de rendre possible une réutilisation des données. Cela suppose d'engager un traitement d'ensemble, qui suppose des modèles et des références, mais aussi un accroissement des capacités d'interopérabilité à tous les niveaux d'accès et d'analyse.

Dans cette conception, le travail d'un laboratoire comme le CORAL ou LLL ne représente qu'une fraction de ce que le Service des Documents Sonores du Département de l'Audio-Visuel de la BnF doit traiter mais il peut avoir valeur de test pour la définition concertée d'une politique raisonnée de l'archive sonore. Qu'il s'agisse des travaux effectués à Orléans, ESLO 2 ou l'Encyclopédie palikur par exemple, ou de collectes effectuées par d'autres laboratoires, l'accompagnement à ses différentes phases de l'élaboration d'un corpus permet aux conservateurs et aux personnels de documentation de suivre *in situ* les segments de la fabrique des données sonores auxquels ils n'auraient pas accès et aux chercheurs de profiter de leurs compétences en matière de saisie, d'indexation ou d'exploitation. Dans tous les cas, les études ne sauraient avoir vocation à se limiter aux apports de l'équipe.

Dans cette collaboration, la contribution attendue du CNRS s'apparente à celle dont bénéficient les équipes qui se sont mises en place autour des documents écrits, que ce soit dans le cadre d'un partenariat avec la BnF, :

- l'Institut de Recherche et d'Histoire des Textes (IRHT) et
 - le Centre de Recherches sur la Conservation des Documents Graphiques,
- ou dans des conventions avec les Universités :
- Analyse et Traitement Informatique de la Langue Française (ATILF, UMR CNRS/Nancy 2)
 - Institut des Textes Et Manuscrits modernes (ITEM, UMR CNRS/ENS)

Le spectre d'activités couvert par ces laboratoires, en matière d'archives, n'inclut pas les ressources orales. Il s'agit donc d'assurer une extension des champs documentaires, confinés jusqu'à présent dans les *scripta* et les *realia*, afin d'y inclure les *vocalia*, de les rendre accessibles à la recherche en rapprochant le traitement des corpus sonores des modalités sanctionnées par le travail des équipes qui se consacrent aux recueils de textes et de manuscrits. C'est l'occasion d'opérer un rapprochement entre l'enquête et l'archivistique à l'intérieur d'un domaine qu'elles partagent sans se rencontrer, en dépit d'initiatives associatives comme l'AFAS (Association Française des Détenteurs de Documents Audiovisuels ou Sonores) ou de projets comme le CRDO (Centre de Ressources pour la Description de l'Oral).

Linguistique et traitement du document sonore

La synergie attendue de l'organisation en UMR répond à quatre objectifs :

- expliciter les critères de constitution de corpus, de traitement, d'archivage...
- représenter sur un mode déclaratif les connaissances qui sous-tendent les catégorisations effectuées par les balises ou les procédures d'indexation,
- modéliser les procédures d'archivage, avec intégration des liens entre les différents documents constitutifs du corpus, notamment les méta-données,
- contribuer à élaborer une notation formelle afin d'exercer un retour réflexif sur les documents pour en contrôler le contenu et faciliter l'accès aux contextes situés.

Le traitement du document sonore est conditionné par une conception du corpus qui a pour ambition de ne pas restreindre son usage à ceux qui l'ont produit mais qui anticipe l'accessibilité auprès d'autres communautés, celles des chercheurs français et étrangers, mais aussi, pour certains enregistrements, à destination du grand public. Cela suppose que l'archive soit paramétrée de façon à assurer la consultation par différents systèmes informatiques, sur des supports fixes ou à distance, et que la structure d'interrogation soit conçue en fonction des normes, standards et usages pratiqués selon le type de diffusion choisi et les destinataires potentiellement concernés. L'impératif d'interopérabilité implique une confrontation des requêtes formulées à l'ensemble des sources en catalogue et à la diversité de leurs formats.

On postule que l'archive sonore constitue le document pivot autour de quoi s'organisent les documents associés et l'ensemble du travail du conservateur et du chercheur. Pour être exploitable et disponible, à tous les moments de sa consultation, en particulier dans une perspective patrimoniale de transmission, il est prévu la conservation des protocoles et des outils qui en assurent le stockage, le transport et la lecture dans un format accessible de façon autonome. La croissance des données et l'évolution rapide des techniques qui assurent le conditionnement des documents sonores, aussi l'imprévisibilité des demandes futures de consultation, tant dans leur approche que dans leur mode opératoire, induisent la poursuite d'une réflexion pour que les compétences de manipulation et d'interrogation de corpus sonores soient portées au niveau atteint par les études sur les corpus textuels.

Ainsi, le point d'application d'une recherche engagée en commun ne concerne ni la strate informatique au principe des langages électroniques et de la documentation automatique, ni la gamme des exploitations assurées par les destinataires ultimes qui peuvent être issus de différentes disciplines : historiens, sociologues, anthropologues, linguistes... Le but n'est pas de se substituer aux communautés scientifiques mais de les accompagner dans l'ouverture de leurs données à d'autres sollicitations, de dépasser la correspondance terme à terme d'un corpus sonore et d'une enquête thématique.

Les études programmées dans le cadre de l'UMR se concentreront donc sur l'interface entre les matériaux traités comme archives – supports, discours enregistrés et méta-données –, et une élaboration qui met à contribution les sciences du langage pour augmenter l'accessibilité et l'enrichissement des corpus traités. Elles se prolongeront jusqu'à l'observation des usages et des exploitations dont la connaissance est indispensable à la modélisation des procédures.

L'accès aux contenus

On part du principe que la reconnaissance du corpus est réalisée au moyen d'un système de descripteurs et de relations, représentés par des mots ou des algorithmes et organisés de façon hiérarchique. Leur efficacité est appréciée en fonction du degré de réduction du temps d'accès à l'information définie comme pertinente par la requête. Par ailleurs, une entité n'est constituée comme telle dans le corpus qu'à partir du moment où elle a été l'objet d'une description et d'une indexation. La structuration ne porte pas directement sur les données mais sur les informations qui leur sont associées, descripteurs formalisés et entités taxinomiques organisées en « métadonnées ».

C'est sur des systèmes de représentation et d'indexation que peuvent être établis une terminographie, des index, des thesaurus, et même aujourd'hui des « ontologies ». Quelle que soit l'approche privilégiée, elle demeure tributaire d'une connaissance approfondie des corpus pour laquelle les procédés d'enregistrement et de conservation et la définition des objets éligibles sont

des pré-requis, sachant que les représentations sont construites en fonction des outils de traitement qui déterminent les modalités de l'indexation.

La finalité de la collaboration réside dans cette conjonction d'une signalétique, réalisée par l'archiviste, et d'une analyse du document effectuée à partir de ses caractéristiques langagières. Le langage XML, qui fonctionne aujourd'hui comme le standard retenu par la plupart de ceux qui s'engagent dans le balisage des corpus sonores, assure l'interface entre requêtes et contenu des données. De la convergence entre l'exploration du contenu et la définition signalétique procède la modélisation de méthodes efficaces aussi bien pour la création que pour l'exploitation des corpus. En particulier, à partir d'un descripteur formulé au moyen d'un ou plusieurs termes extraits des données documentaires et/ou des transcriptions, on peut définir les « pointeurs » qui retrouvent les séquences du fichier sonore ou les séquences de texte correspondant à la recherche formulée.

La contribution de l'UMR au développement international des ressources électroniques sur les langues la désigne comme un partenaire important de l'*Open Language Archives Community* (OLAC) pour la création, en français et dans d'autres langues, d'une phonothèque virtuelle mondiale, en cours de développement collectif dans un cadre associatif. Après la réaction concertée des bibliothèques européennes pour la protection et la mise à disposition de leur patrimoine écrit, c'est à une initiative d'envergure comparable qu'il est prévu d'apporter une dimension de recherche et une visibilité sur la Toile.

Maniabilité des corpus sonores : contraintes et solutions

L'exploitation rationnelle des masses de données disponibles suppose d'automatiser et de fiabiliser la structure et les liens inscrits du corpus. Au centre des relations ordonnées entre les différents supports, il y a les corrélations établies entre les fichiers sonores et les fichiers de transcription correspondants, dont le format informatique diffère, rendant plus complexe l'interconnexion entre les pages du fichier sonore et la transcription textuelle qui la convertit en chaînes de caractères. La pluralité de formats de stockage (WAV / MP3, ASCII / XML) accroît les risques d'impédance quand les enregistrements n'ont pas été effectués en suivant les prescriptions.

Un étalonnage chronométrique permet d'assurer un premier repérage, à la fois homogène et indistinct. Il est intégrable à une copie de l'enregistrement, soit par des découpes en pages, soit par des signaux intercalés (pas forcément perceptibles à l'écoute). Dès lors que la valeur temporelle reste déterminante, en tant qu'elle est la seule à être universelle et fixe, il importe que le calage de la durée, l'invariabilité du défilement et la stabilisation du format soient assurés.

Un fichier sonore ne vaut, pour le chercheur, que par sa disponibilité effective et par la possibilité de l'interroger à partir de ses transcriptions et des métadonnées. L'accès au document procède des documents reliés : connaissance de la littérature grise, formalisation des métadonnées exploitation de la source par des travaux de laboratoire et des publications, systèmes de renvois, descripteurs. La massification des données et l'automatisation des procédures informatiques requièrent une sophistication des relations entre le document pivot et ses documents annexes dont la relation est rendue d'autant plus complexe que sera prise en compte la mémoire des états du corpus.

L'ensemble des acteurs contribuant à la fabrication du corpus (de la conception à la diffusion) s'accordent sur le rôle que jouent l'archivage et l'exploitation du document sur sa conception et sa

représentation. Après qu'ont été exposées les attentes sociolinguistiques, la définition des terrain, la sélection des témoins et la constitution des enquêtes, les métadonnées conditionnent la teneur et la structuration des corpus, leur degré de transparence ou au contraire d'opacité aux sollicitations dont ils sont l'objet. Un déplacement intervient qui conduit à distinguer, au centre des requêtes, l'archive sonore comme document référentiel d'un côté et, de l'autre, les annotations textuelles qui, au cours de leur évolution, sont établies comme document d'accès. Or, les outils d'indexation, de classement, de recherche et de gestion, quels que soient leur principe et leur finalité, sont toujours en dernière instance des modes linguistiques d'organisation de la connaissance, en particulier des éléments du lexique des langues naturelles, dans leur distribution, leur hiérarchie et leur agencement interne. A ce titre, les sciences du langage sont, par définition, contributives pour la conservation et l'exploitation des corpus sonores.

Objectifs et moyens de l'UMR

Grâce au rassemblement de nos compétences, nous pourrions déterminer les apports d'une démarche linguistique dans la gestion des corpus avec la conviction que poser des questions linguistiques permet d'anticiper les procédures de conservation. L'objectif est de modéliser les démarches : un savoir acquis lors la création d'un corpus doit pouvoir être réexploité de façon contrôlée sur des archives du même type et contribuer à l'élaboration de protocoles renouvelables.

L'enjeu est double. D'abord, enrichir les corpus. Ensuite, prolonger leur pérennité en conservant les savoirs et les expertises à même de les valoriser et de les lire. Ainsi, la généralisation de l'usage du magnétophone dans les pratiques scientifiques, il y a une quarantaine d'années, a coïncidé avec un cycle d'activité de chercheurs et de techniciens. Il nous appartient de veiller à la conservation et à la disponibilité des données qu'ils ont constituées, comme le LLL en a fait la démonstration à partir du corpus ESLO.

À partir de la numérisation, on réaccorde les deux approches complémentaires qui ont fait des corpus un des outils fondamentaux des sciences humaines et sociales : l'archivistique et l'éditionnalisation du contenu par son traitement linguistique, en mutualisant l'expertise des analyses sur l'expression et le contenu et le traitement documentaire des données.

Document de travail (2007-2009)

CAT COD

Appel à colloque - 2008

The number of oral recordings which are digitized and available for the study and the description of languages remained quite small for a long time and they were confined within specialized agencies which distribute them. With the development of the Web and of the storage, diffusion and processing capacities, it is now affordable for smaller structure (e.g. research laboratories) to distribute themselves their oral resources.

We are henceforth at a key stage where the capitalization of the efforts of linguistic data digitalization becomes possible, this in order to form the empirical basis for new research projects.

The observation of the coding and cataloguing standards of this type of resources within the various existing projects in France and Europe shows a great heterogeneity among practices. Whereas these very projects had been set up with the aim of facilitating the access, the sharing or the preservation of the data, we notice that the diversity within the formats, the codings and the protocols used, precisely limits these objectives.

In this symposium, we wish to report on the major initiatives within the field of digital documents management which would potentially have an important influence on standardization. To do so, we would like to stress two specific aspects which are: the cataloguing of an oral resource and its coding.

Cataloguing

The recent and fast growth of the number of oral recordings available on the Web needs to be accompanied by an important effort of description and referencing in order that these data should be easily accessible, should not be buried in the mass of available data and that their management should be facilitated.

Thus, it is necessary to distinguish the cataloguing practices whose purpose is to follow up the life cycle of a project of constitution of resources from those whose purpose is to guide the exploitation or the preservation and the distribution of these resources.

This cataloguing activity is all the more important and urgent as the handling operations of this mass of data increase as well (exchanges, maintenance, research, etc.). Some communities already got themselves strongly organised and have sometimes established standards such as the Dublin-Core as far as the resources on the Web are concerned, or dating from an earlier period, the MARC standards maintained by the Library of Congress in order to describe bibliographical resources. Finally more recently, smaller linguistic communities have established cataloguing proposals (OLAC, IMDI). People have now had enough experience in the use of these proposals to be able to criticize them, propose some improvements, give rise to new thinking and establish a minimal charter that should be respected by those who wish to get involved in the distribution of linguistic oral resources in order to facilitate their exchange and more generally their scientific use.

Coding

If cataloguing is essential for the identification of the resources and for a fast comparison between them, coding is itself essential to the description of the interpretation of the content of the resource itself and also essential to its exploitation for specific studies.

If coding the transcription of video or audio material is indeed the clarification of an interpretation, then one notes here as well a great variety in the practices.

The inventories made during the EAGLES, MATE and ISLE successive initiatives have demonstrated the difficulty there is to apprehend the extent of the various coding systems. The specification of a universal software tool for annotation can be an access path to resources coded in a standardized way, as was suggested by ISLE.

But this does not make it any the less necessary for us to undertake a communal and consensual work to categorize, name and structure the phenomena found within the resources if we wish a true interoperability of the data with a view to multiple and future exploitations. Thus, we must now start standardizing the coding of oral corpora similarly to what has already been undertaken on written corpora with the TEL.

If we suppose that a compromise on the coding and cataloguing of oral data has been obtained, it is then necessary to give ourselves rules and tools to check the compliance of specific data with established principles. Therefore, in this symposium, we also wish to report on the practices used to control the quality of the resources.

Topics of interest / Thématiques

- | | |
|---|---|
| - description and cataloguing of oral resources | - description et référencement des données orales |
| - diffusion | - diffusion |
| - tools design and architecture | - spécification d'outils |
| - scientific exploitation of data | - exploitation scientifique |
| - archiving | - conservation |
| - corpus editing | - édition de corpus |
| - annotation | - annotation |
| - versionning | - versionning |
| - cataloguing and coding standards | - standards de catalogage et de codage |
| - ressources comparison | - comparaison des ressources |
| - multimodal and multimedia transcription | - transcription multimodale et multimédia |
| - annotation schemes | - schémas d'annotation |
| - interoperability | - interopérabilité |
| - evaluation, quality control | - évaluation, contrôle qualité |

Program Committee / Comité de programme

Jean-Yves Antoine (Université de Tours)
Claude Barras (LIMSI-CNRS, Paris XI)
Steven Bird (University of Melbourne & LDC University of Pennsylvania)
Lou Burnard (Oxford University Computing Services)
Pascal Cordereix (BnF, Paris)
Benoît Habert (ENS-LSH, Lyon)
Serge Heiden (ENS-LSH, Lyon)
Nancy Ide (Vassar College)
Michel Jacobson (Ministère de la Culture, Paris)
Laurent Romary (MPI Berlin - INRIA)
Emmanuel Schang (Université d'Orléans)
Richard Walter (CNRS, Orléans)
Peter Wittenburg (Max-Planck-Institute for Psycholinguistics, Nijmegen)

Organisation Commitee / Comité d'organisation

Serge Heiden (ENS-LSH, Lyon)
Michel Jacobson (Ministère de la Culture, Paris)
Emmanuel Schang (Université d'Orléans)
Richard Walter (CNRS, Orléans)

REVUE DE SEMANTIQUE ET PRAGMATIQUE

RSP 21-22 (2007)

« La place du sens en linguistique »

Pierre Larrivé La source de la dimension qualitative des indéfinis de sélection arbitraire	9
Neiloufar Family Verbal Islands in Persian	23
Hélène Margerie Etude d'un adjectif qui se grammaticalise « grave »	49
Mirjana Miskavic-Lukovic Procedural semantics at work: the case of the English perception verbs <i>look, listen and see</i>	63
Daciada Vlad Du discours polémique à la polémique dans le discours	81
Pauline Merlet La Pragmatique Intégrée : application au discours lexicographique (le cas des articles <i>femme</i> et <i>homme</i> dans le GR et le TLF)	95
Veran Stanojevic & Tijana Asic One gift or two? The opposition distributive-collective and the syntax-semantic-pragmatic interface	105
Alfredo M. Lescano Vers une grammaire argumentative de <i>LE</i> et <i>UN</i>	119
Hugues Picavez Un exemple d'intégration sémantique de l'information pragmatique : les marqueurs modaux lexicaux	137
Bruno Courbon Usage(s) d'une approche pragmatique en sémantique diachronique : des lieux d'apparition et de stabilisation de l'innovation sémantique lexicale en français moderne	149
Nazaire Mbame Principes de l'adduction et abduction dans la catégorisation dynamique du langage.	175
Pascal Amsili L'annulation des implicatures et des présuppositions	193

RSP 23 (juin 2008)

Anne Theissen <i>Ledit N</i> : des emplois anaphoriques originaux	7
Laurence Kister & Evelyne Jacquey Traits sémantiques et anaphores pronominales	23
Patricia C. Hernández L'emploi locatif <i>sur</i> + ville	35
Pierre Cadiot & Abdelaali Talmenssour Dynamique sémantique du texte proverbial.	55
Vanda Mikšić Les actes de silence à la lumière de la théorie des actes de langage	75
Catalina Fuentes Rodríguez Invalidación, dialogismo y cortesía: Es un decir	95
Bénédicte Laurent Signifiante du nom de marque et de produit Ressorts sémantiques pour un effet pragmatique	121

RSP 24 (décembre 2008)
« Sémantique de la coordination »

Christian Bassac Sémantique de la coordination : une brève introduction	7
Frédéric Lambert & Lionel Clément Présentation	25
 <i>Sémantique des marqueurs de la coordination</i>	
Carlotta Viti The meanings of coordination in the early Indo-European languages	35
Katja Jasinskaja & Henk Zeevat Explaining Additive, Adversative and Contrast Marking in Russian and English	65
Anna Orlandini & Paolo Poccetti Liens de coordination : une approche sémantique à travers les langues anciennes	93
Frédéric Lambert Les intermittences du <i>ou</i> en français	115
 <i>Sémantique des éléments conjoints dans les structures coordonnées</i>	
Caterina Mauri The parallelisms of clausal coordination	145
Anne Abeillé & François Mouret Quelques contraintes sur les coordinations elliptiques en français	177
Mehmet Cicek & Christian Bassac Factorisation des affixes de pluriel et de possessif en turc.	207
Lionel Clément Zeugme sémantique	231

Numéro 25-26 (année 2009)

« Espace / Temps

(ou interprétations temporelles / interprétations spatiales ?) dans les langues »

Jean-Claude Anscombre

« Pour une autre classification des états »

Denis Le Pesant

« Modes d'action, diathèses et portée des ajouts temporels dans les verbes dits de déplacement »

Andrée Borillo

« La notion de "dynamique des forces" dans un emploi non-spatial de *sur* et *sous* en français »

Pierre Jalenques

« Valeurs spatiales et non spatiales dans le domaine des verbes »

Anne Le Draoulec & Denis Vigier

« *Dans* suivi d'un nom de partie de la journée : au croisement de l'espace et du temps »

François Nemo

« Interprétations temporelles et signification (non temporelle) : de *toujours* à *tout* »

Aude Rebotier

« Représenter le temps dans l'espace : les schémas pour la temporalité de l'énoncé »

Marcel Vuillaume

« *Maintenant*, un indexical pur et dur ? »

Samira Moukrim

« L'expression du 'présent actuel' en arabe (marocain), berbère (tamazight) et français parlés »

Adeline Patard

« L'imparfait dans le tour [(et) si *IMP?*] : entre temporalité et modalité ? »

Agnès Provot

« Le « futur dans le passé » en français et en allemand »

Inès Saddour

« The expression of progressivity in Tunisian Arabic: A study of progressive markers in oral retellings of simultaneous situations »

Pierre-Patrick Haillet

« Repérages temporels construits par le discours et approche polyphonique »

Mongi Kahloul

« *Après tout* : d'un *Après* temporel à un marqueur discursif de clôture »

Georgeta Cislaru & Michelle Lecolle

« Noms propres de lieux habités, espace et temporalité »

Numéro 27 (juin 2010)
« Des linguistes et des outils : au-delà des corpus »

Le Gallois & Leroy

« Présentation »

J. François

« Le repérage des segments répétés et la polysémie lexicale »

Ferrari & Legallois

« Modélisation d'un jeu de langage en vue d'explorations textuelles... »

Loiseau & Gréa

« Dictionnaires, théorie des graphes et structures lexicales »

Bertels, Speelman & Geeraerts

« La corrélation entre la spécificité et la sémantique dans un corpus spécialisé »

Recherches actuelles

Duguine & Irurtzun

« Opérateurs d'exclusivité dans les questions et réponses en basque »

Rudel & Mazaleyrat

« A propos d'un curieux adjectif : approche sémantico-cognitive de l'adjectif *curieux* »

Sonia Gomez & Jordana Ferary

« Énonciation proverbiale en français (XVI^e-XVII^e siècles) »

CORELA

Cognition, Représentation, Langage

Présentation de la revue en ligne et du site

CORELA (*Cognition, Représentation, Langage*) est la revue du *Cercle Linguistique du Centre et de l'Ouest* créée en 2000 et dirigée depuis par Gilles Col, Professeur à l'Université François-Rabelais. C'est une revue semestrielle et en accès libre, hébergée par la MSHS de l'Université de Poitiers.

L'une des originalités de CORELA est son caractère non thématique puisque y sont publiées des contributions sur des thèmes ouverts, qui abordent des questions de langue (grammaire, lexicque, syntaxe, phonologie, etc.), mais aussi de langage, de représentation, de modélisation, de sémantique. Dans la mesure du possible, sont regroupés des articles complémentaires par leur approche, leur méthodologie, leur corpus sans que soit imposée une thématique.

CORELA assure également la diffusion de travaux de jeunes chercheurs, ou de chercheurs confirmés sur des thématiques « marginales » ou complémentaires de la série des Travaux Linguistiques du CerLiCO (actes du colloque annuel de l'association CerLiCO publiés aux PUR). Le choix du support électronique répond à la volonté de confronter rapidement des points de vue, d'accroître l'accessibilité et d'augmenter la périodicité des publications.

CORELA comporte une base de données qui regroupe des contributions sous un autre format que des articles, dans une collection de numéros spéciaux (actes de colloque ou de journée d'études, comptes rendus de manifestations scientifiques ou d'articles, pré-publications, textes de conférences, annonces de colloques, activités doctoriales...). Il n'y a pas de publications de thèses en l'état mais les auteurs sont encouragés à soumettre un article fondé sur leur recherche doctorale.

L'objectif de CORELA est de promouvoir les discussions scientifiques et théoriques les plus larges possibles. Le support électronique, qui facilite publication et diffusion, a permis d'en faire plus qu'une revue en ligne, une banque de connaissances et un lieu de confrontation des méthodes et des résultats.

CORELA a depuis le début associé étroitement des enseignants-chercheurs de l'Université François-Rabelais à ses activités, tant au niveau de la direction éditoriale (directeur : Gilles Col, co-directeur : Jean-Michel Fournier) que du travail d'évaluation des propositions d'articles ou de numéros spéciaux soumis aux différents comités (voir organigramme).

A ce jour, CORELA a publié 8 volumes de deux numéros chacun, ainsi que 10 numéros thématiques (dont trois issus de travaux ou de colloques tenus à l'Université de Tours, en 2005 et 2006). Le site de la revue a été visité par plus de 20.000 lecteurs en 2009. Elle est répertoriée par le portail du Directory of Open Acces Journals qui regroupe à l'Université de Lund plus de 5.000 revues en accès libre couvrant tous les domaines scientifiques.

Organigramme de la revue CORELA

Directeur de publication

Gilles Col (Université François-Rabelais, Tours).

Co-directeurs de publication

Jean-Michel Fournier (Université François-Rabelais, Tours)

Daniel Roulland (Université de Haute-Bretagne, Rennes)

Secrétaire éditorial

Franck Zumstein (Université Paris 7).

Comité international

Dirk Geeraerts (Université Catholique de Louvain, Belgique)

Simos Grammenidis (Université Aristote, Thessalonique, Grèce)

Hans Petter Helland (Université d'Oslo, Norvège)

Bert Peeters (Université Macquarie, Australie)

Len Talmy (Université d'Etat de New York, USA)

Comité scientifique

Mario Barra-Jover (Paris 8), André Borillo (Toulouse 2), Didier Bottineau (CNRS, Paris X-Nanterre), Philippe Bourdin (Toronto, Canada), Philipp Carr (Montpellier 3), Jean Chuquet (Poitiers), Gérard Deléchelle (Tours), Jean-Louis Duchet (Poitiers), Jacques François (Caen), Georges Kleiber (Strasbourg 2), Frédéric Lambert (Bordeaux 3), Nicole Le Querler (Caen), Jean-Jacques Lecercle (Paris X-Nanterre), Denis Maurel (Tours), Salah Mejri (Paris-Nord), Claude Muller (Bordeaux 3), Peter Roach (Reading, UK), Anne Trévisse (Paris X-Nanterre), Claude Vandeloise† (LSU, Bâton-Rouge, USA), Bernard Victorri (CNRS, ENS).

Comité de lecture

Jean Albrespit (Bordeaux 3), Nicolas Ballier (Paris-Nord), Hélène Chuquet (Poitiers), Catherine Collin (Nantes), Alain Delplanque (Tours), Hamida Demirdache (Nantes), Alain Deschamps (Paris 7), Nathalie Garric (Tours), Françoise Gayral (Paris-Nord), Thierry Grass (Strasbourg 2), Stéphane Gresset (Paris 7), Daniel Kayser (Paris-Nord), Jean-Charles Khalifa (Poitiers), Anne Lacheret-Dujour (Paris X-Nanterre), Florence Lefeuvre (Paris 3), Dominique Legallois (Caen), Michel Paillard (Poitiers), Gérard Mélis (Paris 7), Amina Mettouchi (EPHE), Véronique Rauline (Paris X-Nanterre), Stéphane Robert (CNRS, INALCO), Sylvester Osu (Tours).

CORELA

(Présentation des numéros)

2007, V, 1

Le premier numéro du volume 5 de CORELA s'ouvre par un article de Laurence Kister et Evelyne Jacquey qui proposent une manière d'acquérir des informations sémantiques pour résoudre les anaphores où l'expression référentielle est de la forme NdeN et l'expression anaphorique est un pronom relatif sujet. Le second article, de Viviane Arigne, aborde le problème de la modalité itérée dans la double perspective de la grammaticalisation et de la polysémie au travers de la description sémantique de divers emplois de l'auxiliaire de modalité *should* en anglais contemporain.

2007, V, 2

La revue CORELA fête son cinquième anniversaire avec la publication de ce nouveau numéro, et c'est encore une fois la diversité de la recherche en sciences du langage qui caractérise cette nouvelle publication puisque sémantique, traitement automatique des langues et grammaire sont représentés, parmi d'autres domaines de notre discipline. Vous pouvez effectivement lire un article de P.-A. Somé sur la polysémie et la recherche d'invariants du verbe « di » en dagara, un autre d'A. Dziadkiewicz qui présente un modèle théorique de la traduction automatique de phraséologismes pragmatiques, et enfin un article d'E. Roussel qui met en évidence certaines propriétés pragmatiques d'une marque aspectuelle de l'anglais (*past perfect*) qui font passer au second plan les propriétés habituellement attribuées à cette forme.

2008, VI, 1

L'objectif de CORELA est d'ouvrir la revue à tous les domaines de la linguistique. C'est ce que ce nouveau numéro tente de faire en publiant quatre articles qui touchent à des questions assez variées comme celles des locutions (article de Jeanne Aptekman sur « si »), des stratégies argumentatives (article de Sara Boutouhami et Daniel Kayser sur les descriptions d'accidents de la route), des structures argumentales (article de Malika Kaheraoui sur « faire + infinitif ») et enfin la question des prépositions (article de Laurence Vincent-Durroux sur l'utilisation des prépositions par de jeunes sourds oralistes).

2008, VI, 2

Outre deux articles qui abordent les phénomènes de discordance entre morphologie et syntaxe (Blandine Pennec) ou les difficultés communicationnelles posées par les féminisations des noms de métiers (Béatrice Fracchiolla), les autres articles de ce nouveau numéro de CORELA s'ouvrent à différents domaines du langage et la cognition. Ils portent effectivement sur la place du langage dans la cognition humaine, à travers l'étude de la compréhension d'énoncés arithmétiques et son impact sur la résolution des problèmes posés (Sylvie Akiguet-Bakong), et à travers la question plus générale de la place du langage dans la biologie humaine (Nicolas Bourguignon).

2009, VII, 1

C'est à différents aspects de la phonologie que ce nouveau numéro de CORELA est en grande partie consacré : l'article de Sylvester Osu et Nick Clemens traite de la question des nasales en ikwéré, celui de Riitta Välimaa-Blum présente la question des contraintes phonotactiques envisagées d'un point de vue cognitif, et enfin celui de Véronique Abasq étudie deux préfixes anglais (pre- et post-) dans une perspective morpho-phonologique. Les deux autres articles de ce numéro sont consacrés à la presse sous l'angle de la modélisation de la dynamique énonciative (Marie Chagnoux) et de la dénomination événementielle et du nom propre (Laura Calabrese Steimberg).

2009, VII, 2

Les deux études présentées dans ce second numéro du volume 7, portent respectivement sur les langues sourdes et l'émergence de nouveaux aspects linguistiques, tels que les commentaires métalinguistiques, les emplois métaphoriques, l'humour et les énoncés complexes (L. Vincent-Durroux) , et sur la relation entre style et identité à travers l'étude de la variation stylistique envisagée comme une stratégie verbale à part entière (L. Buson et J. Billiez).

2010, VIII, 1

CORELA publie dans son nouveau numéro deux articles qui portent sur la langue anglaise. Dans l'un, Mathilde Pinson montre que le GN « the way » est en passe de devenir une locution conjonctive à part entière ; dans l'autre, Mickaël Vallée analyse la structure PI for P2, et tente de déterminer si l'argumentation qui consiste à définir P2 comme une cause et / ou comme une proposition secondaire, voire optionnelle, est fondée.

CERLICO

Réseau des universités du *Cercle Linguistique du Centre et de l'Ouest*

Le CerLiCO est composé d'une douzaine d'universités et de laboratoires de recherche du Grand Ouest implantés à :

- Angers
- Bordeaux (ERSàB)
- Brest (CRBC)
- Caen (CRISCO)
- La Rochelle
- Le Mans
- Limoges (CERES)
- Lorient-Vannes
- Nantes (LING)
- Poitiers (FORELL)
- Rennes (ACE)
- Orléans (LLL)
- Tours (LLL)

Les objectifs du CerLiCo :

- développer, dans le domaine de la linguistique, la communication scientifique entre ses membres, à quelque école de pensée linguistique qu'ils appartiennent,
- faciliter l'intégration des jeunes chercheurs à la vie scientifique de la discipline,
- organiser des colloques et rencontres entre linguistes en collaboration avec les universités et centres de recherches, notamment dans le centre et l'ouest de la France,
- publier les actes de ces colloques et rencontres ainsi que d'autres travaux utiles à la recherche linguistique et à la diffusion de ses résultats,
- faire connaître au niveau international la recherche linguistique menée en France et susciter la participation de chercheurs étrangers aux activités de l'association.

Dans ce contexte de coopération et d'échanges, l'Université François-Rabelais participe activement depuis plus de vingt ans à la vie de l'association :

- en accueillant tous les ans l'assemblée générale d'automne,
- en ayant organisé trois colloques de l'association (1996, 2002, 2010), les deux premiers publiés aux PUR (Réseau des Universités de l'Ouest Atlantique dont font partie les Presses Universitaires François-Rabelais), le troisième (voir annexe 24^e colloque du CerLiCO) devant suivre,
- et en ayant participé à la fondation et à l'animation de la revue CORELA – revue du CerLiCO – depuis 2003.

La dynamique de fusion du LLL a conduit la partie orléanaise à rejoindre le Cercle à partir de 2008, en même temps que s'effectuait le rapprochement entre les deux universités.

Le 24^e colloque international du CerLiCO

Transcrire, Ecrire, Formaliser

Le 24^e colloque

Le colloque annuel du *Cercle Linguistique du Centre et de l'Ouest* s'est tenu à Tours les 4 et 5 juin 2010. Il a réuni une soixantaine de participants, de différents pays : France, Suisse, Grèce, Italie, Royaume-Uni, Japon. Le colloque a confirmé l'implantation tourangelle du CerLiCO grâce à l'engagement des enseignants-chercheurs et des doctorants du laboratoire, ainsi que des étudiants du master de sciences du langage « Linguistique avancée et applications », dans la préparation du colloque. Cet engagement s'ajoute à celui existant depuis de longues années, à tous les niveaux de l'association CerLiCO.

Le colloque accueillait un collègue de l'université Aristote (Thessalonique, Grèce), Simos Grammenidis, pour une conférence plénière sur la traduction : « Traduire : transcrire, ré-écrire ou ré-énoncer ? ». Sa conférence, ainsi qu'une sélection des communications présentées au colloque, seront publiées dans le Volume 24 des *Travaux Linguistiques du CerLiCO*, à paraître en juin 2011. Les Actes des colloques du CerLiCO paraissent depuis 1988 aux Presses Universitaires de Rennes (23 volumes parus à ce jour, dont le dernier en juin 2010) qui anime le Réseau des Universités de l'Ouest Atlantique dont font partie les Presses Universitaires François-Rabelais. Les articles issus des communications sélectionnées pour le colloque font chacun l'objet, avant publication, d'une double relecture.

Le thème traité cette année à Tours sera, conformément à la tradition du CerLiCO, le même que celui du prochain colloque de l'association qui se déroulera à Orléans en juin 2011, renforçant ainsi les activités communes des deux équipes du LLL. Outre le soutien financier du CerLiCO, de l'Université et du LLL, la région Centre s'est engagé et a soutenu notre manifestation.

Organisation du colloque

Président : Gilles Col (Tours)

Co-responsables : Jean-Michel Fournier (Tours), Sylvester Osu (Tours)

Comité scientifique :

Gilles Col, PR, Université François-Rabelais, Tours / EA 3850 LLL

Véronique Abasq, MC, Université François-Rabelais, Tours / EA 3850 LLL

Michel Achard, Professeur Assistant, Rice University, Houston, Texas, USA

Gabriel Bergounioux, PR, Université d'Orléans / EA 3850 LLL

Catherine Collin, MC, Université de Nantes / EA 3827 LING,

Jean-Louis Duchet, PR, Université de Poitiers / EA 3816 FoReLL,

Jean-Michel Fournier, PR, Université François-Rabelais, Tours / EA 3850 LLL,

Jacques François, PR, Université de Caen / EA 4255 CRISCO,

Simos Grammenidis, MC, Université Aristote, Thessalonique, Grèce,

Frédéric Lambert, PR, Université Bordeaux 3 / UMR 5610 ERSàB,

Sylvester Osu, MC, Université François-Rabelais, Tours / EA 3850 LLL,

Véronique Rauline, MC, Université Paris Ouest Nanterre / EA 370 CREA,

Noëlle Serpollet, MC, Université d'Orléans / EA 3850 LLL

Appel à communications

TRANSCRIRE ÉCRIRE FORMALISER

Pour se constituer comme science, la linguistique opère une rupture avec la philologie. En abandonnant l'édition et le commentaire des textes, elle établit son champ d'observation en dehors des écritures vernaculaires et compare des formes sonores. Elle ouvre ainsi un accès à toutes les langues (et à tous les états de langue), qu'elles aient été ou non consignées par écrit. Le geste fondateur de la discipline est présent aussi bien dans la reconstruction des comparatistes, qui proposent une notation de formes conjecturées, que dans le travail de terrain de Boas, Bloomfield et Sapir en Amérique du Nord, de Troubetzkoy dans le Caucase.

TRANSCRIRE, c'est interroger quelle signification revêt la conversion d'un signal sonore en une représentation graphique. Se trouvent concernés, en plus des questions de phonétique et de phonologie, les aspects figuratifs (sémiologie, langue des signes...), acoustiques (capture et traitement du son, visualisations...) et anthropologiques (introduction de l'écriture dans les civilisations à tradition orale, les graphies comme mode privilégié de la connaissance savante). Cette transcription recourt à des moyens variés (pictogrammes, syllabaires, alphabets...) dans une relation complexe avec la mise en forme des langues. Les idéogrammes imprègnent la culture chinoise dont ils assurent l'unité malgré la dialectalisation alors que la notation de l'arabe préserve la transparence de ses mécanismes apophoniques. L'écriture a des répercussions dans les représentations cognitives des auditeurs comme dans la structure sociale qui s'y trouvent inscrites en retour.

ECRIRE, c'est poser la question des relations entre un système conventionnel de représentation graphique et la structure (phonologique, morphologique, syntaxique, sémantique...) des langues. Les correspondances, phonie/graphie, les conventions orthographiques, jusqu'à la création de l'API se situent dans une notation analogique et descriptive qui a permis la production d'instruments tel que le dictionnaire sans expliquer le fonctionnement des langues. La linguistique s'affranchit de ces écritures reçues au moment où elle opère une distinction radicale entre ce qui s'entend et la structure des langues. Celle-ci n'est plus la consignation d'une observation mais le produit d'une analyse qui requiert à son tour une notation spécifique : une formalisation.

FORMALISER, c'est répondre à la nécessité d'un principe d'explication pour construire un système (et non plus un inventaire) ayant ses règles propres et qui vaut universellement dans le temps, dans l'espace et dans la société. Linguistique « algébrique » (Saussure) dont une partie des propriétés est justiciable d'un traitement automatique.

TRANSCRIRE, ECRIRE, FORMALISER constituera le programme des colloques du CerLiCO qui auront lieu à Tours en juin 2010 et à Orléans en juin 2011 afin d'appréhender l'objet de la linguistique au moment de sa constitution, dans sa genèse et ses effets.

(appel publié en juin 2009, sélection des communications en novembre)

Programme

Vendredi 4 juin 2010

9h00-9h15 **Accueil / discours**

9h45h – 10h25 : *Ecrire les nombres : quelles analyses pour la linguistique ?*

Sophie Saulnier, Modyco (CNRS / Université Paris 10-Nanterre)

10h25-11h05 : « *Formaliser* » *ACTUALLY: Reformulation, Formalisme, ou autre ?*

Lucie Gournay (Université Paris 12)

11h05 – 11h25 **pause**

11h25 – 12h05 : *Dans quelle langue transcrire les noms propres ?*

Sarah Leroy (Modyco CNRS / Université Paris 10-Nanterre)

14h00-14h40 : *Transcrire sans interpréter ? Illusion, nécessité ou possibilité*

Elisabeth Delais-Roussarie, H. Yoo (CNRS / UMR 7110-LLF, Université Paris-Diderot)
et B. Post (University of Cambridge, RCEAL (Research Centre for English and Applied
Linguistics))

14h40-15h20 : *Écrire une langue sans forme écrite: réflexions sur l'écriture et la transcription de la Langue des Signes Italienne (LIS)*

Claudia Savina Bianchini, Université de Paris 8, Saint-Denis, Università degli Studi di
Perugia, Istituto di Scienze e Tecnologia della Cognizione (ISTC), CNR, Roma

15h40-16h20 : *Productions langagières bilingues et bimodales d'enfants sourds : quels outils de transcription, pour quelles analyses ?*

Agnès Millet et Isabelle Estève (Université Stendhal Grenoble 3/ Lidilem)

16h20-17h : *'Textos sourds' en Métropole et à La Réunion : « Zones flottantes » du français ordinaire et spécificités du contexte de surdité*

Marion Blondel (UMR7023, CNRS-Paris 8), Jeanne Gonac'h, LIDIFRA, Université de
Rouen, Gudrun Ledegen, Université de St Denis de La Réunion, Julia Seeli, Université de
St Denis de La Réunion

17h00 **Assemblée générale de l'association CerLiCO**

Samedi 5 juin 2010

9h30-10h30 : *Traduire: transcrire, ré-écrire ou ré-énoncer?*

Conférence de Simos Grammenidis

Université Aristote de Thessaloniki

10h40-11h20 : *À la poursuite d'outils spécifiques, ou : comment transcrire l'acoustique ?*

Chantal Rittaud-Hutinet (Paris 3 / EA 1483)

11h20 – 12h00 : *Transcription de corpus d'apprenants multilingues de FLE et analyse interphonologique : enjeux méthodologiques*

Isabelle Racine, Genève, (Suisse), Sylvain Detey, Waseda University (Japon), SILS et Université de Rouen, Françoise Zay Université de Genève (Suisse) Yuji Kawaguchi, Tokyo University of Foreign Studies (Japon)

14h00-14h40 : *La controverse sur la transcription de l'anglais australien : une question identitaire*

Marjolaine Martin (Université François-Rabelais /LLL)

14h40-15h20 : *De la transcription à l'écriture : comparaison entre les cas du picard et du valaque*

Marijana Petrovic et Jean-Michel Eloy (Université de Picardie)

15h20-16h00 : *Quels systèmes de transcription phonétique et phonologique adopter dans le continuum dialectal italien ? Comment transcrire le polymorphisme dans les aires dialectales de l'Emilie occidentale (Italie du nord) ?*

Louis Begioni (Université Lille 3)

16h - 16h30 Clôture du colloque

Collaboration avec le laboratoire LaTTiCe UMR 8094 (CNRS, ENS & Université Paris 3)

Un groupe de travail est animé au LaTTiCe par un chercheur du LLL, Gilles Col, autour de la modélisation de la construction dynamique du sens. Ce groupe réunit des chercheurs du CNRS essentiellement, Bernard Victorri (DR), Jeanne Aptekman (post doc), Stéphanie Girault (IR) et Thierry Poibeau (CR). L'objectif est de proposer une formalisation compositionnelle gestaltiste pour décrire la construction du sens au fur et à mesure du déroulement du discours.

Les recherches du LaTTiCe portent principalement sur la langue française dans une perspective synchronique, mais aussi diachronique et comparative interlangues. Les données linguistiques – incluant une large part d'études sur corpus – sont appréhendées dans une perspective de description, de modélisation et d'expérimentation. Ces recherches s'inscrivent dans les débats cognitifs et anthropologiques actuels sur les rapports du langage et de l'esprit.

L'orientation cognitive des travaux conduits en partenariat portent sur la diversité des formes linguistiques qui permettent d'exprimer de grandes catégories notionnelles telles que l'espace ou la comparaison. Ils concernent également la structuration des discours qui contribuent à l'ancrage des énoncés dans leur contexte et signalent les relations entre les situations dénotées par ces énoncés. Ces études sont menées dans le but (entre autres) d'explicitier et de modéliser les instructions interprétatives codées par ces expressions et de mesurer / expliquer la façon dont les sujets les traitent dans la compréhension en temps réel. C'est cette partie des recherches conduites au LaTTiCe qui est au centre de la collaboration avec le LLL.

Le groupe de travail s'attache effectivement à mettre en évidence les instructions fournies par les unités linguistiques d'un énoncé lors de leur perception dans le déroulement du discours. Ces instructions dynamiques sont formulées par leurs formes schématiques qui reposent sur le principe de convocation / évocation. Elles répondent à la double question : qu'est-ce qu'une unité *convoque* sur la scène verbale pour pouvoir construire du sens et qu'est-ce que cette unité apporte à la construction de la scène – ce qu'elle *évoque* – en fonction des éléments qu'elle a convoqués et sur lesquels elle agit ? La finalité est de proposer une modélisation de la construction du sens qui donne une place centrale à la polysémie (Victorri & Col 2007, Col 2008a, 2008b, Col, Aptekman, Girault & Victorri, 2010, Aptekman & Co, 2010).

Un projet d'ouvrage consacré à l'exposé des principaux résultats ainsi qu'à la présentation d'une « grammaire instructionnelle » est actuellement en cours. La collaboration avec le LaTTiCe se poursuivra lors du prochain quadriennal (cf. le projet de l'axe de recherche SEC, « Sémantique énonciative et cognition ») en s'élargissant à d'autres chercheurs des deux laboratoires.

THEMATIQUE TRANSVERSALE « CONSTRUCTION DU SENS »

Rapport développé

Responsable : Sylvester N. Osu

Participants : V. Abasq, C. Agafonov, A. Cambourian, G. Col, G. Deléchelle, A. Delplanque, J.-M. Fournier, T. Gallèpe, N. Garric, D. Maurel, S. N. Osu, A. Piel, J.-P. Régis, F. Toupin, J. Constanza, P. Fournier, M. Martin, A. Ndione, E. Lecuit, J. M. Nagy, G. Potjans, I. Girard, S. Vanhoutte, L. Perbost, J. Longhi, E. Sauquet.

Participation des étudiants de M2.

Autres participations : Feifei Chen (sinophone et sinologue. Université de Franche Comté, Besançon), Samira Moukrim (LLL-Orléans).

1. Objectifs

La spécificité, et l'originalité, de ce programme réside dans son objectif de faire collaborer à l'analyse de la construction du sens des spécialistes de disciplines linguistiques très différentes, soit non seulement des chercheurs habituels de l'énonciation (dans l'optique d'A. Culioli) mais aussi les chercheurs en TAL, en Analyse du discours, en diachronie de l'anglais ou encore en morphophonologie, et ce dans des langues diverses dans un souci comparatiste double par nature : unité du langage, diversité des langues. En d'autres termes, on cherche ce qui varie de langue à langue et ce qui peut être considéré comme invariant, puis à quel niveau d'analyse (structurel ou abstrait) se situe un tel invariant. Il est ainsi né de la volonté des membres de LLL Tours de se constituer un domaine d'action et d'interaction scientifique commun, à la fois donc lieu d'échange et de convergence, mais aussi ambition d'explorer les potentialités d'une linguistique plus intégrative.

2. Problématique

La plupart des recherches actuelles en linguistique intègrent d'une manière ou d'une autre, la dimension sémantique du langage. C'est le cas en particulier de la grammaire cognitive dont les représentants les plus connus sont R. Langacker, L. Talmy, C. Vandeloise, A. Wierbiczka ; de la Functional linguistics avec M.A.K Halliday, E. Sweetser, Bybee, K. Lambrecht, T. Givon ; de la Formal semantics à la Montague, de la Pragmatique avec O. Ducrot, J. Anscombre ; de l'Analyse du discours avec P. Charaudeau ; de la Sémantique générale de B. Pottier, de la Generative semantics de G. Lakoff, Postal et Ross ; de la Functional syntax à la Kuno, puis de la Dynamic syntax de R. Kempson et ses collègues de King's College London, pour ne citer que celles-là. Le point convergent de toutes ces approches, c'est que chacune tente d'une manière spécifique de rendre compte du sens en tant qu'entité produite dans et à travers le discours.

Les travaux de la thématique « Construction du sens » s'inscrivent dans cette perspective. Toutefois, cette thématique centre son investigation non pas sur le résultat, autrement dit le produit, mais sur ce qui 'engendre' ce produit. Ainsi, elle s'inscrit dans l'optique théorique de la linguistique de l'énonciation telle qu'elle a été formulée et élaborée par A. Culioli et son équipe (cf. Culioli 1990, 1999 et ailleurs). En effet, non seulement la construction du sens se trouve au centre même des préoccupations de la linguistique de l'énonciation mais aussi cette approche

offre un outil théorique rigoureux pour la description fine, la caractérisation et l'explication des phénomènes. De plus, dans le cadre de la linguistique de l'énonciation, l'objet 'sens' est considéré comme le résultat de l'interaction de diverses opérations linguistiques, lesquelles sont représentées en surface par des formes appelées également des marqueurs (au sens de marqueurs d'opérations). Ces formes ou marqueurs peuvent être la prosodie, les mots, les syntagmes, les locutions ou les propositions dans des énoncés. Ces représentations en surface sont effectuées par le sujet énonciateur relativement à la situation d'énonciation. Toute recherche portant sur la construction du sens se doit désormais de prendre ces deux derniers paramètres en considération.

3. Diversité des langues

La recherche au sein de cette thématique transversale est abordée à travers plusieurs langues qui y sont représentées à savoir, l'allemand (langue germanique de la branche indo-européenne), l'anglais (langue germanique de la branche indo-européenne), le dagara (langue gur de la branche Niger-Congo), l'espagnol (langue latine de la branche indo-européenne), le français (langue latine de la branche indo-européenne), l'ikwere (langue igboïde de la branche Niger-Congo) et le russe (langue slave de la branche indo-européenne). De ce fait, le travail qui y est mené se veut contrastif et en même temps, s'inscrit dans une optique typologique. Il est à rappeler que la linguistique de l'énonciation vise à appréhender le langage à travers la diversité des langues naturelles. Ce qui signifie que c'est en allant jusqu'au bout de la singularité des langues qu'on arrivera à cerner les mécanismes généralisables de l'activité langagière. Il a semblé opportun de tenter sur la base de cette diversité des langues, de cerner le mécanisme de la construction du sens à travers des phénomènes précis. Bref, il s'agit de rendre compte de la façon la plus précise possible, des manières spécifiques par lesquelles chacune des langues représente des phénomènes particuliers en rapport avec la construction du sens.

4. Réalisations

L'équipe a abordé dans les cinq années écoulées deux thèmes majeurs à savoir la dénomination (2005-2007) et la temporalité (2008-2010). Un résumé de notre investigation sur chacun des deux thèmes est présenté ci-dessous.

4.1 De la dénomination à la construction d'identité

Pour la période 2005-2007, l'équipe a choisi de travailler sur la problématique de la dénomination. D'abord, cette problématique nous semblait, à défaut d'être un prolongement, être en lien avec les travaux précédents sur les Noms propres par certains membres de l'équipe, et sur les verbes d'apparence par certains autres membres de l'équipe. Ensuite et surtout, la dénomination apparaissait comme un lieu où se manifeste l'interaction entre le lexique (noms, noms propres), le discours (catégorie, catégorisation ou classification, construction d'identité), la morphologie (processus de formation des noms), la morphophonologie (voir en particulier, la prononciation des noms propres dans les différentes langues), la syntaxe et l'énonciation (processus d'identification, la négociation du sens). Le postulat était que seule la prise en compte de cette interaction permettrait de mieux rendre compte des phénomènes observés et ainsi définir plus précisément la dénomination.

La première phase d'exploration nous a permis de constater que la dénomination renvoie à la catégorisation des êtres et/ou des objets en ce sens que catégoriser, c'est identifier un élément à un groupe tout en marquant sa singularité. Cela revient à marquer une différence entre des objets ou des personnes, et donc, à construire une identité. Dès lors, la dénomination semble

construire l'identité à travers un processus de différenciation et inversement, marquer la différence à travers la construction d'une identité. Ce renvoi à la catégorisation nous a fourni un moyen de rapprocher la dénomination de la réduplication (voir par exemple en français, « je veux un café café ») tandis que la mise en jeu de la différenciation dans son fonctionnement rappelle les verbes d'apparence (cf. numéro spécial du *CORELA* coordonné par Delplanque, A. 2006). Par la suite, notre investigation a porté en parallèle sur la dénomination et sur la réduplication comme l'illustrent les calendriers de travail suivants :

2005-2006

19 janvier 2006: « Autour de l'identification » par G. Deléchelle

9 février : « Réduplication : état des lieux » par S. Osu

9 mars : « Classification et tabou » par J.P. Régis

13 avril : « Réduplication en Dagara » par A. Delplanque

11 mai : « Noms propres et expressions définies en tant que supports de la notion de personne » par T. Grass

8 juin : « Réduplication en russe » par C. Agafonov

2006-2007

12 octobre 2006: « Traitement phonologique des noms propres anglais » J.M. Fournier

9 novembre : « Constitution et exploitation d'un dictionnaire relationnel multilingue des noms propres » D. Maurel

7 décembre : « La représentation syntaxique des prépositions faibles du berbère de Chemini » M. Haiden

18 janvier 2007 : « Dérivés adjectivaux en -ish des noms d'animaux » Toupin & Lagorce

15 février : « Terminologie juridique : modèle bilingue canadien » T. Grass

15 mars : « Un tsunami médiatique et linguistique : de la dénomination à l'événement discursif » N. Garric

12 avril : « Réduplication en ikwere » S. Osu

10 mai : « Alternance morphologique systémique dans les noms propres en espagnol » A. Piel

21 juin : « Réduplication en chinois » Fei Fei Chen (Invitée, Université de Franche-Comté, Besançon)

Un peu mieux éclairée par les premiers travaux sur les deux phénomènes, l'équipe décida de reformuler son thème de recherche. De la dénomination l'on est passé à l'identité. Il s'agissait concrètement de cerner la part de la (L)angue dans la construction de l'identité d'une personne ou d'un objet. En d'autres termes, quels sont les procédés linguistiques qui rendent possible une telle construction d'identité. Mais la question s'est vite révélée très complexe aussi bien par les diverses acceptions (sens commun et sens technique confondus) qu'on attribue au terme d'identité, par sa diversité d'emplois (identité religieuse, identité politique, papier d'identité, etc.) que par la diversité de procédés qu'il recouvre. Nous l'avons tout de même abordée à travers les procédés de la dénomination et la réduplication dans les langues représentées dans notre équipe.

4.2 Bilan

Nous avons cherché à préciser les mécanismes spécifiques mis en jeu par chacun des phénomènes en prenant en compte des propriétés relevant du domaine lexical, de la phonologie, de la morphologie, de la syntaxe et de l'énonciation.

4.2.1 Dénomination

La dénomination a été abordée sous deux angles. Une première approche concerne le traitement automatique et le traitement phonologique des noms propres. Dans le premier cas de figure,

prenant comme exemple un dictionnaire relationnel multilingue, on s'est interrogé entre autres choses, sur les relations que les noms propres aussi bien dans une langue donnée que dans le passage d'une langue à une autre (la traduction), peuvent entretenir entre eux. Il s'agissait alors, de tenter de répondre à la question de la variation de noms que l'on peut rencontrer dans des textes lors d'une traduction. Par exemple, qu'est-ce qui dans un texte, permet de lier Mozart à Créateur de la flûte enchantée ? Ce type de travail exige que l'on précise les notions telles que synonymie, métonymie, hypéronymie. Or les relations entre noms propres permettent d'illustrer un système de catégorisation. Quant au traitement phonologique des noms propres, il s'agissait de montrer que l'accent final en anglais par exemple, n'existe que dans des mots provenant historiquement du français. La conséquence immédiate, c'est que l'identité linguistique des noms avec un accent terminal ou final en anglais peut facilement être établie. Il est à rappeler qu'une telle analyse a nécessité la connaissance de l'histoire du contact entre les deux langues. Une seconde approche se concentre sur les conséquences de l'acte de re-nommer. Or, re-nommer, cela consiste à balayer une unité pour la remplacer par une autre. Ainsi, on touche à la valeur identitaire de la personne, chose ou événement re-nommé. C'est ainsi par exemple que tsunami utilisé à la place de raz de marée à la suite de l'événement du 26 décembre 2004, puis étendu à d'autres domaines de la vie dans les discours médiatiques permet aux journalistes qui l'emploient de désigner un événement hors de propos. On parle désormais de tsunami politique pour signifier que l'événement politique est d'une grande ampleur, tsunami d'informations pour signifier que l'on est envahi par les informations et tsunami médical pour parler d'un problème médical qui fait trembler les pouvoirs en place.

4.2.2 *Réduplication*

A l'occasion de nos travaux sur la réduplication, nous avons mis en évidence le fait que ce procédé peut être distingué formellement et sémantiquement de la répétition ; que poser une forme (la base) et la reprendre (la copie) soit partiellement soit entièrement revient à introduire ce que dans la cadre de la TOPE, l'on appelle occurrence quelconque d'un domaine notionnel (base) et occurrence type d'un domaine notionnel (copie) ; puis que la juxtaposition des deux formes se traduit en termes d'identification. Ainsi, l'occurrence quelconque est identifiable à l'occurrence type. C'est ce mécanisme qui engendre non seulement une interprétation de type 'haut degré', 'le vrai', 'le parangon' mais aussi permet de construire une sous catégorie d'un domaine de référence. De ce fait, « un parfum pour les femmes femmes » (un spot publicitaire), c'est un parfum qui une fois acheté par une femme quelconque, affecte cette femme à une sous catégorie de femmes *par excellence*. On comprendra qu'un tel spot publicitaire n'a pour but que de pousser les femmes à acheter ledit parfum afin d'appartenir à la classe réservée aux femmes *par excellence*. Il a été également mis en évidence dans certaines langues que la réduplication à travers le processus d'identification d'une occurrence quelconque à une occurrence type, construit une propriété définitoire ou caractéristique d'un être humain ou d'un objet. Ainsi, lorsqu'en ikwere (langue du Nigeria), l'animal Ekeni dit que la nature lui donne « œil œil », il tend à dire que sa propriété définitoire se ramène à œil.

4.2.3 *Colloque international*

Encouragés désormais par nos observations, nous avons voulu voir s'il existait d'autres procédés dans des langues non représentées au sein de notre équipe et/ou comment et en quoi d'autres approches méthodologiques et théoriques permettraient d'aborder cette problématique que nous avons faite nôtre. Cela a débouché sur un colloque international intitulé « Construction d'identité et processus d'identification » et organisé à l'Université François Rabelais, Tours les 29 & 30 novembre, 2007. Organisé en quatre sessions parallèles, ce colloque a accueilli 66 communications orales plus 2 conférences plénières. Les diverses communications se sont

portées sur 12 langues à savoir, anglais, arabe marocain, arabe jordanien, espagnol, ewe, français, français de Côte d'Ivoire, ikwere, japonais, khmer, roumain et russe. Mais les communications se sont déroulées soit en français, soit en anglais. Quant au profil des participants, il était assez hétérogène : Américains, Canadiens, Nigériens, Belges, Italiens, Sénégalais, Espagnols, Français, Marocains, Japonais, Russes, Ivoiriens, Argentins et Jordaniens.

Ce colloque a révélé d'autres procédés par lesquels les langues du monde marquent et construisent l'identité. Ainsi, à côté de la dénomination et la reduplication, il y a aussi entre autres, l'hypallage, la paradiastole, l'antanaclase, le genre, la structure X entre les X, X chez les X, la structure X comme/tel que Y, le stéréotype, rupture de la chaîne anaphorique...

4.2.4 *Archivages*

Les deux conférences plénières ont été filmés par le service audio-visuel de l'Université François Rabelais, Tours et archivées sur le site de l'université :

http://lettres.univ-tours.fr/1190190229151/0/fiche___actualite/&RH=1192087652523

4.2.5 *Ouvrage*

Vu l'importance du résultat de ce colloque, nous avons décidé de recueillir la version rédigée et anonyme des communications présentées. Et après l'avis d'un comité de relecture que nous avons sollicité et mis en place, nous avons retenu 39 contributions, soit celles qui ont reçu un avis favorable de la part d'au moins de relecteurs. Revues et corrigées, ces contributions sont réunies dans un ouvrage de 623 pages, sous le titre « Construction d'identité et processus d'identification », publié chez Peter Lang (Berne, Suisse) en 2010. Les deux dernières pages de cet ouvrage présente la liste des relecteurs.

4.3 *Temporalité*

Pour la période 2008-2010, l'équipe a choisi de travailler sur la temporalité. Le programme comportait deux volets : 1) comment les langues et discours permettent-ils de construire le temps ? Dans ce cas, la temporalité est envisagée comme objet de description linguistique ; Concrètement, il s'agit de cerner les unités dans les langues qui marquent la temporalité. 2) comment le facteur temps est-il utilisé dans la description linguistique ? La temporalité est alors vue comme moyen de description linguistique. En effet, la temporalité a été largement abordée relativement à certaines langues de l'aire linguistique européenne mais il n'en est pas de même des langues relativement peu documentées et décrites telles que le dagara et l'ikwere. Nous voulions donc à travers une approche contrastive, faire une mise au point sur cette notion en linguistique. Les exposés pouvaient se situer dans l'un ou l'autre volets. L'exploration des concepts de forme schématique, de repérage et de situation d'énonciation, concepts au cœur de la TOPE, s'est avérée très utile à la réflexion. La puissance explicative de ces concepts tient précisément au fait qu'ils font de la variation un facteur indissociable de l'invariant et de ce fait, ils permettent un traitement unitaire et d'emblée syntaxico-sémantique des phénomènes. Ils permettent également de restituer la complexité de la temporalité souvent simplifiée à travers une représentation en axe. Les temps forts de notre réflexion ont été illustrés par quelques interventions. G. Col propose une analyse des prépositions « by » et « until » en anglais, soutient que ces éléments sont loin d'avoir un « sens premier » qui serait spatial ou temporel, mais qu'ils ont surtout un rôle instructionnel dans la construction du sens d'un énoncé, un rôle qu'il analyse comme leur forme schématique. A. Delplanque s'interroge sur la construction de la temporalité dans la conjugaison verbale en anglais comme en dagara, et S. Moukrim, dans une perspective contrastive, s'interroge sur la forme du présent en arabe marocain, berbère tamazight et français.

Mais si le fil conducteur de notre réflexion reste le rôle de la temporalité dans la construction du sens, certains exposés, en conformité avec les objectifs de l'équipe, ont toutefois été l'occasion de revisiter des travaux fondamentaux relatifs à la temporalité (c'est le cas, entre autres, du temps opératif chez Guillaume présenté par V. Mariscal et P. Bunte) alors que d'autres ont revisité des travaux antérieurs (c'est le cas de D. Maurel sur l'interprétation du temps verbal dans le TAL, puis sur la reconnaissance automatique des dates). Certains exposés se sont également demandés ce qui relève de la temporalité dans la grammaticalisation des unités (F. Toupin) tandis que d'autres ont voulu comprendre ce qu'il y a dans le préfixe « pre- » en anglais par exemple, qui peut permettre de renvoyer au locatif et au temporel (V. Abasq); mais aussi, comment le temps travaille la forme et le sens (J. Longhi) et enfin, le temps en tant que durée, c'est-à-dire longueur vocalique (M. Martin).

Sans prétendre avoir répondu à toutes nos questions, les divers exposés ont été en tout cas, l'occasion d'affiner nos questions et hypothèses, de porter un regard différent sur la temporalité, et ainsi renforcer notre objectif de faire collaborer à l'analyse de la construction du sens des spécialistes de disciplines linguistiques très différentes.

4.3.1 *Journée d'étude sur la temporalité*

En rapport avec ce programme, a été organisée à Tours le 5 mars 2010, une journée d'étude qui avait pour intitulé Journée d'étude Temporalité, sous la responsabilité de S. Osu et F. Toupin. Nous avons voulu à travers cette journée d'étude souligner à la fois la perspective typologique de nos travaux et la pluralité des approches comme l'illustre le programme ci-joint.

4.3.2 *Archivage*

L'ensemble des communications de la journée a été enregistré par le service audio-visuel de l'Université François Rabelais, Tours, et archivé sur le site de l'université, accessible via : http://lettres.univ-tours.fr/1266596192086/0/fiche___actualite/&RH=1179214274997

Journée d'étude sur la temporalité

Tours, 2009

Daniel Lebaud

Université de Franche Comté, LASELDI

Etude sur le français

« Les formes verbales du français : régimes énonciatifs et mises en scènes énonciatives des procès »

J'exposerai une recherche en cours, située dans le cadre de la Théorie des Opérations Prédicatives et Énonciatives, qui vise à articuler modes de repérages (régimes énonciatifs : "identité"(=), "différence"(≠), "rupture"(ω) et "ni identité ni différence ou identique ou différent"(*)) et modes de représentations des procès (mises en scène énonciatives : nécessaire, possible, contingent et intentionnalité, visée, Il y a). Il s'agit en conséquence, à partir d'une réflexion sur la distribution des formes verbales dans différents genres de textes (modes d'emploi, articles de presse, biographies, ...), d'une tentative de généralisation aux formes verbales, en rapport avec une réorganisation systémique, des notions subsumées sous l'appellation mises en scène énonciatives.

Stéphane Robert

LLACAN (UMR 8135, CNRS-INALCO)

Etude sur le wolof

« Le temps dans les langues africaines : l'exemple du système verbal du wolof »

Du point de vue grammatical, les langues africaines présentent des systèmes temporels assez divers que l'on ne peut ramener à un modèle unique. On présentera d'abord quelques caractéristiques remarquables relevées dans les systèmes temporels de ces langues : indications métriques, auxiliaires à valeurs aspecto-temporelles variées, amalgames entre temps et négation, entre temps et focus, orientation variable de la flèche du temps... On s'attachera ensuite à décrire plus précisément le système verbal d'une langue africaine parlée au Sénégal : le wolof. On montrera notamment que, dans cette langue, les indications aspecto-temporelles sont amalgamées à d'autres déterminations : il en résulte un système assez complexe à deux niveaux d'organisation et dans lequel la valeur d'accompli est construite de trois manières différentes. On soulignera également le rôle des types de procès dans la construction de la valeur temporelle des conjugaisons, ainsi que dans les glissements de valeur de certaines formes, du temps à la modalité.

Lotfi Abouda & Samira Moukrim

LLL (Université d'Orléans, EA 3850)

Etude sur le français

« Deux groupes de formes temporelles en français »

La morphologie du conditionnel, en tant que temps de l'indicatif, suggère l'existence dans ce mode de deux paradigmes de morphèmes « temporels » susceptibles de se concaténer. Une lecture possible de ces données suggère que la catégorie « temps », telle qu'admise habituellement, serait une notion générique synthétisant deux catégories dont la distinction implique la bipartition des temps de l'indicatif en deux sous-groupes distincts, l'un centré sur le présent, le second sur l'imparfait.

Silvio Liuzzi

CELIA (CNRS)

Etude sur le guarani

« Le temps en guarani »

Cette communication présentera le guarani dans la famille linguistique Tupi-guarani à l'époque pré-colombienne, coloniale, post-coloniale et actuelle ; une liste lexicale d'expressions temporelles ; les rapports temps / aspect / mode.

Laurent Gosselin

LIDIFRA (EA 4305) Université de Rouen

Etude sur le français

« De l'itération à l'aspect itératif »

Cette communication portera sur la nature de l'aspect itératif. Partant d'un examen critique de l'analogie, présentée dans quelques travaux récents, entre la quantification dans le domaine nominal et l'analyse de l'itération comme quantification plurielle sur les événements, j'examinerai la spécificité des propriétés aspectuo-temporelles de l'itération

Gary German

Université de Brest

Etude sur le gallois

« La temporalité et le système verbal du gallois : perspective diachronique »

Cette communication retracera l'évolution diachronique du système verbal du gallois depuis la période du moyen-gallois en soulignant notamment le passage d'un système synthétique à un système analytique, le développement de constructions périphrastiques liées (le plus souvent) à la progressivité, ainsi que les formes « bod » en B et en S, et, plus généralement, les conséquences que ceci a pu avoir sur la perception de la temporalité dans la langue moderne.

VARILING

La réponse à l'appel ANR Corpus 2006 du LLL et de ses partenaires

LE RÉSUMÉ

TRAITEMENT DES VARIATIONS LINGUISTIQUES DANS LES CORPUS (VARILING)

Le projet VARILING répond à une demande récurrente en linguistique : accroître une expertise sur les langues - en l'occurrence le français – qui intègre les variations et en permette l'analyse, qu'il s'agisse de la reconnaissance des données, de leur étiquetage ou des différents traitements dont elles peuvent faire l'objet. Les technologies ont été développées principalement à partir de corpus écrits et pour que de nouvelles connaissances, applicables à l'oral, puissent être élaborées, la communauté scientifique s'est souciée d'établir des recueils de parole circulante constitués en ressource.

Le plus important témoignage sur le français parlé avant 1980 est l'Enquête Socio-Linguistique à Orléans (ESLO 1), une collection de 200 interviews référencées (caractérisation sociologique des témoins, identification de l'enquêteur, date et lieu de passation de l'entretien), complétées par des captures sonores en contexte professionnel ou privé, soit au total 300 heures de parole et un corpus estimé à 4 500 000 mots. Une nouvelle enquête, ESLO 2, a été mise en chantier par le CORAL (Centre Orléanais de Recherche en Anthropologie et Linguistique, EA 3850) afin de constituer, à quarante années de distance, un corpus comparable dans les modalités de collecte et dans l'archive. L'objectif a été fixé à 400 heures environ de documents sonores, soit 6 000 000 de mots environ.

ESLO 2 a pour caractéristique d'être le seul programme en France (il en existe un équivalent au Québec) qui se propose de raisonner la dynamique sociolinguistique du français en bénéficiant d'un horizon de rétrospection de quatre décennies. L'étude est à même d'intégrer la variation diachronique et sociale et leur interaction ; elle est ouverte à une comparaison dialectale avec l'enquête de Lormont conduite par le CELITH-MODYCO.

Dès sa conception, ESLO 2 a été conçue pour préfigurer une référence dans un domaine qui, à l'échelle internationale, est en structuration et où l'adoption d'un format standardisé de *collecte*, de *conservation*, de *traitement* et d'*analyse* est confrontée à la multiplicité des développements et des normes. La certification est construite d'abord en tenant compte des pratiques en usage concernant la fabrication des corpus oraux en linguistique. A partir de l'échange entre les acteurs de la recherche, une synthèse des recommandations sera effectuée par le consensus des usagers et des experts, déterminant une conception du traitement d'ESLO 1 et de fabrication d'ESLO 2 qui permette d'exemplariser ces deux fonds pour de futurs corpus (et leurs traitements).

Si ESLO 2 a une visée cumulative (accroître la quantité de données pour assurer des comparaisons avec d'autres), l'enquête est aussi réflexive (accompagner l'enquête, le traitement et l'exploitation d'une analyse de l'expérience pour contribuer à la définition des normes). Cette conception concerne :

- une prospective sur l'exhaustivité des usages avec un calcul de représentativité,
- un inventaire des techniques de collecte (formats d'enregistrement et numérisation),

- une politique de formation des enquêteurs et d'information des témoins afin d'intégrer dans les critères de variation celle liée à l'enquêteur avec pour projet l'organisation d'une école d'été sur ce thème,
- un recueil des données concurremment à l'enrichissement en méta-données,
- un codage et un catalogage anticipant les principales requêtes émergeant en linguistique, mais aussi en sociologie, en anthropologie, en histoire, en info-com...
- une transcription avec alignement dans une perspective de normalisation,
- un étiquetage, avec catégorisation et lemmatisation (en particulier, recherches sur les problèmes rémanents de disfluences de l'oral et de co-référence anaphorique en situation de parole spontanée),
- une procédure d'anonymisation (l'identification des questions posées par l'anonymisation permettra la confection d'un vade mecum des éléments à prendre en considération à cette étape du travail à partir d'une recherche sur la détection des entités dénommantes (et pas seulement des entités nommées)),
- un stockage, avec archivage et indexation,
- une procédure de mise à disposition : la construction et la maintenance du site doivent assurer une libre consultation sur Internet (avec une convivialité et une ergonomie des applications, si possible dans une version multilingue),
- des données partagées : interopérabilité et protections, en liaison avec les propositions formulées dans le cadre du programme pour le catalogage et codage des corpus CatCod qui prolonge le travail de l'EPML50, « Corpus d'interaction langagière ». Les spécifications retenues seront transmises au consortium « Text Encoding Initiative » (TEI) à titre de proposition.

Au-delà, seront dessinées la mise en place du suivi (maintenance, jouvence et sécurité) et les applications. Il ne s'agit pas d'anticiper les analyses, mais de les rendre possibles.

Ce projet témoigne d'une ambition forte. Constituer en prototype, à toutes les étapes de sa réalisation, un corpus qui puisse se situer au même niveau, qualitatif et quantitatif, y compris par sa dimension patrimoniale, que les grands corpus oraux fabriqués, ou en cours de fabrication, en Europe et dans le monde. En trois ans, l'exécution de ce projet doit contribuer de façon décisive à la structuration de la communauté, en partenariat avec l'ensemble des laboratoires, dans une démarche associant l'accumulation des connaissances et la critique de leur constitution, l'analyse des outils et la disponibilité des données, le dialogue des chercheurs et l'échange de corpus devenus interopérables autour d'une thématique : la prise en compte des variations dans le langage.

L'EXPOSÉ DES MOTIFS

1. Présentation des corpus

On se propose d'exploiter et constituer des corpus qui permettent de mettre en évidence, en jouant sur la comparabilité des données, les configurations de la variation et les solutions de traitement. Pour ce faire, on procédera de quatre façons, concurremment et complémentaires :

- en traitant un corpus ancien de français oral à « récupérer » (ESLO 1) par différence avec un corpus de français contemporain à constituer suivant les standards en cours d'élaboration, auxquels il est prévu qu'ils servent de banc d'essai (ESLO 2) ;
- en éprouvant ces deux corpus pour tester l'alignement et le traitement parallèle de données commensurables quoique hétérogènes ;
- en rapprochant les corpus ESLO 1 et ESLO 2 élaborés à Orléans d'autres corpus du français (Lormont / CELITH ; Phonologie du Français Contemporain) ;
- en confrontant ces corpus de français à des corpus de langues non écrites (corpus LACITO, créoles à base portugaise) et d'Amérique (langues de Guyane).

1.1 *ESLO 1* : un travail pionnier

L'Enquête Socio-Linguistique à Orléans (désormais : ESLO 1) a été conduite en 1968 par des universitaires britanniques avec une visée didactique : l'enseignement du français langue étrangère dans le système public d'éducation anglais. Elle comprend environ 200 interviews, toutes référencées (caractérisation sociologique des témoins, identification de l'enquêteur, date et lieu de passation de l'entretien), soit au total plus de 300 heures de parole incluant pour moitié des interviews en face à face et pour moitié une gamme d'enregistrements variés (conversations téléphoniques, réunions publiques, transactions commerciales, repas de famille, entretiens médico-pédagogiques, etc.).

ESLO 1 couvre l'ensemble des catégories socio-professionnelles, hommes et femmes, avec plusieurs locuteurs originaires de différentes régions. C'est un échantillon des formats de la communication, des tâches linguistiques, des types de discours selon une approche essentiellement dialogique. Ce corpus représente, par son ampleur, sa rigueur et sa cohérence, le plus important témoignage disponible sur le français parlé avant 1980 (corpus estimé à 4 500 000 mots). Si les fins de sa constitution étaient linguistiques, ESLO 1 est un témoignage unique sur les jugements concernant mai 68 vu de la province ou sur les représentations collectives de la cité à cette époque.

Dans sa version première, transférée d'Angleterre il y a une quinzaine d'années et déposée aux Archives du Loiret, ESLO 1 réunit les bandes magnétiques, un catalogue dactylographié, quelques centaines de feuillets de transcription manuscrites (d'une qualité inégale) et les fiches d'identification des locuteurs. Actuellement, les documents sonores ont été recolligés et complétés (la conservation avait été défectueuse), numérisés à partir des enregistrements et une indexation et un premier catalogue informatisé a pu être réalisé. Parallèlement, l'exploitation exhaustive d'un sous-ensemble a été entreprise au point de rencontre de données linguistiques variationnistes et cognitives (description d'une tâche). Il reste à transcrire et baliser l'intégralité du corpus selon des principes ayant valeur de normalisation afin de mettre l'ensemble des données à la disposition de la communauté scientifique dans un format qui en permette une exploitation fiable, optimale et intensive, y compris pour des applications industrielles après sélection des contenus (en particulier concernant l'anonymisation).

1.2 *ESLO 2* : construire de la comparaison

En partant des acquis d'ESLO 1, une nouvelle enquête, dénommée ESLO 2, a été mise en chantier par le CORAL (Centre Orléanais de Recherche en Anthropologie et Linguistique, EA 3850). Il s'agit, à quarante années de distance, de constituer un corpus comparable dans le produit attendu et dans les modalités de la collecte : l'objectif a été fixé à 400 heures environ de documents sonores qui totaliseraient approximativement 6 000 000 de mots. Réunis, ESLO 1 et ESLO 2 formeront une

collection de 700 heures d'enregistrement, soit plus de 10 000 000 de mots, ce qui est considéré aujourd'hui comme une valeur repère pour les investigations projetées.

ESLO 2 a été conçu pour préfigurer la référence attendue dans un domaine qui en est encore à se structurer et dans lequel se manifeste de manière récurrente une demande de définition pour un format standardisé de *collecte*, de *conservation*, de *traitement* et d'*analyse* :

- la *collecte* sur le terrain est première, non seulement dans ses aspects techniques, aujourd'hui bien maîtrisés, mais dans la définition du profil de l'échantillon représentatif et dans la problématisation des interactions entre les témoins et les enquêteurs ;
- la *conservation*, qui inclut la préservation des supports, l'indexation des contenus et l'accessibilité (c'est-à-dire la protection) des données, conditionne le partage des sources à des fins d'étude scientifique ou didactique ;
- le *traitement*, en lien étroit avec le développement des matériels et des langages informatiques, suppose la maîtrise d'une chaîne d'opérations, depuis la conversion numérique des enregistrements jusqu'à une transcription balisée et ouverte à l'ensemble des interrogations pertinentes pour les demandes du linguiste, du sociologue ou des décideurs, des didacticiens voire du grand public ;
- l'*analyse* constitue l'épreuve des théories (et des logiciels) puisqu'elle compare les formalisations et les opérations et qu'elle valide ou infirme les hypothèses en prenant argument de leur compatibilité aux faits.

Les acquis en matière de conservation, de traitement et d'analyse seront reportés sur ESLO 1 comme le requiert la comparabilité attendue.

C'est le second objectif : évaluer la dynamique sociolinguistique du français en bénéficiant de la constitution d'un observatoire constitué par le renouvellement de l'enquête à quarante années de distance. Les usages de la langue ont évolué (vitesse d'élocution, contours prosodiques, tours syntaxique, choix des métaphores et des modalisateurs, stock lexical et modes de dérivation, termes d'adresse ou d'insulte, locutions et allusions...) : il reste à apprécier l'importance, la cohésion et l'orientation des phénomènes avant d'en proposer des principes d'explication. Le changement affecte aussi les jugements sur l'emploi : perception des « accents », tolérance au français familier et aux termes argotiques, relativisation des jugements scolaires. Autant qu'une pratique linguistique qui se transforme, il y a, dans l'usage de la parole, le modèle prégnant des régies de discours qui formatent de nouvelles normes et de nouvelles relations aux usages (modèles des animateurs télévisions et radios, des acteurs ou des professionnels de la parole publique, voix synthétiques des automates et boîtes vocales...) et qui tendent à établir les modèles de référence pour une communauté d'auditeurs en interaction.

La disponibilité des informations concernant les locuteurs d'ESLO 1 permet de prendre en compte la diversité des changements en les rapportant à leur époque et aux paramètres sociaux qui les déterminent, dessinant l'inégalité des résistances ou des propensions à la transformation de la langue, mais aussi une typologie et une dynamique de ses évolutions.

1.3 Lormont : variations sociolinguistique, dialectale et diachronique

Entre 1980 et 1981, une grande enquête sociolinguistique sur le français parlé a été menée à Lormont, communauté urbaine située au nord-est de l'agglomération de Bordeaux. L'une des variables analysées concernait l'élision ou le maintien du schwa en fin de polysyllabe, phénomène particulièrement pertinent dans le parler méridional.

Avant de procéder à une analyse sociolinguistique de faits de variations avérés comme celui du e muet, le recueil des données a été précédé d'une enquête du contexte social et pragmatique de Lormont au moyen d'une méthode statistique et ethnographique (constitution d'un échantillon de locuteurs représentatif de la structure socio-démographique de la communauté et recherche concernant la structure objective de l'espace social avec les représentations que se faisaient les agents sociaux de leur environnement)..

Cinquante-six personnes, natives de l'agglomération bordelaise, ont été retenues pour l'analyse sociolinguistique menée sous forme d'interviews de deux heures chacune, avec fiche biographique, questionnaire sociologique (suivant le modèle de Bourdieu) et linguistique.

Recueillir des données en reprenant le même échantillon de population et le même format d'interview que ceux de l'enquête de 1980 permettrait de mener une véritable enquête longitudinale, ce qui n'a jamais été fait dans l'histoire de l'étude du français parlé. Les avancées scientifiques d'une telle enquête sont potentiellement considérables : elle permettrait de mieux comprendre les processus d'évolution des variables étudiées et, par là-même, certains aspects du fonctionnement du langage comme les conditions d'émergence de patterns grammaticaux, pragmatiques, et phonologiques avérés.

Aujourd'hui, il importe d'élargir les variables observées grâce à de nouvelles modalités d'enquêtes centrées sur l'acquisition de données interactionnelles en intégrant des aspects interactionnels et conversationnels. Ainsi de certains phénomènes conversationnels comme l'autocorrection ou les pré-séquences, les écarts entre les styles formels et familiers. On obtiendra ainsi un deuxième corpus pour la comparaison dialectale avec une variation diachronique intégrée.

2. Les fonctions d'un corpus : standard, norme, benchmark

L'inscription d'ESLO 2 dans une démarche qui vise à définir les propriétés d'un corpus de référence dans la méthodologie de constitution de données implique la confrontation des différentes étapes de l'élaboration à celle d'autres programmes, aboutis ou en construction, afin d'anticiper les contraintes de l'interopérabilité tout en évaluant leur impact sur l'analyse et les théories linguistiques.

2.1 Vers la conception d'un prototype

La réalisation du corpus de référence est effectuée en trois temps :

- 1) une recension des pratiques en vigueur dans la communauté scientifique pour ce qui a trait aux corpus oraux (avec une participation au programme CNRS concerné, à l'ILF et à la coordination du *Guide des bonnes pratiques*),
- 2) une synthèse des recommandations élaborées par l'échange et le consensus à l'intérieur de la communauté des usagers et des experts exposés au projet, une démarche calquée sur les procédures de standardisation internationale.
- 3) une conception de traitement d'ESLO 1 et de fabrication d'ESLO 2 qui puisse permettre de les établir comme corpus prototypique.

Le premier principe retenu revient à considérer qu'un corpus numérique n'existe que par l'exploitation qui peut en être faite, aujourd'hui et dans l'avenir (ce qui suppose d'éviter toute opération irréversible qui n'autoriserait pas une récupération des sources dans leur état premier). Il faut déterminer au préalable et *a priori* l'ensemble non dégradant d'interrogations et de sollicitations, quels que soient les auteurs des demandes et leurs requêtes. Le conditionnement le plus aisément manipulable est celui obtenu par numérisation, c'est-à-dire par une transformation des data qui permet de tracer des graphes de relations. L'accessibilité, élargie au-delà des producteurs, requiert une paramétrisation qui autorise la consultation par des systèmes informatiques différents, sur support fixe ou à distance. La question de la structuration de la base pour l'adapter aux normes, standards et routines pratiqués est déterminante.

C'est dans cette phase qu'intervient la chaîne des actions qui vont de la capture du matériau sonore à la réponse à une demande qui peut être formulée par n'importe quel acteur de la recherche, qu'il interroge les propriétés physiques du signal ou l'emploi des connecteurs, le témoignage du consommateur ou les états mentaux du locuteur. La disparité des requêtes prescrit le choix de systèmes ouverts, évolutifs et interopérables, c'est-à-dire qui respectent les normes de transport et de lecture en vigueur et qui anticipent les normes futures. L'association avec d'autres équipes confrontées aux mêmes problèmes, et partant le choix des laboratoires, ont été dictés par cette perspective.

Un corpus numérique doit être associé aux protocoles et aux outils qui en permettent le stockage, le transfert et l'utilisation : que le rendu se présente sous forme de texte (transcription) ou de son (signal), il s'agit dans les deux cas de concevoir la structuration spécifique d'un objet. Les corpus numériques ont ceci de particulier que, par leur taille exponentielle et leurs possibilités rapides d'évolution ou d'obsolescence, ils ne se prêtent pas immédiatement à une exploitation spontanée et pérenne. Dans le cas d'un corpus sonore, les difficultés augmentent avec la spécificité des données à

manipuler. Comment articuler le système documentaire propre à l'exploitation du corpus et les objets dont il est constitué ? Les découpages ou les formats de requête, qui sont d'évidence sur du texte, ne le sont plus dès que le domaine d'investigation est un fichier sonore.

La finalisation de la recherche vers la définition d'une structure versatile et utilisable d'un corpus suppose au préalable une bonne représentation des objets à indexer, aussi bien en termes de typologie que de contenu. Un corpus numérique ne saurait se concevoir indépendamment des outils de traitement spécifique des métadonnées, en renfort des outils destinés au traitement même des objets. Avec un corpus de dix millions de mots, les opérations de structuration et de mise à disposition exigent que soit établi en amont un système d'indexation des données au moyen d'un étiquetage par métadonnées. Ce sont les normes et les standards qui se mettent en place dans ce domaine, autour du XML, langage de structuration des données, et de ses dérivés.

2.2 Les corpus linguistiques

Dans les disciplines linguistiques, les corpus oraux représentent à la fois une méthode de travail et un outil d'étude. Au fur et à mesure des développements informatiques, les techniques de numérisation et les possibilités de diffusion par le Web rendent possible une gestion raisonnée et à faible coût des corpus. Le nombre croissant de laboratoires qui ont fait le choix de pratiques orientées vers une exploitation automatique des ressources et la nécessité de transférer les données stockées sur support analogique a décidé la mise en place de vastes programmes de numérisation.

Les corpus sont devenus la norme de fabrication et de constitution des équipements de la langue (système phonologique, grammaires et dictionnaires, répertoires de locutions, analyseurs morphologiques...) en même temps que les premiers auxiliaires de la recherche en sciences du langage. Il s'en est suivie une multiplication, à l'échelle internationale, des programmes consacrés à la confection de corpus représentatifs, avec toutes les interrogations qui découlent d'une réflexion sur la qualité de la correspondance entre l'échantillon et la masse de données que serait l'impossible projet d'une collation exhaustive des échanges langagiers. Ainsi, la multiplication des entreprises de fabrication de corpus oraux pour la seule discipline que constitue la linguistique a en quelques années abouti à une importante disparité des codages, des formats et des méthodes de catalogage, de référencement et de diffusion.

Cette variété de procédures aboutit à un gaspillage d'investissement, matériel et humain, dans la mesure où l'effort de codage, de formatage et de développement est accru d'autant par la quantité d'initiatives, un coût encore augmenté par le nombre de passerelles à définir afin de rendre interopérables différentes initiatives incompatibles dans leur philosophie et leurs technologies alors même que certaines applications nécessitent la compilation de ressources provenant de différentes sources pour obtenir les volumes critiques nécessaires. À plus long terme, le cumul des connaissances comme le transfert vers d'autres domaines, aussi bien académiques qu'industriels, est rendu aléatoire.

Les descriptions et les besoins sont en effet différents pour des sociolinguistes, des linguistes de terrain, des dialectologues, des phonéticiens, des spécialistes de TAL ou d'ingénierie linguistique, sans parler d'exploitation en sciences humaines et sociales voire en sciences formalisées. Les analyses linguistiques et donc les codages varieront selon les investigations, allant d'études dédiées à des phénomènes phonologiques particuliers jusqu'à des exploitations qui se satisfont de transcriptions orthographiques aménagées, ou qui au contraire incluent des transcriptions phonétiques, des découpages morpho-syntaxiques, des lemmatisations, des alignements texte/oral, etc.

3. Phasage

On distinguera, dans le programme soumis à l'ANR, deux types de corpus : d'une part, ceux qui sont déjà constitués et qu'il s'agit de reprendre pour les réinscrire dans les standards d'aujourd'hui et les insérer dans les circuits d'échange télématiques. D'autre part, un corpus spécifique à élaborer à partir de principes qu'il convient de tester afin d'en vérifier la pertinence et la faisabilité.

Concernant les corpus constitués, on mentionnera les fonds disponibles auprès des laboratoires associés au projet : outre ESLO 1 (voir infra), le corpus de Lormont et ceux qui pourront être mis à

disposition par Modyco (Université Paris X), ceux du LACITO et du CORAL (créoles à base portugaise, palikur, wayana et wayampi en Guyane).

Le projet central autour duquel s'articulent les différents modules est néanmoins la réalisation d'un corpus spécifique, ESLO 2, dans sa visée cumulative (ajouter des données nouvelles à d'autres qui leur sont comparables) et réflexive (accompagner l'enquête, le traitement et l'exploitation d'une analyse de l'expérience pour participer à la définition des normes).

Bien que chaque étape suppose une anticipation sur le déroulement du processus dans son entier et que toute réalisation ne puisse être conçue sans un effet en retour sur les travaux réalisés dans les phases antérieures, on distinguera à la suite les différents temps de fabrication d'ESLO 2 en rappelant que l'objectif premier est la mise au point d'un corpus oral qui puisse être rapproché d'ESLO 1 tout en servant de prototype aux initiatives qui sollicitent des données orales pour l'élaboration de connaissances.

Phase 1 : Exhaustivité, représentativité, proportionnalité

A la différence de l'écrit public (imprimé) qui est centralisé par de grandes institutions en sorte que la conservation est en théorie identique à la production, aucune instance ne peut préserver l'intégralité des productions orales. De plus, alors que l'écrit privé suppose un fort degré de contrôle et une certaine mise en situation liée à la décontextualisation de la production et de la réception, l'oral se situe en dehors des processus d'appropriation contrôlé, même si l'école exerce une influence sur son emploi, et les conventions d'emploi n'y ont pas la même incidence. L'intérêt pour les écritures ordinaires, les chats ou le SMS a pu fragiliser des distinctions établies, elle ne les a pas abolies.

A défaut d'une exhaustivité des productions, c'est une exhaustivité des usages qui doit être envisagée, dans les différentes dimensions où ils s'inscrivent, c'est-à-dire au cœur même de la première difficulté que soulèvent ces corpus : la variation. ESLO 1 avait exclu deux catégories de locuteurs : les enfants (0 – 20 ans) et les non francophones. Si le contact de langues est un cas très particulier qui suppose une enquête spécifique, la nécessité d'un groupe témoin enfant/adolescent est indispensable.

De façon générale, poser la question de la représentativité, c'est d'abord procéder à l'inventaire des causes de variation. On les distribuera en trois catégories :

- celles qui sont liées à des propriétés intrinsèques des locuteurs, avec une ventilation par âge, par sexe, par CSP, par trajectoire sociale et par origine géographique, sur le modèle de ce qui avait été fait avec ESLO 1 et qui bénéficie depuis des apports cumulés de l'anthropologie et de la sociologie (on intégrera, en face du CSP, la définition donnée par le témoin lui-même de son statut social) ;

- celles qui relèvent des situations de discours et des niveaux de langue exploités, y compris les actes performatifs et la construction des identités relationnelles, la définition du cadre, l'organisation des tours de parole, la construction du point de vue, saisis dans la variété des situations de collecte et des modes de relation entre enquêteur et enquêté ;

- celles qui articulent formats cognitifs (type de tâche à accomplir : récit, description, suggestion, plaisanterie, demande d'assentiment) et exploitation des ressources linguistiques (figures du discours, opérations syntaxiques, choix lexical, deixis et procédures anaphoriques...).

Ces trois versants, distingués pour la commodité de l'exploitation, sont corrélés. Ils se caractérisent par leur caractère obligé (tout discours en tant qu'il est oralisé est assignable à un locuteur, un contexte de production et une tâche cognitivo-linguistique). A cette étape, il s'agit de préfigurer une grille de propriétés, comprenant un nombre fini d'entrées, ouvrant à une normalisation des types de requêtes possibles. Si un siècle et demi d'enquêtes sociologiques ont abouti à un ensemble réglé de conventions concernant le premier point, il n'en va pas de même pour les deux suivants qui ont fait l'objet de nombreuses études sans converger vers un catalogue fini d'items. La démarche retenue consiste à partir de cette interrogation : quelle question peut être formulée qui éclairerait les données de les appréhender à partir de telle caractérisation ?

ESLO 2 a fait appel à une sociologue, M. de Saint-Martin (EHESS), pour évaluer ses propositions de questionnaire et les échanges ont mis en évidence que la diffraction des modes d'opérabilité de la langue en situation ne se prêtent pas si facilement à une homogénéisation ou à une codification. A partir d'un éventail de situations et d'emplois, de façon prédictive et rétroactive, il est

prévu de procéder à une extension catégorielle permettant la détermination générique d'un ensemble fini de propriétés exploitables pour le catalogage.

La question d'une représentativité quantitative (présence dans le corpus d'une proportion homologue de types d'emplois et de locuteurs, conforme à celle des usages quotidiens et des rapports entre groupes sociaux) est pour le moment indécidable, sinon par la négative : on peut être assuré qu'il existe moins de déclamation poétique que de demandes de renseignement dans les magasins d'informatique. Une esquisse de ces mesures est néanmoins envisagée dans la sollicitation d'une reconstruction des successions d'interaction quotidiennes.

Phase 2 : Techniques de collecte : formats d'enregistrement et numérisation

Cette phase est aujourd'hui la mieux maîtrisée encore qu'elle soit susceptible d'évolutions extrêmement rapides. Les équipements proposés par les fabricants - qui cherchent souvent à imposer leurs propres normes dans une logique commerciale d'exclusivité - permettent un échantillonnage du signal, un stockage et des formes de restitution dont la qualité dépend plus souvent des conditions d'enregistrement que des performances électroniques. Les moyens de capture (micros) et de conversion ne sont pas au centre de nos interrogations même s'ils feront l'objet d'un examen attentif, en particulier avec les institutions (BnF) et les associations (AFAS) impliquées dans une conception académique de l'acquisition de données sonores.

A cette étape, il s'agit plus d'une sollicitation de compétences pour mesurer les avantages et les difficultés inhérents à chaque solution technique que d'une véritable normalisation. On trouve dans les fiches techniques du *Guide des bonnes pratiques* des indications sur ces points, en particulier sur l'obsolescence des formats et la durabilité attendue des produits, leur capacité de transfert vers d'autres supports et la fiabilité de leur stabilisation (qualité du signal) dans le temps.

Phase 3 : Formation des enquêteurs et information des témoins

Alors que les pratiques d'enquête sont aujourd'hui de règle dans tous les secteurs des sciences humaines et sociales, il n'existe pas véritablement de définition satisfaisante concernant ce que devraient être les critères d'une enquête valide. La formation se fait sur le tas et les praticiens apprennent autant et plus de leurs échecs (appareil d'enregistrement défectueux, questions oubliées, destruction des données au cours des manipulations, mauvais étiquetage...) que des conseils qui leur ont été prodigués.

ESLO 1 a réuni des collaborateurs de bonne volonté. Les enquêtes du CELITH et du LACITO ont été réalisées par des semi-professionnels (étudiants en thèse) ou des linguistes de terrain. Au CORAL, certaines études ont été conduites par des chercheurs immergés in situ sur de longues périodes (Guyane, Guinée Bissau et Cap-Vert, France), d'autres par des missions de durée variable.

Il paraît essentiel d'intégrer dans les critères de variation celle qui se trouve liée à l'enquêteur, à ses propriétés d'âge et de niveau social, de familiarité avec le témoin et de connaissance des objectifs de l'enquête. Dans un module d'ESLO 2, consacré au réenregistrement d'une dizaine de témoins d'ESLO 1 quarante années plus tard, l'enquêtrice est elle-même une retraitée de l'enseignement qui, pour des raisons d'âge, est plus à même de dialoguer avec des personnes âgées de plus de soixante ans.

On distinguera quatre types d'enquêteur :

- enquêteur « professionnel », rompu aux techniques de l'interview et de la collecte des documents, généralement par une pratique éprouvée au fil des ans ;
- enquêteur « apprenant », correspondant à la situation d'un étudiant en cycle Master ou doctorat ;
- enquêteur « médiateur », sollicité pour aller recueillir des données auprès de locuteurs auxquels l'enquêteur ne peut avoir un accès direct pour des raisons de langue ou d'organisation sociale (public féminin quand l'enquêteur est masculin par exemple) ;
- enquêteur « involontaire » lorsqu'une conversation est intégrée à un corpus sans que la finalité première ait été conçue au moment de la confection de l'enregistrement (lorsqu'un étudiant apporte une cassette où il parle avec quelqu'un de sa famille par exemple).

S'il est difficile d'imaginer un protocole qui permettrait de reproduire exactement la même forme d'entretien en variant seulement l'enquêteur, il est concevable de rechercher des indices, dans les formes d'adresse, le registre lexical, les marques de connivence ou de distance, qui répercuteraient la diversité des relations d'enquête.

L'objectif de cette phase serait, en partenariat avec les équipes impliquées sur différents terrains, et si possible à l'échelle européenne, de prévoir une école d'été qui exposerait les différentes pratiques, avec des exemples précis de résultats, qui réunirait les informations documentaires, en particulier électroniques et qui assurerait en parallèle une formation de haut niveau dans l'enquête. C'est une perspective ambitieuse et qui suppose l'appui des institutions (CNRS, Max-Planck Institut, ESF...) : elle a été anticipée par des séminaires et des sessions de travail qui acquerraient une autre dimension.

Phase 4 : Recueil des données

On ne traitera pas dans cette partie des données déjà collectées qu'il s'agit de récupérer pour les inscrire dans des formats exploitables par la communauté. L'enquête constituée en prototype est ESLO 2. Il s'agit, en deux années maximum, à partir d'un questionnaire réfléchi et d'un échantillon de locuteurs et de situations (phase 1), avec un matériel fiable d'enregistrement et de stockage (phase 2) et des enquêteurs préparés à leur tâche (phase 3), de recueillir 400 heures de parole, quarante pour cent par entretiens semi-directifs, quarante pour cent par collecte dans des situations variées, le restant par l'adjonction de modules spécifiques. Au nombre de ceux-ci, on citera :

- un module « témoins ESLO 1 » : cf. supra.
- un module « enfants » qui n'a pas d'équivalent dans ESLO 1 et qui pourrait faire l'objet d'une saisie vidéo,
- un module « personnalités » constitué par des interviews de personnalités de la société civile, de politiques et de publicistes qui donneront une image sonore de la parole officielle, commune à tous les auditeurs (ce sont ceux qu'on entend dans les médias, les cérémonies officielles, etc.).

A chaque élément d'enquête sera associé une méta-donnée concernant les circonstances de l'enquête (mode de contact du témoin, négociation de l'entretien, caractéristiques sociales de l'enquêté et de l'enquêteur, format d'échange, incidents...) qui feront l'objet d'une saisie séparée avec une possibilité d'interrogation distincte.

La collection des données et des méta-données constituera le matériau d'ESLO 2 qui, pour devenir le corpus de référence attendu, devra faire l'objet de la série d'opérations énumérées dans les phases 5 à 12, parallèlement à des traitements similaires effectués sur ESLO 1 afin d'assurer la comparabilité des data et, autant qu'il sera possible, sur les autres corpus apportés à ce programme de l'ANR en provenance d'Afrique, d'Amérique et de France.

Phase 5 : Codage et catalogage

Ces deux opérations sont liées aux propositions formulées dans le cadre de CatCod, créé en 2005 par une communauté d'enseignants-chercheurs, de chercheurs et d'ingénieurs travaillant sur des corpus de parole afin de préfigurer des solutions efficaces de partage des données, c'est-à-dire la définition de standards et de formats d'échange. La première initiative a consisté dans un échange d'informations sur les différentes pratiques en usage, structuré sous la forme d'une structure d'Equipe Projet Multi-Laboratoire du CNRS, l'EPML-50, « Corpus d'interaction langagière ». Des outils collaboratifs ont été conçus afin d'organiser et d'animer la communauté autour de ces thèmes. Il y a à présent une demande pour que le travail ne soit plus seulement descriptif et rétroactif mais prescriptif et prospectif. En ce sens, l'acquis de l'EPML-50 est directement associé au traitement d'ESLO 2 et au recyclage des autres corpus associés à cette demande d'ANR. Terrain d'expérimentation idéal, parce qu'il associe dès le début de sa conception les acteurs de l'EPML-50, ESLO 2 est déclarée comme membre participant du nouveau centre de ressources CRDO au sein d'un réseau de besoins applicatifs.

Cette phase doit aboutir, en partenariat avec les équipes impliquées au CRDO et à l'EPML, à la rédaction d'un document de spécifications qui propose des normes de standardisation, avec une large

consultation et une recherche du consensus entre les participants. Elles auront été utilisées et éprouvées, à titre expérimental, dans la reprise d'ESLO 1 et la confection d'ESLO 2 en matière de codage et de catalogage.

Afin de stabiliser les acquis de cette recherche, le document de spécifications sera transmis au consortium « Text Encoding Initiative » (TEI) à titre de proposition pour le chapitre transcription de l'oral. La prise en compte de nos travaux dans le cadre normalisateur de la TEI apporterait une garantie de pérennité et l'assurance d'une diffusion, participatif à la réduction d'une diversité des usages préjudiciable à l'échange et à la conservation.

Phase 6 : Transcription et alignement

Dans la continuité du travail de codage, la transcription pose une question spécifique pour laquelle des réponses différentes ont été apportées. En partant des acquis de tous les laboratoires, y compris ceux du LACITO, du CELTH, du LI et du CORAL, il sera fait, de la transcription des quatre cents heures de parole, un benchmark des solutions qui sera indéfiniment réutilisable pour d'autres applications.

Un problème particulier concerne l'alignement du signal sur la transcription, dans le prolongement des études entreprises à Leuven. Le format d'analyse du signal requiert une réflexion sur les outils disponibles, les manipulations effectuées, les distorsions qui peuvent en résulter et l'exploitabilité des représentations et des affichages. La possibilité de retrouver, à partir d'un item ou d'une séquence, la représentation temporelle ou spectrographique, le contour mélodique ou les variations de l'énergie est possible mais sur des formats qui exigent de grosses capacités de mémoire puisque interfèrent une recherche dans les données, un alignement de deux fichiers et du traitement graphique. La mise au point de systèmes à différents degrés de granularité, et la possibilité d'une sélection des informations, devrait permettre d'accélérer les temps de traitement et d'affichage en éliminant ce qui, pour une interrogation spécifique, n'est plus de l'information mais du bruit.

En revanche, l'opération inverse (retrouver l'ensemble des séquences présentant un certain type de variation en F0, ou tous les moments réunissant les formants d'une voyelle donnée) reste encore très complexe puisque la requête ne porte plus sur une chaîne de caractères mais sur la numérisation de données physiques. C'est un aspect qui ne sera pas traité dans cette étude.

Phase 7 : Etiquetage, catégorisation et lemmatisation

L'étiquetage n'est pas seulement un problème technique. S'il existe des logiciels performants, avec des taux de reconnaissance assez satisfaisants, leur entraînement a surtout été éprouvé sur des corpus écrits. Une caractéristique du français parlé est la présence d'unités non standard (apocopes et aphérèses, diminutifs, expression argotiques, troncations, « fautes de français », duplications, néologie dérivationnelle, antonomase...) et les attestations de dérivation impropre et de polycatégorialité qui brouillent les caractérisations taxinomiques.

A côté d'un travail technique de balisage des corpus et de lemmatisation, c'est donc une réflexion sur la morpho-syntaxe du français qui est à conduire. Une autre application concerne la détection des entités nommées dont la définition est à étendre à celle d'une détection des entités dénommantes.

Phase 8 : Anonymisation

L'un des obstacles récurrents qui se présente dans l'exploitation collective des corpus oraux concerne la protection des personnes qui ont accepté d'être sollicités et de répondre. Comme le prévoient les engagements pris par le chercheur, sous forme contractuelle, l'enquêté a le droit de conserver son anonymat ce qui suppose l'effacement ou le bipage de toutes les indications qui permettraient de lever l'anonymat. Il en va de même pour les personnes mentionnées au cours de l'interview qui sont en droit d'exiger la protection de leur vie privée.

Dans ce cadre, et en prolongement des recherches entreprises dans la base 7, ESLO 1 et ESLO 2 doivent servir de test aux programmes développés dans le cadre du LI concernant la reconnaissance des

entités nommées. En recourant à CasSys, un outil qui traite des corpus Unitex et qui permet, entre autres, de détecter les noms propres avec un taux d'efficacité supérieur à 90 %, on soumettra les résultats obtenus aux corpus oraux avec deux objectifs :

- tester sur du français oral transcrit des outils élaborés sur du français écrit,
- accroître la définition des entités nommées à des catégories indépendantes du nom propre (le maire d'Orléans en 2000 est l'exact équivalent d'un nom propre d'un point de vue juridique et, peut-être, philosophique).

Sur les centaines d'heures à traiter, on peut concevoir que tous les problèmes de l'anonymisation seront représentés ce qui permettra la confection d'un vade mecum des éléments à prendre en considération à cette étape du travail :

- détection et sélection des noms propres (tous ne sont pas à anonymiser : la Loire et Jeanne d'Arc ne sont pas à inclure dans l'effacement),
- repérage des éléments d'identification hors nom propre (profession, lien de parenté, participation à un fait divers),
- coordination à distance des éléments d'identification (un professeur ne permet pas d'identification, il n'en va pas de même s'il est précisé par ailleurs que c'est un professeur d'Université, spécialisé en électronique et que c'est une femme, auquel cas on peut arriver à un singleton).

Phase 9 : Stockage, archivage et indexation

La conservation des données, comme leur saisie, a fait l'objet d'un travail de réflexion dont on trouve la synthèse dans le *Guide des bonnes pratiques* dont les prescriptions seront suivies. Les questions du choix de l'institution d'archivage (le laboratoire lui-même, un centre spécialisé à créer sous l'égide du CNRS, les Archives départementales ou nationales, la BnF, un serveur dévolu à cette tâche) et des autorisations et des modes d'accès seront tranchées à la fin de l'ANR avec une justification de la solution retenue. Les discussions en cours à l'intérieur de chacun des établissements ne permettent pas d'anticiper une décision même si l'engagement des chercheurs dans les initiatives CNRS leur fait apprécier positivement une solution d'attente qui préfigure une réflexion en chantier à l'intérieur du Département des archives sonores de la BnF.

La question de l'indexation est autrement plus complexe. En effet, à la différence d'un ouvrage que l'on peut feuilleter, d'un fichier texte que l'on peut consulter par appel d'items, l'indexation des fichiers sons pose un problème dès lors qu'on ne se contente pas d'en traiter la partie transcription mais qu'on se propose d'en relever les caractéristiques de signal. Une proposition sera faite concernant la production de différents niveaux d'indexation :

- un code alphanumérique d'identification,
- une fiche de métadonnées (voir phase 4),
- un descriptif court du contenu (une centaine de signes),
- une synthèse d'un millier de signes recensant (i) les caractéristiques de l'enregistrement, (ii) les principaux éléments de contenus avec une liste de mots-clés réunis dans un lexique séparé, (iii) les problèmes que présente l'enregistrement,
- une fiche évolutive recensant la série des interventions opérées sur le fichier et l'historique des requêtes,
- un dossier associé concernant les exploitations du document dans la littérature scientifique et les applications.

Phase 10 : Mise à disposition : construction du site et traduction

Pour que ESLO 1 et ESLO 2 puissent effectivement fonctionner comme références dans la constitution des corpus oraux, en France et à l'étranger, il faut que soient rendues simultanément disponibles les données et l'explicitation des opérations qui les ont configurées dans les formats auxquels elles seront accessibles.

L'accès suppose l'installation (et la maintenance) d'un site librement consultable par Internet, avec une convivialité et une ergonomie des applications. Il ne s'agit pas seulement d'empiler des heures de transcription mais de réfléchir à la façon dont les interrogations, et pas seulement celles des linguistes, seront adressées. Le principe de construction du site sera fondé sur la notion d'usage, en conciliant, dans une réflexion sur l'interface graphique et les parcours possibles, l'hétérogénéité des sollicitations :

- celle des chercheurs en science du langage à la recherche d'exemples, d'occurrences ou de statistiques (quelle proportion de propositions négatives, quel pourcentage de métaphores, quelle corrélation entre telle catégorie de noms et telle catégorie d'adjectifs (« tous les noms d'animaux » ou « toutes les couleurs » suppose la possibilité de classements hyperotaxiques), combien de hiatus par tranche de mille voyelles réalisées.. ?),
- celle des chercheurs en sciences humaines qui, pour des interrogations sur l'interaction, les groupes sociaux, les types de communication ou les représentations mentales auront recours soit à des indices lexicaux, soit à une requête sur les méta-données,
- celle des étudiants qui pourront trouver, grâce à la dimension réflexive du travail, des moyens d'autoformation,
- celle des décideurs, en particulier en matière de politique linguistique,
- celle des développeurs dans les industries de la langue et les applications informatiques et didactiques, de FLE comme de FLM.

Ce site, s'il est prioritairement destiné à la communauté des chercheurs de langue française, est aussi voué à servir de vitrine internationale. La question d'une traduction des présentations et des opérations est d'ores et déjà posée, d'abord en anglais mais aussi en allemand et en espagnol pour intégrer les compétences et les collaborations déjà présentes au sein des équipes. Il ne s'agit pas seulement de traduire des mots mais d'articuler à l'expérience et aux acquis de chaque communauté scientifique les résultats propres produits par ESLO dans les paramètres particuliers de la situation du français en France.

Phase 11 : Données partagées : interopérabilité et protections

Parallèlement à ce que peut assurer une mise à disposition la plus ouverte possible (avec les restrictions qui s'imposent concernant la protection des personnes et des fonds), ESLO 1 et ESLO 2 doivent pouvoir être inter-opérables avec les corpus linguistiques élaborés par les chercheurs. Une première étape de l'interopérabilité sera réalisée dans la confrontation avec des corpus présents dans les laboratoires participants : corpus palikur, corpus de créoles, corpus de français d'autres régions. Ces rapprochements auront valeur de test pour une extension à l'ensemble des données sonores recueillies à des fins d'analyse linguistique. Ils représentent d'ores et déjà cinq états de différenciation :

1. corpus de français issus de la même région et recueillis dans les mêmes conditions à quarante années de distance (ESLO 1 et 2),
2. corpus de français issu d'une autre région et recueilli avec des méthodes et des perspectives variationnistes différentes (Corpus Lormont),
3. corpus de français écrit (journaux),
4. corpus de langue orale à base romane avec variation interne forte (créoles de Guinée Bissau, du Cap-Vert et de Sao Tome et Principe),
5. corpus de langue orale sans apparemment avec le français sinon par contact (palikur, wayana, wayampi).

Les questions qui se poseront dans l'interopérabilité de ces corpus auront valeur d'exemple pour les autres corpus qui pourront y être associés ultérieurement.

Phase 12 : Applications et développements

Les études sur corpus concernent directement le Traitement Automatique des Langues Naturelles (TALN). Comme l'ont montré par exemple (sur le français) les travaux du LIUM (Le Mans), du DELIC (Aix-en-Provence), du CLIPS-IMAG (Grenoble), du LORIA (Nancy) ou du LI (Tours-

Blois), la recherche d'une description linguistique de l'oral sur des corpus pilotes permet d'orienter les recherches et le prototypage des systèmes de traitement automatique du langage parlé (compréhension de parole, dialogue oral homme-machine, traduction parole-parole etc.). C'est cette perspective, à l'interface entre sciences du langage et TALN, que nous comptons en profitant des caractéristiques des corpus ESLO.

Dans le cadre de ce projet, nous allons plus précisément nous intéresser à la description de deux phénomènes langagiers en situation d'élocution oral spontanée :

- disfluences de l'oral, hésitations, répétitions et auto-corrections,
- co-référence anaphorique en situation de parole spontanée.

Le traitement de ces procédés est essentiel dans le développement d'applications orales interactives.

Les disfluences de l'oral spontané sont susceptibles de perturber fortement la structure syntaxique des énoncés oraux, ce qui rend difficile leur analyse automatique. Il est donc important de les décrire linguistiquement dans l'objectif d'atteindre une modélisation suffisamment robuste pour être implémentée sur des systèmes opérationnels. Des recherches linguistiques sont menées dans cette optique depuis maintenant une vingtaine d'années. En terme d'application, celles-ci se réduisent cependant le plus souvent au dialogue oral finalisé sur des tâches bien précises (réservation aérienne par exemple) et la communauté scientifique ne dispose pas de connaissances précises sur le degré de généralité de ces résultats. Il semble que le contexte d'interaction (degré d'interactivité, familiarité des interlocuteurs, degré de spécialisation du dialogue) influe d'une manière sensible sur la distribution des procédés utilisés. Les corpus ESLO offrent des garanties en terme de significativité des données et surtout de contrôle des variables pouvant influencer sur ces procédés. A partir de ces nouvelles ressources, on se propose de répondre à ces deux questions : (1) quelles sont les variables sociolinguistiques, notamment les différentes caractéristiques du contexte d'interaction, qui influent sur le taux et le mode de réalisation des disfluences orales, (2) les résultats obtenus sur de la parole conversationnelle peuvent-ils être étendus à d'autres genres / registres de langue, au sens de Biber

La coréférence anaphorique est un autre problème qui touche aussi bien à l'écrit qu'à l'oral et le calcul de la référence est ainsi un problème de première importance pour le dialogue oral homme-machine. Dans le cas d'un serveur vocal d'information, il est par exemple essentiel d'accéder à l'ensemble des objets de l'application auxquels réfère l'utilisateur dans sa requête avant de pouvoir interroger la base de données. Consacrées avant tout à l'anaphore pronominale, les recherches actuelles sont limitées par leur application restreinte à l'écrit ou au contraire à un dialogue oral très finalisé. L'étude des anaphores pronominales sur le corpus ESLO traitera les questions suivantes : (1) dans quelle mesure les heuristiques de résolution développées pour l'écrit sont-elles applicables en oral spontané ? (2) Dans quelle mesure la théorie du centrage est-elle vérifiée sur l'oral conversationnel ? (3) La pertinence de cette théorie est-elle variable suivant le genre oral considéré ?

Les réponses à ces interrogations seront envisagées dans une perspective quantitative faisant appel à une analyse différentielle / variationniste entre différentes sous-collections du corpus ESLO.

Phase 13 : Mise en place du suivi (maintenance, jouvence et sécurité) et applications

La pérennité du corpus est liée à la capacité des participants de l'équipe à concevoir une maintenance du site, à une intégration progressive de nouveaux enrichissements des données et à l'adaptation aux nouvelles normes, notamment pour ce qui a trait à la sécurité. Dans le calendrier prévu, les applications et les développements, même s'ils font l'objet d'une réflexion en amont, ne pourront guère avoir dépassé le stade de projets. Ils sont néanmoins mentionnés pour ce qu'ils contribuent à leur façon à la définition du corpus, dans les sollicitations qu'ils représentent et dans les attractions qu'ils exercent. On pense en particulier aux applications dans le domaine du TAL (reconnaissance de la parole, interface homme / machine, analyseurs syntaxiques, constitution automatique de dictionnaires, traduction...) et à celles dans le champ de la didactique, les chercheurs du CORAL participant à des formations de Master et de doctorat associées à ces problématiques.

L'ensemble de ce travail s'inspire des apports de la linguistique qu'il doit à son tour enrichir. Le champ d'application va de la phonologie au TAL, en passant par la morpho-syntaxe, la lexicologie, la

sémantique et pragmatique, l'analyse de la conversation et du discours. Produit d'une formation discursive, le discours y sera envisagé comme ensemble de régularités morphosyntaxiques et verbales fonctionnant comme autant d'indices d'une distribution en différents sous-genres. Les formats d'échange et d'argumentation, les expressions figées et les emplois allusifs (polyphoniques) seront également intégrés dans une contribution orientée corpus à la théorie du langage.

Conclusion

Ce projet manifeste une ambition forte. Il était en discussion entre les partenaires depuis que ESLO 2 a été relancé et l'ANR peut apporter des moyens pour sa réalisation. Constituer un corpus prototype à toutes les étapes est une tâche ambitieuse qui prend sa véritable dimension dans le défi qu'elle relève : établir le plus grand corpus oral francophone diffusé, comparable, à l'échelle du français, aux états premiers du BNC et appelé à devenir, comme celui-ci, un élément du patrimoine historique.

Il s'agit d'apporter, à une base de corpus oraux, des matériaux dont la collecte et le traitement pourront soutenir la comparaison avec les exigences apportées par l'INaLF aux corpus écrits. Dans un délai de trois ans, et sous réserve de disponibilité des financements, VARILING doit (1) accumuler des heures de parole et développer une expérience critique sur leur mode de collecte et d'analyse, (2) créer des matériaux exploitables par les chercheurs et structurer la communauté « corpus », (3) ouvrir des champs d'investigation à l'ensemble des sciences humaines et apporter les attestations requises par la lexicographie ou l'élaboration de grammaires de référence. Le projet est aussi destiné à inaugurer une nouvelle politique de conservation du document scientifique sonore non musical et à préparer des collaborations dans les applications, didactiques et informatiques.

B-3 – Bibliographie et état de la question

La description des langues, dans la tradition linguistique, a été effectuée à partir de la langue écrite, normalisée, avec une valorisation des textes littéraires. Le premier changement apporté par les technologies informatique a été la constitution de grands corpus, le plus souvent étayés par de l'imprimé (cf. en France l'importance accordée au journal *Le Monde* dans les analyses et les corpus d'exemples). Lorsqu'il a paru nécessaire d'étudier la langue parlée, ce sont généralement des corpus de radio et de télévision qui ont été exploités, en dépit des contraintes de format et de sélection des locuteurs inhérentes à ces médias.

La constitution de grands corpus de français parlé non médiatisé s'est heurtée à quatre difficultés :

1. un problème de représentativité (en lien avec une sociologie des locuteurs et des situations de communication),
2. des interrogations concernant la collecte (difficultés d'enregistrement en milieu bruyé, statut juridique des documents, protection de la vie privée...),
3. des solutions à inventer pour le traitement de la variation qui, d'un paramètre, devenait un obstacle à la représentation, au classement et à l'exploitation des données,
4. des incertitudes concernant la recherche sur les corpus obtenus.

Certains pays ont pris une avance en ce domaine et si l'exemple le plus souvent cité est celui du BNC (British National Corpus), il existe d'autres travaux d'une qualité équivalente en Allemagne, Espagne (Catalogne), Pays-Bas... En français, il y a :

- soit des corpus partiels (Aix-en-Provence, Lyon...),
- soit des corpus thématiques (PFC),
- soit des corpus variationnistes extra-métropolitains (Québec).

Il s'agit de conjoindre des expériences et des approches qui existent ailleurs, ou autrement ou encore sur d'autres langues à la situation du français, en partant d'une collecte dans une région peu dialectalisée et moins soumise à la pression internationale des langues que la région parisienne.

Le travail sur l'analyse de corpus en français parlé dans une triple dimension de variation (sociale, diachronique et dialectale) interagit avec l'élaboration d'outils et d'instruments du TAL. Il ne semble pas qu'un programme comparable existe, en France, que ce soit à des fins d'expertise ou d'application. La perspective qui se dessine est une coordination de la recherche à l'échelle des corpus de français en Europe (variantes régionales, Belgique, Suisse...) et dans le monde (en particulier, comparaison avec le Québec) et aussi une confrontation et un cumul des expériences et des entreprises engagées en Europe (avec une continuation possible sur des crédits européens).

ESLO

Formulaire témoin

Les informations nominatives ci-dessous ne sont pas publiques, elles seront stockées dans une base à part.

Nom :

Prénom :

Nom de jeune fille :
.....

Adresse :
.....

Année de naissance :
.....

Sexe : M – F **Lieu de naissance** (ville + région/département) :
.....

NIVEAU D'ETUDES

- Primaire
- Certificat d'études
- Secondaire (collège, middle school)
 - Complet Incomplet
- Lycée
 - Complet Incomplet
- Supérieur
 - +1 +2 +3 +4 +5+

Commentaires

.....
.....

PROFESSION

1- Statuts :

- Professionnel actif
- Retraité
- Chômeur
- Autre

2- Catégorie professionnelle (Code INSEE 1) :

- agriculteur exploitant
- artisan, commerçant ou chef d'entreprise
- cadre ou profession intellectuelle supérieure
- profession intermédiaire
- employé
- ouvrier

- retraité
- autre personne sans activité professionnelle
-

3- Spécifiez la profession en vos propres termes :

.....

LANGUE(S)

Langue(s) parlée(s) [Langue(s) parlée(s) (passive(s) ou active(s)) autre(s) que le français] :

.....

Commentaire(s), niveau de langue(s) :

.....

SITUATION FAMILIALE

- Marié
- Célibataire
- Concubinage (marital, pacs, etc.)
- Autre (veuf(ve), divorcé(e), en couple sans vivre sous le même toit).

Précisez.....

Année d'arrivée à Orléans (si depuis la naissance, indiquez l'année de naissance) :

.....

Domicile actuel (Ville + quartier) :

.....

ENFANT(S)

Nombre :

Sexe, année de naissance, scolarisation et autres commentaires :

.....

REMARQUE(S) DIVERSE(S) :

.....

LCO / PANGLOSS

Programme d'Analyse Géographique et Linguistique à Orléans : Sciences et Société

Objectifs et finalités

L'objectif est de construire, sous forme d'une monographie associant géographes et linguistes, une enquête au centième de la grande enquête INSEE / INED sur les langues parlées en France. La taille de l'échantillon et l'articulation au travail d'un laboratoire de linguistique permettront d'assurer une collecte et surtout un traitement qui affineront et approfondiront les résultats globaux déjà obtenus à l'échelle nationale. Le travail, qui comprendra une phase d'enquête et de transcription, une phase de constitution de bases de données et de cartographie et une phase d'analyses et de prestations, a pour objectifs, au-delà d'une connaissance réflexive des flux de population,

- de recenser une diversité patrimoniale,
- de raisonner l'apport culturel et linguistique des populations allophones,
- de mesurer la dynamique sociale des langues (transmission, intermariage, déplacements intra- et internationaux...)
- de reconstruire les réseaux d'échange et de production pour lesquels la langue constitue un facteur explicatif.

Programme de travail

- Préparation de l'enquête (constitution de questionnaire, canevas d'entretien, etc.)
- Enquête et identification des langues (recueil de textes en différentes langues et transcriptions)
- Analyse des relations et représentation cartographique détaillée de leur présence et de leurs relations à travers les réseaux économiques, culturels et sociaux.
- Constitution et organisation des données (parallèlement à l'enrichissement de la Phonothèque prévue dans le programme CRITERR), construction de corrélations (incluant les représentations cartographiques)
- Analyse théorique sur les questions du traitement de l'information concernant des propriétés attachées à des personnes plus qu'à des territoires et sur les problèmes linguistiques posés par les pratiques des agents et la transformation des systèmes (phonologiques et morpho-syntaxiques).
- Phase d'explication et d'application (enseignement, formation en langue des primo-arrivants, aide à la gestion et à la décision des partenaires sociaux et des administrations).

Retombées scientifiques, techniques, socio-économiques et environnementales

Ce projet concerne l'accueil et l'intégration des publics migrants et des populations non sédentaires. Il devrait permettre d'améliorer leur insertion, leur accompagnement, leur formation en fournissant de nouvelles bases à la formation des formateurs du secteur public ou associatif, notamment pour tout ce qui touche à la question linguistique. Il permettra aussi, en fournissant des données fiables et représentatives, d'apporter une aide à la décision politique et à l'évaluation des actions publiques. Du point de vue scientifique, il permettra de nouvelles approches du contact des langues, de ses conséquences sur chacune des langues en présence (émergence de nouvelles variétés : phénomènes de pidginisation, créolisation...) ainsi que sur leur acquisition et transmission. Il permettra de poser le problème de la représentation cartographique de propriétés non territoriales (modélisation des

transferts, banque de données...) et, croisant le point de vue du linguiste et celui du géographe, tout en prenant en compte les pratiques réelles et les représentations des personnes, de reconsidérer des notions comme celles de *territoire* ou de *communauté linguistique*. Projet pionnier, il est destiné à servir de prototype à des enquêtes similaires conduites dans d'autres régions de France ou d'Europe.

III-1- Objectifs et contexte

Parce que chaque fois qu'un individu transmet, apprend ou emploie une langue « il accroît l'utilité de cette langue » (*loyauté linguistique*) aux dépens des autres langues, la mondialisation économique restreint tendanciellement l'offre linguistique au bénéfice de l'anglais. Nombreux sont ceux, linguistes, pédagogues, mais aussi politiques, qui s'interrogent sur la disparition des langues, dont le rythme tend à s'accélérer, ce qui représente la perte d'un patrimoine non reconstituable. C'est, entre autres, dans cette perspective que le Conseil de l'Europe a impulsé un ensemble d'initiatives pour le développement d'une véritable éducation plurilingue et pluriculturelle. Ces actions ont été accompagnées de la commande de matériaux pédagogiques et didactiques (Cadre Européen de Référence des Langues, portfolio européen, etc.). Si ces dispositions vont dans le sens de la construction d'identités plurielles et sont appréciées pour le rôle positif qu'elles ont dans les prescriptions scolaires et universitaires, on constate parallèlement qu'elles profitent essentiellement aux langues officielles des états-nations.

Dans le même temps, les flux de population à l'échelle mondiale et la grande diversité des types de migrants (retraités des pays riches de l'Europe du nord vers l'Europe méridionale, personnels qualifiés et chercheurs en entreprises et grands organismes, réfugiés politiques, gens du voyage, émigrés économiques, etc.) et la variété des pays d'origine et de destination ont pour conséquences de nouveaux contacts humains, l'élargissement du champ de diffusion de certaines langues (langues asiatiques ou africaines parlées en Europe) et de nouveaux contacts linguistiques, impensables il y a deux ou trois décennies, du moins à cette échelle.

De même que l'expansion européenne au 16^e et au 17^e siècle a eu pour conséquence la diffusion des langues d'Europe et l'apparition de nouvelles langues – les créoles –, avec les flux migratoires contemporains apparaissent de nouvelles formes de plurilinguisme qui ont pour conséquence l'émergence de nouvelles variétés linguistiques. Si certaines comme le spanglish des Hispano-Américains ou le chiac (mixte de français et d'anglais, cf. travaux de M.-E. Perrot) parlé au Nouveau-Brunswick (Canada) ont pu donner lieu à des recherches et même être valorisées (publications de livres, émission de radio, presse...), le plus souvent ces embryons de langues nouvelles sont soit totalement méconnus, soit stigmatisés (stigmatisation qui reflète souvent celle qui frappe les locuteurs).

La France, et la région Centre en particulier, est concernée au premier chef par cette problématique (<http://www.oecd.org/home/>). L'arrivée de migrants et l'affaiblissement des variétés dialectales du français (liée à l'allongement des temps de scolarisation et à la rurbanisation) a bouleversé le paysage sociolinguistique au point que certains chiffres, lus sans précaution, pourraient laisser croire que sur le territoire métropolitain, l'arabe a pris la place de l'occitan dans le Midi ou le chinois celle du breton en région parisienne. Dans le même temps, les acteurs sociaux engagés sur le terrain témoignent de l'apparition de formes du français qui leur sont difficilement compréhensibles.

Si la réflexion sur ces thèmes est aujourd'hui avancée aux Etats Unis (hispano-américains) et au Québec où existe une tradition d'enquête langue par langue (cf.

<http://www12.statcan.ca/francais/census01/Products/Analytic/>), en France la prise de conscience est récente et la première initiative publique correspond à l'enquête INED, (cf. www.ined.fr/en/resources_documentation/publications/) initiative relayée par la Délégation Générale à la Langue Française et aux Langues de France (DGLFLF – Ministère de la Culture et de la Communication) et l'ANR (Agence Nationale de la Recherche) qui soutiennent le LLL dans son programme « Langues en Contact à Orléans », l'un des modules associés au projet ESLO2 (Enquête Socio-Linguistique à Orléans n° 2).

Dans la continuité de ces actions, PANGLOSS (**P**rogramme d'**A**nalyse **G**éographique et **L**inguistique à **O**rléans : **S**ciences et **S**ociété) se fixe comme objectifs de répertorier les pratiques linguistiques et culturelles sur l'agglomération d'Orléans, d'analyser les modes d'acquisition, de transmission et de diffusion des différentes langues, de décrire certaines des variétés émergentes (par exemple le peul parlé à Orléans, le français des Voyageurs etc.), de mettre à disposition des échantillons sonores des différentes langues en contact dans un état de préservation (numérisation des données et métadonnées, copies de sauvegarde, double dépôt avec la Phonothèque Nationale, déclaration des formats d'archive, initiative OLAC et utilisation de logiciels Open Source), d'utilisation (autorisation des ayants droit, déclaration CNIL) et d'identification linguistique et sociologique correspondant à l'état le plus récent des connaissances et des exigences en la matière. Au-delà des pratiques en usage en linguistique, l'un des objectifs les plus novateurs de PANGLOSS est l'association avec un laboratoire de géographie reconnu dans ce domaine et qui a déjà conduit une réflexion et abouti à des réalisations en matière de cartographie des pratiques sociales et culturelles.

La cartographie des pratiques linguistiques et culturelles associées à l'usage d'une langue est un défi scientifique du plus haut intérêt dans la compréhension de l'inscription territoriale d'une société. La langue est l'un des éléments essentiels pour identifier une communauté, à la fois dans la reconnaissance interne entre les locuteurs de la langue et dans leur relation aux locuteurs des autres langues. Son inscription dans les territoires permet de mesurer le degré de fragmentation de l'espace social, qu'il importe de saisir dans toute sa complexité pour qu'une politique de la ville aille au-delà des actions sur le foncier et invente des propositions pour obtenir un véritable effet de levier sur le corps social dans le dialogue entre les cultures et la création d'un lien collectif qui transcende les cultures tout en les préservant.

En effet, si la langue crée un espace d'intercommunication qui dessine les échanges à l'échelle d'un quartier, d'une ville, d'un pays, elle peut être aussi un facteur d'isolement social et devenir une source de précarité : difficultés scolaires, difficultés d'insertion professionnelle, problèmes de santé publique, etc. On s'intéresse particulièrement à la concentration relative des populations allophones dans des « enclaves ethniques » qu'il convient de rapprocher d'autres éléments comme la présence de réseaux de chalandise constitués par solidarité culturelle et d'indicateurs de fragmentation sociale du territoire urbain tels que le chômage pour montrer le lien qui peut exister entre telle ou telle communauté et les fractures urbaines. L'intérêt de l'analyse est de contrer les réactions de stigmatisation en ciblant de façon optimale les moyens d'une stratégie destinée à favoriser l'insertion scolaire des enfants, l'insertion professionnelle des adultes ou des actions de promotion de la santé publique.

La représentation cartographique des résultats d'enquête soulève d'intéressantes questions relatives à la distinction entre les espaces de résidence et les quartiers fréquentés quotidiennement. Elle interroge, dans une société de croisement des données, les possibilités de préserver la confidentialité des renseignements individuels, la représentativité de l'échantillon ou encore la mesure du degré de ségrégation spatiale entre différents groupes. Pour répondre à ces questions, l'équipe du CEDETE entend mettre en œuvre ses compétences au travers de :

1. la constitution d'un SIG urbain permettant d'intégrer les résultats de l'enquête aux échelles pertinentes et de les rapprocher d'indicateurs du contexte démographique et social des quartiers, ainsi que d'autres éléments tels que la localisation de certains types de commerce (qualifiés de communautaires) etc.
2. la représentation cartographique des résultats de l'enquête mettant en œuvre des procédés de traitement innovants à l'aide des méthodes de la statistique spatiale comme l'analyse des proximités et des associations et le lissage cartographique qui permettront de traiter les problèmes de représentativité de l'échantillon et de confidentialité des données individuelles.
3. l'analyse des résultats de l'enquête en termes de ségrégation à l'aide des méthodologies (Massey et Denton,1988).

Les résultats attendus doivent constituer un outil d'aide à la décision pour la politique de la ville. Ils constitueront pour les opérateurs, comme l'ACSE ou l'Agence Nationale pour la Rénovation Urbaine, avec lesquels des contacts sont d'ores et déjà établis, des instruments précieux dans l'accomplissement de leurs missions. Ils apporteront de nouvelles perspectives aux différentes formations de la Région (IUFM en formation initiale des maîtres, formation continue et formation de formateurs, différents Masters des domaines linguistique, Sciences humaines, diplôme universitaire « Apprentissage, langues et interculturalité », formation de travailleurs sociaux), ainsi qu'aux associations engagées dans la formation des migrants. C'est par là qu'elles pourront avoir un impact direct ou indirect sur le marché de l'emploi, dans la qualification des opérateurs notamment.

III-2- Description du projet et résultats scientifiques et techniques attendus

PANGLOSS est un projet particulièrement novateur dans son approche des questions linguistiques liées à l'immigration. En effet, si depuis quelques années, l'idée commence à faire son chemin que la France aussi est une terre de multilinguisme, aucune étude à ce jour n'a proposé d'aborder la question en liant à la fois la description externe (géographie linguistique, sociologie du langage) et description interne (analyse des structures phonologiques, morpho-syntaxiques, sémantiques des langues en contacts et des nouvelles variétés linguistiques). Cette approche permettra de surmonter deux limites inhérentes aux projets de recherche centrés sur cette thématique : un commentaire à visée sociologique qui recense l'utilisation des langues en présence en restant circonscrite sur le plan externe des représentations ; une description linguistique qui par manque d'explication du contexte se révèle difficilement exploitable. Elle vise, par l'éclairage de la géographie, à passer d'une présentation statique des données à une figuration dynamique.

Le poids des représentations personnelles et/ou collectives sur toutes les questions qui touchent aux langues et à leur utilisation est tel que toute donnée doit être remise en perspective du point de vue de ces représentations. Celles-ci apparaîtront en particulier au travers de la comparaison entre le « déclaratif » et les pratiques réelles, ce qui implique que les enquêtes seront menées selon trois points de vue :

Le « macro déclaratif » : Enquête à base de questionnaires distribués massivement et sélectivement, recueil d'indications quantitatives et qualitatives auprès d'associations et ou institutions, étant admis que l'absence de données ou l'impossibilité d'en fournir représente *per se* une donnée. On soulignera que les données collectées ne peuvent être interprétées que si elles sont lues à la lumière des résultats obtenus grâce à l'approche « micro déclarative » et à une observation construite des pratiques réelles.

Le « micro déclaratif » : entretiens semi-guidés avec les membres des différentes « communautés linguistiques », en les positionnant à des degrés divers de compétence (médiateurs de la communauté, personnalités du groupe, néo-immigrés, couples « mixtes »...). Au cours de ces entretiens les thèmes abordés tourneront autour de l'utilisation des langues en contact dans les différents types de situation, y compris les accès médiatiques (chaînes par satellite).

Les pratiques réelles : Analyse interne des entretiens semi-directifs ; recueil et analyse de productions orales diverses (conversations, textes...) dans et à propos des différentes langues (y compris le français), avec une part d'observation dans des situations non contrôlées (grandes surfaces, marché, sortie d'école...).

Le réinvestissement, au cours des enquêtes et de l'analyse des données, des acquis méthodologiques et scientifiques de travaux menés par différents membres de l'équipe sur d'autres terrains est un autre défi du projet. Plusieurs chercheurs du LLL sont les auteurs de travaux sur des situations de contacts de langues dans d'autres pays : Afrique lusophone (créoles de Guinée Bissau, de Casamance, de Sao Tomé et Príncipe et du Cap-Vert, variété populaire du portugais en Afrique et au Brésil); Guyane (langues amérindiennes, créole français), Burkina-Faso (langues africaines et français), Canada (chiac du Nouveau-Brunswick). Confronter les résultats de ces travaux avec ceux obtenus lors des enquêtes à Orléans permettra de dépasser l'opposition entre l'universel et le particulier. Les résultats attendus concerneront donc non seulement la situation orléanaise mais seront aussi des éléments pour une analyse à plusieurs niveaux (local, régional, national, international) des phénomènes de différentes natures liés aux contacts de langues : acquisition, transmission familiale (*linguistic loyalty*), diffusion, scripturisation, emprunts lexicaux, alternance codique (code switching), véhicularisation/vernacularisation, pidginisation, créolisation...

Un des grands défis relevé par PANGLOSS est aussi de croiser sur la question des langues le regard du linguiste et celui du géographe. Ainsi sera posé le problème d'une représentation cartographique qui s'attachera surtout aux personnes, à leur trajectoire et à leurs pratiques (modélisation des transferts, banque de données...). S'attachant aux pratiques réelles et aux représentations des personnes, PANGLOSS œuvrera pour la déconstruction de notions comme celles de territoire, de communauté linguistique, ou même d'identité. La principale ambition réside dans le caractère pionnier d'une recherche qui puisse servir de prototype à des enquêtes similaires dans d'autres régions de France ou d'Europe.

Ce projet est une première réalisation correspondant aux orientations proposées par un groupe de laboratoires de l'Université d'Orléans dans la mise en place du Centre de Ressources et d'Innovation sur les TERRitoires en Région Centre, qui constitue le pôle orléanais de l'axe de recherche régional CREVIT « Villes, Territoires, Civilisations » représenté à Tours par le CREVILLES. Il constitue le premier élément d'un rapprochement entre le laboratoire de linguistique LLL et le laboratoire de géographie CEDETE.

Les recherches du LLL incluent une visée anthropologique qui figure en toutes lettres dans le sigle du laboratoire. Cette approche s'est concrétisée dans l'étude des transformations de langues en situation de transfert et de mélange de populations (créoles à base portugaise d'Afrique), dans la relation des langues aux cultures dont elles sont inséparablement solidaires (Encyclopédie des langues de Guyane : palikur, wayana,

wayampi) mais aussi dans la différenciation du français hors de France (Burkina) et en France.

Reprenant une enquête conduite en 1968-1971 par une équipe d'universitaires anglais afin d'obtenir un « portrait sonore » de la ville, le LLL a conçu une enquête comparative dont il réunira les premiers matériaux au cours du premier semestre 2008. A la recherche entreprise il y a quarante ans sous l'intitulé ESLO (Enquête Socio-Linguistique à Orléans, désormais cité comme ESLO1), les linguistes de l'Université d'Orléans se proposent d'apporter le renouvellement d'une enquête du même type, dénommée ESLO2, qui apprécierait les variations et les changements survenus dans la dynamique du changement linguistique. PANGLOSS figure dans ce cadre comme l'un des modules complétant les entretiens sur échantillon représentatif qui constitueront ESLO2, celui qui intégrera à la dynamique du français parlé les transformations et les enrichissements induits par le contact de langues.

Parallèlement, les travaux conduits par le CEDETE en matière de systèmes d'information géographique et de représentation cartographique, en particulier dans la résolution des problèmes que posent des propriétés des agents qui, pour être présentes sur certaines portions de l'espace, n'y sont pas intrinsèquement liées, a fait de la question de la représentation des langues un premier terrain d'expérimentation pour une collaboration qui ira en s'approfondissant. La dimension d'information, du public et des décideurs, et l'aide à la gestion raisonnée des territoires et des politiques urbaines, constituent l'une des finalités de ces programmes.

Quant au troisième laboratoire fondateur du CRITERR, le LEO, il s'associera lors de l'extension d'échelle de cette recherche, en particulier dans la relation entre langue et travail, dans une collaboration qui impliquera le CEREQ et l'équipe TOTEM (Territoire, Organisation, Travail Et Mondialisation) du LEO, laquelle est destinée à réunir les chercheurs en "économie internationale" "économie du travail" et "économie industrielle".

C'est pourquoi il est prévu de loger dans les infrastructures immobilières du CRITERR l'équipe qui se consacrera particulièrement au PANGLOSS, d'abord pour conforter la synergie des équipes, ensuite pour assurer les transferts de savoir et de savoir-faire, enfin pour assurer les relations avec les projets logés dans le Centre au titre du LLL (COSLA régional, Phonothèque, Observatoire des Langues, Encyclopédie des langues de Guyane...), du CEDETE et du LEO.

A l'échelle de la région, PANGLOSS est conçu comme une première proposition destiné à tester la faisabilité d'un projet qui étendrait à terme son périmètre à l'échelle de la région tout entière et la pertinence de collaborations qui sont en construction, l'accord rapide entre les parties sur des objectifs et des méthodes étant du meilleur augure.

Le rapprochement en cours entre une équipe de linguistique de Tours et le LLL participe de cet effort de constitution d'un pôle de référence à l'échelle européenne en région Centre, au même titre que la réunion, en une fédération de laboratoires (FELINCO) d'équipes linguistiques de Poitiers (FORELL), de Tours (Langage et Représentation), de Limoges (CERES) afin de développer des synergies en matière de recherche, de formation (création d'un Master inter-régional) et de transfert de technologies. Cette collaboration régionale est destinée à s'ouvrir aussi au laboratoire de sociologie de Tours qui a fait de l'immigration l'un de ses thèmes de recherche privilégiés.

En termes de développement du potentiel de recherche de la Région Centre, il s'agit de faire émerger, à côté du pôle cosmétique et de grands laboratoires déjà reconnus (sciences de la terre et énergétique à Orléans), des compétences en sciences humaines à même de faire apparaître et de soutenir une activité économique dans des domaines émergents : traitement de l'information et systèmes d'information géographiques, industries de la langue, génie linguistique, traitement automatique des langues. L'extrême faiblesse en particulier des industries de la langue en Région Centre (à la différence, en dehors de Paris, de ce qui s'est passé à Grenoble, Nancy, Aix ou Toulouse) suppose une stratégie volontariste de renforcement d'équipes positionnées dans ce domaine et capables d'attirer, promouvoir ou développer une activité économique en ce domaine.

Dans l'immédiat, le défi auquel répond PANGLOSS correspond à la nécessité de rassembler des compétences afin d'assurer la maîtrise scientifique d'objets d'études qui ont des implications, au-delà des langues, sur l'occupation des territoires, les organisations sociales (cf. l'histoire des migrations) et qui requièrent des méthodes particulières de traitement (banques de données) et d'analyse (cartographie). Une proposition de travail avec l'ANRU (Agence Nationale de Rénovation Urbaine) est en cours d'instruction dans leurs services et c'est dans une telle connaissance des situations et des personnes que se révèle le caractère crucial des études de terrain, dans le passage de la description à l'explication et de l'explication à la prescription par les pouvoirs publics et les administrations.

GUYANE

Encyclopédie de trois peuples amérindiens

Conseil scientifique

Serge Bahuchet, Professeur d'ethnobiologie au Muséum National d'Histoire Naturelle, Paris

Gabriel Bergounioux, Professeur de linguistique à l'Université d'Orléans

Gérard Collomb, anthropologue, chargé de recherche au CNRS, Paris

François Nemo, Professeur de linguistique à l'Université d'Orléans

Dominique Tilkin-Gallois, Professeur d'anthropologie à l'Université de São Paulo, Brésil.

Collection « ENCYCLOPÉDIES »

Soucieux d'articuler travail linguistique, prise en compte des dimensions sociale et historique ainsi que restitution d'une culture dans son intégralité, des spécialistes français et étrangers ont entrepris la réalisation de trois encyclopédies dédiées à trois peuples amérindiens de Guyane, les Palikur, les Wayana, les Wayâpi.

L'originalité de notre projet se définit par son ambition, présenter l'ensemble des savoirs existants chez ces trois peuples et à leur sujet. Pour réaliser cette ambition, nous avons choisi une double voie : d'abord, nos encyclopédies prendront forme dans un ensemble de fascicules thématiques paraissant progressivement et couvrant tous les aspects de la culture, des grands témoignages du passé aux synthèses thématiques d'aujourd'hui ; puis, au terme de chacune des trois séries, les données publiées seront collationnées dans deux ouvrages de synthèse, un dictionnaire culturel et un dictionnaire de langue.

À l'heure où le savoir est littéralement émietté et ce faisant, très largement inaccessible, ce format est une innovation qui nous semble aujourd'hui devenue indispensable : les savoirs n'existent que s'ils ne restent ni méconnus ni isolés, et ils ne vivent véritablement que lorsqu'il est possible à tous de se les approprier. Cette forme de restitution répond à une forte demande sociale. En effet, une longue pratique du terrain et des hommes nous a conduit à la certitude que, conscientes des dangers de l'intrusion, souvent brutale, de la modernité dans leur genre de vie, les sociétés ayant fondé leur mémoire ainsi que la transmission de leurs savoirs et savoir-faire sur une culture orale, croient désormais aux vertus de la chose écrite.

Chacune de nos trois encyclopédies a pour objet, plus que la description de mots, le portrait du monde par un peuple. Les catégories du vivant, la conception du surnaturel, l'ordonnement de l'univers, le déroulement des grandes étapes de la vie, les arts et les techniques, sont autant de macro-thèmes dont l'approche se fait de l'intérieur, avec pour fil d'Ariane la langue, dans son fonctionnement comme dans ses réalisations. Les ouvrages que nous avons entrepris ne sont pas unilingues. Ils allient la langue vernaculaire en tant que langue source, le français en tant que langue cible et métalangage, les langues de contact en tant que langues de traduction spécialisée.

Nous avons choisi de constituer une collection de publications autonomes à visée globalisante, afin que, sans dépendre les uns des autres, les auteurs publient leurs études à leur rythme, au fur et à mesure de l'achèvement de leurs travaux. Chapeautant notre entreprise, un fascicule zéro intitulé *Langue, milieu et histoire*, offre une présentation de l'ethnie pour chacune des trois encyclopédies, *palikur*, *wayâpi*, *wayana*, et énonce la philosophie de notre démarche.

Françoise Grenand, directrice de la collection

1. PLAN D'ENSEMBLE

Même lorsqu'il leur arrive d'être menées de conserve par certains des auteurs (et bien sûr des éditeurs scientifiques), les trois encyclopédies demeurent des aventures qui se doivent, par respect envers les traits culturels propres aux trois populations, de garder une certaine individualité. Chacune des trois étant profondément ancrée sur la langue, le milieu et l'histoire, l'agencement des connaissances qui fondent la culture revêt une importance capitale.

1.1. Grands domaines des fascicules

Les fascicules se développent autour de plusieurs thèmes majeurs. Il est du ressort des éditeurs scientifiques de veiller à ce que la collection, une fois terminée, les aient fait apparaître sans en omettre. Afin de ne laisser aucune donnée sur le bord du sentier, il a été décidé qu'elle pourrait être amendée ; la décision du découpage et de l'organisation restant fonction de la compétence des chercheurs, la liste qui suit demeure purement indicative.

Domaines majeurs	Thèmes	
• histoire	généalogies histoire des communautés grands témoignages	guerre et paix biographies peuples voisins
• géographie	toponymes spatialisation orientation	topographie astronomie cosmologie
• sciences naturelles	agriculture et tenure foncière faune flore sols	classification indigène couleurs
• technologie culturelle	art et artisanat techniques art musique	jeux parures architecture et habitat aspects de la modernité
• vie sociale	personne entre corps et âmes maladie et médecine nomenclature de parenté étapes de la vie organisation sociale droit coutumier et morale chefferie	chamanisme cuisine et art culinaire division sexuelle du travail aspects de la modernité : école salarial religions importées
• littérature orale	danses chantées chants poésie	mythes récits proverbes et dictons
• langue	questions de phonologie questions de grammaire questions de sémantique questions de pragmatique usages sociaux du langage	expression de la modernité langues reliques langues secrètes différentes écritures représentation de la langue
• dictionnaires	dictionnaire fondamental bilingue	Vaste dictionnaire encyclopédique final

• littérature enfantine	faune flore	techniques contes...
-------------------------	-------------------	---------------------------

1.2. Cohérence de l'encyclopédie

Chaque encyclopédie trouve sa cohérence dans le déroulement global des grandes sections suivantes en deux étapes :

- 1• la présentation générale,
 - les études,
 - la bibliothèque,
 - des publications pour la jeunesse,
 - un dictionnaire fondamental,
- 2• le grand dictionnaire de langue,
 - les index.

• *La présentation générale*

Chacune des trois encyclopédies s'ouvre par une présentation générale, qui en constitue le fascicule zéro. Elle propose la philosophie et les orientations de notre travail, le plan général de l'encyclopédie, la liste ouverte des collaborateurs potentiels et, de fait, le précis des conseils aux auteurs. Elle contient aussi une présentation de l'ethnie concernée, l'état de l'art, ainsi qu'une bibliographie la plus exhaustive possible.

• *Les études*

Synthèses nouvelles spécialement rédigées pour l'encyclopédie, leur nombre n'a pas été arrêté, mais au contraire volontairement laissé ouvert, le but étant de couvrir les domaines majeurs précédemment définis. Les fascicules peuvent être différents dans leur plan, leur ordre de parution et leur nombre final, mais ils se conforment à l'esprit donné par l'éditeur scientifique, après agrément du conseil éditorial. En nombre variable selon les ethnies, ils constituent le cœur de l'encyclopédie. Chacun est une étude scientifique classique, avec de nombreux exemples numérotés, tentant, à chaque fois, la synthèse des connaissances dans le domaine traité. Il comporte des dessins, des cartes et des photographies chaque fois que le thème les impose. Il porte un titre en français, et, dans la mesure du possible, un équivalent dans la langue vernaculaire. Selon nécessité, certains fascicules comportent un index thématique français, ainsi que deux lexiques (français-langue vernaculaire, langue vernaculaire-français).

• *La bibliothèque*

Nous appelons *bibliothèque* l'ensemble des textes de littérature orale produit par chacun des trois peuples et les textes d'archives produits par ceux qui les ont visités. Le fait qu'elle rassemble littérature orale autochtone et sources écrites allogènes lui confère un statut original et singulier auquel nous tenons. Les textes ici rassemblés s'avèrent être les deux versants de la connaissance sur les peuples dont nous parlons. Cette littérature existe, mais elle est souvent soit dispersée, soit non publiée, enfouie dans des archives ou bibliothèques privées, quelquefois même, non encore collectée. Nous nous proposons de rassembler des textes anciens, des inédits ou des traductions, de les publier et d'ainsi mettre pour la première fois à la disposition tant des membres de la communauté autochtone que d'un public élargi l'ensemble des textes connus, recueillis ou non, découverts ou non par les membres de l'équipe. Chaque fois que cela est possible, est présenté le texte dans la langue vernaculaire et dans une traduction française. Dans le cas de textes d'archives, des présentations et des notes (historiques, critiques, linguistiques...) sont indispensables. Les pièces de ce corpus uniquement disponibles en d'autres langues (en l'occurrence le portugais, l'anglais, l'allemand ou le hollandais) sont données en traduction française. Une des tâches des éditeurs scientifiques sera de veiller autant à n'oublier aucun de ces documents qu'à éviter toute redondance avec d'autres fascicules, et l'on pense ici plus particulièrement à la mythologie.

- *Les publications pour la jeunesse*

La jeunesse des trois ethnies ne cesse de gronder devant le manque criant de matériaux didactiques à son usage, cette demande sociale croissante allant de pair avec la conscientisation politique des jeunes générations.

Ainsi, les jeunes Wayãpi par exemple nous ont demandé de rédiger, dans un français usant de mots simples, mais ce pourrait aussi être un texte bilingue, une histoire de leur peuple pour les enfants, ainsi que des fascicules sur les savoirs naturalistes de leurs parents.

Pour les jeunes Wayana, une autre idée a vu le jour, celle d'ateliers d'histoire, en concertation avec l'école : des enquêtes seraient menées auprès des anciens par les jeunes adolescents sur les récits historiques, dont les résultats seraient ensuite comparés, compilés, amendés et illustrés par les enfants, avant d'être publiés dans l'encyclopédie, sous forme bilingue.

Nul doute que dans ce vaste domaine de la littérature enfantine, d'autres projets jailliront.

- *Le dictionnaire fondamental*

À mi-parcours, faisant le lien d'une part entre le grand dictionnaire de langue final et les index, d'autre part entre les études et la bibliothèque, nous avons prévu la parution d'un dictionnaire fondamental bilingue colligeant les mots les plus usuels à partir de leur fréquence relevée dans les enregistrements et les transcriptions.

- *Le grand dictionnaire de langue*

Nous l'ambitionnons le plus complet possible. Véritable encyclopédie culturelle intégrée, par le choix de ses entrées, il reprend de façon synthétique l'ensemble des notions, des concepts, abordés dans les différents fascicules. Au fur et à mesure de la parution des fascicules, le chef du projet aura, entre autres missions, la tâche exaltante et périlleuse d'y puiser la matière vive, et ainsi d'incrémenter tout au long de l'aventure les trois bases de données lexicographiques et dictionnaires. Lorsque toutes les études seront terminées, viendra le temps de publier, sous la direction d'un maître d'œuvre spécifique, un dictionnaire culturel pour chacune des trois langues.

Tous les mots de la langue apparaîtront selon l'ordre alphabétique strict. On aura écarté aussi bien l'enchaînement logique ou analogique des entrées, tenté dans le *Dictionnaire wayãpi* (F. Grenand, 1989), que l'ordre phonologique adopté par les rédacteurs de l'*Encyclopédie des Pygmées Aka* (Thomas & Bahuchet, 1983 et suiv.) : le premier ne reflète bien souvent que la seule logique du compilateur au moment où il rédige ; le second est fondé sur une structure certes cohérente mais fermée ; tous deux complexifient par trop la lecture du néophyte. Pour arbitraire qu'il soit, l'ordre alphabétique, désormais connu de tous, demeure le plus simple à manier.

Afin de ne pas rester à l'état de squelette, les entrées seront agrémentées d'autant d'exemples qu'il sera jugé bon pour expliciter les différents sens de chaque mot.

Le dictionnaire sera accompagné d'un copieux lexique français-langue vernaculaire, établi avec un soin particulier.

- *Les index*

Chacune des encyclopédies se terminera par un volume d'index servant d'entrée et de repère au questionnement extérieur. Plusieurs index sont nécessaires. Tous se doivent d'être parfaits.

- *Index français raisonné* : les grands champs, en gras, découpés en sous-thèmes, doivent y apparaître, avec mention des volumes et toutes les pages d'occurrence.

- *Index des noms propres* : aussi indispensable que le précédent, il comprendra tous les toponymes, noms de cours d'eau, héros mythologiques ou historiques...

- *Table des illustrations* : pareillement éditée en fin de série, cette table rappellera pour l'ensemble des volumes parus, toutes les occurrences de dessins, photographies, gravures, cartes, plans, planches illustrées...

- *Bibliographie générale* : chaque fascicule comprend sa propre bibliographie. Le fascicule zéro, celui qui annonce tous les autres, en comprend également une. Pour autant, une bibliographie générale finale nous a semblé pertinente, car on la transformera en véritable somme bibliographique de l'ethnie considérée. Placée en fin d'édition de chacune des trois encyclopédies, elle permettra la nécessaire mise à jour, incluant la moisson originale (en particulier documents d'archives ou textes miméographiés) que chaque contributeur n'aura pas manqué de faire.

2. ENCYCLOPEDIE ILLUSTRÉE

Une encyclopédie se doit de ménager une place importante au message iconique (Alain Rey, 1982 : 44). C'est ainsi que, dès le départ, nous avons choisi d'accompagner le projet encyclopédique d'une base de données informatisée¹. Le projet consiste à rassembler et conserver tout document iconographique ou sonore touchant les trois peuples. Cette base de données fait suite à un inventaire général - encore en cours - de fonds iconographiques et muséographiques divers, publics ou privés, dont ceux de l'équipe. Les nombreuses illustrations des fascicules édités y ont été puisées. Chaque fois que cela s'est avéré nécessaire, des autorisations de reproduction furent demandées : pour la faune, la flore, des reproductions d'objets, des gravures anciennes...

2. 1. Photographies, gravures

Le noir et blanc a été, pour d'évidentes raisons de coût, privilégié sur la couleur, laquelle sera réservée à des cas très particuliers (parures de plumes, bibliothèque enfantine...).

Répondant en cela à une demande autochtone maintes fois réitérée, nous avons été particulièrement vigilants à rechercher les *portraits* (gravés ou photographiés) des grands hommes des trois peuples. On peut d'ores et déjà dire que les Wayana, grâce à Coudreau puis Hurault, sont les mieux servis. Pour les Wayãpi, on a aussi cherché à illustrer les lieux de la grande migration du XVIII^e siècle. Pour les Palikur, ceux du berceau de l'ethnie en Amapá.

2.2. Dessins

On y a eu recours aussi souvent que nécessaire. Pour l'artisanat, on a fait dessiner au trait tous les objets (*de visu* ou d'après photographie), si possible en les resituant, par comparaison avec d'autres, dans l'espace et le temps ; ainsi sont visualisées les influences des ethnies voisines (espace) et les transformations internes (temps).

2.3. Bandes dessinées

Pour le déroulement d'un savoir faire, d'une technique, d'un tour de main, on a utilisé la technique de la BD en quelques images : par exemple monter une poterie au colombin ; décrocher un balai fructifère du haut d'un palmier.

2.4. Planches

¹ Elle est constituée suivant le cahier des charges défini par les instances internationales comme OLAC, à partir de logiciels « *Open Source* » et avec contrôle de la diffusion afin de préserver les intérêts des ayants droit. La normalisation des fichiers, des transcriptions et des principes de catalogue permettra une interrogation ouverte et une évolution des archives pour une utilisation large et évolutive.

La modernité, dont on a dit qu'elle occupe une place importante, est elle aussi abordée par le dessin. Par exemple, des planches détaillent les parties des objets modernes simples (une bouteille, un moteur hors-bord) ou complexes (la ville, l'école), avec toutes les parties ou sous-parties nommées en langue vernaculaire.

2.5. Cartes

Sans avoir la prétention de doter les trois encyclopédies d'un atlas, nous avons accordé aux cartes un soin particulier ; elles ont pour but de retracer les chemins de migration, les circuits de commerce coloniaux, de délimiter les terroirs, d'en marquer l'évolution, de figurer les ethnies voisines. Dans le contexte particulier de la Bibliothèque, les cartes anciennes sont aussi incluse.

3. CALENDRIER DE PUBLICATION

3.1. Volumes parus

- Wayana 1 2007. Chapuis Jean. *L'ultime fleur*, ekulunpi tihmelë, *essai d'ethnosociogénèse wayana*, PUO, Orléans.
- Palikur 1 2008. Curt Nimuendaju. Traduction de *Les Indiens Palikur et leurs voisins*, présentée par Pierre Grenand.
- Wayana 0 2009. Eliane Camargo & Pierre Grenand. *Une encyclopédie, un peuple*.
- Wayãpi 0 2009. Pierre Grenand & Dominique Tilkin-Gallois. *Une encyclopédie, un peuple*.
- Palikur 0 2009. Antonia Cristinoi, Pierre Grenand & Artionka Kapiberibe. *Une encyclopédie, un peuple*.

3.4. Volumes en préparation

- Palikur 2 2010. Gabriel Bergounioux. *Phonologie du palikur, avec proposition de transcription*.
- Wayãpi 1 2011. Françoise Grenand. *Histoires dans la forêt, contes pour enfants*, édition bilingue.
- Wayãpi 2 2011. Pierre Grenand & Françoise Grenand. *Ethnobotanique*.

Au total, une quarantaine de volumes prévus sur douze ans.

Séminaire du GORDF

2007-2008

- 17 octobre
« Acquisition de la liaison et mise en place de structures syntaxiques dans le cadre du groupe nominal chez les enfants entre 2 et 6 ans », Céline Dugua, MCF, université d'Orléans, LLL
- 23 novembre
« Penser les techniques intellectuelles-scolaires : entre écriture, école et familles », Mathias Millet, MCF, université de Poitiers, SACO (Savoirs, cognition et rapports sociaux)
- 19 décembre
« Une langue de référence pour entrer dans la lecture », Silvia Lucchini, UCL (université catholique de Louvain-La-Neuve) et université de Liège, Belgique
- 16 janvier
« De quelques fondamentaux de la didactique du français », Emile Genouvrier, Professeur émérite, université de Tours
- 6 février
« Présentation de la méthode de lecture "Bulle" (Bordas) par ses auteurs », Patricia Bucheton, MDPH Orléans, GORDF, Marie-Aude Murail, écrivain pour la jeunesse et Christine Thiéblemont, PE
- 12 mars
« Du rapport à l'écriture à la compétence scripturale : apports et limites du concept de rapport à l'écrit », Suzanne Chartrand, université de Laval, Québec
- 2 avril
« Le rapport à l'écriture et l'apprentissage des formes discursives », Christine Barré-De Miniac, Professeur des universités, IUFM de Grenoble, Lidilem
- 14 mai
« L'éveil aux langues en classe ordinaire », Catherine Brumelot, CASENAV Orléans, GORDF
- 11 juin
« La construction de la compétence d'Interaction Orale (IO) en anglais de spécialité par le TCMO (Travail Collaboratif Médiatisé par Ordinateur) », Cédric Sarré, université d'Orléans, LLL-GORDF

2008-2009

- 22 octobre
« Normes discursives et rapport à l'écrit au début du parcours universitaire : entre implicite et conscientisation », Catherine Deschepper et Francine Thyryon, UCL, Belgique
- 26 novembre
« Didactique du français et image mobile : l'analyse sémio-pragmatique de l'offre éditoriale actuelle pour l'enseignement secondaire », Philippe Bourdier, IUFM Centre Val de Loire LLL-GORDF

▪ 17 décembre : « Les reformulations : des outils linguistiques favorisant l'ajustement et l'inter-compréhension », Blandine Pennec, université de Perpignan

▪ 14 janvier
« Typologie des reformulations dans les conversations de jeunes libanais francophones », Layal Kanaan, doctorante (LLL- GORDF)

▪ 18 février
« L'écriture collaborative à distance », Jean-Claude Bertin, université du Havre

▪ 18 mars
« L'intérêt didactique de la variation dans les textes littéraires francophones », Diane Schwob, (LLL- GORDF)

▪ 8 avril
« La question du sujet dans les textes prescripteurs de l'école primaire », Marie-France Bishop (IUFM de Versailles, Théodiles)

▪ 6 mai
« La stéréotypie un outil d'enseignement et d'apprentissage de l'écriture ? A quelles conditions ? », Bernadette Kervyn, UCL Belgique et université de Bordeaux 1.

2009-2010

▪ 14 octobre
« Apprendre à lire et à écrire des énoncés de problèmes mathématiques », Annie Camenisch et Serge Petit, IUFM d'Alsace, LilPa

▪ 18 novembre
« Les écritures parentales : histoire, pratiques et enjeux », Véronique Francis, IUFM Centre, Val de Loire, CREF

▪ 16 décembre
Thème : le développement du langage en maternelle, Véronique Boiron, IUFM d'Aquitaine

▪ 6 janvier
Thème : la didactique de la littérature en Europe ou une didactique intégrée du français et des langues, Dominique Ulma, IUFM de Lyon, Dynadiv et LLL-GORDF

▪ 17 mars
« Orthographe : nouveaux enjeux didactiques », Danièle Cogis, IUFM de Paris, MoDyCo UMR 7114

▪ 21 avril
« En amont du brouillon rédigé : nouveaux regards sur la planification du projet d'écriture de l'école au lycée », Olivier Lumbroso, IUFM de Bretagne, CENTRE ZOLA - ITEM / CNRS UMR 8132

PROHEMIO : Collation d'ouvrages pour le relevé des figements

Ouvrages traduits dans lesquels les expressions figées ont été récupérées à ce jour (juin 2010) pour la constitution du corpus du *Dictionnaire idéologique comparatif des expressions figées espagnoles et françaises*.

Camilo José Cela : *La familia de Pascual Duarte* (*La famille de Pascal Duarte*, Seuil, Paris, 1997. Tr. Jean Viet)

Camilo José Cela : *La colmena* (*La ruche*, Gallimard, Paris, 2003. Tr. Henri LP Astor)

Miguel Delibes : *Cinco horas con Mario* (*Cinq heures avec Mario*, Verdier, Lagrasse, 2010. Tr. D. Blanc)

Miguel Delibes : *El camino* (*Le chemin*, Gallimard, Paris, 1959. Tr. Maurice Edgar Coindreau)

Juan Marsé : *El amante bilingüe* (*L'amant bilingue*, Christian Bourgois, Paris, 1996. Tr. J.-M. Saint Lu)

Carmen Martín Gaité : *Nubosidad variable* (*Passages nuageux*, Flammarion, Paris, 1995. Tr. Claude Bleton)

Carmen Martín Gaité : *Lo raro es vivir* (*Drôle de vie la vie*, Flammarion, Paris, 1999. Tr. Claude Bleton)

Eduardo Mendicutti : *El palomo cojo* (*Le pigeon boîteux*, Christian Bourgois, Paris, 1995. Tr. D. Laroutis)

Eduardo Mendicutti : *Una mala noche la tiene cualquiera* (*Une sale nuit*, Ch. Bourgois, Paris, 1992. Tr. Denise Laroutis)

Eduardo Mendoza : *Sin noticias de Gurb* (*Sans nouvelles de Gurb*, Seuil, Paris, 2006. Tr. François Maspero)

Eduardo Mendoza : *La aventura del tocador de señoras* (*L'artiste des dames*, Seuil, Paris, 2002. Tr. F. Maspero)

Javier Tomeo : *Diálogo en re mayor* (*Dialogue en Ré majeur*, Editions de la Mauvaise Graine, Lyon, 2001. Tr. Mor Gaye)

Javier Tomeo : *Amado monstruo* (*Monstre aimé*, Christian Bourgois, Paris, 2007. Tr. Denise Laroutis)

Javier Tomeo : *La máquina voladora* (*La machine volante*, Ch. Bourgois, Paris, 2004. Tr. Denise Laroutis)

Javier Tomeo : *El cazador de leones* (*Le chasseur de lions*, Ch. Bourgois, Paris, 1990. Tr. Denise Laroutis)

A ces ouvrages s'ajoute la quarantaine de titres constituant l'œuvre de Javier Tomeo (environ 3000 pages). Les expressions figées y ont également été recensées mais uniquement dans la version originale (tous les romans n'ont pas été traduits).

CAHIERS DU PROHEMIO

***Cahiers du P.R.O.H.E.M.I.O.*, n° 1, 1996, 169 pp.**

"Oralité, histoire, écriture : réalisations et perspectives".

Actes de la Journée d'Etudes du 15 décembre 1995.

Université d'Orléans.

Edition préparée par Angel Iglesias Ovejero avec la collaboration de Cécile Iglesias, Monique Güell, Françoise Giraud.

***Cahiers du P.R.O.H.E.M.I.O.*, n° 2, 1998, 318 pp.**

"Usages de la formule".

Actes du II Colloque International (21-22 novembre 1997).

Université d'Orléans.

Edition préparée par d'Angel Iglesias Ovejero avec la collaboration de Cécile Iglesias, Monique Güell, Françoise Giraud, Daniel Vassivière.

***Cahiers du P.R.O.H.E.M.I.O.*, n° 3, 2000, 453 pp.**

"Expressions figées : idiomaticité, traduction".

Actes du III Colloque International (26-27 novembre 1999).

Université d'Orléans.

Edition préparée par Angel Iglesias Ovejero avec la collaboration de Denise Boyer, Lidio J. Fernandez, Françoise Giraud, Monique Güell.

***Cahiers du P.R.O.H.E.M.I.O.*, n° 4, 2002, 379 pp.**

"Idiomaticité, motivation et onomastique : la fabulation étymologique"

Actes du IV Colloque International (22-23-24 novembre 2001).

Université d'Orléans.

Edition préparée par Angel Iglesias Ovejero avec la collaboration de Denise Boyer, Lidio J. Fernandez, Françoise Giraud, Monique Güell.

***Cahiers du P.R.O.H.E.M.I.O.*, n° 5, 2004, 288 pp.**

***Cahiers du P.R.O.H.E.M.I.O.*, n° 6, 2004, 384 pp.**

"Formas de Vida y Cultura Tradicional en El Rebollar y otras partes".

Actes des I Journées Internationales d'Etudes de *El Rebollar*.

V colloque P.R.O.H.E.M.I.O. (23-27 juillet 2003).

El Rebollar (Salamanque).

Edition préparée par Angel Iglesias Ovejero avec la collaboration de Lidio J. Fernandez, Françoise Giraud, Monique Güell.

CAHIERS DU PROHEMIO 7 & 8

Interferencias lingüísticas y culturales en El Rebollar y otras partes

(Actas de las II Jornadas de Estudio de El Rebollar)

VI Coloquio del P.R.O.H.E.M.I.O.

22-25 de julio de 2005

ÍNDICE GENERAL

I Volumen (*Cahiers du P.R.O.H.E.M.I.O.*, VII)

Presentación 15

Mapa de El Rebollar 21

**I. Interferencias lingüísticas:
analogías y contrastes entre las hablas
de El Rebollar y de otras partes..... 25**

Miguel Á. REBOLLO TORÍO,
*Unidad y divergencias lingüísticas en Extremadura
y sur de Salamanca* 27

Ángel IGLESIAS OVEJERO,
*Interferencias lingüísticas y culturales en El Rebollar:
algunas propuestas para la transcripción gráfica
de conversaciones, relatos y testimonios orales
en la modalidad rebollana* 43

Ángel IGLESIAS OVEJERO,
*Analogías y contrastes entre las hablas de El Rebollar
y las extremeñas* 75

M^a Luisa MONTERO CURIEL,
*Cambios fonéticos en Castilla y León:
del ALPI al ALCL* 101

Pilar MONTERO CURIEL,
*El léxico extremeño y salmantino en el Diccionario
de Uso del Español de María Moliner* 115

Ana SERRADILLA CASTAÑO,
*Morfología y sintaxis extremeñas en el norte de
Huelva. Fronteras lingüísticas vs. fronteras políticas ...* 131

M^a Victoria NAVAS SÁNCHEZ-ÉLEZ,
*El léxico de las ocupaciones, influencia de dos áreas
lingüísticas en contacto:
Barancos (Portugal) y Encinasola (Huelva)* 149

Luc TORRES,

Contribución para la elaboración de un Glosario de las voces y modismos de los dialectos y hablas leoneses (leonesismos, montañesismos, asturianismos, salmantinismos, bablas de Extremadura y El Rebollar) presentes en La Pícara Justina) 173

**II. Interferencias lingüísticas:
traducción, bilingüismo, extranjerismos 193**

Josep GUIA, Maria CONCA,
Análisis contrastivo del frásico de Tirant lo Blanc (València, 1490; Barcelona, 1497) y sus traducciones castellana (Valladolid, 1511) e italiana (Venezia, 1538, 1566, 1611). Entre el calco, la equivalencia y la creación 195

Ramon LLADO, Xus UGARTE,
Multilingüismo y diglosia en Rabelais: lo que las traducciones revelan 217

José ALONSO PASCUAL,
Conflictos, clases, ventajas mitos y efectos adicionales del bilingüismo..... 229

Sylvie FOURNIÉ-CHABOCHE,
Las palabras españolas de la lengua francesa: aspectos formales, semánticos e históricos de la transformación cultural 261

Miguel Ángel GARCÍA PEINADO,
La proliferación inadecuada de coetillas y latiguillos en el español actual (galicismos y anglicismos)..... 281

José Antonio VICENTE LOZANO,
Una lengua de TBO: problemas traductológicos con "Mortadelo y Filemón" 293

II Volumen (Cahiers du P.R.O.H.E.M.I.O., VIII)

**III. Interferencias entre cultura oral y escrita:
fijación léxica.
Fraseología, paremiología, onomástica 335**

Francisco RUIZ GIRELA,
Los refranes: testigos de excepción del contacto entre culturas..... 337

Maria CONCA, Silvia PINTER,
Analogies i contrasts entre unitats fraseològiques antropolímiques de llengües romàniques 351

Ángel IGLESIAS OVEJERO, Françoise GIRAUD,
Cuán necios y cornudos son los otros en

El Rebollar y otras partes:
Portuguesis, Serranus y Farinatus
en el refranero xenófobo rebollano,
extremeño y castellano 363

IV. Interferencias literarias y tradición popular:
Motivos tradicionales en la literatura culta.
Autores regionales 401

José RAMIREZ DEL RIO,
Códigos semióticos e interferencias culturales en
los textos castellanos medievales..... 403

María del Pilar COUCEIRO,
Tradición grecolatina en la obra
de Rosalía de Castro..... 419

Isabel VÁZQUEZ DE CASTRO,
De la ‘tragedia’ a la ‘trigedia’:
interferencia lingüística
y creación de una estética en Los cuernos de
don Friolera de Valle-Inclán 429

Cécile IGLESIAS,
Canciones infantiles revisitadas:
las nanas en la lírica
de Federico García Lorca 443

José Luis PUERTO,
J. M^a. Gabriel y Galán:
tradición clásica y moderna 467

Pilar MAGADÁN CHAO,
Cuatro villancicos de Juan del Encina:
formas populares que sobreviven
en algunas comarcas salmantinas 483

José Ramón CID,
La gaita salamanquina
y el tamboril tradicionales,
en relación a la obra de Juan del Encina 507

Alfonso SAURA,
Vicente Medina, poeta social y
reivindicativo 523

V. Analogías y contrastes en la historia
y el arte iconográfico.
Constitución de la frontera. Figuras históricas locales.
Arte pastoril. Cine y realidad social.
Simbolismo cromático y política 553

Francisco Javier MORALES PAINO,

*Formación y evolución histórica de
la raya luso castellana:
Convivencia y conflictividad social
en torno a las regiones del Alto Coa y
El Rebollar miróbricense (ss. XII-XV) 555*

Emilio GARCÍA LOZANO,
*Fray Martín de Robleda,
primer obispo de Chile (1513-1561) 587*

José Ramón CID,
*Don Serafín Tella Gallego, robledano ilustre
y destacado arqueólogo 611*

Carlos GARCÍA MEDINA,
Arte pastoril del Rebollar 631

Lidio. J. FERNÁNDEZ RODRÍGUEZ,
*Estudio bibliográfico e histórico del documental
sobre Las Hurdes Tierra sin pan
de Luis Buñuel en 1933 645*

James DURNERIN, Catherine VIENET,
Colores y política 675

**VI. Tradición literaria y modalidad de habla rebollana.
Testimonios. Léxico. Romancero. Textos 685**

José Benito MATEOS PASCUAL,
*Razones (razonadas y de peso) para la defensa
del habla de El Rebollar 687*

Yolanda GARCÍA VICENTE,
*Flora rebollana. Plantas útiles y venenosas.
Sus nombres, características y usos 695*

Gabriel CALVO,
*Anotaciones a un romance geográfico
tradicional salmantino 715*

Pedro Javier CRUZ SÁNCHEZ,
*El cuaderno de rezos de Antonio Franco,
de Espino de la Orbada (Salamanca), 1869:
notas a propósito del “Romance del Arado” 733*

Ángel IGLESIAS OVEJERO, Françoise GIRAUD,
*Tres relatos orales rebollanos:
El lobu y los pastoris, Los polvus de las cocas
y La Muerti 757*

PROHEMIO 9 & 10

Memoria histórica - Historia oral : guerra, postguerra y emigración en El Rebollar y otras partes

III Jornadas Internacionales de Estudio de El Rebollar
VII COLOQUIO DEL P.R.O.H.E.M.I.O.
Curso extraordinario de la Universidad de Salamanca
20-22 de julio de 2007

ÍNDICE GENERAL

I Volumen (*Cahiers du P.R.O.H.E.M.I.O.*, IX)

Presentación 17

I. Historia oral y escrita 23

Fernando SÁNCHEZ MARROYO,
*La Guerra Civil. Algunos libros significativos
y ciertas cuestiones pendientes* 25

Julián CHAVES PALACIOS,
*Represión en la Guerra Civil y movimiento guerrillero
de posguerra en la Sierra de Gata y Salamanca* 37

Santiago LÓPEZ GARCÍA, Severiano DELGADO CRUZ,
*Terror de Estado en la retaguardia:
la represión en Castilla y León
en los primeros meses de la Guerra Civil* 87

Ángel IGLESIAS OVEJERO,
*Archivos vivientes: las víctimas del terror de 1936
en El Rebollar y pueblos aledaños* 101

José Luis HERNÁNDEZ SÁNCHEZ, Miguel LIRIO DÍAZ,
*Notas sobre la presencia militar del Tercer Reich
en Salamanca a través de las Causas del
Archivo Militar de Ferrol* 203

Julio FERNÁNDEZ GARCÍA,
Las cárceles del Franquismo..... 215

Floren DIMAS BALSALOBRE,
*Consejo de guerra:
la máquina asesina del franquismo* 233

M^a Luz de PRADO HERRERA,
*Auxilio Social en Salamanca durante la guerra civil:
Organización y desarrollo* 253

Cayetano IBARRA BARROSO,
Recuperación de la Memoria Histórica en Extremadura:

<i>Investigación, localización y exhumación de fosas</i>	271
José Ignacio MONTEAGUDO, Manuel MATELLÁN, <i>Testimonios de la guerra, posguerra y emigración en el Archivo de la Escritura Popular del Bajo Duero</i>	285
Rafael MORALES RUIZ, <i>Entrevistando a mayores. La experiencia del Taller de Historia Oral del Archivo Municipal de Córdoba</i>	303
II. Historia Oral – Testimonios	337
Mar GONZÁLEZ DE LA PEÑA, <i>Interrelación de la historia oral y los archivos en la búsqueda de un desaparecido en la Guerra Civil española</i>	339
Matilde GARZÓN RUIPÉREZ, <i>La Guerra Civil, El Golpe de Estado y la represión franquista a través de mi experiencia</i>	347
Manuel CORRAL BACIERO, <i>Tirar del hilo: La Guerra Civil 1936-39 en Peñaranda de Bracamonte</i>	359
Luis CALVO RENGEL, <i>La represión fascista en Salamanca</i>	371
Isabel MATEOS MATEOS, <i>Hechos ocurridos en los municipios de Robleda y Boadilla en los días 13 y 14 de agosto de 1936 y posteriores gestiones</i>	387
José ALONSO PASCUAL, <i>...Y pudieron contarlos</i>	393
II Volumen (Cahiers du P.R.O.H.E.M.I.O., X)	
II. Historia Oral – Testimonios (cont.)	421
Maria CONCA, Josep GUIA, <i>Muertos y represaliados republicanos en Beneixama (1939-45)</i>	423
Alfonso SAURA SÁNCHEZ, <i>La represión de los vencedores en la Casa del Pueblo del LLano del Beal (Murcia)</i>	443
Luc TORRES, <i>Historia, intrahistoria y visión regionalista de la Guerra Civil: Los días del Alzamiento en Navarra (18 de julio y días próximos)</i>	461
<i>Testimonios de Robleda 1973, 1976 y El Payo 1973</i>	

transcritos y anotados por
Ángel IGLESIAS OVEJERO 473

III. Postguerra y emigración en su contexto 551

Emiliana LEÓN LÓPEZ,
Españoles en Francia. Emigración y exilio
en *La Voulte* 553

Carlos GARCÍA MEDINA,
La cocina del hambre 569

Gabriel CALVO,
La postguerra: racionamiento, estraperlo
y *prostitución* 589

Beatriz SÁNCHEZ VALDELVIRA,
La arquitectura en El Rebollar: el caso de Robleda
en el contexto de la postguerra y la emigración 615

Yolanda GARCÍA VICENTE,
Remedios caseros hechos con plantas
medicinales en la postguerra 625

IV. Visión de la Guerra en la literatura y el arte .. 645

María Pilar COUCEIRO,
La guerra en los Cancioneros populares:
Damaso Ledesma, Bonifacio Gil,
Manuel García Matos 647

Cécile IGLESIAS,
Los cantares de la guerra en el repertorio tradicional:
¿memoria u olvido de una poesía circunstancial? 663

José Luis PUERTO,
La perspectiva de los derrotados
(*arte, música, literatura*) 697

Miguel Á. REBOLLO TORÍO,
La guerra en Baroja 713

Ángeles GARCÍA CALDERÓN,
Lorca y la Guerra Civil en la poesía canadiense:
traducción de poemas al castellano 729

Carmen BALBUENA TOREZANO,
La guerra civil española en la literatura alemana
y *su traducción al castellano* 753

André Michel FEUGAIN,
Análisis contrastivo de los carteles
de propaganda durante la Guerra Civil 765

James DURNERIN, <i>El rescate de la memoria en Maquis</i> <i>de Alfons Cervera</i>	785
Felipe APARICIO NEVADO, <i>Una guerra cinegética</i>	799
Natalie NOYARET, <i>La memoria de la guerra en</i> <i>Soldados de Salamina y</i> <i>La velocidad de la luz de Javier Cercas</i>	811
Isabel VÁZQUEZ DE CASTRO, <i>Memoria y escritura: estrategias de la transmisión</i> <i>de una historia en La voz dormida de Dulce Chacón,</i> <i>Los rojos de ultramar de Jordi Soler y Maquis</i> <i>de Alfons Cervera</i>	829
Lidio J. FERNÁNDEZ RODRÍGUEZ, <i>El maquis antifranquista en la posguerra</i> <i>de Cantabria y su tratamiento en el cine</i>	845

CAHIERS DU PROHEMIO 11

Historia, Habla y Cultura Tradicional

IV Jornadas Internacionales de Estudio de El Rebollar VIII COLOQUIO DEL P.R.O.H.E.M.I.O. 17-19 DE JULIO DE 2009

ÍNDICE

Cahiers du P.R.O.H.E.M.I.O., XI

Presentación	11
I. Historia reciente y testimonios	19
Ricardo ROBLEDO, <i>Reforma agraria y represión</i>	21
Luis CASTRO BERROJO, <i>El embalse del Águeda, los regadíos y la colonización: una reforma técnica frustrada</i>	37
Maria CONCA, Josep GUIA, <i>Anàlisi d'un sumariíssim d'urgència: el núm 7545 d'Alacant, contra Vicent Conca Amorós, alcalde de Beneixama</i>	55
Franciso Javier MORALES PAÍNO, <i>Historia grabada en mis recuerdos. Memoria viva de una supervivencia. "Puchero Zarceño"</i>	95
Luc TORRES, <i>Cuadernos de un desertor del ejército de Franco. Itinerario de un navarro republicano</i>	115
Miguel LIRIO DIAZ, <i>Apuntes sobre el estudio de Fuentes de Oñoro y su Aduana en las postrimerías de la II República y Guerra Civil</i>	171
Ángel IGLESIAS OVEJERO, <i>Ensayo de cronología del alzamiento militar terror y represión de 1936 a 1946 en el Alto Águeda</i>	177
II. Habla y literatura tradicional	321
Xus UGARTE BALLESTER,	

Les vins d'Anjou, les couilles de Lorraine et les Gentilz, qui furent buveurs eternelz: <i>ambivalencia y sociocentrismo lúdico</i> <i>en los blasones populares del Pantagruel de Rabelais,</i> <i>con muestras de versiones en varias lenguas.....</i>	323
Claude CARCENAC, <i>Femme et mariage à travers les sagesses de</i> <i>l'Égypte antique et les proverbes</i> <i>espagnols et français</i>	335
Cécile IGLESIAS, <i>La imagen del pícaro en los pliegos de cordel</i>	349
Ángel IGLESIAS OVEJERO, <i>Apotegmática:</i> <i>hechos y dichos tradicionales de El Rebollar</i>	381
José Benito MATEOS PASCUAL, <i>Riofrío, motes de las dos orillas</i>	411
Pilar TORRES VÁZQUEZ, <i>Son de antes</i>	423
Saturnino RODRÍGUEZ MIGUEL, <i>Intelectus apretatus.....</i>	429
III. Folclore y otras formas de cultura tradicional 441	
José Luis PUERTO, <i>La perspectiva celeste: leyenda, poesía y creencia</i> <i>en torno al cielo en el sur salmantino</i>	443
Pedro Javier CRUZ SÁNCHEZ, <i>Surmegir al santo. Las rogativas "pro pluvia" a San Ginés</i> <i>en Robleda (Salamanca)</i>	459
Lidio Jesús FERNÁNDEZ RODRÍGUEZ, <i>Manuel Llano y la mitología cántabra</i>	479
Rosa LORENZO LÓPEZ, <i>La mujer en la cultura tradicional.</i> <i>El drama de la madre soltera</i>	501
Carlos GARCÍA MEDINA, <i>Los trabajos y oficios desaparecidos</i> <i>de la tierra de Ciudad Rodrigo</i>	519
Gabriel CALVO, <i>Dendrolatrías familiares</i>	531
Ángel IGLESIAS OVEJERO, <i>El folclore fue historia vivida: Que rompan la Jesa Arriba.....</i>	547

UN EXEMPLE DE PROGRAMME REGIONAL : CRITERR

Rapport du Laboratoire Ligérien de Linguistique sur les actions inscrites dans le programme CRITERR – 2008 –

1. Observatoire du français et des contacts de langue en Région Centre

Ce programme, bénéficiant d'un soutien significatif de la part du Centre de ressources, a connu un développement très important au cours de l'année 2008, mobilisant une part essentielle des financements disponibles, soit :

- en matériel : achat d'équipement d'enregistrement et logiciels de traitement consommant les 9.000 € de crédits portés au crédit de cette action ;
- en personnel par la rémunération d'étudiants en vacation, chargés des opérations de saisie, transcription et relecture.

Pour mémoire, l'objectif est de constituer Orléans comme centre de référence pour la constitution d'un grand corpus de référence de français parlé (10.000.000 de mots) dont les applications concernent à parts égales la didactique (français langue maternelle / français langue seconde) et le Traitement Automatique des Langues (reconnaissance de la parole, fouille de données orales, indexation automatique...).

Le principal résultat du travail conduit cette année par le laboratoire concerne la constitution d'une base de données qui est d'ores et déjà la principale ressource en langue française parlée et qu'il convient d'incrémenter pour la porter au niveau requis par les exploitations visées. Afin d'assurer une accessibilité optimale aux données tout en réduisant les coûts de traitement, le LLL s'est rapproché du département de linguistique de l'Université de Louvain pour le formatage de données et du Laboratoire d'Informatique Mathématique et des Sciences pour l'Ingénieur (LIMSI - Paris XI/CNRS) pour mettre au point des techniques d'automatisation des procédures.

La phase technique de préparation des deux cents heures d'enregistrement est terminée, mais la nécessité d'une anonymisation interdit la mise en ligne qui sera l'aboutissement d'une démarche de partage des données sur le modèle de ce qui a pu être réalisé par l'ATILF à Nancy pour l'écrit. Pour le moment, et avec un accès réservé, les documents sont déposés à l'Université d'Orléans, dans la base ESLO (Enquête Socio-Linguistique à Orléans). Ils peuvent être présentés à des fins de démonstration et rendus accessibles pour des consultations contrôlées afin de respecter les exigences juridiques inhérentes à ce type d'enquête.

Cette base de données comprend en réalité trois enquêtes corrélées :

- une restitution archivistique d'une enquête conduite à Orléans à la fin des années 60,
- un renouvellement de cette enquête à des fins comparatives en cours de constitution (dynamique des langues),
- une étude des langues parlées à Orléans.

Chaque enregistrement est documenté, transcrit sous Transcriber et correspond aux prescriptions Dublin Core et OLAC. Il est associé à une documentation papier et à une mémoire des opérations conduites pour parvenir à l'élaboration de la base.

En termes de collaboration, la recherche associée, en France :

- la Bibliothèque nationale de France, Département des Documents Audiovisuels et Sonores (ex Phonothèque Nationale),

- l'École des Hautes Études en Sciences Sociales (Centre de Linguistique Théorique),
- l'Institut de Linguistique Française qui fédère dix équipes CNRS travaillant sur la langue française et le LLL,
- l'Observatoire des Pratiques Linguistiques de la Délégation Générale à la Langue Française et aux Langues de France auprès du Ministère de la Culture et de la Communication,
- le LIMSI et
- le Laboratoire d'Informatique de l'Université de Tours (une collaboration qui doit être étendue au LIFO dès cette année).

A l'étranger, participent au projet des chercheurs des universités de Bonn (Allemagne), Louvain (Belgique), Exeter (Grande-Bretagne) et Limerick (Irlande).

Dans l'étude des langues parlées à Orléans, des études spécifiques ont été entreprises concernant les langues arabes, tamaghit, peul, anglaise, kriol... Une recension générale est en cours afin de mesurer d'une part la diversité des langues en usage dans l'agglomération et les influences réciproques entre le français et ces langues, dans les deux sens.

Des travaux ont été présentés au cours de cette année

- à Bâle (colloque sur les corpus romans, mai 2008, publication en cours, I. Eshkol),
- à Paris (Congrès Mondial de Linguistique Française, juin 2008, publication de G. Bergounioux consultable sur le site du colloque),
- à Lausanne (formations doctorales en linguistique de la Suisse romande, septembre 2008, publication de G. Bergounioux prévue en juin 2009 dans les *Cahiers de l'Institut de Linguistique*, Lausanne),
- à Oxford (*French Language Studies* Colloquium, septembre 2008, à paraître à OUP, contribution d'O. Baude), à Swansea (*Corpus Linguistics*, septembre 2008) et à
- à Oviedo (Congrès International de Linguistique, septembre 2008, à paraître dans les actes du colloque, contributions de L. Abouda, B. Hamma, F. Nemo et M. Petit).

D'autres présentations sont prévues dans les mois à venir.

2. Actions collaboratives (Guyane et Burkina Faso)

L'assistance apportée aux chercheurs en linguistique du Burkina Faso a conduit le laboratoire à accueillir deux doctorants issus du CNRST burkinabé dans l'objectif d'assurer, en plus d'un approfondissement des contenus disciplinaires, une formation aux techniques de collecte et de traitement des données. L'objectif est d'apporter une assistance scientifique et technique aux acteurs nationaux pour qu'ils puissent constituer progressivement une phonothèque nationale au standard international avec des copies de sauvegarde qui seraient hébergées par la BnF. Une part de l'accroissement de l'équipement informatique du laboratoire a été consacrée à ce projet.

Les travaux sur la Guyane sont en cours d'exploitation, sous la conduite d'un chercheur du laboratoire qui a été nommé à la direction du CNRS en Guyane. Une demande forte concerne la préparation des travaux qui seront publiés dans l'*Encyclopédie des langues et cultures de Guyane*, en partenariat entre les Presses Universitaires d'Orléans et le Comité des Travaux Historiques et Scientifiques (Ministère de l'Éducation Nationale). Sur les vingt ans à venir, c'est une soixantaine d'ouvrages qui seront publiés.

3. Comité pour la Simplification de la Langue Administrative en Région Centre

Décliné à partir de son homologue national, le COSLA, service du Premier Ministre destiné à l'orientation et à la prescription en matière de réforme de l'état, le COSLA-Centre a vocation à fonctionner comme un observatoire des pratiques administratives à même de vérifier à son niveau la mise en application des directives et comme un organe de recommandation. En 2008, un premier inventaire des besoins et des attentes a été dressé. La Région, par la voix de son Président, et le COSLA national ont manifesté leur intention de soutenir résolument cette initiative.

RESEAU FRANÇAIS DE PHONOLOGIE

Appel à communication pour le

Colloque du Réseau Français de Phonologie

In memoriam Nick Clements

Université d'Orléans

Judi 1, vendredi 2 et samedi 3 juillet 2010

Afin de prolonger la série des rencontres annuelles impulsées par le GDR « Phonologies », et dans le même esprit, le Réseau Français de Phonologie (RFP) lance un appel à communication pour un colloque qui se tiendra à l'Université d'Orléans du 1 au 3 juillet 2010, avec le soutien des deux fédérations de recherche du CNRS, Institut de Linguistique Française (ILF / FR 2393) et Typologie Universaux Linguistiques (TUL / FR 2559).

Toutes les écoles et tous les courants de la phonologie sont invités à participer. Les problèmes traités peuvent concerner la phonologie générale comme l'analyse phonologique d'un phénomène linguistique ou d'une langue particulière. Seront prises en considération les soumissions concernant l'analyse du signal, la perception, l'acquisition, la diachronie, la dialectologie, les formalismes, l'épistémologie et toutes les questions qui s'inscrivent explicitement dans le domaine de la phonologie et de ses interfaces.

Format de soumission et sélection

Les propositions de communication, rédigées en français ou en anglais, seront présentées sous forme d'un résumé de deux pages maximum (format A4, police Times taille 12), incluant les éléments essentiels de bibliographie, et transmises à rfp@univ-orleans.fr en fichier Word, Latex ou PDF. Ce fichier ne fera aucune mention de l'auteur qui indiquera ses nom, civilité, adresse, institution et le titre de sa communication sur un fichier séparé. Les propositions seront transférées pour évaluation à deux rapporteurs et la sélection finale fera l'objet d'une discussion collective.

Calendrier

Appel à contributions: 1^{er} décembre 2009

Envoi des contributions avant le 1^{er} mars 2010

Retour des avis aux auteurs : 15 avril 2010

Colloque : du jeudi 1^{er} au samedi 3 juillet 2010, *in memoriam* Nick Clements.

(Les prochains colloques du RFP se tiendront à Tours en juin 2011 et à Paris en juin 2012)

Organisation

Les communications seront de 30 minutes, réductibles à 20 si le conférencier en fait la demande au moment de la soumission, à quoi s'ajouteront dans tous les cas dix minutes de discussion. Pas de session posters prévue. Deux conférenciers invités ouvriront les journées et une fin d'après-midi sera réservée à une assemblée générale du RFP.

Des sessions thématiques peuvent être proposées. L'organisateur est invité à donner un titre général avec un argumentaire de quelques lignes et la liste des collègues qui participeraient à la session avec un bref résumé de la thématique que chacun développerait. L'évaluation se fera lors de la réunion finale du comité de lecture sur proposition de deux rapporteurs et sans anonymat.

Le colloque ne prend en charge aucun frais, le coût de l'inscription couvrant les dépenses courantes. Aucune publication n'est prévue mais la question sera rediscutée au moment du colloque.

Comité de lecture

J.-P. Angoujard, S. Bendjaballah, G. Bergounioux, J. Brandao de Carvalho, Ph. Carr, M. D'Imperio, C. Dugua, J. Durand, S. Ferré, J.-M. Fournier, S. Herment, M.-P. d'Imperio, M. Lahrouchi, B. Laks, J.-L. Léonard, J. Lowenstamm, N. Nguyen, R. Noske, C. Patin, A. Rialland, R. Ridouane, M. Russo, P. Sauzet, T. Scheer, Ph. Ségéral, A. Tifrit, N. Vallée, S. Wauquier.

Comité d'organisation (Orléans & Tours)

V. Abasq, G. Bergounioux, C. Dugua, S. Ferré, J.-M. Fournier, N. Lampitelli, X. Luo, M. Martin

COLLOQUE RFP
Orléans 2010

Jeudi 1er juillet

10:00 - Allocutions de bienvenue

10:20 - Hommage à Nick Clements

10:40 - A. Rialland & M. Embanga Aborobongui : Le système intonatif de l'embosi

11:20 - T. Scheer : What the initial CV is initial of

13:30 - J. Pater : Systemic Simplicity

14:10 - S. Wauquier : Erreurs de liaison en acquisition du français L1 et L2

14:50 - C. Patin : De l'intérêt des domaines pour l'analyse tonale: tons et intonation en shingazidja

15:50 - N. Ballier & Ph. Martin : Corrélats prosodiques et acoustiques de la syllabification

16:30 - M. L. Mazzola : Aspects segmentaux du français régis par la prosodie

17:30 - R. Noske : Contrast Preservation Theory, Glottalic Theory and the Germanic Sound Laws

18:10 - M. Lahrouchi & Ph. Ségéral : Le gabarit en tagnawt

Vendredi 2 juillet

9:00 - E. Dresher : Conférence plénière: Feature Hierarchies Halle (1959) to Clements (2009)

10:00 - J. Bucci : La réduction vocalique en Coratin

11:00 - B. Laks : La phonotactique saussurienne : loi de la valeur relationnelle et positionnelle

11:40 - E. Caratini : "Divorce" in the history of German

11:40 - A. Fournet : La fusion de phonèmes en proto-arabe

13:40 - S. Ben Si Saïd : L'interaction entre la structure et la mélodie

13:40 - N. Arbaoui : Les multiples secrets de la forme II de l'arabe classique

14:20 - O. Piot : The prosody of complex intonation phrases in French

14:20 - N. Yamaguchi : La fréquence des traits peut-elle prédire l'ordre d'acquisition des consonnes?

15:00 - K. Hammane : Miroir, miroir, dis-moi à quoi ressemble une consonne syllabique...

15:00 - N. Lampitelli : Du caractère abstrait de la composante phonologique

16:00 - J.-M. Fournier : Accentuation lexicale et poids syllabique en anglais

16:40 - S. Bendjaballah & Ph. Ségéral : Le système vocalique et l'accent en mehri.

17:20 - O. Rizzolo : Une énième ballade sur les chemins du verlan

18:20 Assemblée générale du RFP

Samedi 3 juillet

9:00 - J.-L. Léonard : Classes naturelles, traits, complexité et simplicité segmentale en tu'un savi (mixtec)

9:40 - A. Delplanque : Les consonnes finales du dagara: propriétés des Eléments Phonologiques

10:20 - R. Välimaa-Blum : Why it is so difficult to violate vowel harmony in Finnish: A cognitive view

11:20 - J. Brandão de Carvalho : Je ll'ai vu : the case for abstract candidates

12:00 - J. Berns : Bruits fricatifs en français parisien : le dévoisement vocalique et l'affrication plosive

14:00 - J. Lowenstamm : Cycles et phases

14:40 - J.-M. Beltzung : Dissimilation consonantique par effacement dans certaines langues bantu

15:20 - N. Faust : Où est le [u]? Sur les conséquences théoriques de la décomposition de morphèmes

16:00 Conclusion

1. Présentation du projet

Identité du coordonnateur du projet :

Nom – Prénom : Caron Philippe et Hanote Sylvie

Corps et grade : PR et MCF

Laboratoire d'affiliation du coordinateur : FORELL A EA3816

Courriel : philippe.caron@univ-poitiers.fr ; sylvie.hanote@univ-poitiers.fr

Téléphone : 05.49.45.46.51 (Marie-Christine Mérine, secrétariat FORELL)

Nom du projet : « Parole »

Mots-clés (5 maximum) : Phonétique, prosodie, analyse acoustique, synthèse de la parole

Type de projet

- Lié à une action existante du présent contrat quadriennal (CPER, ACI, ANR,...)
- Déjà élaboré (en cours de soumission auprès d'un organisme financeur)
Précisez lequel :
- En cours d'élaboration

Concernant le volet « phonétique anglaise et analyse acoustique », une ACI Université de Poitiers est en cours (2009-2010). Cette ACI réunit deux laboratoires de l'université de Poitiers : FORELL A (EA3816, MSHS Poitiers) et le laboratoire XLIM-SIC (UMR CNRS 6172), en association avec la fédération FELINCO (Poitiers, Limoges, Orléans, Tours). Cette ACI s'appuie sur le plateau technique « Analyse du signal sonore et traitement de la parole » de la MSHS. (Budget alloué : 12 000 euros.)

Objet d'étude et approche scientifique

Le projet « Parole » met en oeuvre des connaissances (phonétiques, prosodiques) et des outils technologiques au service de deux programmes complémentaires :

- l'analyse acoustique par ondelettes (avec l'anglais oral contemporain comme principale application) permet de pénétrer plus avant dans la réalisation orale et en élucide plus finement les caractéristiques ;
- la restitution de la diction haute du français classique vers 1700 (synthèse de parole, logiciel Kali) est un cas d'archéologie phonético-prosodique intéressant puisqu'il s'agit de faire revivre un objet phonétique et esthétique perdu.

L'objet d'étude commun est la parole, le projet est l'élaboration de corpus oraux et le recours à des outils de traitement du signal pour analyser ou reconstruire la parole. Les deux axes sont donc complémentaires et s'enrichissent mutuellement.

Etat de l'art et référence clefs :

Organisation de 2 journées d'études à la MSHS de Poitiers (Octobre 2008 et octobre 2009).

- o Octobre 2008 : journée « Parole » co-organisée par P. Caron et S. Hanote, où sont intervenus des spécialistes de la mise en place de corpus oraux (présentation du corpus *Eslos*, LLL Orléans), des spécialistes de l'analyse de la parole et de la synthèse de parole.

- Octobre 2009 : journée « Intensité » co-organisée par S. Hanote et S. Moore, où sont intervenus des spécialistes de la phonétique acoustique (chercheur invité : P. Martin Université Paris-Diderot EA3967), avec une table ronde finale sur la question « L'intensité, est-elle une mesure pertinente en analyse de la parole ? »

La publication des travaux de ces deux journées d'études est prévue dans la revue électronique *CORELA – Cognition, Représentation, Langage*. <<http://revue-corela.org>>. La première paraîtra en juin 2010, la seconde à l'automne 2010.

Problématique et méthodologie :

- Le programme 'restitution de la diction haute du français vers 1700' s'impose actuellement car il répond à un besoin chez les artistes, musiciens et acteurs, qui sont demandeurs de connaissances avérées. D'autre part il est le seul de son type actuellement et il permettra d'éviter que se développent des idées originales mais scientifiquement mal assises, dérivées de l'œuvre pionnière mais entachée d'erreurs d'Eugène Green (*La Parole baroque*, Desclée de Brouwer 2002).

La méthodologie dans ce programme consiste à collationner dans les manuels (grammaires, dictionnaires généraux et dictionnaires de rimes, manuels de français oral etc.) et les remarques du temps un maximum de données et d'indices relatifs à la phonétique et à la prosodie de cet habitus, de les comparer, d'essayer de les comprendre et de les interpréter. Puis l'enregistrement par un acteur d'un corpus oral simulé va permettre d'extraire à la fois les segments minimaux de parole (les diphtongues), segments qui seront ensuite réagencés librement par la synthèse de parole à la demande d'un usager qui demandera l'oralisation-type d'un texte de son choix. Puis ce texte, vers par vers, sera analysé et sa structure métrico-syntaxique se verra dotée, sur la base de ses propriétés, d'un patron prosodique le plus proche possible de sa structure.

- En ce qui concerne le versant « analyse acoustique et phonétique anglaise », le travail commencé dans le cadre de l'ACI (2009-10) a d'ores et déjà permis :

- le début de la mise en place d'un corpus d'anglais oral authentique (enregistrements de locuteurs d'une part et enregistrements radiophoniques d'autre part),
- le développement du logiciel de traitement de la parole *Speech Wave Analyser* dans le cadre de la coopération avec le laboratoire XLIM-SIC.

Il s'agira de poursuivre le travail commencé sur ces deux versants et de valider le logiciel sur des problématiques de phonétique anglaise (la préfixation, la liaison, la prononciation des diphtongues...). L'objectif est de faire ressortir des signatures acoustiques en accord avec les phénomènes phonétiques / phonologiques.

Programme (organisation temporelle des étapes) :

Analyse acoustique et phonétique anglaise :

- a) Extension du corpus d'anglais oral et mise en place d'un mode d'interrogation de ce type de corpus. Pour affiner les contours scientifiques et techniques de cette étape, des contacts sont pris avec des laboratoires qui travaillent déjà sur la mise en place d'un corpus oral (corpus Aix-MARSEC Machine Readable Spoken English à Toulouse Le Mirail, Corpus PFC (Phonologie du Français Contemporain) et maintenant PAC (Phonologie de l'Anglais Contemporain : Jacques Durand et Bernard Laks, entre autres, à Toulouse également). De plus, pour la mise en place d'un mode d'interrogation efficace de ce type de corpus, l'intervention d'un ingénieur d'étude est nécessaire. A ce titre, un dialogue est établi avec Julien Eychenne, qui a travaillé sur la plateforme PFC pour qu'il nous fasse bénéficier de son expérience sur le corpus PFC. (24 mois)
- b) Test du logiciel *Speech Wave Analyser* sur des phénomènes de phonétique anglais pertinents. Dans sa thèse (2009-) « Préfixation et phonologie de l'anglais : analyse lexicographique, phonétique et acoustique » (direction : Jean-Louis Duchet, co-direction

Sylvie Hanote), Nicolas Videau va travailler sur la prononciation des préfixes en anglais oral contemporain. Actuellement, le logiciel a été développé par Philippe Carré (responsable équipe de recherche ICONES, laboratoire XLIM-SIC), puis par Raphaël Soulard (Doctorant XLIM-SIC) dans le cadre de l'ACI. Maintenant que les bases du logiciel sont posées, et en parallèle du travail de recherche fondamentale sur les phénomènes de phonétique anglaise, une partie technologique doit être assurée par un ingénieur pour faire évoluer le logiciel et assurer sa diffusion au sein de la communauté. En effet, nous souhaitons, à moyen terme, proposer une solution logicielle aux laboratoires travaillant sur des sujets connexes. (24 mois)

Ce projet d'analyse acoustique et phonétique anglaise s'articule avec un autre projet avec le Laboratoire Ligérien de Linguistique (LLL Tours) sur la numérisation de données phonétiques (J.M. Fournier (LLL) et J.L. Duchet (FORELL-A)).

Restitution de la diction haute du français

2010 : journée Parole III, destinée à agréger autour des laboratoires cocontractants (FORELL Poitiers, CRISCO de Caen et CESR de Tours) des partenaires étrangers spécialistes des manuels de français langue étrangère au XVIIe siècle et des experts (Yves-Charles Morin, Université de Montréal et Benoit de Cornulier, Université de Nantes). Concertation en vue de la rédaction d'un livre collectif qui devrait faire autorité.

2011 : enregistrement du corpus de base par Olivier Bettens (Lausanne) et/ou Pierre-Alain Clerc (Genève). Prélèvement des diphtonges et ajout de logatomes pour les combinaisons absentes du corpus

2012 : création d'un glossaire phonétique des mots spécifiques du registre grave (tragédie dramatique et lyrique)

2013 Premiers tests du logiciel de synthèse de parole. Correction des erreurs phonétiques et prosodiques.

2014 Création de l'environnement scientifique de l'outil en ligne (c'est-à-dire les pages fixes qui constitueront la mise en perspective des résultats acquis par le programme et qui s'incarnent dans la synthèse de parole).

2015 Installation de la base en ligne, en libre accès.

2. Partenaires, ressources humaines et matérielles

Partenaires du projet :

- Scientifiques (labos – MSH – universités...)
 - FORELL-A EA3816 (MSHS Poitiers)
 - CRISCO (MSHS de Caen)
 - X-LIM SIC UMR CNRS 6172, Université de Poitiers / Limoges
 - CeReS (Université Limoges)
 - LLL (future intégration dans la MSHS Tours)

Ressources humaines :

Nom et prénom	Statut (PU, MCF, CR, DR, ...)	Laboratoire	Organisme d'appartenance (CNRS, Université de Poitiers, ...)	Taux de participation (ETP) ¹
Caron Philippe	PR	FORELL -A	Université Poitiers	0,6
Duchet Jean-Louis	PR	FORELL -A	Université Poitiers	0,2
Col Gilles	PR	LLL	Université Tours	0,3
Hanote Sylvie	MCF	FORELL -A	Université Poitiers	0,6
Moore Susan	MCF	CeReS	Université Limoges	0,3

Videau Nicolas	Doctorant	FORELL- A	Université Poitiers	0,6
Morel Michel	Ingénieur	CRISCO	Université Caen	0,4
Laurenti Jean-Noël	MCF	CESR	Université Tours	0,3
Delente Eliane	MCF	CRISCO	Université Caen	

En outre, des partenaires étrangers participent au projet et viendront à la journée PAROLE III, destinée à rendre leur participation plus organique :

- Yves-Charles Morin (Université de Montréal)
- Olivier Bettens (chercheur indépendant, spécialiste de musique vocale et de phonétique historique (Lausanne))
- Elmar Schafroth (Université de Düsseldorf)
- Sergueï Vlassov (Université de Saint Petersburg)

Ressources matérielles :

Utilisation des plateaux techniques MSHS :

Oui Non

Si oui lequel :

Ressources numériques

Analyse du comportement humain en temps réel

Plateau technique « Parole » pour l'enregistrement de locuteurs et le traitement des données.

3. Valorisation

Productions envisagées (revues, ouvrages, base de données...)

Corpus oral anglais contemporain en discours avec mode d'interrogation automatique.

Logiciel de traitement de la parole 'Speech Wave Analyser'.

Publications dans revues.

Colloques et manifestations

Journée PAROLE III (automne 2010), réservée au volet « restitution de la parole ».

4. Nature du soutien demandé à la MSHS

Projet porté par la MSHS (soutien financier, technique et administratif)