

## **De l'opportunité de la création d'un Institut de Recherche sur les Langues du Monde**

*Dans le contexte des évolutions en cours depuis deux ans au CNRS, et à l'instigation de notre direction scientifique adjointe, le CRLAO, le CELIA, le LLACAN, le LMS, le LPP et le LACITO travaillaient à l'élaboration d'un projet de Laboratoire de Recherche Commun (LRC) en sciences du langage. En juin 2008, au moment, pour les quatre premiers nommés de ces laboratoires, de déposer nos projets pour le quadriennal de la vague D, nous apprenions que des incertitudes pesaient sur la poursuite des créations de LRC, mais qu'en tout état de cause, les regroupements ne pourraient se faire que sur la base d'un partenariat commun avec un établissement d'enseignement supérieur, éliminant de facto, les trois derniers laboratoires susnommés. Nous avons donc poursuivi notre réflexion dans ce cadre plus restreint. La Direction des Partenariats nous confirmait en janvier 2009, la suspension des structures LRC mais nous encourageait à continuer dans la voie d'un regroupement.*

En prévision de la création à court terme (en principe à l'horizon 2012) d'un pôle recherche sur le site de Tolbiac 2 intitulé « Pôle de recherche sur les langues et civilisations du monde », auquel sont associées nos tutelles (INALCO, EHESS, CNRS) dans le cadre d'un GIP, nous souhaitons proposer la constitution d'un Institut de linguistique des langues regroupant trois UMR, le CRLAO, le LLACAN, le CELIA, ainsi que les enseignants-chercheurs du CLI initialement prévus dans la proposition d'une UMR intitulée SEDYL<sup>1</sup>. L'Institut serait alors sous la tutelle partagée du CNRS, de l'INALCO, notre tutelle universitaire commune, de l'EHESS, et en partenariat avec l'IRD. Ce laboratoire, regroupant plus de 70 personnels statutaires, chercheurs, enseignants-chercheurs et ITA, serait centré sur la linguistique des langues (terrain, typologie, linguistique historique), avec une large assise géographique et des possibilités d'interfaces vers l'informatique, la cognition, la génétique des populations, l'anthropologie et l'archéologie. Les spécificités thématiques et aréales des trois laboratoires trouveraient ainsi à s'intégrer dans un tel projet. Soulignons que ce projet s'appuie sur une tradition de collaborations qui s'est trouvée renforcée par la création de la Fédération Typologie et Universaux Linguistiques.

La création d'un Institut permettrait donc de fédérer un potentiel important, et reconnu au plan international, en linguistique des langues. Il est conçu pour être largement ouvert à d'autres laboratoires ayant les mêmes méthodes et les mêmes objectifs, tel le LACITO, ainsi qu'à d'autres compétences aréales, qui souhaiteraient, à terme, rejoindre ce projet centré sur la diversité linguistique.

Rappelons que cette diversité se présente comme une donnée de départ de l'étude du langage : avec quelque 6000 langues dans le monde, le langage ne se présente jamais « à nu » mais toujours sous la forme d'une variété. Cette diversité définitoire constitue un enjeu scientifique considérable pour la modélisation du langage et ses implications à la fois cognitives et applicatives. Le champ de la diversité linguistique n'est pas infini mais contenu dans des limites et cette variation est régulée par des principes organisateurs dont la connaissance est nécessaire pour comprendre la nature du langage mais aussi pour pouvoir espérer, par exemple, améliorer la traduction automatique ou comprendre l'impact de la diversité linguistique sur l'acquisition du langage.

---

<sup>1</sup> CRLAO = « Centre de recherches linguistiques sur l'Asie orientale » (UMR 8563 CNRS, EHESS, INALCO) ; LLACAN = « Langage, langues et cultures d'Afrique noire » (UMR 8135 CNRS, INALCO) ; CELIA = « Centre d'études des langues indigènes d'Amérique » (UMR 8133 CNRS, INALCO, Paris 7, IRD) ; CLI = « Cercle de linguistique de l'INALCO » (PPF du Ministère de l'ESR), initiateur du master « Linguistique et diversité des langues » de l'INALCO et de la publication *Faits de langues : les Cahiers* ; SEDYL = « Structure et dynamique des langues » (UMR proposée par le CELIA et le CLI).

Les compétences actuelles des chercheurs, enseignants-chercheurs et doctorants du futur Institut couvrent plus de 130 langues appartenant à 22 phylums linguistiques :

### Liste des langues par phylum

AFRO-ASIATIQUE : afar, amharique, arabe, bata, bedja, berbère, dahlik, hakmi, haoussa, hobyot, kotoko, maltais, mbara, mehri, munjuk, saho, soqotri, somali, tigre, zaar - ALTAÏQUE et FINNO-OUGRIEN : finnois, mongol, dongxiang, manchou - ARAWAK : wayuu, añun, arawak-lokono - AUSTRONESIEN : malgache, tagalog - CARIBE : kali'na, wayana - CHOCO : embera - COREEN : coréen - CREOLES : capverdien, casamançais, palenquero, ndyuka, pamaka - HMONG-MJEN : miao-yao - GUAHIBO : sikuni - INDO-EUROPÉEN : arménien, grec moderne, hindi, letton, lituanien, polonais, roumain, russe, singhalais - JAPONIQUE : japonais (ancien, moderne), langues ryukyu - KATUKINA : katukina-kanamari - MACRO-GE : xavante, rikbaktsa - MAYA : maya - MIXTES : afro-brésilien - NIGER-CONGO : aiki, aka, bainunk, bak (langues), balante, bambara, banda, bagiro, bangala, bayot, bijogo, chamba, niellim, defoïdes (langues), fongbe, fulfulde du diamaré, gbaya bodeo, gbaya kara, gbaya-manza-ngbaka, gengbé, kabiye, kimbundu, koalib, lehar, ngbaka minagende, ngbugu, pepel, peul, sango, sotho, swahili, tira, tupuri, wolof, yakoma, yendang, yoruba, zande, zulu - NILO-SAHARIEN : afiti, bongo, daajo, debri, fer, gula, méroïtique nara, yulu, nubien, nyimang, runga - QUECHUA : quechua - SINITIQUE : chinois ancien, chinois médiéval, chinois moderne, mandarin, min méridional, gan-hakka, cantonais, xiang, wu, tangwang - SINO-TIBÉTAÏN et TIBÉTO-BIRMAN : bai, tibétain, baima, shixing, rgyalrong (japhug), tangoute - TUPI : nhenngatu, wayampi - LANGUES NON CLASSEES : laal, guambiano, purepecha.

Le projet d'Institut comprend **3 équipes** :

- Equipe 1. Description, documentation et typologie linguistique (34 pers.)
- Equipe 2. Histoire des langues et changement linguistique (16 pers.)
- Equipe 3. Oralité, écriture et politique linguistique (11 pers.)

et **4 centres aréaux** transversaux (voir les rapports des laboratoires concernés) :

- Afrique Noire (LLACAN <http://llacan.vjf.cnrs.fr/>)
- Amérique (CELIA <http://celia.cnrs.fr/Fr/Celia.htm>)
- Asie Orientale (CRLAO <http://crlao.ehess.fr/>)
- Eurasie (CLI – PPF INALCO – [http://a.donabedian.free.fr/index\\_cli.htm](http://a.donabedian.free.fr/index_cli.htm))

Au plan technologique, l'Institut s'articulerait à une **plate-forme numérique** de corpus et bases de données (textuels, oraux, lexicaux, thématiques) déjà constitués ou en cours de constitution et dont certains sont déjà accessibles en ligne.

Les trois directeurs actuels (Redouane Djamouri, Francesc Queixalos, Martine Vanhove) ainsi que la directrice pressentie pour le projet d'UMR SEDYL (Anaïd Donabédian) assureront collégialement le montage et le suivi du projet jusqu'à son éventuelle acceptation.

L'**organigramme** prévu est le suivant (voir aussi l'annexe) :

- 1 directeur (Martine Vanhove, DR CNRS LLACAN)
- 1 directeur adjoint (Hilary Chappell, DE EHESS CRLAO)
- 1 secrétaire général (Jeanne Zerner, IE CNRS LLACAN)
- 3 secrétaires-gestionnaires (Magali Diraison, AI CNRS LLACAN, Hugues Feler, T CNRS CRLAO + 1 recrutement)
- 1 ingénieur informaticien (Christian Chanard, IE CNRS LLACAN)

+ 1 IE ingénieur informaticien à recruter (celui du LLACAN est déjà occupé plus qu'à plein temps par les projets du laboratoire)

1 technicien en saisie de bases de données (Benoît Legouy, T CNRS Llacan)

2 documentalistes (Michèle Abud et Ying Cheng, IE EHESS, CRLAO)

1 secrétaire de rédaction (Suzanne Chane-Kon, IE EHESS, CRLAO)

1 coordinatrice des projets en Guyane (Duna Troïani, IE CNRS CELIA)

1 coordinateur des projets en Afrique du Sud (Michel Lafon, IE CNRS LLACAN)

1 coordinateur des projets au Soudan (Pierre Nougayrol, IR CNRS LLACAN)

1 coordinateur plate-forme numérique maya (Jean-Michel Hoppan, IE CNRS CELIA)

4 responsables des centres aréaux : Nicolas Quint (Afrique Noire, LLACAN), Claudine Chamoreau (Amérique, CELIA), Redouane Djamouri (Asie orientale, CRLAO), Anaïd Donabédian (Eurasie, CLI).

Le nombre des **personnels statutaires** qui composeront cette future unité s'élève à **73** (35 chercheurs, 26 enseignants-chercheurs, 12 ITA et BIATOSS) auxquels s'ajoutent à l'heure actuelle 58 doctorants (LLACAN 24, CRLAO 20, CELIA 9, INALCO 5). Les personnels statutaires concernés sont les suivants :

LLACAN (28) :

- 16 chercheurs CNRS : Raymond BOYD (CR), Pascal BOYELDIEU (DR), Bernard CARON (DR), Fathi DEBILI (DR), Marcel DIKI-KIDIRI (CR), Yves MOÑINO (DR), Didier MORIN (CR), Elsa OREAL (CR), Nicolas QUINT (CR), Claude RILLY (CR), Stéphane ROBERT (DR), Paulette ROULON-DOKO (DR), Guillaume SEGERER (CR), Marie-Claude SIMEONE-SENELLE (DR), Henry TOURNEUX (DR), Martine VANHOVE (DR)

- 6 enseignants-chercheurs : Ursula BAUMGARDT (Pr INALCO), Jean DERIVE (Pr émérite), Gérard DUMESTRE (Pr INALCO), Sylvie GRAND'EURY (MCF Univ. Nancy), Aliou MOHAMADOU (Pr INALCO), Konstantin POZDNIAKOV (Pr INALCO)

- 6 ITA CNRS : Christian CHANARD (IE), Magali DIRAISON (AI), Michel LAFON (IE), Benoît LEGOUY (T), Pierre NOUGAYROL (IR), Jeanne ZERNER (IR)

CRLAO (25) :

- 12 chercheurs CNRS : Raoul BLIN (CR), Françoise BOTTERO (CR), Ekaterina CHIRKOVA (CR), François DELL (DR émérite), Redouane DJAMOURI (CR), Robert ILJIC (DR), Chrystelle MARECHAL (CR), Barbara NIEDERER (CR), Waltraud PAUL (CR), Alain PEYRAUBE (DR), Laurent SAGART (DR), Thekla WIEBUSCH (CR)

- 9 enseignants-chercheurs : Anton ANTONOV (MCF INALCO), Hilary CHAPPELL (DE EHESS), Zhitang DROGOURT-YANG (MCF INALCO), FENG Li (MCF INALCO), Monique HOA (MCF Univ. Paris VII), Guillaume JACQUES (MCF Univ. Paris V), Sumikazu NISHIO (MCF INALCO), QI Chong (MCF Univ. Paris VII), XU-SONG Dan (Pr INALCO)

- 4 ITA : Michelle ABUD (IE EHESS), Suzanne CHANE-KON (IE EHESS), Cheng YING (IE EHESS), Hugues FELER (T CNRS)

CELIA (11) :

- 4 chercheurs CNRS : Claudine CHAMOREAU (CR), Isabelle LEGLISE (CR), Marie-France PATTE (CR), Francisco QUEIXALOS (DR)

- 3 chercheurs IRD : Odile RENAULT-LESCURE (CR), Sophie ALBY (CR1 en accueil IRD), Pascal VAILLANT (CR1 en accueil IRD)

- 2 enseignants-chercheurs : André Cauty (Pr Univ. Bordeaux I), Bernard POTTIER (Pr émérite),

- 2 ITA : Jean-Michel HOPPAN (IE CNRS), Duna TROÏANI (IE CNRS)

Cercle de Linguistique de l'INALCO (9) :

- 9 Enseignants-Chercheurs : Fida BIZRI (MCF), Christine BONNOT (Pr), Tatiana BOTTINEAU (MCF), Anaïd DONABEDIAN (Pr), Outi DUVALLO (MCF), Alexandru MARDALE (MCF), Henri MENANTAUD (MCF), Annie MONTAUT (Pr), Sophie VASSILAKI (Pr)

**Locaux à prévoir sur le site du « Pôle de recherche sur les langues et civilisations du monde » :**

Surface totale nécessaire : 1118 m<sup>2</sup>

Bureaux : 73 x 9 m<sup>2</sup> = 657 m<sup>2</sup>

Bureaux pour 6 invités : 3 x 12 m<sup>2</sup> = 36 m<sup>2</sup>

1 salle de réunion = 100 m<sup>2</sup>

3 salles d'archivage = 300 m<sup>2</sup>

1 salle de bibliothèque pour les fonds documentaires nécessaires aux recherches en cours et pour les ouvrages de référence = 100 m<sup>2</sup>

1 espace détente = 50 m<sup>2</sup>

1 salle de reprographie (photocopieuse, fax, relieuse...) = 15 m<sup>2</sup>

5 salles de 30 m<sup>2</sup> pour les doctorants = 150 m<sup>2</sup>

## **Contenu scientifique des équipes**

### **Equipe 1 : Description, documentation et typologie linguistique**

Responsables : Claudine Chamoreau, Anaïd Donabédian

La description, la documentation et la typologie linguistique ont pour but, d'une part de contribuer à l'inventaire de la diversité linguistique (y compris, et en urgence, celle des langues en voie de disparition) qui constitue un préalable incontournable à toute généralisation sur les propriétés du langage humain, d'autre part à établir une typologie aréale permettant de définir les apparentements structurels (ou les divergences) entre les langues d'une même aire géographique.

Les travaux de l'équipe 1 visent donc (1) à contribuer à la collecte sur le terrain de la diversité linguistique dans un but de documentation, de connaissance et de préservation, (2) à fonder sur cette documentation fine et diversifiée, un travail de généralisation et de modélisation sur le langage appuyé sur le cadre théorique de la typologie linguistique. L'ambition qui réunit ces chercheurs est de définir le lieu et la portée de la variation linguistique et de la diversité culturelle à l'intérieur des mécanismes généraux qui caractérisent le langage par rapport aux autres capacités cognitives. En outre, l'équipe 1 doit participer à l'élaboration de *ressources numériques* pour la linguistique : bases de données, corpus en ligne exploitables pour l'analyse.

Pour ce quadriennal, l'équipe 1 organise son travail en 3 axes dans lesquels sont repris les projets scientifiques développés initialement dans les dossiers des unités constitutives. Dans le cadre du nouvel Institut, cependant, les projets qui ont initialement une portée aréale s'ouvriront aux autres aires de l'équipe. Ces trois axes étudieront dans une perspective typologique trois aspects complémentaires des composants linguistiques : (1) la nature et l'organisation des catégories, (2) les constructions syntaxiques, (3) les interactions entre les catégories et les niveaux structurels dans le fonctionnement des langues (axe Complexité). Cette équipe s'appuie sur un champ de compétences qui permettra de couvrir un vaste ensemble d'aires linguistiques : Afrique, Amérique, Asie, Europe.

### **AXE 1 : CATEGORIES** (Responsables : Hilary Chappell, Stéphane Robert)

La question de la catégorisation est centrale dans cette approche : par nature, les systèmes linguistiques *sélectionnent* en effet certaines dimensions de l'information qu'ils *encodent* de manière variable dans différentes catégories linguistiques. La catégorisation opérée conditionne alors le fonctionnement des langues et les processus de construction du sens. Elle est donc intéressante scientifiquement, d'une part pour définir les traits retenus par les langues dans cette sélection (ce qui est conçu comme saillant dans une langue et suffisant pour permettre l'accès à la référence), ainsi que le système qu'ils forment dans chaque langue et l'impact cognitif de ce choix ; et d'autre part, pour étudier le mode de construction du sens dans l'énoncé à partir de cette sélection. On se concentrera pour cette étude des catégories sur les langues africaines, le chinois et le japonais. On s'intéressera plus particulièrement dans cet axe à :

- (1) la question de la modélisation de la description grammaticale, avec pour objectifs de résoudre les tensions entre des modèles permettant la comparaison et la généralisation d'un côté, et les spécificités des langues ou familles de langue décrites, de l'autre, et de construire des descripteurs morphosyntaxiques pour l'analyse automatique,
- (2) l'étude de la catégorie du temps (dans quelles parties du discours est-elle implémentée, s'agit-il d'une catégorie pure ou est-elle amalgamée à d'autres catégories ? Comment s'organisent les systèmes temporels à travers les langues ? Leur typologie est-elle liée aux familles linguistiques ou marquée par des effets aréaux ?),
- (3) pour ce qui est de la sémantique lexicale, à la morphologie du lexique du chinois et du sino-japonais et, en particulier, à la question de la compositionnalité sémantique,

### **AXE 2 : SYNTAXE** (Responsables : Waltraud Paul, Sophie Vassilaki)

Les structures syntaxiques représentées dans les diverses langues du futur Institut fournissent un échantillon remarquable de la diversité des langues du monde, et permettent d'aborder des questions-clés de la linguistique théorique et de la typologie des langues. La composition du laboratoire permet que se rencontrent des chercheurs formés à des approches théoriques variées (typologie, fonctionnalismes, grammaires formelles, théorie de l'énonciation, etc.) et que soient confrontés les résultats issus de ces différentes approches, tous appuyés sur des données validées (terrain ou corpus, notamment dans le cadre de la plateforme de ressources commune au laboratoire).

Dans cet axe, on abordera l'étude de la syntaxe à différents niveaux structurels : de l'étude de la structure argumentale de la phrase simple à celle de la hiérarchisation des prédicats dans la phrase complexe. Les différentes approches mises en œuvre doivent permettre, d'une part, d'établir une nouvelle classification pour les groupes de langues étudiés dans une perspective typologique (langues sinitiques et tibéto-birmanes), d'autre part, de fournir une analyse des différents composants impliqués dans la structuration syntaxique (structure argumentale et constructions syntaxiques, construction de la continuité référentielle et des anaphores, rôle des TAM dans la hiérarchisation des prédicats) et de leurs effets sémantiques et pragmatiques, afin de comparer les systèmes des différents groupes de langues étudiés (langues amérindiennes, langues africaines, langues sinitiques, japonais, coréen, langues d'Europe et de l'Inde appartenant à différents domaines linguistiques). Enfin, ces travaux doivent permettre de construire un analyseur syntactico-sémantique pour le japonais et le coréen.

### **AXE 3 : COMPLEXITE** (Responsables : Anaïd Donabedian, Annie Montaut)

Les travaux de cet axe portent sur le langage conçu comme un *système complexe* marqué par un emboîtement de niveaux d'organisation qui présentent des caractéristiques telles que les propriétés d'un niveau ne sont pas strictement déductibles à partir de celles du niveau

inférieur. Les recherches portent en particulier sur les phénomènes d'interaction entre les différents niveaux structurels et leur inter-dépendance.

On s'intéressera ainsi aux différents composants linguistiques dans lesquels est implémentée une dimension linguistique encore mal étudiée, celle de la saillance : en effet, l'information linguistique véhiculée par les énoncés n'est pas seulement marquée par une organisation morphologique et syntaxique mais également par l'indication d'une saillance attentionnelle variable sur ses différents composants. Cette dimension pragmatique ne recouvre pas seulement la question classique du focus et de la structure informationnelle de la phrase, mais aussi différentes catégories sémantiques qui peuvent être implémentées dans des composants morphosyntaxiques divers. Ainsi, un pronom de première personne peut être considéré comme plus saillant qu'un pronom de troisième personne, une conjugaison à valeur mirative a une force interlocutive plus grande qu'un simple inaccompli, tout comme la définitude, le trait +humain ou +animé peuvent être considérés comme plus saillants que l'indéfini, le trait -humain ou -animé. Il s'agira donc ici de construire une définition unifiée de cette catégorie de saillance et de décrire son fonctionnement dans une perspective inter-linguistique.

On étudiera également la cohérence des systèmes linguistiques et l'inter-dépendance entre certains types morphologiques et les règles syntaxiques des langues concernées. On commencera par l'étude des langues agglutinantes (comme l'arménien) et le lien entre leurs structures et l'absence du marquage de l'accord (coindexation). On cherchera à établir la validité d'un universal implicatif qui lie l'absence de genre grammatical au type agglutinant et à l'étendre au phénomène plus général de la rection.

Enfin, l'axe Complexité étudiera les interactions entre le niveau syntaxique et le niveau prosodique à partir d'une analyse instrumentale de corpus oraux dans diverses langues, afin d'établir une typologie de l'articulation entre syntaxe et prosodie. Il est prévu pour cela de travailler sur deux types de corpus : l'un, élaboré par des spécialistes de la phonologie de laboratoire et produit par élévation, l'autre constitué de textes oraux spontanés. La comparaison entre ces deux corpus doit permettre de déterminer :

- quelles sont les unités de l'oral spontané (par opposition à celles de l'écrit ou obtenues par élévation). Sont-elles comparables d'une famille de langues à l'autre, d'une langue à l'autre ? Existe-t-il réellement des universaux en la matière ?
- Ces unités sont-elles de nature différente selon les systèmes prosodiques (accentuel/tonal) ?
- Comment s'articulent les unités prosodiques et les unités morphosyntaxiques, en particulier au niveau de la structure de l'information.

Les corpus des différentes langues étudiées annotés, indexés et découpés en unités prosodiques seront mis en ligne à l'issue du projet.

## **Equipe 2 : Histoire des langues et changement linguistique**

Responsables : Guillaume Jacques, Isabelle Léglise

La diversité linguistique résulte naturellement en partie de diversifications anciennes et continues, renforcées par un morcellement accentué des groupes ethniques. Mais, et le paradoxe n'est qu'apparent, elle résulte également, comme le montrent certains travaux récents (Thomasson et Kaufmann 1988, Nicolai 1990, Manessy 1990, 1995, Lefebvre 2004), d'importants phénomènes de contacts, qui viennent sans cesse remodeler l'identité des langues, au point parfois de remettre sérieusement en cause leur filiation 'classique'. L'influence que peuvent exercer différentes communautés les unes sur les autres est aussi déterminante pour ces interférences que la pratique individuelle des locuteurs qui maîtrisent, souvent, plusieurs systèmes linguistiques auxquels ils ont recours selon leurs interlocuteurs.

Cet état de fait détermine les deux approches, nécessairement complémentaires, qui structurent les travaux de recherche qui seront menés dans l'Equipe 2 et qui se développeront selon trois axes:

- Axe 1. Mécanismes et motivations des changements morphosyntaxiques
- Axe 2. Reconstruction phonologique, morphologique, lexicale, parenté des langues
- Axe 3. Contact des langues, changement linguistique et changements sociaux

## **AXE 1 : MÉCANISMES ET MOTIVATIONS DES CHANGEMENTS**

### **MORPHOSYNTAXIQUES** (Responsable : Redouane Djamouri)

Cet axe de recherche rassemblera des spécialistes de linguistique historique travaillant sur les changements morphosyntaxiques. On s'intéressera notamment aux mécanismes majeurs du changement grammatical que sont l'analogie et la réanalyse. Le changement analogique commande la création ou la modification d'une forme à l'image d'une autre plus régulière pour maintenir l'équilibre du système linguistique (*exemplar-based analogical changes* dans la formulation de Kiparsky). Ce type de changement comprend la « dégrammaticalisation » qui correspond la plupart du temps à une lexicalisation et qui contredit le principe d'unidirectionnalité généralement associé à la grammaticalisation. Le changement par réanalyse comprend quant à lui la grammaticalisation proprement dite, à savoir, comme le formulait Meillet (1912), l'« *attribution du caractère grammatical du caractère grammatical à un mot jadis autonome* » mais aussi les cas d'« exaptation » où l'on a affaire à l'apparition et l'évolution d'une forme grammaticale complexe à partir d'une forme initiale simple.

Au-delà des mécanismes, on étudiera aussi les motivations du changement. En ce qui concerne le changement analogique, plusieurs facteurs motivants seront examinés. Considérant un changement du type A > B, on s'intéressera à la complexité/simplicité relative de A et de B (*pull/push model*). On intégrera par ailleurs dans les motivations d'un tel changement les facteurs sémantico-pragmatiques (extension métaphorique).

En ce qui concerne les motivations pour la réanalyse, on inclura, outre les facteurs sémantico-pragmatiques (telles l'extension anaphorique, la métonymisation et la subjectification), les facteurs liés à une exigence structurale mais aussi ceux résultants d'une tension typologique.

Pour ce qui est des changements dus aux évolutions phonologiques et ceux dus au contact entre langues ils seront traités plus spécifiquement dans le cadre des deux autres axes de cette équipe.

Le but principal des recherches qui seront menées au sein du groupe est de vérifier nombre d'hypothèses concernant les changements morphosyntaxiques, aussi bien dans le cas concret de langues données qu'en linguistique générale. Il s'agit d'examiner les états antérieurs de diverses langues avec la méthodologie développée pour l'analyse en synchronie et de tenir compte du contexte structural dans lequel les changements ont lieu, afin de dégager aussi précisément que possible les contraintes portant sur ces changements

Divers phénomènes pourront être utilement appréhendés : le changement d'ordre des constituants, la réanalyse de termes lexicaux en items grammaticaux, les cas de dégrammaticalisation et d'exaptation, la simplification ou la complexification des structures...

Cet axe réunira des spécialistes de plusieurs domaines linguistiques : pour le domaine asiatique, des spécialistes du chinois et de langues sino-tibétaines, pour le domaine africain, des spécialistes de langues atlantiques, oubanguiennes, méroïtique, soudaniques et sara-bongo-baguirmiennes.

**AXE 2 : RECONSTRUCTION PHONOLOGIQUE, MORPHOLOGIQUE, LEXICALE, PARENTÉ DES LANGUES** (Responsables : Laurent Sagart, Pascal Boyeldieu)

L'histoire des langues est le produit de la superposition de deux mécanismes distincts : transmission verticale, par filiation, et transmission horizontale, par contact. La linguistique historique a mis au point des méthodes qui permettent de distinguer les effets de ces deux mécanismes dans les langues particulières, notamment l'établissement de correspondances phonétiques régulières entre mots apparentés de langues différentes. L'importance respective de ces deux mécanismes varie selon les situations.

Dans cet axe de recherche, nous nous intéressons plus particulièrement au mécanisme de transmission par filiation (nous sommes conscients que l'évolution des langues ne se réduit pas à sa dimension verticale, qui peut dans certains cas jouer un rôle secondaire), à travers l'étude détaillée de plusieurs familles de langues, principalement en Asie orientale et en Afrique, et dans ses diverses dimensions : phonologique, morphologique et lexicale.

Réduite à sa dimension verticale, l'évolution et la diversification des langues peuvent être saisies par un modèle arborescent, qui définit les relations de parenté au sein d'un même groupe.

En inversant le sens des changements postulés à l'intérieur d'un groupe de langues, il est possible de reconstruire des éléments (de la phonologie, de la morphologie et du lexique) du dernier ancêtre commun au groupe de langues modernes considéré. Il est possible ensuite de déduire du vocabulaire reconstruit des informations importantes sur la localisation de la langue reconstruite dans l'espace et dans le temps, ainsi que le mode de vie de ses locuteurs. Ceci nous donne la possibilité de contribuer, avec l'archéologie, la génétique des populations, et la génétique des plantes et animaux domestiques, à l'effort interdisciplinaire d'élucidation de l'histoire non écrite du peuplement humain.

**OBJECTIFS**

Etude des relations entre groupes de langues d'Afrique et d'Asie : contact ou parenté (en particulier coréen et japonais, turc, mongol et mandchou, sino-tibétain et austronésien, austronésien et tai-kadai, relations au sein des ensembles niger-congo et nilo-saharien). Elucidation de l'histoire de ces groupes linguistiques et de leur généalogie. Progrès dans la reconstruction des proto-langues ; datation, localisation, mode de vie en liaison avec d'autres disciplines.

Nous chercherons aussi à contribuer à améliorer les méthodes de la discipline :

- Etablissement de l'inventaire, de la distribution et de la fréquence des changements phonétiques attestés dans les langues dont nous sommes spécialistes.
- Utilisation des méthodes statistiques, pour :
  - a) permettre de déterminer si des ressemblances entre plusieurs langues sont explicables par le hasard ou demandent une autre hypothèse (contact ou parenté)
  - b) servir à analyser les déséquilibres distributionnels dans le système phonologique, ce qui peut contribuer à la reconstruction interne d'une famille de langues.
- Meilleure prise en compte des changements internes par analogie (en s'inspirant notamment des travaux sur l'indo-européen).

**AXE 3 : CONTACT DES LANGUES, CHANGEMENT LINGUISTIQUE ET CHANGEMENTS SOCIAUX** (Responsables : Isabelle Léglise, Claude Rilly)

Suivant une tradition linguistique historique intéressée par le changement linguistique et la comparaison (des langues et des situations), les linguistiques du contact – contact linguistics – (Thomason, 2001, Winford, 2003) proposent des typologies des situations de contact s'appuyant sur des paramètres tels que le type de situation concerné (maintien ou perte de la langue), les facteurs sociaux (durée et intensité du contact, degré de bilinguisme), les



mécanismes (emprunt, transfert) et résultats du contact (variations, changement, réorganisation au sein d'un système, etc. Heine & Kuteva, 2005). Si la plupart des travaux avec une orientation typologique a été réalisée en diachronie, l'essentiel des travaux en synchronie concerne l'alternance de langues et l'étude des parlors bilingues (cf. Poplack, 1980, Myers-Scotton, 1993, Muysken, 2000).

On sait que le changement linguistique, en situation de contact de langues, a presque toujours des causes multiples (Thomason, 2007). Toutefois, il a encore rarement été tenu compte de cet aspect pluri-factoriel dans les explications des phénomènes de contact qui sont proposées dans la littérature. Une exception notable concerne les travaux sur le changement linguistique, en diachronie, qui, pour rendre compte de modifications (phonétiques ou syntaxiques) advenues s'appuient d'une part sur des tendances internes aux langues et, d'autre part, citent souvent le « contexte socio-historique » qui permet de justifier un potentiel effet du « contact de langues ». Si ces différents facteurs sont mentionnés, ils ne sont que trop rarement intégrés, ensemble, dans l'explication de l'évolution des langues. Les travaux descriptifs, en synchronie, laissent quant à eux peu la place à des facteurs explicatifs contextuels – se concentrant sur des facteurs linguistiques ou typologiques. Quant aux travaux sociolinguistiques, ils ne s'intéressent généralement pas à expliquer les conséquences linguistiques du contact de langues – mais se concentrent plutôt sur le contexte ou l'usage.

Les objectifs de notre programme sont de contribuer à mieux expliquer le changement linguistique (en train de se faire ou advenu) en situation de contacts de langues, en Amérique, en Asie du Sud et en Afrique, en (i) complexifiant le niveau de prise en compte des « facteurs sociaux », en tirant profit des travaux anthropologico-linguistiques, pragmatiques et sociolinguistiques, (ii) affinant les « facteurs linguistiques » pris en compte dans la littérature, en s'appuyant en particulier sur les travaux en typologie linguistique, (iii) articulant enfin ces différents facteurs explicatifs.

Les moyens mis en œuvre concernent :

- a) la réalisation d'une méthode de travail commune, pluri-factorielle. Les données, issues de situations de contact actuelles, linguistiquement productives, seront abordées à travers cinq niveaux d'analyse (lexical, morphosyntaxique, pragmatique, sociolinguistique, typologique),
- b) la réalisation d'outils informatiques dédiés. Au-delà des fonctions de recherche classiques sur corpus, un certain nombre d'outils de repérage spécifiques pouvant être appliqués sur les corpus hétérogènes seront créés, comme par exemple : le repérage de patrons de structures alternantes, de patrons de calques, de variations dans l'ordre des constituants, etc. (approches émergentes dans le domaine de la classification automatique ou *manifold learning*).

### **Equipe 3 : Oralité, écriture et politique linguistique**

Responsable : Ursula Baumgardt

L'équipe 3 propose un travail sur les problèmes d'utilisation de la langue dans des contextes multiples, et dans la perspective d'une collaboration pluridisciplinaire rapprochant notamment la linguistique, la littérature, l'anthropologie, la géographie et l'histoire. Il s'agit de définir plusieurs types de contextes qui seront abordés selon trois problématiques :

- l'influence institutionnelle sur les usages de la langue ;
- les théories de l'écriture ;
- le rôle de l'environnement territorial et social dans la pratique discursive.

- Dans le cadre de la première problématique, sera étudiée la question de la régulation des pratiques langagières par des institutions historiques ou contemporaines, nationales ou supra-

nationales, qui mènent ou ont mené des politiques linguistiques différentes susceptibles d'interagir avec ces pratiques, notamment à propos de la transmission et de la systématisation des savoirs, y compris dans le cadre traditionnel ou historique.

- La deuxième problématique sera centrée sur l'étude de l'épigraphie et de la lexicographie des langues concernées.

- Enfin, la troisième problématique, portant principalement, mais non exclusivement, sur l'oralité, analysera les discours suivant le lieu et l'énonciateur.

L'équipe sera donc partagée en trois axes, chacun articulé à deux thématiques :

• **AXE 1 : ETUDE DES POLITIQUES LINGUISTIQUES :**

1. Pratiques langagières et politiques linguistiques

2. Aménagements linguistiques et transmission.

• **AXE 2 : ECRITURE ET COGNITION :**

3. Epigraphie et déchiffrement

4. Classification et lexicographie

• **AXE 3 : ANALYSE DES DISCOURS LITTÉRAIRES :**

5. Textes et territoires

6. Autour de l'énonciateur : voix, figure, posture.

• **AXE 1 : ETUDE DES POLITIQUES LINGUISTIQUES** (Responsables : Sophie Alby, Michel Lafon)

Cette thématique traverse l'ensemble des aires linguistiques couvertes par le laboratoire.

Cet axe se propose d'examiner, à partir de plusieurs études de cas, (1) les politiques linguistiques menées par certains Etats, en particulier dans l'éducation, pour le développement des langues concernées en Afrique, dans les Balkans, en Chine et en Guyane, avec une participation ponctuelle des spécialistes de singhalais et d'arménien ; dans ce cadre seront aussi étudiées les interactions entre pratiques langagières et politiques linguistiques ; (2) de développer, pour deux langues africaines, une plate-forme expérimentale pour la reconnaissance automatique de la parole.

Sera également étudié le choix des langues de références (nationales et officielles), de la ou des variantes dans les processus de standardisation, de fixation des références normatives (élaboration des systèmes d'écriture et des orthographe). La standardisation linguistique s'effectue par étapes et met en jeu plusieurs notions qui renvoient à l'ensemble des opérations de définition et de formation d'une langue commune (koinè) dont le code de référence est l'écrit, au service de l'éducation et de l'administration d'un Etat (moderne ou ancien). On sait par exemple qu'une étape fondamentale dans le processus de standardisation représente la constitution d'une variété "légitime", à partir des différentes variétés effectivement parlées par les locuteurs. Cette variété va servir de fondement sur lequel sera édifée, codifiée et diffusée la nouvelle langue, avec la mise en place d'un système d'écriture et de règles d'orthographe et l'élaboration d'outils linguistiques de référence (grammaires, dictionnaires, systèmes de reconnaissance automatique).

Enfin, nous aborderons la problématique de la transmission des connaissances, dans les domaines formel et informel. On se propose d'envisager de quelle façon certaines expériences éducatives en Afrique tentent d'établir un pont entre les domaines de la modernité et les techniques et connaissances traditionnelles, par le recours aux langues africaines. Cela implique à la fois leur passage à l'écrit mais aussi la prise en compte des manifestations et modes de transmission oraux. Il s'agit donc d'examiner les conditions d'un aménagement linguistique raisonné, basé sur les réalités locales.

• **AXE 2 : ECRITURE ET COGNITION** (Responsable : Françoise Bottero)

Les langues chinoises et mayas s'écrivent toutes au moyen de systèmes graphiques mixtes originaux, combinant des éléments sémantiques et phonétiques qui font intervenir un certain nombre de mécanismes cognitifs qui n'ont pas leur raison d'être dans les systèmes dits purement phonétiques (alphabets, syllabaires). Elles présentent de surcroît l'avantage de posséder de grands corpus d'écrits anciens sur lesquels fonder l'analyse.

Cet axe s'intéressera, premièrement, à des questions liées à l'épigraphie et au déchiffrement. L'approche épigraphique combinée aux données linguistiques, historiques et archéologiques permettra, d'une part, d'étudier les mécanismes généraux et la variation qui sont à la base de la formation, de l'utilisation et de l'évolution des systèmes graphiques, et d'autre part, de procéder à l'identification des caractères et variantes archaïques. Une attention particulière sera portée aux contextes culturels dans la création des graphies pour dégager les motivations non phonétiques à la base des changements graphiques.

La seconde problématique de cet axe concerne les systèmes de classification des caractères et des éléments sémantiques dans les dictionnaires. Elle a pour objectifs de comprendre la conception de l'organisation du monde sous-jacente aux systèmes et de dégager les principes d'organisation sémantiques et cognitifs qui sont implicites dans l'écriture.

Cet axe développera en parallèle une interaction avec les comparatistes et spécialistes de reconstruction de l'équipe 2 du laboratoire.

Pour mener à bien ses recherches, l'axe 2 s'appuiera sur plusieurs bases de données existantes pour le chinois et le maya : le vaste corpus de textes chinois et de variantes graphiques du *Thesaurus linguae sinicae* et sa traduction analytique en anglais (développé par Christoph Harbsmeier : [tls3.uni-hd.de](http://tls3.uni-hd.de)), la base de données pour l'étude de l'écriture, des dictionnaires et la philologie chinoise (interrogeable en ligne, développée au CRLAO), le dictionnaire numérique des signes de l'écriture maya (développé au CELIA).

Des ouvertures vers d'autres écritures (île de Pâques, hiéroglyphes égyptiens, méroïtiques) sont envisagées afin d'entreprendre une réflexion plus générale sur les théories de l'écriture.

• **AXE 3: ANALYSE DES DISCOURS LITTÉRAIRES** (Responsable : Ursula Baumgardt)

Cette problématique concerne principalement mais non exclusivement, les productions discursives en Afrique. Celles-ci sont souvent abordées isolément, soit du point de vue de l'oral ou de l'écrit, soit en termes de "passage" de l'un à l'autre, sans prendre en considération de façon globale la complexité des processus de leur genèse.

Nous proposons de travailler sur les productions verbales tant orales qu'écrites à partir d'une mise en variation de la notion de contexte discursif. La notion de contexte sera entendue à la fois du point de vue de la séquence énonciative, dans le sens de son inscription textuelle et intertextuelle, et du point de vue de la production et du produit discursifs, dans leurs implications culturelles, sociales, historiques, géographiques.

Deux types de contexte seront privilégiés dans deux pôles distincts mais complémentaires : l'articulation du texte et du territoire, ainsi que sa production par l'énonciateur.

Dans la première thématique, textes et territoires, la notion de contexte d'énonciation est prise dans sa dimension géographique et spatiale. L'identification de lieux pour les productions verbales et l'examen de leurs modalités d'influence sur les discours seront au cœur de nos analyses. Les lieux seront pris comme entités énonciatives réalisant l'agencement de productions discursives diverses. On s'interrogera sur la possibilité de considérer les lieux comme des agents énonciatifs pertinents, susceptibles de déclencher des dynamiques discursives ou d'infléchir les formes de discours.

La deuxième thématique est construite autour de la notion d'énonciateur. Le discours est alors compris comme une interaction langagière accomplie dans une situation d'énonciation

impliquant des participants, une institution, un lieu, un temps et les contraintes d'une langue donnée.

L'analyse des discours littéraires sera approchée en privilégiant les énonciateurs en tant que personnes physiques (auteur/performateur), mais également par rapport à leur figuration dans le discours, à leur degré d'adhésion à celui-ci et dans leur interaction avec les destinataires (par exemple : les délégations de l'énonciateur au travers de l'étude des modélisateurs, des discours direct, indirect et rapporté ; la deixis ; l'examen des genres à énonciation plurielle et des modalités d'existence des polyphonies intradiscursives ; la prise en charge des discours par les énonciateurs et leur portée idéologique ...).

Cette thématique s'appuie sur des réflexions menées dans le domaine de la linguistique discursive, énonciative et textuelle.