

Vague B (2012-2015)  
Unité de recherche : dossier unique

PROJET  
UMR 6039 "Bases, Corpus, Langage"  
(Partie I : Projet scientifique)

## Table des matières

### I. Projet du laboratoire

Résumé .....	4
Summary .....	5
<b>1. Auto-analyse.....</b>	<b>6</b>
1.1. Rythme de changement soutenu.....	6
Installation physique et géographique .....	6
Périmètre scientifique .....	6
Restructuration de la charpente des équipes .....	6
Equipe émergente: LiDida .....	6
Nouvelle équipe: Langage et Cognition .....	7
Départs à la retraite, pyramide d'âge.....	8
Qualité et quantité de la production scientifique .....	8
Vie de laboratoire, politique de chercheur sur site .....	8
1.2. Points forts .....	9
Cohérence et signature niçoise .....	9
Leader national en dialectologie et en lexicométrie .....	9
Consolidation de la dialectologie.....	9
Attractivité .....	9
Doctorants relativement peu nombreux mais bien placés.....	9
Fonctionnement collectif, consensuel et réglémenté .....	9
Dynamique de renouvellement .....	10
Nombre de HDR soutenues, potentiel à venir .....	10
1.3. Points faibles .....	10
Pyramide d'âge.....	10
Faibles effectifs dans trois équipes .....	11
Peu d'activité sur programmes financés, absence de projets ANR (où BCL est porteur).....	12
1.4. Risques .....	13
Thématiques peu porteuses dans un "environnement piloté" .....	13
1.5. Opportunités.....	13
Soutien de l'UNS.....	13

<b>2. Projet et objectifs scientifiques</b> .....	<b>15</b>
2.1. Synopsis objectifs.....	15
2.2. Première priorité: consolidation de la logométrie .....	15
2.3. Développement de deux activités applicatives.....	15
2.4. Intégration des psychologues .....	16
2.5. Maintenir la cohérence malgré la désunion géographique entre enseignement et recherche.....	16
2.6. Organigramme.....	17
2.7. Chercheurs associés et chercheurs adjoints.....	18
2.8. Vision à 4 et à 8 ans .....	19
2.9. Un quadriennal de transition .....	20
<b>3. Mise en œuvre</b> .....	<b>21</b>

## II. Projet des équipes

### Equipe 1

#### Logométrie et corpus politiques, médiatiques et littéraires

<b>1. Composition</b> .....	<b>23</b>
<b>2. Vers une logométrie co-occurentielle</b> .....	<b>23</b>
<b>3. Domaine d'application : discours politique et discours littéraire</b> .....	<b>25</b>
3.1. Corpus politiques .....	25
3.2. Corpus littéraires .....	26
<b>4. Hyperbase</b> .....	<b>26</b>
<b>5. Objectifs stratégiques</b> .....	<b>27</b>

### Equipe 2

#### Linguistique de l'énonciation

<b>1. Composition</b> .....	<b>28</b>
<b>2. Projet scientifique</b> .....	<b>28</b>
2.1. Axe 1 : L'altérité en langue et en discours (resp. Sylvie Mellet).....	29
2.2. Axe 2 : Figures et contextualisation (resp. Geneviève Salvan).....	29
Comment la contextualisation éclaire-t-elle la question des limites du champ des figures ? .....	30
Comment les variations de contextualisation peuvent-elles influencer la réception des figures ? .....	30
Qu'apporte la contextualisation à l'analyse des figures ? Les figures peuvent-elles construire et promouvoir un contexte ? Quel rôle jouent-elles dans les déterminations génériques ? .....	31
Inversement, la décontextualisation d'une figure peut-elle conduire à des effets de « défiguralisation » ?.....	31
Certaines figures peuvent-elles échapper au contexte, à la contextualisation, en étant reconnaissables « hors contexte » ? .....	31
<b>3. Objectifs généraux de l'équipe</b> .....	<b>31</b>

### Equipe 3

#### Dialectologie, diachronie et phonologie

<b>1. Composition</b> .....	<b>32</b>
<b>I. Dialectologie</b> .....	<b>32</b>
1. Le "Thesaurus occitan" (THESOC) .....	33
Enrichissement de la base de données .....	33
Ingénierie de la base de données.....	33
Retombées attendues.....	33
2. Reconstruction.....	33
Lexique .....	33
Syntaxe .....	34
Morphologie .....	35

<b>II. Phonologie</b> .....	<b>35</b>
1. Gabarits .....	36
2. Jeux et erreurs de langage .....	37
3. Diachronie du français et dialectologie .....	38
4. Interface, modularité, Sciences Cognitives .....	38
5. Slave et Sciences Cognitives .....	39
6. Théorie phonologique .....	40

#### Equipe 4

##### Contact de langues et français d'Afrique

<b>1. Composition</b> .....	<b>41</b>
<b>2. Approche théorique du contact des langues</b> .....	<b>41</b>
<b>3. Le terrain africain : contact des langues, nouvelles pratiques et nouvelles identités</b> .....	<b>42</b>
3.1. Français en contact en Afrique sub-saharienne et en Algérie .....	42
3.2. Pidgin-english du Cameroun .....	43
<i>Variation et changement linguistique</i> .....	43
<i>Projet européen TypoCorp</i> .....	43
<b>4. Activité éditoriale</b> .....	<b>43</b>
<b>5. Perspectives</b> .....	<b>44</b>

#### Equipe 5

##### Langage et Cognition

<b>1. Composition</b> .....	<b>45</b>
<b>2. Structuration de la thématique Langage et Cognition</b> .....	<b>45</b>
A1. Organisation lexico-sémantique .....	46
A2. Organisation syntaxique .....	47
B1. Acquisition lexico-sémantique .....	48
B2. Acquisition de la syntaxe à l'écrit, articulation entre grammaire et écriture.....	48
C1. Traitements lexico-sémantiques (base textuelle).....	49
C2. Traitements syntaxiques : Modélisation des règles syntaxiques .....	50
A2 et C2. Architecture cognitive de la grammaire.....	50

#### Equipe 6

##### Linguistique et Didactique (LiDida)

<b>1. Composition</b> .....	<b>52</b>
<b>2. Projet scientifique</b> .....	<b>52</b>
2.1. Démarche heuristique.....	53
2.2. Axes de recherche .....	53
A propos d'interactions verbales .....	54
A propos de productions langagières monogérées.....	54
Prise en compte des certifications.....	54
<b>3. Animation, diffusion</b> .....	<b>54</b>

## Résumé

Pour son projet quadriennal 2012-15 L'unité *Bases, Corpus, Langage* sera toujours membre de la fédération de recherche I.L.F. (Institut de Linguistique Française, FR 2393) et participe activement à la construction du nouveau projet de la MSH niçoise (resp.. R. Arena). Le projet rassemble 5 chercheurs, 28 enseignants-chercheurs, 5 ingénieurs, techniciens et administratifs (dont 3 CNRS), 16 doctorants et un certain nombre de personnels non permanents (ATER, post-doc, associés etc.). Le laboratoire présente six équipes, soit deux de plus qu'au quadriennal précédent: 1) logométrie, 2) énonciation, 3) dialectologie, diachronie et phonologie et 4) contact de langues et français d'Afrique, 5) langage et cognition, 6) linguistique et diadactique (LiDida), les deux dernières étant des créations.

Les principales évolutions au niveau de la charpente du laboratoire sont donc l'ouverture de deux nouvelles équipes, et le changement de périmètre du laboratoire induit par l'arrivée de 5 psychologues 16<sup>e</sup> section CNU (dont un IE) et un médecin 44-02<sup>e</sup> CNU, qui forment le noyau de l'équipe Langage et Cognition (la psychologie fait partie des Sciences de la Vie au CNRS: INSB, section 27 du CN).

Ceci coïncide avec d'autres événements qui ont lieu indépendamment de la volonté du laboratoire (démographie: départs massifs à la retraite ; nouveau déménagement pendant l'été à la demande de l'UNS et de son Président, sur un autre campus au centre ville, 228 m<sup>2</sup> > 421 m<sup>2</sup>), induisent un rythme soutenu de changements et font que de tout point de vue le quadriennal 2012-15 sera un quadriennal de transition. Mais il existe aussi des évolutions qui ne sont pas subies. Elles concernent les deux nouvelles équipes, dont la création correspond à une logique interne-externe: interne parce que dans les deux cas il existe une dynamique scientifique au sein du laboratoire (émergence de la didactique, collaborations avec les psychologues), externe étant donné la demande d'adhésion des psychologues, et le souci de mieux répondre à la pression institutionnelle (locale et nationale) en faveur des activités applicatives pour ce qui est de LiDida.

Sur ce fond, l'auto-analyse fait état, du côté positif (opportunités) 1) d'une production scientifique à peu près stable en quantité, mais en progrès significatif en qualité (support, langue de publication) par rapport au quadriennal précédent ; 2) du fait que le laboratoire est leader national dans deux domaines (dialectologie, lexicométrie) ; 3) du soutien constant de l'UNS (moyens financiers et humains) ; du côté négatif (risques): 1) l'absence de recherche sur programmes financés due en partie aux thématiques peu applicatives et non-porteuses dans un environnement "pilote" (impact sociétal etc.) ; 2) la mise en danger de la cohésion sociale et scientifique du laboratoire par la désunion géographique entre enseignement et recherche provoquée par le dernier déménagement ; 3) les faibles effectifs de trois équipes (logométrie, LiDida, contact des langues).

Le laboratoire propose de réagir aux trois derniers points par 1) le développement de deux activités applicatives susceptibles d'améliorer la situation au niveau de l'impact sociétal et de la recherche sur projets (LiDida, analyse automatique des textes en logométrie), 2) un travail avec les Départements d'enseignement visant à aménager les emplois du temps des EC afin de libérer des plages pour la recherche, 3) la croissance de deux équipes (logométrie, LiDida) et la décélération du Contact des langues, qui cessera son activité en 2015 à la fin du quadriennal.

Compte tenu de ceci, parmi les objectifs mentionnés que se fixe le laboratoire, il y en a un qui a la priorité absolue : rendre l'équipe logométrie viable en termes d'effectifs et valoriser son potentiel, au niveau scientifique ainsi qu'au niveau technique. LiDida, de son côté, croîtra "naturellement" par le recrutement d'EC dans le secteur langues. Durant le quadriennal 2012-15, sur les six équipes, deux seront ainsi stabilisées et pourront fonctionner en vitesse de croisière (DDP et énonciation), alors que quatre équipes seront en transition: logométrie (consolidation), LiDida et langage et cognition (rampe de lancement), contact des langues (décélération).

Le laboratoire consent donc à une prise de risque sur plusieurs dossiers, mais il n'a guère le choix : reconduire toujours la même chose n'est pas une marque de bonne santé scientifique, ne pas répondre aux sollicitations des tutelles et du paysage institutionnel n'est pas opportun, et ne pas favoriser et soutenir l'émergence de dynamiques scientifiques (psychologie, LiDida) serait une erreur.

Il est donc vrai que le développement du laboratoire passe par une concentration des risques durant le quadriennal 2012-15 ; il s'agira de les canaliser afin de les maîtriser. Les facteurs favorables à cela et à la mise en œuvre sont 1) le soutien de l'UNS, 2) une situation confortable, stable et prévisible des ressources financières du côté UNS, 3) un matelas de postes d'EC vacants suite aux départs à la retraite, 4) une bonne maîtrise du changement et de la canalisation des risques: le laboratoire a montré dans la période 2006-10 qu'il est capable a) de se restructurer de manière très significative (8 > 4 équipes), b) de mener à bien un déménagement, c) de développer une vie de laboratoire, d) de consolider une équipe menacée de disparition (DDP), e) d'opérer avec succès le rapprochement de deux disciplines qui sont a priori éloignées: fusion de la dialectologie et de la phonologie autour de la diachronie.

## Summary

For the project period 2012-15, the laboratory *Bases, Corpus, Langage* continues to be a member of the research federation ILF (Institut de Linguistique Française, FR 2393), and it actively participates in the unearthing of the new project of the Nice MSH (coord. R. Arena). The project accommodates 5 researchers, 28 teachers-researchers, 5 engineers, technicians and administrative personnel (of which 3 are CNRS), 16 Ph.D students and a number of non-permanent personnel (ATER, post-doc, associates etc.). These work in six research teams: 1) logometrics, 2) énonciation, 3) dialectology, diachronics and phonology, 4) language contact and French spoken in Africa, 5) language and cognition, 6) linguistics and didactics (LiDida), the two latter being freshly created.

The main evolution of the lab's architecture is thus the opening of two new teams, as well as the enlargement of the scientific fields covered that is induced by the arrival of 5 psychologists (16<sup>th</sup> section CNU, among which one IE) and a medical doctor (section 44-02 CNU), who are the heart of the new team Language and Cognition (psychology is part of biology at the CNRS: INSB, CN section 27).

This coincides with other evolutions that occur independently of the lab's action (demography: massive number of retirements; on demand of Nice University and its President, the lab has moved to a different campus downtown during the summer 2010, 228 m<sup>2</sup> > 421 m<sup>2</sup>). The global situation is thus one where a strong wind of change makes the project 2012-15 a project of transition. There are also elements of change, however, which do not have a foreign origin. These concern the two new teams, which come into being following both an internal and an external logic: internal because in both cases there are relevant scientific dynamics within the lab (emergence of didactics, collaboration with the psychologists), external given the demand of the psychologists to join the lab, and the (country-wide as well as local) pressure to develop applicative activities (LiDida).

On this backdrop, the auto-evaluation of the lab records on the positive side (opportunities) 1) a nearly stable scientific production quantitatively speaking, but which has made significant progress regarding quality (type of publication (journals), language of publication (English)); 2) the fact that Nice is leader in France in two areas (dialectology, lexicometrics), 3) the support of Nice University (financial and human facilities). On the negative side (risks), one notes 1) the absence of project-based (and hence project-funded) research, which is partly due to the fact that the themes developed in Nice have little added value in terms of application in a "piloted" environment (societal impact etc.), 2) the danger regarding the social and scientific cohesion of the lab due to the geographical disunion of the places where people teach, and where they do research, 3) the underdimensioned size of three teams (logometrics, LiDida and language contact).

The lab proposes to react on the three latter points by 1) the development of two applicative activities that ought to help improving its position regarding societal impact and project-based research (LiDida, automatic analysis of text in logometrics), 2) working out timetables for teachers with the relevant Departments in order to facilitate their continuous presence at the lab at least on some days, 3) the growth of two teams (logometrics, LiDida), while language contact decelerates until 2015 and will not be renewed beyond that date.

On this backdrop, among the goals mentioned that the lab defines, one is of absolute priority: to consolidate the team of logometrics by developing its personnel and scientific as much as technical potential. On the other hand, LiDida will "naturally" grow by the successive hiring of EC in the area of individual languages. Among the six teams that are proposed for the period 2012-15, two will thus be in a cruising state (DDP and énonciation), while four will be in transition: logometrics (consolidation), LiDida and language and cognition (getting out of the starting blocks), language contact (deceleration).

The lab thus takes risks on a number of occasions, but there is not really any alternative: proposing identical things without any change is not an indicator of good scientific health, not responding to demands of the authorities and the institutional landscape is not a good idea, and not favouring the emergence of new scientific dynamics (psychologists, LiDida) would be an error.

Hence it is true that the development of the lab supposes a concentration of risks during the period 2012-15 – the goal is to be aware of them and to channel them in order to master their evolution. Factors favouring this strategy and the implementation of the project are 1) the support of Nice University, 2) a comfortable, stable and predictable situation regarding funding from Nice University, 3) a stock of positions that will be able to be reoriented when colleagues retire, 4) a good experience in conducting change. The lab has indeed shown in the past that it is able a) to operate a complete restructuring (8 > 4 teams), b) to handle moving to another place, c) to develop laboratory life, d) to consolidate a team under existential threat (DDP) and e) to accompany the convergence of two disciplines that are not really predestined to collaborate (fusion of the teams of phonology and dialectology).

# 1. Auto-analyse

## 1.1. Rythme de changement soutenu

Dans la période du bilan (2006-10), les modifications de l'existant ont été plutôt rapides au sein de BCL, et le projet 2012-15 continue à évoluer à ce rythme, voire à l'accélérer. Ceci implique la prise d'un certain nombre de risques, qui pour certains sont sans alternative. Le projet expose en section 3 comment le laboratoire entend les canaliser et maîtriser. Voici les éléments de changement qui caractérisent le projet 2012-15.

### Installation physique et géographique

Le laboratoire a déménagé une première fois fin 2007 au sein du campus Carlone (98 m<sup>2</sup> > 228 m<sup>2</sup>) et de nouveau, à la demande de sa tutelle universitaire, pendant l'été 2010 de Carlone au campus Saint-Jean d'Angély (centre ville, cf. le descriptif détaillé en section 2.5). Surface dans sa nouvelle installation: 421 m<sup>2</sup>.

### Périmètre scientifique

Depuis 2008 des collaborations avec des psychologues ont été développées. Pour le nouveau quadriennal 2012-15, quatre psychologues et un IE ITRF rejoindront BCL. Cette évolution est moins significative en nombre qu'en termes de périmètre disciplinaire et institutionnel: il y aura désormais une équipe construite autour de la 16<sup>e</sup> section CNU (qui comportera également un médecin MCF-PU qui relève de la section 44-02 CNU), et il est à noter qu'au niveau du CNRS la psychologie relève de l'Institut des Sciences Biologiques (INSB, section 27 du CN).

### Restructuration de la charpente des équipes

En 2006-07, les 8 équipes d'alors ont été resserrées autour de 4 équipes thématiquement cohérentes et ayant une carte à jouer au niveau national et international: DDP (dialectologie, diachronie et phonologie, resp. JP Dalbera), logométrie (resp. D. Mayaffre), énonciation (resp. S.Mellet), contact des langues et français en Afrique (resp. C. de Féral). L'objectif était d'augmenter leur visibilité et de créer une identité niçoise qui soit perçue au niveau au moins national. Cette restructuration a été acquise en début 2008. Depuis lors, le laboratoire a soutenu de manière volontariste les deux équipes qui étaient leader national ou en passe de le devenir: celles de dialectologie et de lexicométrie.

Sur ce socle de quatre équipes, trois mouvements importants s'annoncent pour le quadriennal 2012-15:

- rattachement des quatre psychologues sus-mentionnés et ouverture d'une nouvelle équipe Langage et Cognition, en concert avec l'ouverture d'une filière d'enseignement en Sciences Cognitives à l'UNS (L et M) et la constitution d'un pôle en Sciences Cognitives au sein du nouveau projet MSH niçois.
- émergence, au sein de BCL et à la demande de l'UNS, d'une nouvelle équipe Linguistique et Didactique (LiDida).
- difficulté structurale de l'équipe Contact des Langues et Français en Afrique qui, déjà petite, vient de perdre son leader (R. Nicolai, parti à la retraite en novembre 2009) et comptera en 2012 trois membres permanents qui auront respectivement 61, 63 et 64 ans. Cette équipe a encore un patrimoine à défendre durant le quadriennal 2012-15, elle décélèrera progressivement et cessera son activité à la fin de ce contrat en tant qu'équipe (ce qui ne préjuge pas de l'activité individuelle de ses membres qui ne seraient pas partis à la retraite).

### Equipe émergente: LiDida

L'émergence de LiDida répond à une logique à la fois interne et externe. Les facteurs externes sont d'ordre institutionnel (demande forte de l'UNS) et applicatif (demande sociétale au niveau national), mais le projet est également porté par une dynamique interne à BCL: l'investissement et les compétences en didactique au sein de deux équipes (énonciation et logométrie).

L'UNS a en effet besoin d'une structure de recherche dans un laboratoire reconnu à laquelle elle puisse adosser des postes d'EC en langue (anglais, allemand, espagnol, portugais etc.). Or le secteur langue est défaillant au niveau de la recherche: le seul laboratoire (EA) construit autour d'une langue particulière (l'anglais) a été fermé voici deux ans, et les besoins en enseignement créent une pression au niveau de l'adossement à la recherche. Il va de soi que LiDida n'est pas simplement un garage pour EC qui ne font qu'enseigner et ont besoin d'un paravent de recherche. LiDida se propose 1) d'apporter un éclairage à la linguistique à partir de questionnements pédagogiques, et d'améliorer la pratique pédagogique par la prise en compte de concepts linguistiques (cf. le projet d'équipe), et ainsi 2) de peser sur le recrutement des futurs EC en langue. Au niveau national et international, la didactique est un domaine foisonnant, à forte valeur applicative et répondant à une demande sociétale soutenue, surtout dans le domaine de l'apprentissage des langues. Contrairement à d'autres activités dans ce domaine d'avenir, l'objectif de LiDida est de ne pas céder

à l'utilitarisme pur et simple en proposant un véritable partenariat avec la théorie linguistique qui enrichisse les deux partenaires, didactique et linguistique. C'est par ce positionnement qu'on peut espérer que LiDida acquière une visibilité nationale – c'est en tout cas le pari que fait le laboratoire.

Par ailleurs, deux équipes actuelles – énonciation et logométrie – renferment un potentiel en didactique qui s'est fait jour progressivement: venant d'horizons a priori différents, des EC anglicistes, germanistes et travaillant en FLES ont compris qu'ils s'occupent des mêmes problématiques, et ont commencé à organiser des séminaires transversaux (par-delà les frontières des équipes). S'ajoute à cela le problème qu'évoque le bilan de l'équipe énonciation, à savoir la taille trop grande que cette équipe a atteinte: il est difficile de faire des recherches ciblées et collectives avec 15 personnes. Il est donc tout à fait naturel que des sous-groupes se créent, et le fait qu'un groupe prenne son autonomie complètement est positif pour la vie de l'équipe. Il est vrai que la situation n'est pas la même pour l'autre équipe fournissant des EC pour LiDida, i.e. l'équipe de logométrie, qui n'est pas riche en effectifs, et il s'agit là d'une prise de risque (cf. section 1.3).

L'émergence de LiDida est favorisée par les deux porteurs du projet, Simona Ruggia, MCF FLES, et Laurent Rouveyrol, MCF anglais: il s'agit de jeunes EC (38 et 40 ans, respectivement) très actifs et ayant l'un et l'autre déjà acquis une première expérience dans l'administration (S. Ruggia en tant que directrice du Département de FLES et de l'Ecole d'Eté Internationale en FLES de Nice, L. Rouveyrol en tant que directeur de la Maison des Langues de l'UNS qui a à charge l'organisation de l'apprentissage des langues étrangères de l'ensemble des étudiants et des personnels de l'Etablissement). Les deux sont à quelques années de leur HDR (4 à 5 ans environ), et constituent donc des porteurs adéquats d'une équipe qui devra faire ses preuves (le laboratoire compte la créer dès 2011), mais qui, en cas de succès scientifique et de positionnement national, représente certainement un investissement d'avenir.

Enfin, le succès du projet LiDida est aussi fonction de son attractivité et de sa capacité à structurer le paysage niçois de la didactique. Hormis les deux porteurs qui sont eux-mêmes encore enracinés dans leurs équipes d'origine (à 25% chacun dans le projet 2012-15), les forces de l'équipe se recrutent pour l'instant uniquement au sein du laboratoire: pour ce qui est des permanents, un DR à 25% (75% en Langage et Cognition), une MCF germaniste à 25% (75% en énonciation) et un MCF angliciste à 50% (50% en logométrie). A ceci s'ajoute une MCF angliciste qui rejoint le laboratoire, mais n'ayant pas de production scientifique suffisante n'aura que le statut de chercheur adjoint (cf. section 2.7) et devra donc faire ses preuves. Même si le maillage avec d'autres équipes est souhaité et efficace pour la cohésion du laboratoire et l'ouverture d'esprit, la charpente que présente LiDida en juillet 2010 est clairement insuffisante au niveau des effectifs. Une mesure de la viabilité du projet LiDida sera donc sa capacité à augmenter ses effectifs en attirant des EC extérieurs au laboratoire, et en réussissant des recrutements dans le secteur langue.

### **Nouvelle équipe: Langage et Cognition**

Cette nouvelle équipe répond également à une logique interne-externe: les facteurs externes sont d'ordre institutionnel (création d'un pôle en Sciences Cognitives au sein du nouveau projet MSH niçois, création d'une filière d'enseignement en Sciences Cognitives à l'UNS), mais reposent avant tout sur une collaboration naissante entre psychologues et linguistes ; six psychologues (1 PR, 1 MCF HDR, 2 MCF, 1 MCF-PU et 1 IE docteur) rejoignent le laboratoire. Le projet est également porté par une dynamique interne à BCL: il a été préparé et rendu possible par une orientation des générativistes vers l'architecture cognitive de la grammaire et l'appréciation de la place de la linguistique dans le système cognitif (notamment T. Scheer: livre sur l'interface morpho-syntaxe / phonologie à paraître à l'automne 2010, création du cours de Master "architecture cognitive de la grammaire" il y a deux ans). Outre avec les générativistes (qui appartiennent tous à l'équipe DDP: T. Scheer, M. Oliiviéri, X. Barillot, O. Rizzolo), il s'est avéré que les psychologues ont un autre terrain important de collaboration avec l'équipe logométrie autour de la sémantique co-occurrence (activation d'un mot dans un texte par un autre, mise en parallèle à l'activation d'un neurone par un autre, cf. projet de l'équipe Langage et Cognition).

Du côté des psychologues, le rapprochement avec BCL a été naturel étant donné qu'ils travaillent tous sur le langage, et qu'ils ont jugé les collaborations avec des linguistes plus propices à leur développement que celles qu'ils ont pu avoir avec leurs collègues en psychologie sociale. Leur EA d'origine, le Laboratoire de Psychologie Cognitive et Sociale LPCS (dir. L. Chanquoy), disparaîtra en 2012. Le LPCS était la mise en commun des psychologues cognitivistes et sociaux, ces derniers rejoignant l'EA d'anthropologie LASMIC (qui s'appellera alors LAPCOS). Il est à noter qu'il y a encore un potentiel de croissance pour BCL du côté des cognitivistes, dont certains pour des raisons institutionnelles ont suivi leurs collègues de psychologie sociale en s'intégrant au projet LAPCOS.

En plus de ses forces propres, la nouvelle équipe Langage et Cognition peut donc compter sur un maillage avec deux autres équipes: DDP (T. Scheer à 50%, X. Barillot, O. Rizzolo et M. Oliiviéri à 25%

chacun) et logométrie, où deux psychologues feront une partie de leur activité (F. Lavigne et L. Dumercy à 25% chacun).

La charpente de cette équipe est donc assez solide, et il a été mentionné qu'elle a encore un potentiel de croissance parmi les EC psychologues (un ou deux EC) qui peut être concrétisé pour le début du nouveau quadriennal en 2012. La participation de deux psychologues dans l'équipe logométrie renforce celle-ci, qui a un souci d'effectifs (cf. section 1.3).

S'agissant de son potentiel scientifique et institutionnel ainsi que de sa visibilité nationale et internationale, l'équipe Langage et Cognition doit faire ses preuves. C'est ici que se situe la prise de risque pour le laboratoire: il faut réussir 1) l'intégration des psychologues dans le laboratoire et leur collaboration avec les linguistes ainsi que 2) le positionnement au moins national de cette nouvelle activité.

Au niveau institutionnel local, le rapprochement des psychologues et des linguistes autour des Sciences Cognitives a créé une émulation qui est favorablement accueillie par l'Etablissement. L'originalité du projet autour des Sciences Cognitives que l'UNS développe est la mise en commun de la cognition "pure" et de la cognition sociale. La première, qui à Nice est ancrée dans le langage (côté linguistique comme côté psychologie) considère donc le système cognitif en tant que tel sans prendre en compte les paramètres sociaux (c'est un postulat de base de la linguistique moderne depuis Saussure). Elle est complémentaire de la seconde, qui est basée sur l'anthropologie et la psychologie sociale (il a été mentionné que ces deux secteurs seront regroupés dans la même EA en 2012: Laboratoire d'Anthropologie et de Psychologie Cognitives et Sociales, LAPCOS). L'ensemble des forces en Sciences Cognitives de l'UNS s'articulera également autour d'un projet de filière en Sciences Cognitives pour ce qui est de l'enseignement (Licence "Sciences Cognitives et Sciences Sociales", Master "Langage et Sciences Cognitives") et du nouveau projet MSH. Celle-ci aura son siège dans le bâtiment de l'Horloge du campus Saint-Jean d'Angély, bâtiment qui hébergera également BCL et le LAPCOS dès la rentrée 2010 (cf. section 2.5).

Le paysage institutionnel niçois est donc très favorable au projet de l'équipe Langage et Cognition, qui sera attendue sur deux terrains: sa visibilité nationale d'une part, la collaboration réussie entre psychologues et linguistes ainsi que l'intégration à BCL d'autre part.

### **Départs à la retraite, pyramide d'âge**

La pyramide d'âge étant très déséquilibrée (ballon important dans les 55-65 ans, trou dans les 45-55 ans, bonne base dans les 35-45 ans, cf. la section 4 du bilan général, section 1.3 infra), le laboratoire a déjà été confronté, et le sera massivement durant le quadriennal 2012-15, aux départs à la retraite, et donc à son renouvellement. A cette transition générationnelle correspondra la réorientation thématique partielle ci-dessus évoquée (développement des Sciences Cognitives et de LiDida, décélération de Contacts des Langues et Français en Afrique qui arrivera au terme de son activité en 2015).

### **Qualité et quantité de la production scientifique**

La production scientifique ayant été quantitativement stable et de niveau satisfaisant en comparant les périodes 2002-06 et 2006-10, le laboratoire a répondu aux objectifs qui lui ont été fixés par les diverses évaluations antérieures en augmentant significativement la part des publications dans des revues (plutôt que dans des ouvrages collectifs) et le taux de publications en anglais. Par ailleurs, le taux de publications dans des revues référencées (par rapport aux revues non-référencées) a également augmenté significativement (cf. le bilan général, sections 3.7 et 3.8). Le changement de périmètre scientifique induit par l'arrivée des psychologues et l'activité de la nouvelle équipe construite autour d'eux augmentera encore la qualité des supports des publications (ouvrages collectifs > revues référencées) ainsi que le taux des publications en anglais. En dehors de cet effet mécanique prévisible, le laboratoire doit poursuivre ses efforts dans ce sens.

### **Vie de laboratoire, politique de chercheur sur site**

Les nouveaux locaux (228 m<sup>2</sup>) disponibles à partir de la rentrée 2007 ont rendu possible une vie de laboratoire qui a été développée de manière volontariste depuis lors (journée du laboratoire, journées des doctorants, cycle BCL, cf. bilan). La nouvelle donne géographique où BCL sera localisé entièrement sur le campus de Saint-Jean d'Angély au centre ville fait que la vie de laboratoire devient un enjeu encore plus important car tous les enseignants-chercheurs auront tout ou partie de leur service d'enseignement sur un, voire deux autres campus et devront se déplacer dans la ville pour venir fréquenter le laboratoire. Ce que BCL gagnera en termes de conditions de travail (421 m<sup>2</sup>), il risque donc de le perdre en cohésion sociale et scientifique (cf. section 2.5).



## 1.2. Points forts

(éléments internes qui vont aider au choix et à la réalisation du projet)

Compte tenu des éléments d'analyse mentionnés, les points forts du projet 2012-15 sont les suivants.

### **Cohérence et signature niçoise**

Le laboratoire s'est forgé une identité originale qui assure visibilité et cohérence: 1) travail sur corpus, 2) application de la méthodologie "consusubstantialité de l'empirie et de la théorie" (cf. bilan), 3) thématiques choisies.

### **Leader national en dialectologie et en lexicométrie**

Le laboratoire est leader national dans deux domaines, la dialectologie et la lexicométrie (cf. le descriptif plus détaillé en section 1.5). Il fait entendre sa voix dans deux autres domaines, le contact des langues et l'énonciation, la spécificité niçoise dans ce dernier étant fondée sur la prise en compte de la diachronie, la variété des langues étudiées et le travail sur corpus soutenu par des compétences en statistique. Enfin, la phonologie niçoise est de toutes les équipes, et peut-être des centres français qui font de la phonologie, celle qui est le plus visible dans le monde anglo-saxon.

### **Consolidation de la dialectologie**

Le recrutement récent (juillet 2010) d'E. Carpitelli en tant que PR à la tête de la dialectologie niçoise, combiné avec le recrutement de P-A Georges en tant que IR CNRS en développement informatique, a significativement consolidé l'équipe de dialectologie, qui a désormais toutes les armes pour se développer et pour renforcer sa visibilité. Le pari était loin d'être gagné il y a encore deux ans et demie de cela, où l'équipe, vieillissante et dépourvue de développement informatique, était menacée de disparition.

La capacité d'encadrement a été restaurée durablement (recrutement d'E. Carpitelli, soutenance HDR de M. Oliviéri), les projets de développement sont bien identifiés et en bonne voie (nouvelle Etymologie, Module Morpho-Syntaxe du Thesoc, partenariat renforcée avec l'Atlas Linguistique Roman ALiR par l'arrivée à BCL de la coordinatrice scientifique E. Carpitelli), et le renouvellement par les doctorants est vivant (A. Malfatto, S. Casagrande, à venir en 2010: M. Hochart).

### **Attractivité**

Un corollaire, et donc une mesure de la visibilité nationale et internationale du laboratoire est son attractivité pour des chercheurs et doctorants étrangers qui viennent à Nice avec leur propre financement (ou l'acquièrent du fait de leur venue). La section 3.3. du bilan général en dresse la liste, et les cas sont en forte augmentation depuis 2009.

### **Doctorants relativement peu nombreux mais bien placés**

Les doctorants sont relativement peu nombreux par rapport à la capacité d'encadrement (15 inscrits en juin 2010 pour 15 HDR, 9 thèses soutenues en 2006-10), mais les docteurs formés sont assez bien placés dans l'enseignement supérieur et la recherche, cf. section 5.7 du bilan général qui décrit la politique en faveur des doctorants et docteurs (notamment le statut de post-doc interne). Sur les neuf thèses soutenues en 2006-10, trois sont trop récentes (décembre 2009: L. Revest, K. Palasis, E. Caratini) pour que l'on puisse se prononcer ; sur les six cas restants, on note un recrutement MCF (J. Dalbera), une post-doc rémunérée (C. Philippard), une créatrice d'entreprise (en rapport direct avec sa thèse: FLES, S. Brandt), un retour au pays avec placement dans le supérieur (D. Costea, Roumanie), un emploi stable de fonctionnaire dans le secondaire (B-M Tranh) et une mère au foyer (D. Seigneur-Froli). A cela s'ajoutent trois recrutements en tant que MCF (M. Kastberg: Besançon, M. Pecman: Paris 7, O. Rizzolo: UNS) en 2006.

### **Fonctionnement collectif, consensuel et réglementé**

Le fonctionnement interne du laboratoire est bien rôdé: il repose sur des décisions collectives et très largement consensuelles qui sont prises au Conseil dans le cadre d'un règlement assumé (cf. section 5 du bilan général). Par ailleurs, la cohésion du laboratoire au travers de ses équipes et malgré des orientations individuelles qui peuvent être assez divergentes est assurée par 1) le maillage entre les équipes, i.e. les pourcentages croisés de participation de beaucoup de chercheurs entre plusieurs équipes et 2) la politique volontariste de favoriser une vie de laboratoire (cf. section 3.4 du bilan général).

### **Dynamique de renouvellement**

Un indicateur de bonne santé (mais aussi de prise de risque) est la capacité du laboratoire de faire émerger deux nouvelles équipes, sur la base de logiques interne et interne/externe: tout en développant ses fondamentaux et en assurant la continuité, le laboratoire ne reconduit pas simplement l'existant.

### **Nombre de HDR soutenues, potentiel à venir**

Le nombre de HDR soutenues dans la période du bilan, quatre (V. Magri-Mourgues en 2006, D. Mayaffre et M. Olivieri en 2010, M. Virolle soutiendra en novembre 2010), est relativement élevé par rapport au nombre de MCF/CR, 18 en juin 2010 – soit 22%.

Le potentiel en matière de HDR demeure conséquent:

- G. Olivier est inscrit en HDR depuis 2009 et soutiendra en décembre 2010.
- C. de Féral a obtenu l'autorisation de s'inscrire en HDR et travaille sur son mémoire de synthèse. On peut penser à une soutenance en 2011 ou 2012.
- G. Salvan, L. Gaudin et V. Montagne préparent activement leur mémoire de synthèse et demanderont l'autorisation de s'inscrire en HDR dès la rentrée 2010 pour la première, en 2011 pour les deux autres.
- B. Guillaume et S. Ruggia ont aussi programmé la même démarche pour 2012.

Cette situation atteste de la vivacité des membres du laboratoire qui appartiennent à la tranche 35-45 ans (C. de Féral et M. Virolle étant des cas à part: soutenance à 60 ou 61 ans).

### **1.3. Points faibles**

(éléments internes qui vont pénaliser le projet)

Compte tenu des éléments d'analyse mentionnés, les points faibles du projet 2012-15 sont les suivants.

#### **Pyramide d'âge**

Le point faible principal du laboratoire est sa pyramide d'âge, qui comporte un ballon important tout en haut qui a commencé à partir à la retraite, mouvement qui frappera le laboratoire massivement durant le quadriennal 2012-15. En dessous, dans la tranche d'âge 45-55, le laboratoire déplore un trou générationnel, alors que la tranche d'âge 35-45 est bien fournie et en bonne santé au niveau de l'activité et de la production (un indicateur en est le nombre de HDR soutenues, cf. section 1.2). Le renouvellement concerne les enseignants-chercheurs (10 EC auront 60 ans ou plus en 2015, dont 8 plus de 65 ans) comme les chercheurs (trois auront 60 ans ou plus en 2015, dont 2 plus de 65 ans) (cf. bilan, section 4).

Ce problème est pointé dans l'auto-évaluation statutaire (janvier 2010, jointe au dossier), dont il constitue la principale conclusion. La façon d'y répondre est différente selon qu'il s'agit d'EC ou de chercheurs. Pour ces derniers, le renouvellement est traditionnellement plus aléatoire, et peu prévisible: le CNRS n'a pas de politique bien affirmée en la matière. Le laboratoire cherche à recruter un jeune CR: pour l'équipe de logométrie par le passé (la candidature de Céline Poudat, post-doc CNRS, a été bien appréciée en 2010 et aurait pu aboutir en 2011, mais l'intéressée a été recrutée en tant que MCF à Paris 13 au printemps 2010), probablement en Sciences Cognitives à l'avenir.

S'agissant des EC, l'auto-évaluation expose la stratégie: étant donné l'environnement institutionnel d'aujourd'hui, on ne peut plus compter sur des recrutements basés sur un profil enseignement "à l'ancienne", i.e. purement disciplinaire (c'est vrai notamment pour les langues, vivantes et anciennes). Il faut donc diversifier le profil enseignement et l'adapter aux réalités des formations, tout en veillant que le lien avec la recherche ne soit pas rompu par le caractère par trop applicatif des demandes en enseignement: l'idéal est toujours l'unité de l'enseignement et de la recherche, et ni la société ni l'Université ni les individus partie prenante ne seront gagnants si des disciplines SHS étaient abaissées à de pures prestataires de service.

La stratégie pour le renouvellement des EC est exposée en sections 2.2 et 2.3 infra: le laboratoire développera deux activités applicatives, LiDida et l'étude automatique des textes (lexicométrie). De son côté, l'UNS soutient le laboratoire et fait ce qu'elle peut: par exemple, elle a inscrit dans sa lettre de cadrage pour la recherche le principe qu'un laboratoire classé A ou A+ ne peut diminuer en effectifs, et elle vient d'assurer le renouvellement à la tête de la dialectologie niçoise sur un poste PR au fil de l'eau.

Enfin, outre le renouvellement avec maintien de la section CNU et adaptation aux besoins d'enseignement, le laboratoire cherche à se renouveler thématiquement et à attirer de nouvelles forces par ce biais. LiDida est appelé à jouer ce rôle d'aimant à l'avenir (la période de constitution du projet n'a pas suffi à attirer des EC extérieurs), et le rapprochement avec les psychologues autour de l'ouverture de la nouvelle

équipe Langage et Cognition est un exemple de ressourcement réussi au niveau du projet – reste à en faire un succès scientifique et d'intégration.

La comparaison des fichiers .xls bilan et projet qui établissent la liste nominative des EC du laboratoire est parlante concernant l'importance du ressourcement apporté par le rapprochement avec les psychologues: le nombre l'EC est quasiment identique (27, dont 12 HDR pour le bilan ; 28, dont 10 HDR pour le projet) malgré l'apport de 5 EC psychologues. Sans ce développement, les départs à la retraite auraient mis le laboratoire dans une perspective de déclin. Et encore, le chiffre de 28 EC présents dans le projet est tributaire de trois EC nés en 1947 et d'un EC né en 1948 qui ont affirmé être encore en activité en 2012, mais dont le départ sera imminent en 2012 (ils sont marqués en rose dans la colonne D du fichier .xls projet).

### Faibles effectifs dans trois équipes

Une donnée structurale de trois des six équipes présentées sont leurs faibles effectifs: c'est le cas de l'équipe Contacts de Langues et Français en Afrique (3 emplois pleins), logométrie (3 emplois pleins) et LiDida (2,5 emplois pleins).

#### *Logométrie*

Le projet de cette équipe présente un seul personnel à 100% (D. Mayaffre, chercheur), une chercheuse à 50% (S. Mellet), deux EC à 50% (P. Follette, V. Magri-Mourgues), deux EC à 25% (F. Lavigne, L. Rouveyrol), un IR à 40% (PA Georges) et un IE à 25% (L. Dumercy), soit un total de trois emplois pleins (300%) de chercheurs/EC et 0,65 d'un emploi plein en ITA/BIATOS.

La situation de l'équipe logométrie s'explique par 1) son histoire (activité initiée et entretenue par des "anciens" dont le départ à la retraite n'a pas été compensé: E. Brunet, X. Luong), 2) sa non-implantation à l'Université et dans l'enseignement (cf. sections 2.2. et 2.3. infra), enfin 3) en partie par l'émergence de LiDida, qui a provoqué le départ de L. Rouveyrol (avant 100% logométrie, dans le projet 75% LiDida, 25% logométrie) et de Peter Follette (avant 100% logométrie, dans le projet 50% LiDida, 50% logométrie). Le laboratoire a fait des efforts soutenus par le passé pour renforcer cette équipe, avec un échec du côté EC et un succès concernant un poste de post-doc CNRS rémunéré. Celui-ci a été occupé par C. Poudat en 2008-09 et 2009-10 et a aidé à maintenir l'activité à flot, mais l'équipe retrouve sa situation antérieure dès la rentrée 2010 avec le départ de C. Poudat (recrutée en tant que MCF à Paris 13).

Entre 2005 et 2008, les efforts pour construire un profil d'EC ont été orientés vers une collaboration avec les mathématiciens statisticiens du laboratoire niçois Dieudonné (UMR 6621): des séminaires communs ont été organisés, mais *in fine* il s'est avéré que l'orientation SHS du profil nécessaire pour poursuivre la collaboration ne s'intégrait pas suffisamment bien aux projets de l'UMR 6621 ; parallèlement, bien que l'Université ait un besoin constant en cours de statistique (histoire, sociologie, psychologie, biologie etc.), la construction d'un service d'enseignement sur cette base rencontre l'obstacle des disciplines particulières, qui jugent qu'un statisticien pur, i.e. non-historien, non-psychologue, non-sociologue etc. ne saura pas suffisamment orienter ses cours au bénéfice de la discipline. Cette piste a donc cessé d'être explorée depuis 2009.

La situation de cette équipe est particulièrement grave pour le laboratoire puisqu'il s'agit d'un des deux secteurs où Nice est leader national – malgré donc le faible tissu en personnel. En d'autres termes, la lexicométrie niçoise vit en grande partie grâce aux collaborations externes de longue durée qu'elle a su développer tant au niveau scientifique (relations suivies avec le LASLA : cf. bilan d'équipe) qu'en ce qui concerne les développements informatiques (partenaire actif du programme ANR « Textométrie »). Mais cette situation n'est pas pleinement satisfaisante, en particulier pour le développement du logiciel Hyperbase, fleuron de la production appliquée de l'équipe (cf. section 3.5 du bilan général, section 3 infra). Le très fort potentiel dans les deux secteurs, scientifique avec la topologie textuelle et la sémantique co-occurentielle, technique avec le logiciel Hyperbase, ne peut actuellement être valorisé faute de forces en personnel stable et dédié uniquement à la logométrie. Le développement de cette équipe est donc la priorité absolue du laboratoire dès à présent (juillet 2010), et la stratégie pour la développer est exposée infra (sections 2.2, 2.3 et 3).

#### *Contact des Langues et Français en Afrique*

La situation de l'équipe Contact des Langues et Français en Afrique est différente: elle est constituée d'une chercheuse de la section 38 CN (anthropologie) qui aura 66 ans en 2015 et de deux EC qui auront respectivement 64 ans et 67 ans en 2015. Cette équipe représente l'héritage africaniste niçois qui a une longue tradition depuis les années 70 (G. Manessy), et dont la tête, R. Nicolaï, est récemment parti à la retraite (novembre 2009). Fait partie de cet héritage la revue *Le Français en Afrique* (cf. bilan général) qui est éditée par le laboratoire. Cette équipe a donc un patrimoine à défendre, mais à terme ne sera plus viable

étant donné sa structure d'âge et ses effectifs. Le quadriennal 2012-15 servira donc à décélérer l'activité: une cessation du son fonctionnement dès 2012 n'aurait pas tenu compte du niveau tout à fait honorable de sa production scientifique et de l'engagement collectif de ses membres dans le laboratoire.

### *LiDida*

Enfin, l'équipe LiDida telle qu'elle se présente dans son projet demeure frêle au niveau des effectifs, et en vitesse de croisière ne serait guère viable. On compte deux EC à 75% – les deux porteurs du projet (S. Ruggia et L. Rouveyrol) –, un EC à 50% (P. Follette) et deux participations à 25% (Chr. Hudelot, DR, et N. Schnitzer, MCF), pour un total de 2,5 emplois pleins (250%). Il faut encore mentionner une MCF à 100%, Shona White, mais qui n'est pas actuellement active en recherche et donc participera au projet en tant que membre adjoint du laboratoire (cf. section 2.7 pour le descriptif de cette catégorie).

La création de cette équipe pour le nouveau quadriennal est donc un pari sur l'avenir et constitue une prise de risque qui se fonde sur 1) la jeunesse et le potentiel de ses deux porteurs (cf. le descriptif de l'équipe en section 1.1), 2) les perspectives de croissance en effectifs par voie de recrutement d'EC dans le secteur langues et le soutien de l'UNS en la matière (cf. le descriptif de l'équipe en section 1.1), 3) le développement d'une didactique véritablement linguistique qui ait une visibilité nationale.

### **Peu d'activité sur programmes financés, absence de projets ANR (où BCL est porteur)**

Dans la période 2006-10, le laboratoire a déposé 4 projets ANR en tant que porteur: deux en dialectologie (2006, 2007), un en phonologie (2008) et un en lexicométrie (2006). Aucun des ces 4 projets n'a été retenu. Certes BCL participe à des participations à des projets ANR (lexicométrie, projet Textométrie piloté à Lyon) ainsi qu'à d'autres projets (cf. bilan des équipes: Procope en dialectologie, Agence de Projets tchèque et projet CNRS-Hongrie en phonologie etc., cf. bilan des équipes), mais le laboratoire n'a pas réussi à fertiliser cette source de financements.

Il convient de distinguer, en la matière, entre les aspects liés 1) à l'évaluation de la recherche, 2) à la mise en œuvre de la recherche et 3) au financement. La position du laboratoire est que la tendance actuelle, à tous les niveaux du système et au niveau national, qui consiste à confondre projets et évaluation, ou à substituer à l'évaluation le décompte des succès à l'ANR, est pernicieuse et indue. L'ANR n'a pas de mission d'évaluation en dehors des projets soumis, et la qualité de la recherche se mesure au résultat publié – et non pas à la déclaration d'intention (les projets retenus n'ont encore produit aucune science). On pourrait également retourner l'argument et considérer qu'à résultat égal, il est plus méritant de mener à bien la recherche sans fonds spécifiques. S'ajoutent à cela les problèmes structurels de la recherche sur projet, et les problèmes spécifiques de l'ANR française: taux trop faible de projets retenus, orientation vers l'utilitarisme ("retour sur investissement (pour le monde socio-économique)", "demande sociétale") explicite des programmes thématiques, présente aussi dans les programmes blancs (la raison d'être de l'ANR affichée ouvertement est bien le pilotage de la recherche), consommation importante d'heures de travail chercheur pour faire tourner la machine (du côté chercheur surtout la partie financière du dossier, très onéreuse en temps, du côté de l'agence l'évaluation consomme beaucoup d'heures chercheur), précarisation systématique des personnels, chercheurs et accompagnement: il n'est pas sûr que l'insécurité de l'emploi soit bénéfique à la performance et au rendement des personnels. Même si le succès à l'ANR était un critère d'évaluation probant, celle-ci serait basée sur une concurrence faussée et non-équitable puisqu'orientée en fonction de paramètres non exclusivement scientifiques (pilotage de la recherche): la byzantologie n'a aucune chance de faire jeu égal avec les recherches sur la maladie d'Alzheimer ou la mobilité autoroutière, quelle que soit la qualité de la recherche.

Enfin, il va sans dire que ce qui précède ne met aucunement en cause l'existence d'une recherche sur projet: les projets ont toujours existé, et il est évident que pour faire avancer un projet dans tel ou tel secteur avec un objectif industriel, applicatif ou sociétal, la meilleure façon de faire est par projet financé. Ce dont il est question ici est 1) la proportion que prend la recherche sur projet, beaucoup trop élevée, et 2) la confusion entre le succès à l'ANR et la recherche de qualité ("il n'y a que la recherche sur projet qui vaille") qui s'est installée dans les esprits: en France c'est l'AERES qui a le monopole de l'évaluation scientifique, et son travail n'est pas moins sérieux que celui de l'ANR ; elle prend en compte des logiques à plus long terme ainsi que des paramètres locaux et nationaux qui sont absents du travail de l'ANR. Elle évalue des bilans avérés et non pas seulement des projets. En étant omniprésente et encensée, la recherche sur projet se place non plus en complément nécessaire et intéressant de la recherche quadriennale, mais en concurrent: les deux logiques sont opposées et irréconciliables. Un projet quadriennal sera toujours abandonné au profit d'un projet qui vient d'être accepté, ce qui empêchera les personnels impliqués d'honorer leur engagement quadriennal.

Tout ceci étant dit, il n'en reste pas moins que l'environnement institutionnel national favorise la recherche sur projet, et que BCL a du mal à suivre ce mouvement et à satisfaire la demande abstraite "il faut que les laboratoires aient des projets ANR".

Considérons à présent les deux autres points mentionnés supra, à savoir la mise en œuvre de la recherche et son financement. Si la recherche s'est aussi faite à BCL en l'absence de financement ANR, elle s'est faite plus lentement et certains résultats, notamment parmi les "livrables" matériels (tel le module de cartographie du THESOC qui était au centre d'un des projets ANR déposés) n'ont pas vu le jour. S'agissant des financements, enfin, il est vrai que le laboratoire pourrait affecter ses moyens à d'autres recherches si des projets venaient financer des programmes particuliers en son sein.

L'absence de projets ANR constitue donc bien une faiblesse du laboratoire dans le paysage institutionnel contemporain. L'arrivée des psychologues modifiera cette situation quelque peu (les psychologues sont habitués à obtenir des projets ANR), mais de l'extérieur.

#### 1.4. Risques

(éléments externes de nature à contrarier la réalisation du projet)

##### **Thématiques peu porteuses dans un "environnement piloté"**

Un risque pour le laboratoire est le maintien, voire l'accroissement de la politique nationale qui valorise ce qui est en rapport avec les réalités, avec la demande sociétale, avec le tissu socio-économique dans le cadre du pilotage de la recherche et du "retour sur investissement" (formule dont la sémantique visée exclut les retours intellectuels et scientifiques: l'intention de ses concepteurs est une interprétation restrictive "retour pour le monde socio-économique"), bref de l'utilitarisme. Quoi qu'en disent les responsables, ces paramètres non-scientifiques pèsent sur l'évaluation et sur l'existence de recherches "fondamentales" ou non immédiatement monnayables en termes socio-économiques.

De par la nature de ses recherches et thématiques, le laboratoire n'est pas bien placé dans cette course à l'utile. Son projet cherche à pallier cette faiblesse en développant deux activités applicatives, la didactique (LiDida) et l'analyse automatique des textes (logométrie) (cf. section 2.3 infra). Dans une moindre mesure, la dialectologie peut prétendre à la valorisation d'un patrimoine matériel et immatériel à travers les dialectes et leur importance pour la culture et l'enseignement locaux.

#### 1.5. Opportunités

(éléments externes qui vont favoriser l'émergence du projet)

##### **Soutien de l'UN**

Le laboratoire est soutenu par l'UN: en témoigne le poste PR au fil de l'eau sur lequel E. Carpitelli a été recrutée en juillet 2010, le renouvellement du poste BIAT0.0007 Tc3 Tc 0.05572 Tw 236TJ0.0009eR18(c Ou, AJT )] se

et au-delà l'Europe sont étudiées à Grenoble, Nice se concentre sur le gallo-roman, i.e. les langues issues du latin qui sont parlées en France et dans des pays voisins (domaine de recherche). Par ailleurs, le centre des activités grenobloises est en géolinguistique, i.e. concerne la production d'atlas (Atlas Linguistique Multimédia de la Région Rhône-Alpes, Atlas Linguistique Roman, Atlas Linguarum Europae, Atlas Multimédia Prosodique de l'Espace Roman), alors que Nice élabore un méta-atlas (atlas des atlas), le THESOC, qui intègre dans une base multimédia les données des atlas régionaux individuels, en les agrémentant de données originales acquises par des enquêtes (type d'activité). Enfin, Nice articule la dialectologie avec la théorie linguistique et la diachronie (reconstruction), alors que Grenoble privilégie une approche classificatoire et typologique dans deux domaines spécifiques, la phonétique et la sémantique motivationnelle (la morphologie et la syntaxe n'étant pas pris en compte par les chercheurs grenoblois). On parle ainsi de l'école de Nice qui est articulée autour de la Nouvelle Etymologie de J-Ph Dalbera (cf. bilan et projet équipe) et, partant de l'étude du lexique, essaime maintenant en morpho-syntaxe, et dans une certaine mesure en phonologie.

Par le recrutement au fil de l'eau sus-mentionné qui est effectif au 1<sup>er</sup> septembre 2010, Nice a réussi à attirer la tête de l'équipe grenobloise, E. Carpitelli, qui conduira désormais la dialectologie niçoise. Il est probable que l'on assistera ainsi dans les années qui viennent à un effet boule de neige: les phares niçois et grenoblois gagneront autant en hauteur, en visibilité et en importance que les autres sources de lumière s'affaibliront, et les forces restantes en France auront le regard tourné surtout vers la Côte d'Azur en raison de son ancrage gallo-roman.

Concernant la lexicométrie, c'est au moment de la fermeture du laboratoire de Saint-Cloud – déjà ancienne – que le paysage national a profondément changé. Les forces en matière de statistique linguistique se sont dispersées et l'équipe niçoise, qui, en la personne d'E. Brunet, avait toujours fait partie des leaders nationaux et internationaux, a su profiter de la circonstance pour affirmer sa présence et son identité. A l'heure actuelle il existe en France quatre centres importants dans ce domaine de recherche : l'ENS de Lyon qui pilote le programme Textométrie et qui assure principalement des développements techniques ; une équipe de l'Université de Franche-Comté très orientée vers la philologie numérique plutôt que vers la lexicométrie à proprement parler ; une équipe de Paris 3, mise en péril par le départ en retraite d'André Salem ; et l'équipe niçoise qui collabore étroitement avec chacune des trois autres (voir bilan d'équipe).

Nice se positionne de manière originale en associant aux traditionnelles analyses de discours (politique, littéraire, scientifique) statistiquement fondées une recherche théorique pointue sur les modèles textuels sous-jacents aux traitements quantitatifs. En effet la statistique linguistique classique qui applique aux textes le traditionnel schéma d'urne doit laisser la place à des méthodes nouvelles qui prennent en compte l'organisation de l'axe syntagmatique du texte saisi à la fois à courte portée (dans les séquences répétitives d'une même forme ou d'une même structure grammaticale et dans les ruptures de ces séquences) et à longue portée (dans la répartition des unités étudiées à travers les différentes parties du texte). Dans ce cadre l'équipe niçoise a eu l'ambition de passer de la lexicométrie à la logométrie ; elle a formalisé un nouveau cadre d'analyse, la topologie textuelle, dans lequel le terme « topologie » n'est pas purement métaphorique, mais dont la pertinence a été validée en collaboration avec des mathématiciens. L'équipe a aussi, très récemment, proposé une réflexion théorique qui renouvelle sensiblement l'analyse des cooccurrences. Elle a activement contribué à une réflexion épistémique sur la constitution des corpus et se donne les moyens logiciels d'outiller sa réflexion avec le développement et la diffusion dans le monde francophone d'Hyperbase 8.0 – 2010. Bref, c'est dans le va-et-vient entre analyse des données empiriques issues des corpus et recherche théorique sur divers types de modélisation que se situe l'originalité de l'équipe par rapport à un paysage national relativement fragmenté.

## 2. Projet et objectifs scientifiques

### 2.1. Synopsis objectifs

Compte tenu de l'analyse supra, les objectifs du laboratoire pour le quadriennal 2012-15 sont les suivants:

1. rendre l'équipe logométrie viable en termes d'effectifs et valoriser son potentiel, au niveau scientifique ainsi qu'au niveau technique.
2. développer deux activités applicatives autour de la didactique et de l'analyse automatique des textes.
3. réussir l'intégration des psychologues dans le laboratoire, au niveau scientifique comme au niveau social.
4. rendre l'équipe LiDida viable au niveau des effectifs ainsi qu'à celui de l'orientation et de la production scientifiques.
5. s'agissant de la production scientifique, continuer à travailler sur la qualité du support (la quantité étant satisfaisante globalement): progresser encore dans le mouvement vers les revues (plutôt que de publier dans des ouvrages collectifs), et en leur sein, vers les revues référencées ; enfin, publier davantage en anglais. Au niveau individuel il y a pour quelques membres du laboratoire (deux ou trois) un problème au niveau de la quantité de la production scientifique, ainsi que concernant sa constance: il faut travailler ici sur l'intégration de ces collègues dans une structure collective qui installe une culture qui ne se satisfait pas de l'existence de recherches et de résultats, mais exige encore leur mise sous forme écrite, puis leur partage avec la communauté scientifique.
6. maintenir la cohérence au sein du laboratoire, cohérence qui à partir de la rentrée 2010 pourrait être mise à mal par la désunion géographique entre sa nouvelle installation (Campus Saint-Jean d'Angély) et le(s) lieu(x) d'enseignement de ses EC (campus Carlone et Trotabas) (sur le déménagement cf. section 2.5 infra).

### 2.2. Première priorité: consolidation de la logométrie

L'objectif prioritaire est le premier mentionné. Les deux équipes stratégiques où BCL est leader national (et dont l'activité repose sur un outil informatique) étaient mal en point au début du quadriennal 2008-11. La dialectologie a pu être stabilisée avec 1) le recrutement d'un IR BAP E (décembre 2007), 2) le passage de témoin à sa tête entre JP Dalbera et E. Carpitelli à la faveur du recrutement de celle-ci (juillet 2010), 3) la pérennisation des capacités d'encadrement avec la soutenance HDR de M. Olivieri (juin 2010) et 4) la construction d'un vivier de doctorants correct (cf. bilan et projet équipe). En revanche, les efforts de consolidation de l'équipe logométrie n'ont pas abouti (d'une part le recrutement d'un EC autour de la statistique et des mathématiques, cf. section 1.3, sous-section faibles effectifs, d'autre part le recrutement d'un informaticien) ou n'ont produit qu'un effet temporaire (post-doc CNRS pendant deux ans).

Puisque l'équipe de dialectologie, stabilisée, peut désormais exploiter son potentiel, le développement de l'équipe logométrie devient le verrou stratégique pour le laboratoire. Sa consolidation passe par deux recrutements, d'un ingénieur informaticien et d'un EC dont l'activité d'enseignement est basée sur l'analyse automatique des textes, au service de disciplines diverses (cf. section 2.3 infra). Le développement de cette activité fera d'une pierre deux coups puisqu'elle contribuera à atteindre l'objectif numéro 2, i.e. l'enracinement de BCL dans des activités applicatives. La stratégie proposée pour atteindre ces objectifs est exposée en dans la section suivante et en section 3 infra.

### 2.3. Développement de deux activités applicatives

Le développement de deux activités applicatives autour de la didactique d'une part et de l'analyse automatique des textes d'autre part poursuit plusieurs objectifs:

- faire le lien avec l'enseignement et donc permettre de construire des profils cohérents de postes d'EC. C'est par ce biais que les effectifs des équipes logométrie et LiDida pourront être consolidés.
- enraceriner les recherches dans des activités applicatives est augmenter les chances du laboratoire, dans le paysage contemporain de la recherche, d'obtenir des financements sur projet.

L'analyse automatique des textes est basée sur les acquis de l'équipe logométrie: toutes les disciplines SHS, et au-delà, sont aux prises avec des textes, et ont besoin, d'une manière ou d'une autre, de "les faire parler", i.e. d'en extraire des informations qui ne sont pas visibles à l'œil nu (même après lecture minutieuse). L'équipe de logométrie reçoit des demandes dans ce sens de collègues de disciplines diverses depuis longtemps, et y répond bien volontiers – mais au cas par cas et sans que cela ne soit visible institutionnellement. A l'avenir il faudra faire comprendre à l'Etablissement et aux collègues historiens,

sociologues, littéraires, psychologues etc. l'intérêt que l'analyse automatique des textes revêt pour leur travail et pour leur enseignement. L'organisation, depuis deux ans (2009-10), de l'école thématique CNRS MISAT (Méthodes Informatiques et Statistiques en Analyse des Textes, cf. le bilan général) est une bonne base et montre l'attractivité de la lexicométrie pour l'usage d'autres disciplines (ainsi l'édition 2010 de l'école MISAT a attiré un anthropologue, un géographe, trois historiens, deux littéraires, deux psychologues, deux sociologues, deux spécialistes de TAL, un didacticien, un statisticien, cinq linguistes).

Il s'agira maintenant d'enraciner cette orientation dans l'enseignement local afin de fournir à cette équipe à fort potentiel et à héritage important les forces en recherche qui lui manquent afin de développer les techniques de pointe d'aujourd'hui, qui seront les bases applicatives de demain: topologie textuelle, sémantique co-occurrence (cf. projet d'équipe).

#### **2.4. Intégration des psychologues**

L'arrivée des psychologues représente un changement du périmètre scientifique du laboratoire et induira le contact avec

- d'autres pratiques de publication (plus proches des pratiques en sciences dures encore que la linguistique, qui dans le concert SHS est déjà considérée proche des mœurs scientifiques),
- une très forte culture de laboratoire (chercheur sur site, doctorants compris),
- une culture expérimentale absente du laboratoire

Le spectre de BCL sera donc un grand arc allant de la stylistique (proche de la littérature) à la psychologie en passant par l'énonciation, le contact des langues, la lexicométrie, la dialectologie et la grammaire formelle générative. Certes cet élargissement représente une opportunité, mais qui demande à être gérée, maîtrisée et réussie. Il faut donc veiller à ce que les cultures s'entrelacent, à ce que le tissu social se développe, et à ce que la réciprocité règne dans les collaborations scientifiques avec les linguistes: il n'est pas indiqué que les psychologues soient de simples prestataires de services expérimentaux, ni que les linguistes de simples fournisseurs de théories.

#### **2.5. Maintenir la cohérence malgré la désunion géographique entre enseignement et recherche**

L'opportunité que constitue l'élargissement du périmètre scientifique et l'évolution afférente du laboratoire qui montre qu'il ne reconduit pas éternellement la même chose a pour contrepartie le risque de mettre à mal sa cohésion interne. Ce risque est accru par le déménagement tout récent du laboratoire du campus Carlone (Magnan, Nice-Ouest) au campus Saint-Jean d'Angély (centre ville). Ce déménagement s'est fait physiquement pendant l'été au mois d'août, et au moment de la rédaction du présent projet (début septembre 2010) le laboratoire est en train de s'installer dans ses nouveaux locaux en défaisant ses cartons et en attendant du nouveau mobilier. Si par cette évolution les conditions de travail auront été significativement améliorées (228 m<sup>2</sup> > 421 m<sup>2</sup>, incluant toutefois les six psychologues absents des anciens locaux, proximité d'autres laboratoires et notamment du LASMIC/LAPCOS, anthropologie-psychologie), le danger encouru en contrepartie vient de la rupture géographique entre l'enseignement et la recherche: tous les EC du laboratoire feront tout ou partie de leur enseignement sur un ou deux autres campus (Carlone, Trotabas) et de ce fait devront faire un effort supplémentaire pour venir fréquenter le laboratoire.

Le déménagement s'est fait à la demande de l'UNS et de son Président A. Marouani, qui a souhaité regrouper dans un nouveau bâtiment (dit SJA 3 ou de l'Horloge: substance ancienne 19<sup>e</sup> siècle et entièrement rénové: il s'agit du bâtiment de commandement de l'ancienne caserne qui se trouvait sur le site et a été démolie pour laisser place au développement de l'Université) le gros de ses forces en SHS au niveau de la recherche, et notamment ses UMR. Dans le projet de l'Etablissement, ce bâtiment est appelé à devenir le siège de la nouvelle MSH, qui est en cours de constitution.

Il s'agira donc de redoubler d'effort pour organiser une vie de laboratoire: cela implique notamment des aménagements dans les emplois du temps des EC afin de leur permettre de libérer du temps recherche qu'ils passent à SJA 3. On peut espérer que la forte culture de laboratoire qu'apportent les psychologues et les nouveaux locaux attractifs aideront à développer cet esprit, et à le faire vivre au quotidien.



## 2.6. Organigramme

Considérons d'abord la répartition de l'activité des membres du laboratoire sur les équipes en fonction du pourcentage de leur participation.

NB:

- contrairement au fichier .xls, le tableau infra ne fait pas la différence entre les emplois pleins de chercheurs et les emplois plein d'EC: les deux valent 100% (les chiffres indiquent des pourcentages).
- la somme des emplois pleins, 32,25, n'est pas un entier parce que l'activité de recherche d'E. Kotler, VP CEVU de l'UNS, est évaluée à 25% seulement.

### Chercheurs et enseignants-chercheurs

statut	nom	prénom	corps	sec-tion	HDR	énon-ciation	logo-métrie	DDP	con-tact	Langage et Cognition	LiDida
EC	MAGRI-MOURGUES	Véronique	PR2	09	oui	50	50				
ch	MAYAFFRE	Damon	CR1	34	oui		100				
ch	MELLET	Sylvie	DR2	34	oui	50	50				
EC	GAUDIN	Lucile	MCF	09	non	100					
EC	GUILLAUME	Bénédicte	MCF	11	non	100					
EC	MONTAGNE	Véronique	MCF	09	non	100					
EC	SALVAN	Geneviève	MCF	09	non	100					
EC	SCHNITZER	Nathalie	MCF	12	non	75					25
EC	WATINE	Marie-Albane	MCF	09	non	100					
EC	BIRAUD	Michèle	PR1	08	oui	100					
EC	JAUBERT	Anna	PR1	09	oui	100					
EC	KOTLER	Eliane	PREX	09	oui	25					
EC	BARILLOT	Xavier	MCF	07	non			75		25	
EC	LIBROVA	Bohdana	MCF	07	non	50		50			
EC	MOLINU	Lucia	MCF	07	non			100			
EC	RANUCCI	Jean-Claude	MCF	73	non			100			
EC	RIZZOLO	Olivier	MCF	07	non			75		25	
EC	DALBERA	Jean-Philippe	PR1	07	oui			100			
EC	CARPITELLI	Elisabetta	PR2	07	oui			100			
EC	OLIVIERI	Michèle	MCF	07	oui			75		25	
ch	SCHEER	Tobias	DR2	34	oui			50		50	
EC	DE FERAL	Carole	MCF	07	non				100		
EC	QUEFFELEC	Ambroise	PR1	09	oui				100		
ch	VIROLLE	Marie	CR1	38	oui				100		
EC	MAGNIE-MAURO	Marie-Noële	MCF	44 02	non					100	
EC	LAVIGNE	Frédéric	MCF	16	non		25			75	
EC	NEGRO	Isabelle	MCF	16	oui					100	
EC	CHANQUOY	Lucile	PR1	16	oui					100	
EC	OLIVIER	Gérard	MCF	16	non					100	
ch	HUDELOT	Christian	DR2	34	oui					75	25
EC	FOLLETTE	Peter	MCF	11	non		50				50
EC	ROUVEYROL	Laurent	MCF	11	non		25				75
EC	RUGGIA	Simona	MCF	07	non	25					75
						975	300	725	300	675	250
<b>TOTAL</b>						<b>3225</b>					

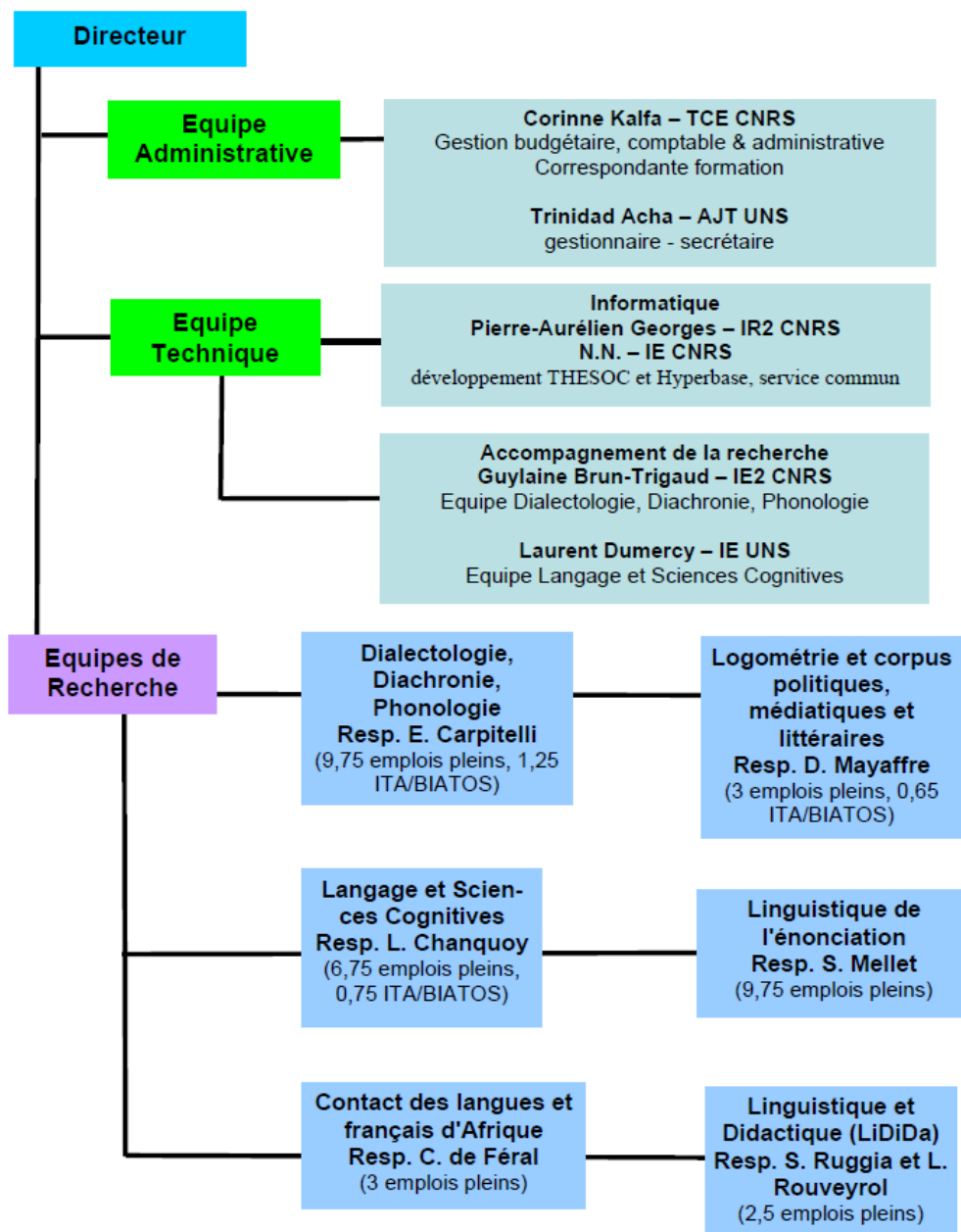
### ITA, BIATOS

statut	nom	prénom	corps	BAP	énon-ciation	logo-métrie	DDP	con-tact	Ige & Cogn	LiDida	service commun
ITA	GEORGES	P-A	IR2	E		40	40				20
ITA	BRUN	Patricia	IE2	D			100				
ITA	KALFA	Corinne	TCE	J							100
BIATOS	DUMERCY	Laurent	IE	E		25			75		
BIATOS	ACHA	Trinidad	AJT	J							100
					0	65	140	0	75	0	220

Compte tenu de ce cela et de ce qui précède, l'organigramme fonctionnel de BCL se présente de la manière suivante pour la période 2012-15.

### Organigramme BCL (2012-15)

5 chercheurs (dont 5 HDR), 28 EC (dont 10 HDR), 3 ITA, 2 BIATOS



## 2.7. Chercheurs associés et chercheurs adjoints

Outre la liste nominative des personnels déclarés (cf. fichier .xls), le laboratoire souhaite reconduire le statut de "chercheur associé" qui a fait ses preuves: il s'agit de personnes dont la collaboration avec une ou plusieurs équipes de BCL est réelle et constante, mais qui pour des raisons institutionnelles ne peuvent faire partie de BCL pleinement (cf. le bilan général: le laboratoire peut consentir à les soutenir financièrement, notamment pour des missions). Les chercheurs associés sont aussi une mesure de l'attractivité de BCL, qui rayonne également dans la société civile (présence d'une journaliste en logométrie). La liste des chercheurs associés telle qu'elle se présente en juillet 2010 apparaît ci-dessous.

**chercheurs associés**

	nom	Etablissement	équipe
1.	Monte, Michèle	PR Univ. Lyon 2	énonciation
2.	Pagani, Cendrine	enseignante secondaire	énonciation
3.	Longrée, Dominique	PR Univ. de Liège / Belgique	logométrie
4.	Raso della Volta, Léa	journaliste	logométrie
5.	Dalbera-Stefanaggi, Marie-Josée	PR Univ. de Corte	DDP
6.	Medori, Stella	MCF Univ. de Corte	DDP
7.	Poli, Muriel	MCF Univ. de Corte	DDP

Enfin, le laboratoire souhaite formaliser un statut qui n'était qu'implicite par le passé: celui de "chercheur adjoint", qui s'applique aux collègues dont la production scientifique décline puisqu'ils sont en partance de la recherche vers d'autres activités, ou au contraire aux collègues qui souhaitent se rapprocher de la recherche mais n'ont pas encore atteint le niveau de croisière nécessaire. Les chercheurs adjoints ne sont pas déclarés en tant que membres statutaires dans la liste nominative des personnels (fichier .xls), mais pourront fonctionner normalement au sein du laboratoire (émargement normal aux moyens). Le contrat sera alors d'évaluer leur activité au bout d'un an ou deux pour déterminer dans quelle direction ils évoluent.

Pour l'instant, un seul cas se présente, mais il se peut que d'autres chercheurs ou EC soient déclarés insuffisamment productifs par le comité d'évaluation, auquel cas ils auront leur place ici.

**chercheurs adjoints**

	nom	Etablissement	équipe
1.	Whyte, Shona	MCF UNS, 11 <sup>e</sup> section	LiDida

**2.8. Vision à 4 et à 8 ans**

Le formulaire AERES demande de prévoir quel sera l'état du laboratoire dans 4 et dans 8 ans, donc en 2014 et en 2018 (à compter de juillet 2010). Voici ce que l'on peut envisager.

équipe	2014	2018
service administratif	un demi ou un personnel supplémentaire pour gérer les psychologues, qui ont leurs propres moyens de gestion jusqu'en 2012.	idem
service informatique	un IE CNRS supplémentaire pour arriver à l'état idéal où chacun des deux grands logiciels, le THESOC et Hyperbase, soit développé par un ingénieur.	idem
accompagnement de la recherche	après recrutement de l'IE informaticien pour Hyperbase (qui est la priorité absolue), besoins en dialectologie. Sollicitation éventuelle de l'UNS, qui ne fournit qu'un seul personnel BIATOS pour l'instant.	idem
énonciation	aura rapetissé: départs à la retraite, en direction de LiDida ; taille plus petite (autour de 8 emplois pleins) optimale et souhaitée. S. Mellet passera la main à la direction de l'équipe vers 2014-15.	taille et composition sont largement fonction des flux en EC (départs, recrutements), surtout en 9 <sup>e</sup> et en langues. Après la saignée du contrat 2012-15, les départs à la retraite seront beaucoup moins nombreux.
logométrie	présence d'un IE BAP E en développement informatique (cf. supra) et d'un EC recruté sur un profil analyse automatique des textes (enseignement) et lexicométrie (recherche).	présence d'un ou deux autres EC, peut-être avec un profil moins spécifique que celui mentionné en 2014.
contact des langues et français en Afrique	décélération de l'activité pour la fin du quadriennal (2015).	–
DDP	déploiement de son potentiel, consolidation de son leadership en dialectologie en France ; promotion de M. Oliviéri en PR, remplacement	idem

équipe	2014	2018
	de l'occitaniste linguiste J-CL Ranucci sur un poste MCF 73 <sup>e</sup> section CNU. Renforcement par un BIATOS en accompagnement de la recherche.	
LiDida	renforcement en effectifs à la faveur de recrutements d'EC en langue (motivés principalement par les besoins d'enseignement, mais orientés vers la didactique en ce qui concerne la recherche). S. Ruggia et L. Rouveyrol HDR.	poursuite de la croissance en EC.
Langage et Cognition	présence de deux ou trois psychologues cognitivistes en poste à l'UNS actuellement. Présence d'un PR en Sciences Cognitives recruté. Présence d'un CR CNRS en linguistique/Sciences Cognitives.	idem

## 2.9. Un quadriennal de transition

Compte tenu des éléments discutés, on peut dire que le quadriennal 2012-15 sera un quadriennal de transition: démographiquement, thématiquement, au niveau du périmètre scientifique et en ce qui concerne le portefeuille des équipes.

Deux équipes seront stabilisées et pourront fonctionner en vitesse de croisière: DDP (y compris les phonologues) et l'équipe d'énonciation. Les quatre autres équipes seront en transition:

- Logométrie: en consolidation pour pouvoir valoriser son potentiel
- LiDida: en rampe de lancement – création d'une activité scientifique et besoin de croissance en EC
- Langage et Cognition: en rampe de lancement – orientation d'une activité existante vers les Sciences Cognitives et vers des collaborations avec les linguistes
- Contact des langues et français en Afrique: décélération

Enfin, on rappellera le projet du laboratoire de s'enraciner dans deux activités applicatives (analyse automatique des textes, didactique), partiellement en réponse au paysage institutionnel (recherche sur projet).

Cette situation de transition est donc le résultat du concours de plusieurs facteurs:

- interne mais non maîtrisable: la démographie (logométrie, contact des langues)
- externe: regroupement des UMR dans un bâtiment recherche par la volonté de l'UNS
- interne-externe: rapprochement avec les psychologues autour des Sciences Cognitives
- interne-externe: émergence de LiDida en raison d'une dynamique scientifique interne, d'une demande de l'UNS et en réponse au paysage institutionnel

Le laboratoire consent donc à une prise de risque sur plusieurs dossiers, mais il n'a guère le choix: reconduire toujours la même chose n'est pas une marque de bonne santé scientifique, ne pas répondre aux sollicitations des tutelles et du paysage institutionnel n'est pas opportun, et ne pas favoriser et soutenir l'émergence de dynamiques scientifiques (psychologie, LiDida) serait une erreur.

Il est donc vrai que les risques existent et sont concentrés dans le quadriennal 2012-15 ; il s'agira de les canaliser afin de les maîtriser. La mise en œuvre infra exposera les stratégies pour y arriver.

### 3. Mise en œuvre

Les facteurs favorables à la mise en œuvre du projet et à la maîtrise des risques pris sont les suivants:

- le soutien de l'UNS
- une situation confortable, stable et prévisible des ressources financières du côté UNS, qui a inscrit dans sa lettre de cadrage recherche que les ressources des laboratoires A et A+ ne peuvent diminuer durant le quadriennal (cf. le bilan général)
- un matelas de postes d'EC vacants suite aux départs à la retraite, postes qui peuvent être réorientés en fonction des besoins du projet si ceux-ci répondent aux orientations de l'UNS
- une bonne maîtrise du changement et de la canalisation des risques: le laboratoire a montré dans la période 2006-10 qu'il est capable
  - de se restructurer de manière très significative (8 > 4 équipes)
  - de mener à bien un déménagement
  - de développer une vie de laboratoire
  - de consolider une équipe menacée de disparition (DDP)
  - d'opérer avec succès le rapprochement de deux disciplines qui sont a priori éloignées: fusion de la dialectologie et de la phonologie autour de la diachronie

Plus spécifiquement, les stratégies pour mettre en œuvre les objectifs sont les suivantes. Concernant la priorité absolue, i.e. la consolidation de l'équipe logométrie, l'objectif à court terme est le recrutement d'un IE CNRS BAP E par voie de NOEMI lors de la campagne d'automne 2010 afin d'organiser le portage du logiciel Hyperbase dans un langage informatique moderne et adapté, et afin de pérenniser son développement. Ce projet semble être en bonne voie et a été bien préparé par l'identification d'un personnel CNRS adéquat (actuellement en poste à Orléans) qui est venu à Nice en avril 2010: le contact a été concluant et l'IE est prêt à candidater pour Nice si le CNRS ouvre une NOEMI. Le VP Recherche de l'UNS JM Lardeaux a appuyé cette demande par une lettre adressée au directeur de l'InSHS P. Bourdelais. L'objectif à moyen terme (2 à 3 ans) est le recrutement d'un EC sur profil recherche spécifique logométrie avec un profil enseignement statistique/analyse automatique des textes. Le travail préparatoire que cela demande est d'une part l'exploration du vivier, d'autre part l'introduction d'enseignements en analyse des textes dans les plans de développement des départements d'enseignement concernés.

S'agissant de l'intégration des psychologues, elle devra se faire d'une part par les collaborations avec les deux équipes DDP et logométrie. Celle-ci sera favorisée par l'installation commune à SJA 3 dans de bonnes conditions de travail. D'autre part on peut compter sur l'effet de la filière d'enseignement (Licence et Master) en Sciences Cognitives que L. Chanquoy et T. Scheer sont en train de créer en collaboration avec le DU du laboratoire d'anthropologie-psychologie sociale voisin (J. Candau, LAPCOS). Il devrait y avoir aussi, dans cette perspective, le recrutement d'un PR qui ne soit pas un linguiste, psychologue ou anthropologue de droit avec quelques intérêts dans des disciplines voisines (ce qui est le cas de toutes les forces niçoises actuelles), mais réponde vraiment au profil de quelqu'un dont le cœur de métier sont les Sciences Cognitives.

Il a déjà été mentionné que la stratégie pour rendre LiDida viable au niveau des effectifs est le recrutement d'EC qui se présentera dans le secteur des langues sur la base des besoins en enseignement, et dont l'adossement à la recherche pourra être offert par cette équipe.

On l'a dit, la cohérence du laboratoire, scientifique et sociale, est mise à l'épreuve d'une part par l'élargissement du périmètre scientifique (psychologie, Sciences Cognitives), d'autre part par la rupture géographique entre l'enseignement et la recherche. La réponse à cette situation est un redoublement dans l'effort de développer la vie de laboratoire: outre la poursuite des mesures déjà en vigueur (journée de laboratoire, cycle BCL, journée des doctorants), la clef d'une vie de laboratoire réussie sera l'aménagement des emplois du temps des EC. Le directeur se concertera à cet effet avec les directeurs des départements d'enseignement.

Concernant la production scientifique, les objectifs (plus de revues, plus de revues référencées, plus d'anglais) seront poursuivis en intégrant les efforts de visibilité et concernant des support de qualité dans le calcul de la bibliométrie incitative.

S'agissant enfin des moyens financiers, ils seront utilisés en priorité, comme par le passé, pour désenclaver la recherche niçoise dont une donnée structurale est de se trouver dans une situation difficile en raison sa géographie excentrée sur le territoire national. Le désenclavement sera promu dans les deux sens, i.e. vers l'extérieur en finançant les missions pour des colloques, des collaborations, des activités éditoriales etc., et vers Nice en favorisant la venue de collègues, conférenciers, collaborateurs etc. (entre autres dans le cadre du cycle BCL). S'ajoutera à cela une pratique relativement récente (2009): le laboratoire peut soulager ponctuellement un service en surcharge de travail par l'embauche de personnel non-permanent sur la base de

vacations ou d'un stage rémunéré. Les deux cas se sont présentés, en 2009 pour la gestion (vacataire) et en 2010 pour l'informatique (étudiant stagiaire). Les deux expériences ont été concluantes.

## Equipe 1

### Logométrie et corpus politiques, médiatiques et littéraires

(resp. Damon Mayaffre)

#### 1. Composition

NOM	Statut	Taux de participation	Date d'arrivée dans l'équipe
<b>Membres permanents</b>			
<b>HDR</b>			
MAYAFFRE Damon	CR HDR	100%	2006
MELLET Sylvie	DR	50% (50% énonciation)	2006
MAGRI-MOURGUES Véronique	PR	50% (50% énonciation)	2006
<b>non-HDR</b>			
FOLLETTE Peter	MCF	50%	2010
LAVIGNE Frédéric	MCF	25% (75% cogn. et langage)	2012
ROUVEYROL Laurent	MCF	25% (75% Lidida)	2006
<b>BIATOS, ITA</b>			
GEORGES Pierre-Aurélien	IR CNRS	40%	2010
DUMERCY Laurent	IE UNS	25% (75% cogn. et langage)	2012
<b>Membres associés</b>			
(cf. section 2.7 du projet général sur ce statut)			
RASO DELLA VOLTA Léa	journaliste		2010
BRUNET Etienne	PR retraité		2006
LUONG Xuan	MCF retraité		2006
<b>Membres non-permanents</b>			
BENDINELLI Marion	Doctorante	50% (50% énonciation)	2008
BONNEAU Julien	Doctorant	100%	2007
PARADA José	Doctorant	100%	2008
VERNIERS	Doctorant	100%	2009

Le projet présenté, tel qu'il se dessine dans les plus récents travaux de l'équipe, comporte trois facettes que l'on souhaite intimement liées : une facette théorique, une facette méthodologique et une facette applicative. Le lien entre ces trois dimensions du projet se trouve *interfacé* par les développements logiciels et la refonte d'Hyperbase qui sera, quant à elle, le grand enjeu technique du quadriennal ; l'outil – que l'on doit pérenniser durant ces 4 ans à venir sous risque de le perdre – est considéré selon l'expression de Bachelard comme une *théorie matérialisée* susceptible de mettre à l'épreuve notre modèle du texte, nos options méthodologiques de traitement des corpus et nos hypothèses pragmatiques d'analystes de discours.

#### 2. Vers une logométrie co-occurentielle

Lors du quadriennal précédent, la topologie textuelle niçoise a essayé de mettre au jour l'organisation spatiale des textes [BCL 159, 160, 264, 283, 284, 408, 447, 448, 463, 555, 590, 604, 612, 671, 701]. Nous souhaitons poursuivre aujourd'hui ce travail en repensant la notion cardinale de *colocation* ou de *co-occurrence*, importante par exemple dans la linguistique computationnelle, les études phraséologiques, le traitement automatique des langues ou la traductologie automatique mais toujours marginale dans la linguistique textuelle *stricto sensu*, quand bien même le tissage co-occurentiel est admis comme le facteur premier de la texture ou de la structuration réticulaire du texte.

L'exigence méthodologique, d'essence quantitative, est pour nous centrale. Forte de sa richesse interdisciplinaire et de son dialogue avec les mathématiciens, la logométrie niçoise veut apporter une contribution éclairée dans le domaine : une co-occurrence est, dans notre perspective, non pas seulement un

fait linguistique qualitatif (une co-présence ponctuelle, syntaxiquement ou distributionnellement déterminée dans la chaîne du texte) mais un fait statistique (une co-présence significativement récurrente dans une fenêtre donnée du texte, le paragraphe par exemple).

La calcul des co-occurrences a été l'objet de nombreuses tentatives tant dans la bibliographie internationale depuis Palmer, Firth ou Harris que, plus pertinemment peut-être, dans la tradition lexicométrique française. Notre travail est donc une recherche complexe de génération deux.

L'enjeu affiné est double :

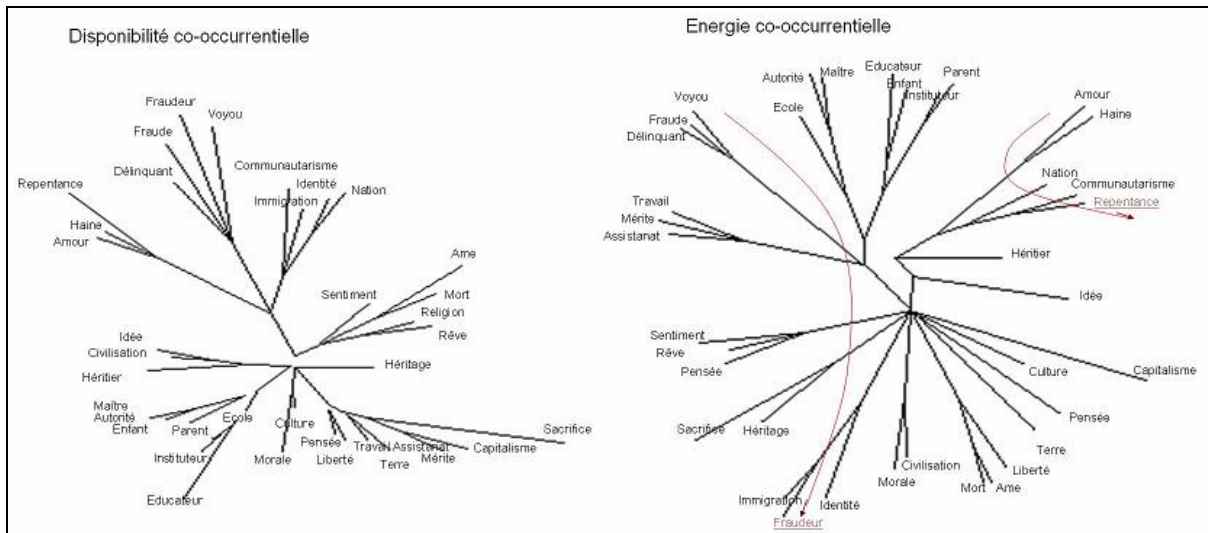
- prendre désormais en considération *l'asymétrie de la relation co-occurrence* entre deux unités linguistiques (deux mots A et B par exemple ou un mot et une catégorie morpho-syntaxique). Les indices *d'attraction mutuelle* de Church and Hanks par exemple, utilisés majoritairement dans le monde anglo-saxon, comme le calcul co-occurrence d'inspiration saint-clousienne utilisé en France, « mutualisent » l'attraction respective de A pour B et de B pour A pour fournir un indice statistique *moyen*. Le but aujourd'hui – dont nous soulignons ci-dessous l'implication linguistique – est de distinguer la part respective de A et de B dans leur association co-occurrence c'est-à-dire dans la paire constituée (AB)<sup>1</sup>. A terme, nous proposerons de distinguer *l'énergie co-occurrence* du mot (la part de ses effectifs que le terme A donne aux autres mots B, C, D, etc.) et la *disponibilité co-occurrence* du mot (la part qui, dans l'effectif de B, C, D, etc., est accueillie ou sollicitée par A) ou, en d'autres termes, la capacité d'un mot à être attiré par d'autres mots versus à attirer d'autres mots. L'hypothèse linguistique, surgie de nos premiers constats et que l'on ne peut, pour l'heure qu'énoncer naïvement, est forte : l'énergie co-occurrence d'un mot témoignerait de *ses relations en langue* avec les autres mots (relation synonymique par exemple, ou regroupement selon un sème inhérent, ou encore famille de mots sur base lexicale commune), lorsque la disponibilité co-occurrence témoignerait de *ses relations en discours* (relation isotopique construite en discours, ou regroupement selon un sème afférent). Si tel était le cas, il deviendrait alors possible de modéliser la dualité fondatrice décrite qualitativement par Sinclair, mais jusqu'ici jamais formalisée par une méthode de traitement, entre *the idiom principle* et *the open-choice principle* des phénomènes collocationnels ; les conséquences seraient importantes tant pour les domaines sus-évoqués des études phraséologiques ou traductologiques que pour la linguistique textuelle et l'analyse du discours.
- prendre désormais en considération, avec des outils mathématiques appropriés, le *profil co-occurrence* global du mot autant que ses relations co-occurrence particulières ou binaires avec tel autre mot<sup>2</sup>. Soit une série de mots dans un texte A, B, C, D, E, etc., le profil co-occurrence complet de A est l'ensemble des relations (AB), (AC), (AD), (AE), etc. Ici aussi la perspective de recherche est d'autant plus forte qu'elle s'enracine dans une tradition linguistique identifiée : renouant avec les travaux niçois pionniers de Pierre Guiraud qui posait, dans une formulation restée célèbre, *que le sens d'un mot est la somme de ses contextes*, nous voulons explorer, plus précisément et de manière outillée, l'idée que *le sens des mots est la somme de ses relations co-occurrence* (i.e. relations co-occurrence asymétriques comme évoqué ci-dessus)<sup>3</sup>. Les principaux outils de la topologie textuelle niçoise, mis en place dans le quadriennal précédent (matrice de dissimilarité, calcul de distance, analyse arborée) semblent tout indiqués sous condition d'être adaptés. Le profil co-occurrence de chaque mot peut être considéré comme un vecteur. Et nous pouvons espérer traiter l'ensemble de la matrice co-occurrence du texte (les profils ou vecteur co-occurrence de tous les mots du texte, consignés dans un tableau) grâce à la distance vectorielle normée. Les premières représentations arborées, selon la méthode de Xuan Luong sont encore expérimentales [BCL 161 ou ACTI 35], mais leur potentiel herméneutique en terme d'analyse du discours apparaît d'ores et déjà important (cf. à titre illustratif les arbres comparés de l'énergie et la disponibilité co-occurrence de 24 mots dans le discours de campagne de Sarkozy-2007.)

<sup>1</sup> Techniquement, cela passe par un traitement mathématique qu'il reste à expérimenter. Pour l'heure, la simple division de la fréquence du couple co-occurrence (AB) par, respectivement, la fréquence de A *versus* la fréquence de B permet la distinction envisagée, et donne des premiers résultats encourageants.

<sup>2</sup> Et par ailleurs, sans qu'on puisse le développer ici, réfléchir aux co-occurrences de deuxième ordre (les co-occurrences des co-occurrences) plutôt que les co-occurrences directes, avec cette simple question si A co-occurrence avec B et B co-occurrence avec C, comment mesurer et qualifier la relation entre A et C ?

<sup>3</sup> De manière grossière, on écrira ainsi que le sens de A,  $s_A = f(AB)+f(AC)+f(AD)$ , ou  $f(AB) =$  fréquence de la co-occurrence de A et de B,  $f(AC) =$  la fréquence de la co-occurrence de A et de C, etc.





Insistons enfin pour conclure que le travail de l'équipe Logométrie sur la co-occurrence, porteur en soi dans la perspective textuelle de l'équipe, se trouve conforté par la dynamique nouvelle du laboratoire. En effet, il apparaît comme un point de rencontre fort avec les collègues psychologues qui viennent renforcer BCL. Frédéric Lavigne et Laurent Dumercy travaillent, d'un point de vue cognitif, sur les réseaux sémantiques et essayent de modéliser l'émergence de réseaux par associations lexicales : la convergence des questionnements, par exemple sur la façon d'évaluer la force des associations ou encore sur les moyens de prouver leur asymétrie, est frappante, même si les cadres épistémologiques sont différents. Pour cette raison, comme nous contribuerons à leur recherche au sein de la nouvelle équipe *Langage et Cognition*, en apportant notre expertise linguistique, ils acceptent de contribuer à notre réflexion en intégrant pour partie l'équipe Logométrie.

### 3. Domaine d'application : discours politique et discours littéraire

Selon le principe revendiqué du laboratoire de la consubstantialité de la théorie et de l'empirie, nous ne concevons pas un développement théorique ou méthodologique sans un champ d'application. Notre travail sur les co-occurrences est motivé par les corpus réels que l'on entend traiter, et sera appliqué à deux domaines privilégiés : le discours politique et médiatique contemporain d'une part qui s'est imposé à Nice lors du quadriennal précédent, et le discours littéraire d'autre part qui reste une signature niçoise depuis Etienne Brunet [OS-ouv 4]. A ces domaines de prédilection, s'ajoute désormais une perspective comparative (centralement français / anglais ; marginalement français / espagnol) que l'on semble pouvoir construire autour de Laurent Rouveyrol (angliciste), de notre nouvelle recrue Peter Follette (arrivé en 2010 qui étudie des corpus d'anglais natif *versus* anglais acquis), de nos doctorants et post-doctorant, Marion Bendinelli qui travaille sur le discours politique américain, Léa Raso della Volta qui travaille sur le discours d'Obama et José Parada qui est, par ailleurs, auteur hispanisant : de fait, la logométrie niçoise *via* la commercialisation de son logiciel Hyperbase cherche désormais à se développer à l'international (voir *infra* les versions française, anglaise, espagnole, allemande, portugaise et latine d'Hyperbase).

#### 3.1. Corpus politiques

Dans le discours politique, plus qu'ailleurs, les mots n'ont pas de signification donnée mais un sens construit dans des parcours interprétatifs calculés au sein de corpus bien formés. Dans cette construction, la contextualisation en général, la co-occurrence – définie comme la forme minimale du contexte – en particulier sont donc centrales. Damon Mayaffre a récemment essayé de montrer, par exemple, que le terme « patrie » très employé dans le discours de Sarkozy durant sa campagne présidentielle 2007 était un terme non pas polysémique mais infrasémique que la définition *en dictionnaire* ne renseigne que modestement. En revanche, le sens de « patrie », plus connotatif que dénotatif, dans le discours sarkozien s'éclaire par le constat co-occurrence statistique de son utilisation conjointe et répétée avec les termes « famille » et « travail » : ce n'est point l'occurrence de « patrie » qui fait sens mais ses co-occurrences.

Les perspectives de recherche sur le discours politique français se feront sur deux axes : un axe diachronique et un axe politique. C'est en diachronie d'abord que les associations co-occurrence doivent être étudiées dans le discours politique. L'équipe Logométrie niçoise dispose aujourd'hui d'un fonds important de plusieurs milliers de discours politiques français sur 2 siècles (de Robespierre à Sarkozy) qu'il

reste à exploiter (fonds en partie librement disponible sur notre site POLITEXT (<http://www.unice.fr/ILF-CNRS/politext/>) qu'il s'agira de restaurer durant le quadriennal après son endommagement par les services informatiques de l'Université lors du quadriennal précédent). Dès lors, il devient intéressant de suivre le régime co-occurentiel d'un item linguistique donné sur plusieurs décennies pour comprendre comment l'unité de la forme peut cacher une pluralité de sens. Cette vision diachronique est au coeur du programme de recherche que Damon Mayaffre a présenté lors de sa soutenance d'HDR au printemps 2010 ; programme actuellement en partie mis en place dans la thèse de Julien Bonneau sur le grand corpus de Pierre Mendes France s'étalant de 1924 à 1980. Très morcelées, et devenues rares depuis la disparition du laboratoire de Saint Cloud, les analyses du discours politique *via* la méthode logométrique, doivent désormais s'inscrire, à Nice, dans une vision chronologique à la fois élargie et synthétique. Ensuite, c'est par contraste politique que la lecture co-occurentielle sera développée. Ici aussi les premiers travaux montrent les avantages d'une logométrie co-occurentielle. S'il n'est pas rare que la gauche et la droite, par exemple, utilisent avec une même fréquence un mot donné, leur différence idéologique transparaît dans la contextualisation minimale desdits mots (*i.e.* de leur co-occurrence). Par exemple, Arlette Laguiller et Nicolas Sarkozy utilisent dans les mêmes proportions le mot « travail », mais, dans leur discours respectif, « travail » ne co-occure jamais avec les mêmes mots [BCL 457, 458, 459]. Fondatrice depuis Foucault, cette vision contrastive des discours ou formations discursives sera évidemment au cœur des études niçoises de logométrie politique.

### 3.2. Corpus littéraires

Dans le discours littéraire, plus qu'ailleurs, le statistique occurrentielle touche à ses limites. Elle permet au mieux d'analyser le « contenu » des textes mais reste impuissante à rendre compte de la richesse stylistique, du maillage textuel, de la complexité de l'oeuvre. La statistique co-occurentielle, en revanche, doit permettre de contribuer à la classification textuelle. Véronique Magri a ainsi développé, au cours du quadriennal précédent, pour son HDR, une recherche qui s'inscrit dans la poétique des genres. A partir d'un corpus contrastif regroupant deux ensembles de textes, qu'un enjeu pragmatique différencie, l'un composé de textes fictionnels, l'autre regroupant des textes d'ordre factuel, elle a induit des paramètres définitoires du genre *récit de voyage* au XIXe siècle. C'est donc sur l'observation d'un corpus établi qu'une théorie générique a pu être énoncée et donner lieu à la publication d'un ouvrage *Poétique du récit de voyage* (Paris, Champion, 2009). Le calcul des co-occurrences a montré en particulier que les constellations lexicales qui s'esquissent caractérisent tel ou tel genre et que, parallèlement, l'environnement d'un même terme diffère selon le genre textuel. Les affinités lexicales affirment leur spécificité. Pour le prochain quadriennal, l'analyse co-occurentielle sera testée comme un outil de caractérisation d'une écriture particulière en rejoignant cette fois la stylistique d'auteur. Les corpus seront choisis parmi la littérature du XXe siècle et regrouperont en particulier les oeuvres d'Aragon et celles de Giono. Au-delà de l'écriture particulière d'un individu, l'étude pourra viser la caractérisation de l'écriture de la modernité.

## 4. Hyperbase

Notre investissement collectif dans le projet ANR Textométrie (2007-2010) qui a rassemblé dans une fédération les grands concepteurs des logiciels lexicométriques français et étrangers, nous a convaincu, de manière raisonnée, de l'importance d'Hyperbase dans le paysage scientifique. Après une comparaison objective avec les logiciels partenaires (Lexico, Xaira, Weblex), Hyperbase nous paraît, sur le marché scientifique, le logiciel de logométrie le plus riche et le plus développé quant aux fonctionnalités proposées, et le plus fragile quant à sa portabilité informatique, ses vitesses de traitement, son langage de programmation, ses interfaces.

Tout en voulant continuer à participer au développement de la plateforme communautaire open source TXM que nous avons contribué à mettre en place (cf. site Textométrie, <http://textometrie.ens-lsh.fr/>), il nous paraît impérieux, à Nice, de rétro-concevoir et de développer Hyperbase ; ou plus précisément, nous comprenons que l'avenir de la plateforme communautaire TXM dépend de notre capacité (et de celle de nos partenaires) à développer continûment les outils les plus en pointe tel Hyperbase.

L'enjeu d'une telle rétro-conception est à la hauteur des ambitions affichées et tenues par Hyperbase lors du quadriennal précédent.

- Hyperbase est actuellement un des rares logiciels qui permet le traitement, à la volée, du texte dans son épaisseur linguistique c'est-à-dire à 4 niveaux : formes graphiques, lemmes, étiquettes morphosyntaxiques, combinaisons syntaxiques. Cette caractéristique majeure a des répercussions en terme de programmation tant au niveau des moteurs de recherche documentaire mis en place (possibilité de rechercher des expressions régulières complexes ou selon l'expression niçoise des *motifs*) qu'au

niveau des traitements statistiques (avec possibilité de croiser mathématiquement les différents niveaux d'analyse et de calculer par exemple la distribution d'un mot ou d'un lemme, d'un présent de l'indicatif ou d'un adverbe).

- Hyperbase est actuellement un des rares logiciels à se donner les moyens de son ambition polyglotte. Articulé au lemmatiseur/étiqueteur universel TreeTagger, il offre la possibilité à l'utilisateur de traiter les grandes langues européennes : français, anglais, espagnol, allemand, italien, portugais. (Et il a été par ailleurs développé pour le Latin, en collaboration avec le LASLA à Liège). Ici aussi la tâche de rétro-conception s'annonce ample.
- Hyperbase enfin propose à côté de la statistique lexicométrique traditionnelle, partagée par d'autres logiciels (calcul de spécificités, richesse lexicale, accroissement du vocabulaire, AFC, etc), des outils mathématiques originaux et plus complexes (distance intertextuelle, topologie textuelle, arbres non plantés, ou, depuis le printemps 2010, grâce à l'aimable donation de Ludovic Lebart, ellipses de confiance et méthode de bootstrap).

C'est dans ces conditions que le laboratoire et l'équipe entendent se donner les moyens de pérenniser Hyperbase. Pierre-Aurélien Georges, ingénieur informatique de recherche, vient d'être rattaché depuis 2010 à l'équipe et planifie la réécriture. Le principe d'un recrutement d'un ingénieur d'étude, susceptible d'aider Pierre-Aurélien Georges, et entièrement détaché à cette tâche, vient d'être acté par le CNRS. En étroite collaboration avec le concepteur et programmeur d'Hyperbase, Etienne Brunet, les exigences de cet HYPERBASE 2 ont été identifiées : portabilité sur internet, modularité des fonctionnalités (susceptibles d'être reversées à la plateforme TXM), accueil des nouveaux formats de textes (XML-TEI notamment), souplesse des partitionnements et de la tokenisation. Après discussion, le langage de programmation Java a été adopté notamment pour sa portabilité et pour la richesse de ses bibliothèques : le développeur qui sera recruté devra disposer de compétences dans ce langage.

## 5. Objectifs stratégiques

Le développement de l'équipe Logométrie bénéficie jusqu'ici d'une articulation forte entre un dynamisme local spécifique, dont l'origine remonte à la création du laboratoire par Etienne Brunet, et un dialogue global, national et international, conséquent. C'est ce dialogue que nous continuerons à développer autour du réseau international des JADTs (Paris, Rome, Montréal, Louvain-la-neuve, etc.) et des liens privilégiés avec nos partenaires étrangers. Notamment : après une décennie de collaboration étroite avec le LASLA à Liège, qui a pris des formes institutionnelles variées (échanges *Tournesol*, convention CNRS/FNRS) et a abouti à de nombreuses publications [BCL 73, 159, 160, 264, 265, 283, 284, 447, 448, 590, 701], un programme Erasmus (2011-2013) entre Nice (resp. D. Mayaffre) et Liège (resp. D. Longrée) vient d'être accepté. Ce programme, qui pérennise un peu plus une collaboration efficace dans la durée, devrait pouvoir être prolongé en Erasmus Mundus entre Nice, Liège et Rome afin d'y agréger, outre les trois partenaires européens contractuels, nos partenaires canadiens (Moncton, resp. Sylvie Kasparian) et égyptiens (Alexandrie, resp. Soheir El Shamy).

Cependant, si la Logométrie niçoise apparaît comme une niche scientifique pertinente au niveau recherche, elle ne semble pouvoir s'affirmer sans un lien plus marqué avec l'enseignement. D'importants efforts ont été produits pour inscrire les axes de recherche de l'équipe dans les maquettes au niveau Master SDL, Littérature et Histoire (4 modules dispensés) mais il reste sans doute à faire entrer ces axes, en amont, au niveau Licence afin de former plus précocement les étudiants aux problématiques de la linguistique textuelle et quantitative et de l'analyse du discours assistée par ordinateur. C'est sous cette condition que le potentiel d'encadrement de thèse de l'équipe (Véronique Magri (PR), Damon Mayaffre (CR-HDR), Sylvie Mellet (DR)) pourra être exploité.

Enfin, la redéfinition des périmètres du laboratoire avec l'apparition de deux nouvelles équipes aura des répercussions dynamisantes pour la Logométrie. Si Laurent Rouveyrol et Peter Follette partageront désormais leur temps entre l'équipe Lidida et l'équipe Logométrie, les collaborations nouvelles avec l'équipe Langage et Cognition (L. Dumercy, F. Lavigne) s'annoncent richement heuristiques autour des projets conjoints *co-occurrences* (cf. supra : *Vers une logométrie co-occurrence*) et *organisation, acquisition et traitement lexico-sémantiques* (cf. infra le projet de l'équipe Langage et Cognition).

## Equipe 2

### Linguistique de l'énonciation

(resp. Sylvie Mellet)

#### 1. Composition

NOM	Statut	Taux de participation	Dates à noter
<b>Membres permanents</b>			
<b>HDR</b>			
BIRAUD Michèle	PR	100%	
JAUBERT Anna	PR	100%	
KOTLER Eliane	PR	100%	
MAGRI-MOURGUES Véronique	PR	50% (50% en logométrie)	
MELLET Sylvie	DR	50% (50% en logométrie)	
<b>non-HDR</b>			
GAUDIN-BORDES Lucile	MCF	100%	HDR prévue en 2012
GUILLAUME Bénédicte	MCF	100%	
LIBROVA Bohdana	MCF	50% (50% en dialectologie)	
MONTAGNE Véronique	MCF	100%	HDR prévue en 2012
SALVAN Geneviève	MCF	100%	HDR prévue en 2011
SCHNITZER Nathalie	MCF	50% (50% en LiDida)	
WATINE Marie-Albane	MCF	100%	
<b>émérite</b>			
VUILLAUME Marcel	PR émérite	100%	retraite en oct. 2010
<b>Membres associés</b>			
(cf. section 2.7 du projet général sur ce statut)			
MONTE Michèle	MCF HDR	Université Lyon 2	
PAGANI Cendrine	ens. second.	Lycée Estienne d'Orves, Nice	
<b>Membres non-permanents</b>			
BENDINELLI Marion	Doctorante	50% (50% en logométrie)	thèse en cours depuis 2008
COMBE Vincent	Doctorant	100%	thèse en cours depuis 2008

#### Commentaire :

Dans le prochain quadriennal, l'équipe « Enonciation » devrait connaître un léger resserrement de ses effectifs principalement dû à la constitution de l'équipe LiDida, co-dirigée par Simona Ruggia. Devraient rejoindre cette équipe, au moins à temps partiel, et donc quitter en tout ou en partie l'équipe « Enonciation », outre Simona Ruggia, Nathalie Schnitzer, Christian Hudelot et Stefanie Brandt (membre associée).

En revanche, le départ en retraite de M. Vuillaume (2010) n'aura guère d'impact puisque ce collègue a obtenu l'éméritat et poursuivra son activité de recherche, et il devrait en aller de même pour les quelques autres départs prévisibles au cours du contrat.

Par ailleurs, Lucile Gaudin-Bordes, Véronique Montagne et Geneviève Salvan devraient soutenir leur HDR avant le début du prochain quadriennal ou dans ses tout premiers mois. G. Salvan est d'ores et déjà responsable d'un des deux axes proposés dans le projet. La relève d'encadrement sera donc assurée au sein de l'équipe.

#### 2. Projet scientifique

Sur la base du constat (voir partie bilan) que le choix d'une seule thématique était un peu trop contraignant pour notre équipe, nous avons donc choisi de retenir deux grands axes de travail collectif pour le prochain

quadriennal. Ce choix, associé à la légère réduction des effectifs qui entraîne aussi une homogénéisation des centres d'intérêt de l'équipe, devrait permettre de relancer la dynamique collective et la rendre plus efficace. Il autorise en outre Sylvie Mellet à participer activement au projet de l'équipe Logométrie sur les cooccurrences et il permet de préparer de manière progressive une relève de la direction d'équipe.

### 2.1. Axe 1 : L'altérité en langue et en discours (resp. Sylvie Mellet)

Cet axe permettra de poursuivre un certain nombre de recherches engagées depuis plusieurs années dans l'équipe (discours rapporté, dialogisme, modalités notamment), d'en assurer l'homogénéité grâce à une même problématisation théorique et d'accueillir simultanément de nouvelles recherches particulières.

Nos récents travaux sur le dialogisme [BCL 278 et 461 notamment] nous ont en effet conduits à nous interroger sur la place qu'il convient d'accorder, en langue et en discours, à la construction des points de vue autres : est-il possible, lorsqu'on s'intéresse aux hétérogénéités énonciatives qui parcourent les textes et les discours, d'articuler le niveau discursif et celui de la langue dans un cadre théorique pertinent ? Y a-t-il, par exemple, entre le concept d'altérité de la Théorie des Opérations énonciatives et prédicatives de Culioli et celui de bivocalité proposé par J. Authier-Revuz, autre chose qu'une simple rencontre de fortune, un air de ressemblance approximatif, ou bien tentent-ils l'un et l'autre de décrire le même fait de langage appréhendé à partir de points de vue différents ? En somme, la langue se contente-t-elle de se prêter, de manière secondaire et indirecte, à l'expression de faits purement discursifs que l'on étudie sous le nom de polyphonie, dialogisme, feuilletage énonciatif, etc., ou bien le système de la langue intègre-t-il en lui-même des opérateurs spécifiques en charge d'une opération énonciative primaire, fondamentale au niveau cognitif et qui serait à la source d'une multiplicité de faits observés en discours<sup>4</sup> ?

On tentera de fournir des éléments de réponse à ce questionnement théorique en approfondissant nos travaux sur les marqueurs de dialogisme, au nombre desquels on privilégiera un certain nombre de connecteurs discursifs étudiés en diachronie (B. Librova, G. Salvan), les temps et modes verbaux (S. Mellet, C. Pagani, M. Vuillaume), les modalités propositionnelles et phrastiques (M. Biraud, B. Guillaume).

Dans ce cadre, une attention particulière sera portée aux processus de grammaticalisation et de pragmatization dont on pense qu'ils pourraient fournir une illustration exemplaire de l'absorption par la langue elle-même du fait dialogique. Ainsi on constate aisément que les divers sens modaux du verbe *pouvoir* sont étroitement associés à sa dialogisation en discours ; ce sont aussi des interactions contextuelles fortement dialogiques qui font basculer l'adverbe *bien* vers le statut de marqueur énonciatif et ponctuant de discours. Mais, simultanément, il convient d'observer que le signifié abstrait de chacun de ces termes (y compris dans leur sens lexical de base : *avoir la capacité de*, *bien* antonyme de *mal*) peut être réduit à une forme schématique qui sollicite le concept d'altérité ou, si l'on préfère, l'existence d'une alternative, d'une bifurcation entre deux branches possibles. On cherchera donc à analyser et, si possible, à modéliser cette interaction remarquable. La pragmatization sera traitée à la fois comme processus diachronique par le biais d'études prenant en compte l'histoire de la langue (analyse de l'adverbe *or* et du présentatif *ez vos* en ancien français par exemple ; évolution sémantique des adverbes en *-ment* dont on a coutume de dire qu'ils acquièrent une valeur énonciative et pragmatique au 18<sup>ème</sup> siècle) et comme exemplification d'un modèle théorique (par exemple la subduction guillaumienne).

L'étude des modalités propositionnelles et phrastiques, quant à elle, permettra d'analyser les contraintes que l'engagement énonciatif fait peser sur la syntaxe : elle fournira donc une autre entrée à la problématique du rapport entre langue et discours. Les différentes études envisagées porteront, entre autres, sur les énoncés exclamatifs en grec ancien, sur la parataxe et la subordination au 18<sup>ème</sup> siècle, sur les interactions entre les valeurs modales d'une part, le statut énonciatif et le niveau prédicatif d'autre part de certaines propositions subordonnées, en anglais et en français.

Cet axe s'inscrivant dans la continuité de travaux antérieurs, l'objectif est ici de produire principalement des travaux de synthèse.

### 2.2. Axe 2 : Figures et contextualisation (resp. Geneviève Salvan)

L'idée de ce nouveau programme est née de la convergence entre l'intérêt que certains membres de l'équipe portent aux figures du discours [BCL 52, 58, 143, 145, 146, 147, 148, 153, 241, 429, 431, 436, 472, 473, 505, 617], en participant au renouveau actuel de leur étude dans une perspective pragma-énonciative, et le souci partagé par l'ensemble de l'équipe de prendre en compte les contextes d'emploi dans l'analyse des faits linguistiques. De « transversales », les figures du discours sont ainsi devenues un objet d'étude représentatif dans l'équipe : ce programme devrait permettre de prolonger le travail entrepris dans les axes « Dialogisme » et « Frontière » et de mettre à profit les résultats déjà obtenus. Enfin, ce programme s'inscrit

<sup>4</sup> Dans l'hypothèse d'une opération cognitive de base, on envisage aussi de confronter nos concepts théoriques avec celui d'« alternatives » de la sémantique formelle.

dans la logique du laboratoire qui consiste à articuler théorie et empirie linguistique dans la mesure où il entend contribuer au rendement scientifique d'une méthodologie qui articule l'analyse des faits linguistiques à une conception théorisée des contextes d'emploi (voir la partie bilan 2007-2011).

En quoi la question du contexte et de la contextualisation est-elle intéressante et pertinente pour l'étude des figures ?

La pratique décontextualisante de certains énoncés remarquables, étiquetés comme « figures » et rangés dans une nomenclature fermée, pratique caractéristique de certains traités des plus anciens aux plus récents, conduit à considérer les figures comme des « morceaux » de discours, détachés de leurs conditions de production et de réception et autonomisés par rapport à leur contexte linguistique et discursif.

L'approche pragma-énonciative que nous défendons à la suite de Détrie 2001, Bonhomme 2005 et Rabatel 2008 privilégie quant à elle l'étude des figures dans leurs contextes, et vise simultanément à dégager le fonctionnement énonciativo-discursif et pragmatique propre à chacune. Considérant les figures comme des saillances discursives qui constituent des schèmes sur le plan communicationnel, nous posons comme fondamental le rôle du contexte aussi bien dans la production que dans la réception des figures. Dans notre approche, le contexte retrouve droit de cité et nous le considérons même comme *une condition de la figuralité* : une configuration discursive n'est pas *a priori* figurale, mais replacée dans son contexte, ou plutôt, *en lien dynamique avec un contexte*, elle peut accéder à la figuralité. Nous pensons en effet qu'une structure n'est susceptible de présenter un relief dans le discours qu'en situation et dans l'interaction avec un contexte.

Une saillance discursive peut ainsi rester au stade de la saillance ou parvenir à la figure dans certains contextes discursifs, et c'est la contextualisation de cette saillance qui permettra de mesurer son statut figural.

Mais si la nécessité de la contextualisation des figures ne fait pas débat, puisqu'elle est finalement à l'origine du renouveau scientifique actuel, il nous semble en revanche que la question mérite d'être posée frontalement et pour elle-même. En effet, notre propos n'est pas seulement d'étudier les figures en contexte, ou de montrer l'avantage de la prise en compte du contexte dans l'analyse des figures, mais il s'agit de poser de nouvelles questions, en essayant d'y répondre, sur *la relation dynamique entre figures et contexte*, ce que le choix du terme « contextualisation » traduit. Prenons un exemple : toute rupture de construction n'est pas une figure d'anacoluthie, et peut même s'apparenter, dans une vision normative de la langue, à une faute. Sur quelles bases syntaxiques, énonciatives, discursives, peut-on donc dire que l'énoncé *En approchant d'Alexandrie, l'air s'allège* est une anacoluthie ? Inversement, peut-on décrire le fonctionnement syntaxico-énonciatif général de l'anacoluthie à partir de l'observation d'un grand nombre de contextes permettant d'identifier des invariants qui rendraient prédictible sa saillance figurale et non seulement discursive ?

Par ailleurs, si la figure gagne indéniablement à être contextualisée, il est plus délicat, mais aussi sans doute plus intéressant, de mesurer en quoi la figure peut être également *contextualisante* : la figure fait-elle le genre de discours par exemple ? En quoi contribue-t-elle à construire un monde de référence qui en retour la légitime et la rend pertinente ? Quelle est sa dynamique textuelle, et contribue-t-elle à faire évoluer un/le contexte ?

Nous partirons d'une conception commune du contexte, sans pour autant nous empêcher d'affiner chemin faisant ce que chacun(e) entend par ce terme : le contexte d'emploi, construit et non donné, à la fois polyvalent et polyfonctionnel, est « conçu comme un ensemble dynamique d'unités en interaction les unes avec les autres, visant à construire un énoncé pertinent et porteur de sens » (S. Mellet, bilan d'activité). Plusieurs axes peuvent être envisagés, qui reprennent les suggestions formulées par les membres de l'équipe souhaitant s'impliquer dans ce programme et ménagent une place à différentes approches disciplinaires. Ces axes pourraient correspondre aux types possibles de contextualisation (linguistique, discursive, générique, socio-historique, esthétique) et répondre aux questions suivantes qui ont émergé :

*Comment la contextualisation éclaire-t-elle la question des limites du champ des figures ?*

Cendrine Pagani s'intéresse notamment aux jugements métalinguistiques qui classent selon les contextes les configurations comme figures ou comme solécismes, posture reconduite depuis l'Antiquité en ce qui concerne les figures de construction comme l'ont montré B. Colombat et M. Baratin. Véronique Montagne remarque quant à elle que les textes théoriques ont, à une certaine époque, essayé de théoriser les critères (génériques, éthiques, discursifs) qui permettraient de circonscrire le champ des figures.

*Comment les variations de contextualisation peuvent-elles influencer la réception des figures ?*

Anna Jaubert note qu'une contextualisation large – prise en compte d'attendus esthétiques et discursifs d'une époque – a promu l'interprétation litotique du célèbre *Va, je ne te hais point*, qu'une contextualisation étroite bat en brèche au profit d'une interprétation « littérale ». Parallèlement à cette minoration de la figure par la prise en compte du contexte, on peut évoquer les cas inverses de maximisation figurale par

recontextualisation. Ainsi, Véronique Magri s'intéresse à la frontière entre absurde et figuralité : certains énoncés sont en effet « absurdes » hors de toute contextualisation et ne prennent de sens figural que recontextualisés. Plus généralement, l'effet figural est-il toujours/seulement contextuel ?

*Qu'apporte la contextualisation à l'analyse des figures ? Les figures peuvent-elles construire et promouvoir un contexte ? Quel rôle jouent-elles dans les déterminations génériques ?*

Si la figure peut être décrite dans sa singularité et son fonctionnement énonciatif propre, quelles sont les interactions entre les propriétés de la figure et le contexte ? En quoi la figure construit-elle sa place dans le discours et en quoi peut-elle activer un contexte ? Lucile Gaudin-Bordes et Geneviève Salvan défendent l'idée que la figure révèle un ajustement perceptif, référentiel, énonciatif et permet à l'énonciateur de dire « au plus près » ce qu'il veut dire, autrement dit que la figure s'ajuste au contexte et l'exploite, voire l'informe. Michèle Monte réfléchit au statut évolutif de la métaphore dans le langage poétique (marque de poéticité, banalisation, relittéralisation), Marie-Albane Watine réfléchit quant à elle aux mutations qui s'opèrent entre figures du parlé et figures du style vocal en contexte littéraire.

*Inversement, la décontextualisation d'une figure peut-elle conduire à des effets de « défiguralisation » ?*

Peut-on reprendre la question de la mort d'une figure à la lumière du contexte, en faisant l'hypothèse qu'une figure « morte » ou lexicalisée serait une figure « dé- ou a- contextualisée », pour laquelle on n'a plus besoin du contexte pour construire un référent et qui aurait passé en langue (voir les métaphores devenues catachrèses, ou d'autres figures telles que l'oxymore *silence assourdissant*, l'hypallage *odeur verte*, le pléonasme *descendre en bas*) ? En quoi les figures, certes « entre langue et discours », éclairent-elles non seulement le passage de la langue au discours, mais également le mouvement du discours à la langue ?

*Certaines figures peuvent-elles échapper au contexte, à la contextualisation, en étant reconnaissables « hors contexte » ?*

Le contexte est-il toujours décisif, quel que soit le type de figure (les figures de mots seraient par exemple moins « impactées » par le contexte que les figures de pensée) ? Ne peut-on revoir la distinction figures de mots / figures de pensée, de même que la distinction entre figures microstructurales et figures macrostructurales (G. Molinié), à la lumière du contexte ?

Le projet « figure et contextualisation » ne vise donc pas à valider l'importance du contexte dans la production et la réception des figures, mais à interroger les relations réciproques entre figures et contexte, à évaluer le processus de fonctionnalisation des figures par leur mise en texte et l'action des figures sur le discours / le texte qui les fait émerger.

On notera que chacun de ces axes s'appuient non seulement sur des forces internes reconnues (nombreuses publications antérieures et en cours), mais aussi sur des collaborations extérieures stimulantes (avec « Praxiling », Montpellier 3, pour l'axe 1, avec « Sens et Texte », Paris 4, pour l'axe 2).

### 3. Objectifs généraux de l'équipe

- Continuer les efforts déjà entrepris pour augmenter, dans notre production bibliographique, la part des articles publiés dans de grandes revues à comité de lecture. Nous restons cependant attachés à la publication d'ouvrages collectifs permettant de rassembler sous une seule entrée bibliographique le résultat de nos travaux d'équipe et de les faire connaître en tant que tels dans la communauté.
- Attirer davantage de doctorants : les soutenances de HDR devraient y contribuer, ainsi que l'adossement au laboratoire BCL et, plus précisément, à l'équipe « Énonciation » d'un parcours de master mutualisé entre le département de sciences du langage et le département de lettres modernes.
- Intensifier le dialogue et les collaborations avec les autres équipes du laboratoire : l'étude des figures et de leur contextualisation, aussi bien que celle de l'altérité énonciative et du dialogisme se prêtent particulièrement bien aux interactions avec l'équipe de logométrie et analyse du discours, d'où la double affiliation de deux chercheuses et d'une doctorante ; mais au lieu de reposer simplement sur la double compétence de quelques individus, les interactions pourraient être généralisées sur la base des séminaires communs qui ont commencé à se mettre en place et seront systématisés. D'autres collaborations sont aussi envisageables avec les psychologues et la nouvelle équipe de cognition, l'étude des figures en discours pouvant notamment apporter un enrichissement certain (mais aussi une complexité pour l'instant difficile à modéliser) au travail sur les associations lexicales.

## Equipe 3

### Dialectologie, diachronie et phonologie

(resp. Elisabetta Carpitelli)

#### 1. Composition

NOM	Statut	participation	dates utiles
<b>Membres permanents</b>			
<b>HDR</b>			
CARPITELLI, Elisabetta	PR	100%	arrivée sept. 2010
DALBERA Jean-Philippe	PR	100%	retraite en 2011 ou 2012
OLIVIERI Michèle	MCF HDR	75%	
SCHEER Tobias	DR	50%	
<b>non-HDR</b>			
BARILLOT Xavier	MCF	75%	
LIBROVA Bohdana	MCF	50%	arrivée oct. 2007
MOLINU Lucia	MCF	100%	arrivée mars 2009
RANUCCI Jean-Claude	MCF	100%	retraite en 2012 ou 2013
RIZZOLO Olivier	MCF	75%	
<b>BIATOS, ITA</b>			
BRUN-TRIGAUD Guylaine	IE2	100%	
GEORGES, Pierre-Aurélien	IR2	40%	arrivée déc. 2007
<b>Membres associés</b>			
(cf. section 2.7 du projet général sur ce statut)			
DALBERA-STEFANAGGI M. José	PR Univ. Corse	associée	
MEDORI Stella	MCF Univ. Corse	associée	
POLI Muriel	MCF Univ. Corse	associée	
<b>Membres non-permanents</b>			
CARATINI Emilie	post-doc interne	100%	soutenance déc. 2009
PALASIS, Katérina	post-doc interne	50%	soutenance déc. 2009
BUCCI Jonathan	doctorant	100%	depuis oct 2009
BEN SI SAID Samir	doctorant	100%	depuis oct 2009
CASAGRANDE Sylvain	doctorant	100%	depuis oct 2008
CURCULESCU, Elena	doctorante	100%	depuis oct 2010
MALFATTO Albert	doctorant	100%	depuis oct 2008

Au niveau institutionnel, l'équipe a deux objectifs majeurs pour la période qui s'ouvre à partir de la rentrée 2010:

1. à la tête de l'équipe et de la dialectologie niçoise, réussir le passage de témoin entre Jean-Philippe Dalbera et Elisabetta Carpitelli qui a été recrutée en juillet 2010 sur un poste au fil de l'eau.
2. réussir la collaboration avec la nouvelle équipe "Langage et Sciences Cognitives" dont le noyau est constitué par les psychologues qui rejoignent le laboratoire en 2012 mais qui puise, au sein des anciennes équipes, surtout dans DDP: les générativistes de l'équipe ont une affinité naturelle et s'engagent dans la nouvelle équipe: T. Scheer (50%), O. Rizzolo (25%), M. Olivieri (25%), X. Barillot (25%).

#### I. Dialectologie

Les projets des dialectologues s'inscrivent dans une double perspective, à la fois de continuité et d'innovation. D'une part, les chantiers engagés, tels que le THESOC et son module morpho-syntaxique seront poursuivis et les grands axes de recherche conservés, et d'autre part, une nouvelle dynamique se met en place avec une plus grande place faite à la morphologie, des développements originaux en reconstruction lexicale



(Nouvelle Étymologie) et syntaxique et une réflexion en collaboration avec les phonologues et les psychologues sur l'architecture cognitive de la grammaire.

## 1. Le "Thesaurus occitan" (THESOC)

### *Enrichissement de la base de données*

Une fois que la saisie des données des atlas sera achevée (ALLY, ALMC, ALG restent encore incomplètement intégrés), nous projetons de compléter les data du THESOC par la saisie des données inédites des carnets d'enquêtes de l'ALLOc et de l'ALLOr, que nous venons de numériser. Ces carnets contiennent en effet des informations de première importance, qui n'ont pu être cartographiées au moment de la publication des atlas et qui nous font défaut aujourd'hui. En ce qui concerne le lexique, les données qui n'ont pas été publiées viendront compléter les cartes actuellement lacunaires du THESOC et permettront de vérifier les hypothèses formulées dans les travaux antérieurs. Un deuxième volet associé à ce nouvel ensemble de datas consistera à lemmatiser les nouvelles données et compléter les références étymologiques.

Mais la partie la plus intéressante et la plus prometteuse en termes d'exploitation des saisies à venir sont les données morphologiques et morpho-syntaxiques: bien que présentes dans ces carnets, elles n'ont pas été considérées par le passé et ne manqueront pas d'apporter de nouveaux éléments à la réflexion sur les changements morpho-syntaxiques. Ces données viendront en complément du corpus de morphologie verbale occitane que S. Casagrande est en train de constituer pour sa thèse et des résultats des enquêtes effectuées avec nos collaborateurs allemands (Université de Konstanz, resp. G.A. Kaiser) au sein du programme Procope dans le nord de l'aire occitane (novembre 2010).

Parallèlement, il s'agira de compléter la numérisation des illustrations des atlas et de les implémenter dans la base, ainsi que la numérisation des bandes sonores de l'ALLOc et de l'ALAL, l'objectif étant une mise en ligne de la totalité des données.

### *Ingénierie de la base de données*

Le premier objectif technologique est de finaliser le travail de refonte informatique de la base lexicale du Thesoc, en vue de son portage sur un environnement OpenSource. Cette opération présuppose plusieurs lourdes tâches préalables, notamment le passage au standard Unicode des transcriptions phonétiques et la réécriture au standard SQL des requêtes du moteur de la base de données. Dans cette perspective de modernisation, nous projetons également d'améliorer l'interface utilisateur et l'ergonomie du logiciel, ce qui nécessitera une réorganisation des menus et de la disposition des nombreux écrans de la base lexicale.

D'autre part, nous voulons implémenter de nouvelles fonctionnalités, à la fois dans le Module MorphoSyntaxique et dans la base lexicale. Ainsi, nous allons par exemple intégrer l'utilisation des formules "LTA" qui viennent d'être ajoutées au THESOC dans les différentes fonctionnalités de recherche et de consultation de la base. Par ailleurs, le site web va être complètement renouvelé. Outre l'apport des nouvelles données, il offrira de nouvelles possibilités, dans la recherche et la consultation.

Enfin, à plus long terme, nous envisageons la création d'un module de cartographie interactive en ligne (accessible à la fois aux internautes sur le site web du THESOC et en interne dans le logiciel THESOC) qui sera fonctionnel tant dans la base lexicale que dans MMS (Module Morpho-Syntaxique).

### *Retombées attendues*

Plus notre outil est performant, plus nous sommes sollicités et notre renommée internationale est croissante. Le THESOC est en effet une base de données très complète et dotée d'outils originaux nés d'une étroite collaboration entre linguistes et informaticiens, qui a la chance aujourd'hui d'être sous la responsabilité technologique d'un ingénieur à la fois informaticien et linguiste (P-A Georges). Le THESOC est en passe de devenir un modèle, nous recevons de nombreuses visites (en projet à court terme celles notamment d'A. Ledgeway, Université de Cambridge, UK, M.-O. Hinzelin, Université de Hambourg, Allemagne, F. Hack, Université d'Oxford, UK, R. D'Alessandro, Université de Leiden, Pays-Bas) et, à chaque présentation dans les colloques internationaux, des collègues français et étrangers d'horizons divers (tels que A. Dagnac, Université de Toulouse II - Le Mirail, B. Vance, Université de Bloomington, Indiana, USA, D. Heap, Université de London, Canada, I. Fernández-Ordóñez, Université de Madrid, Espagne) nous demandent déjà si on peut nous acheter MMS.

## 2. Reconstruction

### *Lexique*

Les recherches initiées par Jean-Philippe Dalbera dans le domaine de la reconstruction du lexique roman selon l'approche de la Nouvelle Étymologie [opérant de manière autonome sur chacun des deux plans du signe linguistique, phonique et sémantique] seront poursuivies en étendant la démarche comparative à

d'autres langues de l'espace roman, avec une attention particulière réservée aux domaines gallo-roman et italo-roman. Le déplacement d'une partie des riches archives multilingues (zoonymie, phytonymie, désignations de phénomènes atmosphériques, de maladies) de l'*Atlas Linguistique Roman* (ALiR) à BCL, grâce au recrutement d'E. Carpitelli qui coordonne ce grand chantier international avec lequel par ailleurs les dialectologues niçois ont une collaboration permanente depuis l'origine du projet, permettra de mesurer la portée de la Nouvelle Étymologie sur un éventail très large et diversifié de matériaux dialectaux encore inédits et pour l'instant presque inexploités dans le cadre de la reconstruction lexicale, provenant des différents domaines linguistiques de la Romania.

L'important réseau de chercheurs de l'ALiR et son lien étroit avec celui de l'*Atlas Linguarum Europae*, projet international de géolinguistique à l'échelle continentale, assureront la promotion des résultats dans le cadre des colloques annuels qui se tiennent dans les différents centres européens de recherche en dialectologie qui participent à l'entreprise. La participation des dialectologues niçois aux deux prochaines rencontres internationales de l'ALiR, organisées par les Universités de l'Aquila (Italie) et de Barcelone (Espagne), est déjà prévue ainsi que la publication de quatre contributions acceptées pour le volume 2.c de l'ALiR, en préparation, essentiellement dédié à l'ornithonymie dialectale. Un cycle de conférences (20h) de Dialectologie et géolinguistique romanes à l'École Pratique de Hautes Études sera assuré par E. Carpitelli à partir du mois de février 2011 – dans la continuité du cycle proposé jusqu'à l'année dernière par J.-Ph. Dalbera – contribuera à promouvoir les recherches niçoises sur la diachronie du lexique dialectal roman notamment dans les secteurs de la phytonymie, du vocabulaire des phénomènes atmosphérique, des maladies mais surtout de la zoonymie. Dans ce dernier domaine s'inscrit notamment la thèse d'Albert Malfatto, financée par une bourse du Conseil Régional PACA, consacrée à l'étude des dénominations des animaux domestiques.

### Syntaxe

Les projets de M. Olivieri et ses étudiants se situent également dans le prolongement des travaux antérieurs. D'une part, les travaux sur les clitiques sujets doivent se poursuivre, en mettant à l'épreuve de nouveaux faits les hypothèses élaborées. Ce projet est l'objet de notre étroite collaboration avec nos collègues étrangers, notamment G. Kaiser de l'Université de Konstanz (Allemagne). Ainsi, les données que nous récoltons actuellement devraient nous permettre, à terme, de vérifier si l'analyse de l'émergence des clitiques en termes de traits est valide ou si elle doit être amendée. L'hypothèse du cycle motivationnel fonctionnel demande elle aussi à être approfondie, de même que le fonctionnement des clitiques explétifs. Le clitique qui apparaît avec les verbes météorologiques doit en particulier être distingué de celui que l'on trouve avec les verbes du type *falloir*, certains parlant utilisant des formes différentes pour les deux contextes. Toutes ces hypothèses de travail sont directement reliées aux questions posées par les travaux menés au sein de la nouvelle équipe "Langage et Sciences Cognitives". En effet, le problème central du statut du clitique (entre morphologie, phonologie et syntaxe) renvoie aux interrogations sur le fonctionnement des différents modules de la grammaire et leur(s) interaction(s) (cf. infra II.4). En outre, il s'agit de déterminer dans le cadre du programme biolinguistique actuel de la grammaire générative (cf. notamment Chomsky 2005) ce qui relève des Principes universels de la Faculté de Langage (et donc est spécifiquement linguistique) et ce qui est une émanation de fonctions cognitives plus générales (i.e. non spécifiquement linguistiques).

D'autre part, les problèmes posés par les clitiques sujets et les réponses apportées ont réveillé des questions concernant les clitiques objets, notamment celle de leur statut. La notion de statut variable (entre affixe morphologique et argument syntaxique) des clitiques sujets peut-elle éclairer le comportement et l'analyse des clitiques objets ? De plus, ceux-ci entrent dans une chaîne, une séquence de clitiques dans laquelle l'ordre des pronoms varie. Dans le cadre de notre collaboration avec l'*Atlas Linguistique Roman* (ALiR) accrue encore par l'arrivée à BCL d' E. Carpitelli, nous avons le projet de développer un volet morpho-syntaxique dans cet atlas que nous allons inaugurer avec l'ordre relatif des clitiques accusatifs et datifs dans la Romania. Ces travaux sont actuellement initiés avec M. Lobo et E. Carrilho de l'Université de Lisbonne, et nous envisageons d'étendre cette collaboration, notamment avec l'Université de Toulouse (A. Dagnac, P. Sauzet), l'Université de Konstanz (G.A. Kaiser) l'Université de Madrid (I. Fernández-Ordóñez) et l'Université de Firenze (L.M. Savoia et M.R. Manzini).

Mais d'autres pronoms retiennent également notre attention. Il s'agit des relatifs, récemment revenus sur le devant de la scène avec l'article de Rizzi & Shlonsky (2005) "Strategies of Subject Extraction", dont les relations avec les clitiques demandent à être approfondies. Dans quelle mesure les propositions émises à propos des clitiques pourront-elles fournir des clés pour l'étude des relatifs ? Et lorsque l'on se penche sur le sort des relatives, on rencontre inévitablement les interrogatives qui présentent en outre des cas de clitique sujet postverbal.

La boucle est en quelque sorte bouclée et la question centrale qui est posée concerne l'ordre des mots. En latin, on prétend traditionnellement que l'ordre est libre mais en réalité, il ne l'est pas complètement et des

contraintes existent, qui restent à déterminer. Cependant, du fait de la richesse de la flexion en latin, l'ordre reste plus variable que dans les dialectes modernes. Comment s'est produit le passage d'un système à l'autre, qu'est-ce qui relève de la Grammaire Universelle, quels paramètres et quels mécanismes sont en jeu ? De nombreuses hypothèses classiques ont été formulées mais qui restent à préciser et à formaliser dans notre cadre théorique et de récents travaux<sup>5</sup> menés par les générativistes permettent de reconsidérer les faits. En outre, nous pensons que l'étude de la variation dialectale apportera des éléments essentiels à cette étude, afin de confirmer ou d'infirmer les analyses proposées, en fournissant la photographie de systèmes diachroniquement intermédiaires.

### *Morphologie*

Le volet Morphologie est plus récent. Il s'inscrit dans la dynamique amorcée et qui doit se confirmer dans les années qui viennent. Plusieurs forces se conjuguent ici.

D'une part, concernant la morphologie nominale, les travaux sur la zoonymie et la phytonymie romanes annoncés supra concerneront également les processus de morphosyntaxe et de morphologie lexicale (notamment dérivation et composition), afin d'identifier et de décrire les stratégies de création spécifiques à ces deux domaines.

D'autre part, un plus vaste chantier sera consacré à la morphologie verbale. Le rattachement récent de L. Molinu à BCL (mars 2009) a permis de compléter le paysage dialectologique niçois avec ses travaux sur le sarde. Ses recherches portent sur l'analyse morpho-phonologique du sarde et sur la description comparative des langues romanes. Là encore, et toujours en collaboration avec l'*Atlas Linguistique Roman*, il s'agira d'examiner la variation diatopique et diachronique afin d'essayer de dégager des constantes, des principes qui puissent rendre compte du foisonnement des réalisations de surface en synchronie et de l'évolution des formes dans un espace fragmenté mais cohérent. A partir de la recherche menée actuellement sur une sous-classe d'infinitifs en sarde d'un point de vue diachronique pour essayer de comprendre les processus phonologiques et les pressions analogiques qui ont déterminé leur formation et leur diffusion, le projet sera consacré à la morphologie verbale du sarde, aspect souvent négligé dans la description de cette langue. Ce travail d'envergure a pour objectif l'élaboration d'un Atlas morphologique du sarde, en collaboration avec Simone Pisano (Université de Pise) qui coordonne les enquêtes de l'*Atlante Toponomastico Sardo* (ATS) à l'Université de Cagliari. Il va de soi qu'un tel projet bénéficiera de l'expérience acquise par l'équipe du THESOC dans le traitement informatique des données dialectales et il permettra surtout de comparer les données occitanes avec les résultats des enquêtes d'un autre espace dialectal, afin d'enrichir la réflexion sur l'organisation des formes verbales à l'intérieur des paradigmes.

En effet, le THESOC aussi s'enrichit de nouvelles données morphologiques et des recherches de S. Casagrande pour sa thèse. Celui-ci étudie les paradigmes verbaux dans une large aire occitane et son travail de reconstruction rencontre des échos, non seulement avec celui de L. Molinu en morphologie, mais aussi avec celui de M. Olivieri en syntaxe. En effet, la présence ou non de clitiques sujets dans les paradigmes verbaux nous intéresse de différents points de vue, leur statut étant – encore – éminemment controversé.

## **II. Phonologie**

Comme pour la seconde moitié de la période de référence du bilan (2006 - juin 2010), l'activité des phonologues pour la période à partir de la rentrée 2010 sera fédérée, au niveau thématique, par le projet sur les gabarits: chacun des phonologues le déclinera à partir des données qui lui sont familières (cf. infra), et l'édition de l'ouvrage collectif issu du workshop sur les gabarits d'OCP-7 (cf. bilan) constituera le fil conducteur. Enfin, les collaborations extérieures seront également définies en bonne partie par le travail sur les gabarits (cf. infra).

Au niveau théorique, c'est l'approche partagée du modèle CVCV (ou CV strict) qui fédère les travaux des phonologues: Nice est une composante importante de la Phonologie de Gouvernement en France, et probablement le centre français en phonologie qui est le plus visible au niveau international.

Par ailleurs, les centres d'intérêt des phonologues se structurent en fonction de leurs terrains d'application respectifs: le slave (O. Rizzolo, T. Scheer), le couchitique (X. Barillot), les jeux de langage (O. Rizzolo), la diachronie du français ou phonétique historique (T. Scheer), l'interface avec la morpho-syntaxe (T. Scheer).

Le pont avec les dialectologues sera fait d'une part par l'intérêt partagé pour la diachronie, d'autre part par des thèses en cours qui conjuguent la variation diatopique et la théorie phonologique. Sont concernées les thèses de J. Bucci (sur des dialectes du Sud de l'Italie, début: 2009), de S. Ben Si Saïd (sur le berbère kabyle,

<sup>5</sup> Cf. entre autres l'ouvrage en préparation d'A. Ledgeway, *From Latin to Romance*, Oxford University Press.

début: 2009), d'E. Curculescu (sur les dialectes espagnols de l'Andalousie, début: 2010) et aussi de S. Casagrande (sur les dialectes occitans, début: 2008). En ce qui concerne la diachronie, c'est en particulier celle du français et des dialectes occitans qui généreront des collaborations: 1) des articles co-signés sur les dialectes et la diachronie du français (T. Scheer); 2) l'exportation du savoir-faire technique (BDD, cartographie) et méthodologique vers le couchitique et les langues sémitiques à la périphérie Est de l'arabe (X. Barillot).

Enfin, précisons que la collaboration des phonologues avec la nouvelle équipe "Langage et Sciences Cognitives" sera située à cheval entre les deux équipes: la base sera phonologique (donc DDP), mais la méthodologie et les buts poursuivis seront situés en psychologie/Sciences Cognitives. C'est la raison pour laquelle les projets afférents apparaissent dans le projet des deux équipes.

### 1. Gabarits

Les gabarits seront abordés par deux angles d'attaque, l'un ayant trait à leur réalité grammaticale et psychologique, l'autre posant la question de savoir à quoi ils servent. Il existe un débat controversé depuis une décennie à peu près, et en rapport direct avec l'avènement de la théorie de l'optimalité, qui s'articule autour de la question de savoir si les gabarits existent vraiment – en tant qu'objet grammatical et cognitif. Traditionnellement les gabarits sont admis en tant qu'instrument de la grammaire du sémitique autant par les grammairens arabes depuis le 10<sup>e</sup> siècle que par les néogrammairiens du 19<sup>e</sup> siècle et les études structuralistes et génératives du 20<sup>e</sup> siècle. Or des auteurs tels que Ratcliffe (1997), Ussishkin (1999) et Bat-El (2003) entre autres cherchent à réfuter la notion de racine/gabarit/schème dans les langues sémitiques et proposent d'autres moyens pour rendre compte du comportement morphologique de ces langues: il s'agit de montrer que les langues sémitiques n'ont rien de bien particulier par rapport aux langues qui ne sont pas réputées être gabaritiques. L'idée est de prendre le contre-pied de ce mouvement en montrant non pas que le fonctionnement des langues sémitiques peut se réduire aux mécanismes trouvés ailleurs, mais au contraire que des langues non-sémitiques présentent également une activité gabaritique. L'universalité de la faculté de langage est ainsi garantie, mais en établissant le gabarit en son sein, plutôt que de le rayer.

Afin de concrétiser cette hypothèse de travail et d'insérer les travaux niçois dans le concert des voix qui cherchent à montrer la validité grammaticale et cognitive des gabarits (p.ex. Prunet et al. 2000, 2008, en France S. Bendjaballah, J. Lowenstamm à Paris 7, M. Lahrouchi à Paris 8), les moyens sont de deux ordres. D'une part X. Barillot est engagé dans un travail empirique de longue haleine qu'il achèvera: sur la base de la saisie systématique du dictionnaire le plus important de l'arabe (Kazimirski 1860, les deux tiers des 83000 entrées que comporte ce dictionnaire ont déjà été implémentés dans une BDD), il reconsidérera le phénomène gabaritique en étudiant la question de savoir dans quelle mesure les gabarits constituent un morphème en arabe: la catégorie grammaticale et la valeur sémantico-grammaticale du schème ne sont-elles pas dérivables des segments vocaliques insérés dans la racine et des éventuels infixes consonantiques?

D'autre part, les phonologues continueront (cf. bilan) à mettre en évidence le fait que des langues à morphologie purement concaténative (le sémitique étant basé sur une morphologie non-concaténative) présentent également une activité gabaritique. X. Barillot explorera les familles non-sémitiques du phylum afro-asiatique, principalement le couchitique et le tchadique. Il a certes déjà mis en évidence des contraintes gabaritiques en somali, à un endroit de la morphologie où on ne s'y attendait pas (des gabarits de types très divers, fixes ou variables, dans des *processus concaténatifs*), mais une meilleure compréhension de ce type de gabarits nécessite d'étudier comment font les langues proches du somali. S'il s'avère que le somali est un cas isolé, il faudra comprendre pourquoi. Ce projet sera soutenu par un important travail de constitution de corpus qui est en cours depuis quelques années (langues afar, oromo, rendille, boni, sidamo, arbore, dasenech, etc. pour le couchitique).

En dehors des langues afro-asiatiques, trois pistes empiriques seront poursuivies: O. Rizzolo travaillera sur l'activité gabaritique dans les jeux de langage, X. Barillot est engagé dans les langues anciennes et leur diachronie, et T. Scheer s'occupe du tchèque et du slovaque.

X. Barillot a déjà commencé à étudier certaines altérations de la racine en latin, grec et donc ultimement en proto-indo-européen, comme par exemple les métathèses (inversions de consonnes à l'intérieur de la racine) ou les infixes, alternances qui pourraient recevoir une explication gabaritique. L'ambition à long terme est de découvrir le rôle possible des gabarits dans l'évolution des langues. S'ajoute à cela la recherche de gabarits dans les noms de personnes ou les noms de lieux, notamment dans les langues celtes et en anglais. Enfin, O. Rizzolo a des raisons de penser que les jeux de langage et notamment le verlan mettent en œuvre des stratégies gabaritiques (cf. point sur les jeux de langage infra).

T. Scheer sera davantage concerné par le fonctionnement de l'activité gabaritique dans l'architecture de la grammaire: les contraintes d'espace et leur association à une catégorie morpho-sémantique (p.ex. "la forme II en arabe (intensif/itératif) a la forme CVC<sub>i</sub>C<sub>i</sub>VC-") sont spécifiées en-dehors de la phonologie. La computation phonologique ne fait que mettre en rapport ("associer") les éléments mélodiques (consonnes et

voyelles) avec la structure gabaritique. Classiquement, les gabarits sont considérés comme des items lexicaux stockés tels quels (McCarthy 1979): en arabe par exemple il va y avoir autant d'entrées lexicales qu'il y a de schèmes grammaticaux (forme I: CVCVC-, forme II: CVC<sub>i</sub>VC- etc.). L'association d'un objet lexical à une catégorie morpho-sémantique ("intensif/itératif") est facile puisqu'il s'agit là de la simple relation entre une structure morpho-syntaxique et son représentant phonologique (Vocabulary Insertion). Or ce système ne peut pas fonctionner avec les gabarits à forme variable que l'on trouve dans les langues non-sémitiques. La distinction entre variable shape templates et fixed shape templates a été mise en évidence par Barillot & Scheer (2008): la dernière catégorie impose un volume global à la forme gabaritique (p.ex. "trois mores" en tchèque), qui peut ensuite être réalisé de plusieurs manières (p.ex. CVCVCV ou CVVCV en tchèque). Par conséquent il n'existe pas d'objet unique que l'on puisse stocker dans le lexique, et que la morpho-syntaxe puisse appeler lors de l'insertion du vocabulaire. Il faut recourir à un système plus complexe où la contrainte globale de volume soit "comprise" par la phonologie, et matérialisable. Il s'agit donc d'un problème au cœur de l'interface entre la morpho-syntaxe et la phonologie. Une piste qui sera explorée est celle de sortir de l'idée que les gabarits sont nécessairement des objets stockés dans le lexique: l'alternative, peu orthodoxe il est vrai, est que l'exponent phonologique de la structure morpho-syntaxique est de la *computation* phonologique. Il y aurait ainsi un "ordre" donné au système computationnel d'infléchir son activité, juste pour la catégorie en question, pour produire le volume "3 mores". Dans un système à contraintes hiérarchisées de type OT, cela revient à ce que la morpho-syntaxe "intervienne" en promouvant la contrainte "3 mores".

Enfin, X. Barillot, O. Rizzolo et T. Scheer poursuivront la piste lancée dans le projet ANR non retenu (cf. bilan) concernant le statut cognitif des gabarits: à quoi servent-ils? Ont-ils un rôle grammatical, expressif, mnémotechnique? Peuvent-ils intervenir n'importe où dans le langage, lors de l'acquisition, dans la morphologie ou la syntaxe, dans la diachronie, etc.? Enfin, ont-ils une existence ou un impact en dehors du langage, au niveau des sens, du fonctionnement de la mémoire par exemple? L'hypothèse ici, en application du programme minimaliste (3rd factor explanations), est que la raison d'être des gabarits est extra-linguistique. Une indication forte dans ce sens est le fait, avéré en acquisition (p.ex. Vihman 2007), que les enfants se fabriquent des gabarits en l'absence complète de stimulus, i.e. en étant exposés à des langues non-gabaritiques. Ce volet a donc trait à des domaines cognitifs voisins et sera abordé dans le cadre de la nouvelle équipe "Langage et Sciences Cognitives".

Il est utile de préciser que le travail de X. Barillot et d'O. Rizzolo repose une base empirique constituée en corpus informatisé qui se veut le plus exhaustif possible (à partir de dictionnaires validés si possible par des informateurs pour X. Barillot, à partir d'un travail d'enquête avec des informateurs pour O. Rizzolo). Cela nécessite un travail conséquent de saisie et de terrain qui est en cours et sera poursuivi.

Les travaux d'un doctorant, S. Ben Si Saïd, ainsi que d'une post-doc interne, E. Caratini, sont en plein dans la thématique gabaritique: le sujet de la thèse du premier est le fonctionnement gabaritique du système de nombre (sg-pl) en berbère kabyle. E. Caratini, prenant le cas de l'allemand comme point de départ de son étude, s'intéressera à la formation des hypocoristiques bisyllabiques de type *Mutti* «maman» (all.) ou *Willi* (anglais) dans les langues germaniques et au-delà. Afin de réaliser la première partie de cette étude (all.), E. Caratini déposera une demande de financement franco-allemand auprès de la Fondation des Maisons des Sciences de l'Homme et de la Fondation Fritz Thyssen, en collaboration avec J. Trommer de l'Université de Leipzig/RFA. Cette étude cherchera à établir des collaborations avec d'autres chercheurs qui ont travaillé sur les hypocoristiques allemands (S. Lappe [U. de Siegen], L. Downing [Zentrum für Allgemeine Sprachwissenschaft, Berlin], B. Alber [U. de Vérone]). Elle vise des retombées dans trois domaines: la typologie des hypocoristiques (restrictions (morpho)phonologiques présidant à leur formation), l'interface entre la phonologie et la morphologie (et donc, la place de la morphologie dans l'architecture de la grammaire) et les gabarits (en lien, notamment, avec l'acquisition de la phonologie).

## 2. Jeux et erreurs de langage

Depuis quelques années O. Rizzolo développe une thématique originale qui est déshéritée en France depuis que Marc Plénat a d'abord changé d'orientation et est ensuite parti à la retraite: les langages ludiques et secrètes, notamment sur la base du français (verlan, louchébem, javanais etc.). Après avoir consacré le gros de la période du bilan (2006-juin 2010) à l'étude de deux jeux de langage serbo-croates, O. Rizzolo fera un retour au domaine français en se focalisant sur le verlan, terrain défriché dans les années 80 et 90. Certaines des hypothèses qu'il a émises pour le serbo-croate ont déjà été confrontées aux données du français, avec des résultats encourageants. Il apparaît notamment que, au contraire de ce qui a été avancé dans les travaux classiques, la syllabe n'est pas le constituant manipulé par ces jeux de langage dont le mécanisme central est l'inversion. La question reste de savoir quelle est cette unité qui est déplacée dans ce type de jeux. Le cadre théorique commun aux phonologues niçois, CV strict, prédit justement que le constituant minimal

manipulable est plus petit qu'une syllabe, i.e. une paire attaque/noyau (CV). Les données du serbo-croate, ainsi que les résultats déjà obtenus sur le verlan français, vont dans ce sens.

Afin de voir si l'analyse des jeux serbo-croates sur la base de l'unité minimale CV est réellement applicable au verlan français, O. Rizzolo constituera une base de données. Un corpus est disponible dans Plénat (1992), mais il s'agit d'un corpus brut: les données ne sont pas organisées, aucun filtre n'a été mis en place et surtout, et les données demandent à être informatisées. Par ailleurs, elles doivent être actualisées: un certain nombre de termes verlan a été créé depuis, d'autres sont hors d'usage.

En complément et toujours dans le domaine des données issues de "déformations" du français, O. Rizzolo étudiera les erreurs de langage. Un corpus conséquent est disponible dans Rossi & Peter-Defare (1998), mais le prisme analytique choisi par ces auteurs n'est pas celui de l'architecture des représentations phonologiques. L'intérêt de ces données est qu'elles reposent, comme le verlan, sur des unités de la chaîne parlée qui sont déplacées – cette fois accidentellement. Leur témoignage devrait donc permettre de jouer un rôle d'arbitre au regard des hypothèses émises à partir du verlan.

Enfin, une question fondamentale et récurrente se pose pour l'étude des activités ludiques et secrètes du langage: les formes "déformées" étudiées sont-elles fabriquées au moyen d'une activité grammaticale à chaque fois qu'elle sont énoncées, ou représentent-elles simplement la prononciation d'une forme stockée dans le lexique? Dans la seconde hypothèse, quand la lexicalisation a-t-elle lieu, et comment faire la différence entre les occurrences d'une forme "déformée" qui sont le résultat d'une activité grammaticale, et celles qui sont simplement le résultat d'un accès lexical? S'il est évident que le degré d'expertise et de pratique d'un locuteur joue un rôle important (des items tels "laisse béton" < laisse tomber ou "barjot" < jobard ont trouvé leur chemin dans le lexique du français de tous les locuteurs, y compris de ceux qui ne pratiquent pas le verlan), le mode de transmission entre locuteurs de verlan reste à élucider, ainsi que le seuil de pratique/expertise à partir duquel un locuteur cesse de fonctionner par accès lexical et produit les formes par activité grammaticale.

Ces questions seront étudiées en collaboration avec les psychologues au sein de l'équipe "Langage et Sciences Cognitives". La stratégie ici sera de mettre au point des tests basés sur le temps de réaction qui soient à même de discriminer entre un accès lexical simple et un accès lexical augmenté d'une activité grammaticale.

### 3. Diachronie du français et dialectologie

Depuis plusieurs années, T. Scheer est engagé dans un projet de grande envergure, la confection de la Grande Grammaire Historique du Français (GGHF), dont il est un des éditeurs aux côtés de Chr. Marchello-Nizia, B. Combettes et S. Prévost (cf. bilan). Outre les tâches de gestion du projet (qui compte plus d'une dizaine de collaborateurs), T. Scheer a à charge, avec Ph. Ségéral (Université Paris 7), d'écrire la partie Phonétique Historique. Il s'agit là de condenser et de faire culminer le travail qu'il a entrepris depuis longtemps avec Ph. Ségéral sur la diachronie du français.

Le projet Phonétique Historique de la GGHF, dimensionné à quelque 250 pages, a trois spécificités: l'énorme héritage séculaire depuis les fondateurs de la discipline (Schuchardt, Darmesteter, Bourciez, Fouché, Straka etc.) sera apprécié 1) à travers le filtre de la théorie phonologique moderne (les grammaires existantes s'arrêtent, au mieux, au phonème), 2) à travers du témoignage des dialectes, 3) au moyen d'instruments informatiques. L'idée de ce dernier point est de constituer en bases de données recherchables le stock de mots qui sert de base pour l'observation de l'évolution de la langue. Ce stock de mots est disponible dans les index des grandes grammaires (Fouché etc.), et notamment du FEW. Il s'agit donc de scanner l'ensemble des sources appropriées, de pratiquer la reconnaissance de caractères sur le résultat, et de produire des BDD qui contiennent, colonne par colonne, l'étymon latin, germanique ou autre, la forme afr. si connue, la forme du français moderne et les éventuelles formes dialectales.

La GGHF fonctionne à la vitesse de croisière depuis un an à peu près (les responsables des chapitres sont embauchés), et on prévoit cinq ans au moins jusqu'à son achèvement. Une cinquantaine de pages de la partie Phonétique Historique est rédigée, et T. Scheer consacrer une partie de plus en plus importante de son temps recherche à la rédaction, à la valorisation des résultats et à la gestion de la GGHF dans les années qui viennent.

### 4. Interface, modularité, Sciences Cognitives

T. Scheer travaillera par ailleurs sur ce qui est le fil rouge de son activité depuis 2001, à savoir les livres qu'il publie sur la théorie CVCV et l'interface avec la morpho-syntaxe. Le premier volume a paru en 2004, et le second est bouclé au moment de la rédaction de cette prospective (juin 2010): sa publication est prévue pour l'automne 2010. Ce second volume a été initialement conçu comme une histoire de l'interface entre phonologie et morpho-syntaxe depuis Troubetzkoy (Parties I et II), augmentée du point de vue de l'auteur (et de CVCV) (Partie III). Or la taille du manuscrit (qui a circulé depuis octobre 2008) s'est révélée prohibitive

pour une publication en l'état. Les 871 pages ont donc été coupées en deux morceaux, et le volume qui paraîtra à l'automne chez de Gruyter à Berlin ne comportera que les Parties I et II. Il s'agira donc "uniquement" d'une histoire de l'interface (Partie I) augmentée d'une introduction pour linguistes à la modularité et aux Sciences Cognitives (Interlude) et, enfin, d'une Partie II qui tire les leçons de l'histoire et la prolonge dans l'environnement minimaliste actuel. Le volume se termine, en conclusion, sur l'idée de l'argumentation intermodulaire, i.e. le fait que la théorie des phases, admise en syntaxe minimaliste, établit un canal par lequel transite l'échange entre morpho-syntaxe et phonologie et permet donc de juger une théorie phonologique, établie à partir de données phonologiques uniquement, par son comportement en morpho-syntaxe (et vice versa).

La Partie III du manuscrit qui n'a pas pu être incluse sera donc publiée en tant que livre indépendant, toujours chez de Gruyter: T. Scheer y esquissera sa propre théorie de l'interface. Le repère de celle-ci est la modularité au sens des Sciences Cognitives (Fodor 1983), et elle est basée sur deux idées: 1) l'espace syllabique (une unité CV) est le seul objet représentationnel vecteur d'information morpho-syntaxiques en phonologie (classiquement: phonèmes de joncture, #, constituance prosodique), 2) tout vecteur d'information morpho-syntaxique a son origine dans le lexique. Ce dernier point se résume par le slogan One Channel Translation, car classiquement le matériau phonologique naît par deux traductions différentes: 1) celle qui l'insère à la place des terminaux syntaxiques (Vocabulary Insertion) et donc concerne le corps des morphèmes, et 2) celle qui insère des # etc., de provenance non-lexicale. One Channel Translation dit donc qu'il n'existe qu'une seule source pour tous les objets phonologiques: le lexique. Par conséquent, les objets qui pour une raison ou une autre ne satisfont pas aux critères des objets stockables (tels # etc.) sont exclus de la traduction.

Comme par le passé, l'activité de T. Scheer s'organisera autour de la progression du travail sur ce Vol.3: il en sortira des présentations à des colloques sur des points particuliers et des ensembles de données cohérents, ainsi que des articles. Comme le squelette du Vol.3 est déjà écrit, on peut penser que sa publication, après révision, durera moins longtemps que les 6 ans qu'aura mis le Vol.2 à voir le jour. La préparation du Vol.3 se fera en collaboration avec Michal Starke, de l'Université de Tromsø (Norvège), à la fois pour les aspects syntaxiques de la théorie de l'interface (sandhi externe) et pour ses aspects cognitifs.

C'est ce travail sur l'interface qui a provoqué la réorientation des intérêts de T. Scheer vers l'architecture de la grammaire, la syntaxe, la modularité et donc in fine les Sciences Cognitives. L'arrivée de la nouvelle équipe "Langage et Cognition" ainsi que le projet de formation (Licence et Master) en Sciences Cognitives sont un aspect de cette évolution, et ouvrent des perspectives. Ainsi T. Scheer, F. Lavigne et L. Dumercy travailleront sur les propriétés des modules cognitifs, qui doivent in fine être compatibles avec leur implémentation dans le cerveau (dans la wetware). L'idée est de confirmer l'unité fonctionnelle des modules (ils ne servent qu'à une seule tâche, ce qui n'implique pas qu'ils sont localisés à un seul endroit du cerveau), i.e. leur spécificité de domaine, mais de mettre en doute leur encapsulation informationnelle, autre propriété classique. L'architecture de la grammaire basée sur la théorie des phases y invite, puisque les modules morpho-syntaxique et phonologique échangent des informations dans les deux directions et à répétition au cours de la dérivation d'une phrase. Il s'agit donc ici de travaux qui seront conduits à cheval sur l'équipe DDP et "Langage et Cognition" (cf. le projet de cette équipe).

## 5. Slave et Sciences Cognitives

T. Scheer a un projet en cours en collaboration avec Azra Ali (Université de Huddersfield, Royaume-Uni) et Markéta Ziková (Université de Brno, Rép. Tchèque). Il s'agit de tester le comportement des locuteurs tchécoslovaques lorsqu'ils sont exposés à un stimulus dichotique. Il est établi depuis Cutting (1975) que les locuteurs anglophones perçoivent XY (dans cet ordre) si X arrive par une oreille et Y par l'autre même si Y est présenté bien avant X (jusqu'à 150 ms, ce qui est énorme). Par exemple, les mots "pay" et "lay" sont perçus (fusionnent) en tant que "play" même si "lay" précède "pay". Ceci est dû au "filtre phonologique" des anglophones: il n'existe pas de groupes liquide-obstruante (p.ex. [lp]) en début de mot dans cette langue, si bien que la grammaire de l'anglais "force" les locuteurs de percevoir [pl] quand l'input sensoriel est [lp]. On sait qu'il ne s'agit pas simplement d'un effet "mot existant" vs. "mot non-existant" puisque Cutting a montré que l'effet est aussi robuste avec des mots inventés (nonce words).

Or le tchèque n'a pas les contraintes sur les groupes initiaux de mots que connaît l'anglais: on trouve [pl] autant que [lp] en tchèque. T. Scheer a travaillé depuis longtemps sur les groupes initiaux de mot dans les langues slaves (p.ex. Scheer 2007), et ses résultats prédisent que n'importe quel groupe, réellement existant ou non dans une langue slave, est toléré par la grammaire. La prédiction est donc que l'effet dichotique de Cutting est absent dès lors qu'on expose un tchécoslovaque au test: un [pl] sera perçu en tant que tel, de même qu'un [lp]. Si l'expérience confirme cet état de fait, la réalité cognitive de la prédiction aura été rendue plausible. Le test n'a jamais été fait avec des locuteurs de langues qui ne présentent pas les contraintes sur les groupes initiaux que sont familières des langues comme l'anglais.

Les travaux sont actuellement en préparation (création des stimuli audio) avec A. Ali, spécialiste des signaux dichotiques (ou incongruents, effet McGurk), et l'expérimentation sera conduite à Brno par M. Ziková sur une population d'étudiants tchéco-phones.

## **6. Théorie phonologique**

La théorie phonologique continuera à être développée par le biais des activités dans les divers champs d'application mentionnés: gabarits, interface, slave, diachronie, jeux de langage.



## Equipe 4

### Contact de langues et français d'Afrique

(resp. Carole de Féral)

#### 1. Composition

NOM	Statut	participation
<b>Membres permanents</b>		
<b>HDR</b>		
QUEFFÉLEC, Ambroise	PR (Univ. de Provence)	100%
VIROLLE, Marie	CR HDR (soutenance nov. 2010)	100%
<b>non-HDR</b>		
FÉRAL, Carole de	MCF	100%
<b>émérite</b>		
NICOLAÏ, Robert	PR émérite	
<b>Membres non permanents</b>		
ABDOULAYE, Moussa	Doctorant (assistant Univ. de Bangui, République Centrafricaine)	100%
KOZIAS, Ekila	Doctorant (assistant Univ. de Kinshasa, République Démocratique du Congo)	100%
NGOIE, Irène,	Doctorante (assistante à l'Univ. de Lumumbashi, République Démocratique du Congo)	100%
THOA-TSAMBU, Gisèle	Doctorante (assistante Univ. de Kinshasa, République Démocratique du Congo)	100%

Le contact des langues, considéré comme un phénomène central dans les dynamiques de leur transformation, continue à être la thématique fédératrice de l'équipe.

#### 2. Approche théorique du contact des langues

Même si les recherches se focalisent sur divers objets linguistiques apparemment bien identifiés (« français », « pidgin-english », « camfranglais/francanglais »...) étant donné les différents projets dans lesquels les membres de l'équipe sont impliqués, il ne s'agit plus tant de décrire ces objets linguistiques que d'observer et de rendre compte de dynamiques langagières dans un espace anthropo-social donné. Ces dynamiques, d'une part, sont initiées par le contact entre les « langues » (perçues comme entités homogènes dans les représentations des locuteurs et présentées comme telles dans certaines monographies) et, d'autre part, elles jouent un rôle essentiel dans la construction de nouvelles pratiques, de nouvelles catégorisations, de nouveaux positionnements identitaires (identité « urbaine » voire « nationale » plutôt qu' « ethnique », par exemple, avec des « vernaculaires interethniques » (Féral) comme le pidgin-english et le camfranglais/francanglais du Cameroun ou encore le nouchi de Côte d'Ivoire).

À partir des données empiriques du terrain africain et de propositions théoriques comme celles de R. Nicolai (*La construction du sémiotique*, à paraître), l'équipe poursuivra sa réflexion sur la nécessaire articulation entre les pratiques (hétérogènes) et les représentations – que ce soient celles des locuteurs ou des linguistes – qui, elles, tendent à créer de l'homogène et à ériger des frontières qui sont absentes dans les pratiques. Ne peuvent donc être écartées de l'analyse des pratiques une réflexion qui porte non seulement sur l'activité des locuteurs mais aussi sur celle du linguiste, les représentations qu'il propose et qui sont susceptibles d'influencer à leur tour les représentations et les pratiques des locuteurs dont il décrit les activités langagières [BCL 49, Féral].

Sur proposition de R. Nicolai, un colloque international : *Rethinking Contact Induced Change / Contact et changement linguistique : nouvelles approches?* se tiendra à Leiden (Pays-bas), du 9 au 11 juin 2011.<sup>6</sup> Il sera l'occasion pour plusieurs membres de l'équipe de présenter leurs nouvelles réflexions sur le changement linguistique initié par le contact des langues et d'en discuter avec les plus grands spécialistes de la « Contact Linguistics ».

### 3. Le terrain africain : contact des langues, nouvelles pratiques et nouvelles identités

#### 3.1. Français en contact en Afrique sub-saharienne et en Algérie

On continuera d'étudier les processus linguistiques qui sont en œuvre dans le français parlé et écrit en Afrique. A. Queffélec poursuivra ses recherches dans le cadre du projet franco-norvégien *Le français contemporain en Afrique et dans l'Océan Indien : usage, variétés et structure* (CFA : *Contemporary French in Africa and the Indian Ocean*), qui vise à la description et à la comparaison d'un ensemble de « variétés » de français parlées au sud du Sahara et dans l'Océan Indien. Le projet CFA se propose également de cerner les caractéristiques communes aux variétés africaines, partant de l'hypothèse qu'il existe un français panafricain qui transcende les différents contextes sociolinguistiques et les différents substrats, et que cette « variété » panafricaine s'inscrit dans un cadre plus vaste, le français panlectal. A. Queffélec participe, dans ce cadre, à l'organisation du colloque « Où vont les français d'Afrique ? » qui se tiendra les 8 et 9 février 2011 à Cologne.

A. Queffélec poursuivra également ses recherches sur la créativité lexicale dans le français parlé en République Démocratique du Congo, au Gabon et en République du Congo dans le cadre de *La Base de données lexicographiques panfrancophone (BDLP)*, projet international (dirigé par C. Poirier, Université de Laval) qui s'inscrit dans l'entreprise du Trésor des vocabulaires français, lancée par Bernard Quemada dans les années 1980. L'objectif est de constituer et de regrouper des bases représentatives du français de chacun des pays et de chacune des régions de la francophonie. Les bases de données sont conçues de façon à pouvoir être interrogées de façon séparée ou comme un seul corpus et à servir de complément au Trésor de la langue française informatisé, implanté au Centre Analyses et traitements informatiques du lexique français (ATILF) de Nancy (<http://www.bdlp.org/>).

Dans le cadre du projet ANR/DGF *CIEL*, qui se propose de développer une analyse comparative du français à travers la francophonie à partir de corpus de type écologique, le corpus camerounais de C. de Féral (1974→), qui est le seul corpus de français de ce type élaboré dans les années 1970 et 1980 dans ce pays, a été numérisé à l'Université de Freiburg (Allemagne). Il est en train d'être retranscrit selon les normes établies pour le projet. Il complète des enregistrements récents, ce qui permettra la prise en compte d'une certaine dimension diachronique (presque quarante ans). Il sera accessible en ligne dans la base *ciel-f*. Les extraits les plus significatifs de ce corpus seront publiés dans un ouvrage accompagné d'un CD (C. de Féral : *Le français des Camerounais dans sa diversité*).

M. Virolle poursuivra son analyse socio-littéraire et sociolinguistique de la nouvelle littérature écrite, algérienne et d'inspiration algérienne, notamment dans le cadre du Projet ANR « Nedjma ». Représentations de la guerre d'Algérie (axes : "Représentations littéraires" et "Mémoire artistique") s'il est accepté. Elle prépare également un ouvrage sur la rhétorique et la poétique de l'oralité traditionnelle et moderne, en Algérie et dans la diaspora.

C. de Féral et A. Queffélec poursuivront leurs recherches complémentaires sur les nouvelles pratiques urbaines en Afrique sub-saharienne (camfranglais/francanglais du Cameroun, nouchi de Côte d'Ivoire, toli bangondo du Gabon) qui mettent à contribution le français et des langues avec lesquelles il est en contact. D'une part, ils continueront d'étudier les processus linguistiques qui sont en œuvre dans les pratiques, et d'autre part, les représentations. On a montré, en effet, que les processus linguistiques mis en œuvre (emprunt, manipulations morphologiques et sémantiques ...) étaient somme toute banals et se retrouvaient dans d'autres langues ainsi qu'en français courant mais qu'ils servaient à catégoriser du discours français comme « autre chose » (camfranglais/francanglais...), dont les locuteurs pouvaient revendiquer la légitimation étant donné les représentations qu'en ont donné les médias et certains linguistes [BCL 49, Féral].

C. de Féral devrait pouvoir finaliser l'élaboration de la « Base de données lexicographiques de camfranglais/francanglais » avec ses collaborateurs camerounais (notamment V. Feussi, Université de Douala et A. Souop, Université de Buéa) grâce à une amélioration notoire de la plate-forme informatique

<sup>6</sup> <http://www.hum.leiden.edu/lucl/research/conferences/upcoming-conferences/language-contact.html>

(<http://www.unice.fr/equipevirtuellecama/>) qui a été élaborée au laboratoire BCL. Elle permet le travail à distance des différents membres impliqués dans le projet sur le camfranglais/francanglais de l'équipe virtuelle internationale *Français identitaires des jeunes en Afrique francophone*, qui avait été créée à la fin 2004 par C. de Féral avec la caution de l'AUF. Cette base de données est un des outils mis en œuvre pour l'analyse des pratiques de l'ouvrage en préparation : *Le francanglais, « parler jeune » des francophones camerounais : pratiques, représentations, stratégies*.

Les doctorants, qui sont également des enseignants-chercheurs dans des universités africaines, travaillent sur des corpus de français parlé (Abdoulaye et Ngoie), littéraire (Kozias), ou journalistique (Thoa-Tsambu), dans une perspective sociolinguistique (Abdoulaye et Ngoie) ou de l'analyse de discours (Kozias et Thoa). Ces chercheurs bénéficient de bourses en alternance (quatre fois trois mois) du Ministère des Affaires Étrangères (gestion Égide). Leur présence régulière à BCL est un apport non négligeable à la dynamique de l'équipe.

### 3.2. Pidgin-english du Cameroun

#### *Variation et changement linguistique*

C. de Féral a constaté, en comparant le corpus qu'elle a élaboré dans les années 1970 et 1980 avec des enregistrements récents (2009), qu'il y a un changement en cours qui concerne la distinction formelle de genre humain /non humain dans les pronoms anaphoriques du pidgin camerounais, phénomène apparemment tout à fait original dans les pidgins et les créoles à base lexicale européenne, qui peut notamment s'expliquer par la présence d'un substrat bantou (Féral 2001<sup>7</sup>). Un questionnaire qui vient d'être diffusé au Cameroun lui permettra de faire une analyse quantitative de type labovien, avec des corrélations entre les variables indépendantes internes et des variables explicatives telles que l'origine (région anglophone/francophone), l'âge, le degré de scolarisation des locuteurs. Cette nouvelle étude empirique va aussi lui permettre de questionner à nouveau le rôle des langues en contact (substrat africain bantou et superstrat anglais, également langue lexificatrice du pidgin) dans le changement linguistique et même le changement de langue (« language shift »).

En tant que membre du « support committee » de WOCAL 7 (World Congress of African Linguistics) qui se tiendra à Yaoundé en août 2012, C. de Féral profitera de ce voyage pour mettre sur pied un réseau de chercheurs qui travaillera sur le changement linguistique dans le pidgin du Cameroun à des endroits du système qui était déjà extrêmement soumis à la variation il y a une vingtaine d'années, notamment les combinatoires aspecto-temporelles et l'expression de « être » (Féral 1989<sup>8</sup>).

#### *Projet européen TypoCorp*

Le projet TypoCorp doit être présenté par Bernard Comrie (Université de Leipzig) et Amina Mettouchi (Université de Nantes et EPHE) à la *European Science Foundation*. Il s'agit de rendre accessible dans une base de données des extraits de corpus qui permettraient de comparer des langues appartenant à des familles différentes. S'il est accepté, la participation de C. de Féral concernerait le pidgin-english du Cameroun en zone francophone et en zone anglophone.

## 4. Activité éditoriale

A. Queffélec continuera d'assumer la direction de la revue *Le français en Afrique* qui est éditée par BCL. Cette revue, téléchargeable sur [www.unice.fr/ILF-CNRS/ofcaf/](http://www.unice.fr/ILF-CNRS/ofcaf/), fédère les travaux de chercheurs travaillant dans d'autres universités africaines ou françaises (plus de 50 000 visites en 2009 authentifiables grâce au logiciel PHP My Visit). Le numéro 26 sera consacré à l'édition des actes du colloque franco-norvégien « autour du verbe en français d'Afrique » qui s'est tenu à Paris en décembre dernier.

<sup>7</sup> « Quand être ne peut plus être : prédication et changement linguistique en pidgin-english camerounais », *Travaux du Cercle Linguistique de Nice* 20, p. 59-71.

<sup>8</sup> *Pidgin-english du Cameroun*, Paris, Peeters/SELAF.

## 5. Perspectives

L'obtention de l'Habilitation à Diriger des Recherches par Marie Virolle (soutenance prévue en novembre 2010) et par Carole de Féral (qui doit soutenir son HDR dans le courant de l'année 2011 et qui reprend dès l'année universitaire 2010-2011 le séminaire de Master Contacts de langues qui était jusqu'alors assuré par R. Nicolai) permettra d'insuffler une dynamique supplémentaire à l'équipe en leur permettant de satisfaire des demandes d'encadrement et de direction de recherches (thèses mais aussi HDR) qui proviennent notamment d'étudiants du Maghreb et de l'Afrique sub-saharienne. Cela permettra de défendre et de valoriser le patrimoine niçois en la matière (G. Manessy, R. Nicolai, revue *Le Français en Afrique*) encore pendant un contrat quadriennal. A la fin du contrat, toutefois, les membres de l'équipe auront respectivement 64, 66 et 68 ans, et l'équipe cessera alors son activité en tant que telle (sans préjuger de l'activité des individus).

## Equipe 5

### Langage et Cognition

(resp. Lucile Chanquoy)

#### 1. Composition

NOM	Statut	discipline	participation
<b>Membres permanents</b>			
<b>HDR</b>			
CHANQUOY, Lucile	PR	psychologie (16 <sup>e</sup> )	100%
HUDELOT, Christian	DR	linguistique (34 <sup>e</sup> CNRS)	75%
NEGRO, Isabelle	MCF HDR	psychologie (16 <sup>e</sup> )	100%
OLIVIERI, Michèle	MCF HDR	linguistique (7 <sup>e</sup> )	25%
SCHEER, Tobias	DR	linguistique (34 <sup>e</sup> CNRS)	50%
<b>non-HDR</b>			
BARILLOT, Xavier	MCF	linguistique (7 <sup>e</sup> )	25%
LAVIGNE, Frédéric	MCF	psychologie (16 <sup>e</sup> )	75%
MAGNIE-MAURO, Mariel-Noële	MCU-PH	médecine (44 02 <sup>e</sup> )	100%
OLIVIER, Gérard	MCF	psychologie (16 <sup>e</sup> )	100%
RIZZOLO, Olivier	MCF	linguistique (7 <sup>e</sup> )	25%
<b>BIATOS, ITA</b>			
DUMERCY, Laurent	IE2		75%
<b>Membres non-permanents</b>			
PALASIS, Katérina	post-doc interne	linguistique	50%
CAMPIGOTTO, Aurélia	doctorante	psychologie	100%
GINDT, Morgane	doctorante	psychologie	100%
LUSSON, Charlotte	doctorante	psychologie	100%
CASTANO, Laura	doctorante	psychologie	100%
DUPONT, Marie-Agnès	doctorante	psychologie	100%
FARTOUKH, Michael	doctorant	psychologie	100%
MARMET, Fanny	doctorante	psychologie	100%
MERCIER, Brunissende	doctorante	psychologie	100%
PRUDENT, Elodie	doctorante	psychologie	100%
DARMON, Nelly	doctorante	psychologie	50%
NGO, Véronique	doctorante	psychologie	50%

Cette équipe intègre des psychologues et des linguistes. Durant le contrat quadriennal 2008-11, les premiers faisaient partie de l'Equipe 1 du Laboratoire de Psychologie Cognitive et Sociale LPCS (EA 1189). Se joint à cette équipe un MCU-PH de la Faculté de Médecine (le Docteur M.N. Magnié-Mauro, Chef de Service d'Explorations Fonctionnelles du Système Nerveux) qui collabore depuis de nombreuses années avec l'équipe notamment pour l'encadrement des étudiants de Master 2 de neuropsychologie et psychologie du développement.

Si la thématique de cette équipe est située en psychologie (Ch. Hudelot est psycholinguiste et de ce fait proche des psychologues), les linguistes apportent une orientation précise dans deux domaines : la sémantique co-occurentielle (D. Mayaffre et S. Mellet qui ne prennent pas de pourcentage officiel pour ne pas fragiliser leur équipe d'origine, Logométrie) et l'architecture cognitive de la grammaire (cette thématique regroupe les linguistes générativistes : T. Scheer, M. Olivieri, X. Barillot, O. Rizzolo).

#### 2. Structuration de la thématique Langage et Cognition

L'équipe centre ses recherches sur le langage en tant que fonction cognitive : ses structures et processus, son fonctionnement et son acquisition. Si les recherches portent typiquement sur des productions écrites,

l'interaction avec les linguistes génératifs vise notamment à faire la part entre ce qui est proprement grammatical dans le processus d'écriture (i.e., les procédés qui sont également à l'œuvre lors de la production orale) et ce qui est le résultat d'autres processus (mémoire visuelle, corrélation graphème-phonème etc.).

Les différents projets envisagés s'inscrivent dans le champ des sciences cognitives en s'intéressant aux relations entre Langage et Cognition avec un rôle important, pour la réalisation de ces projets, du calcul scientifique et de la modélisation numérique (L. Dumercy et F. Lavigne). Les projets de l'équipe peuvent se résumer de façon synoptique de la façon suivante.

		LANGAGE	
		1. Lexico-sémantique	2. Syntaxe
Cognition	A. Organisation	Structure des réseaux sémantiques, associations et champs	Structure en mémoire et propriétés d'organisation des connaissances
	B. Acquisition	Normes, particularités et relations entre cooccurrences et force d'association	Processus de production de textes, formulation et traitement de l'orthographe
	C. Traitement	Dynamiques de processus de rappel contextuel en mémoire	Analyse en temps réel de la production de phrases et de textes et modélisation neuronale des processus de traitement des règles syntaxiques

### A1. Organisation lexico-sémantique

F. Lavigne, L. Dumercy, G. Olivier, S. Mellet, D. Mayaffre, M.N. Magnié-Mauro

Collaborations externes : ModyCo (CNRS Paris Ouest) ; A. Berthoz (Collège de France, CNRS LPPA)

Les réunions de travail entre linguistes et psychologues ont permis de mettre en évidence l'étonnante ressemblance entre certains concepts étudiés par l'équipe de logométrie et par les psychologues cognitivistes. Cela va de pair avec des approches très différentes qui poursuivent des intérêts complémentaires. Il s'agit notamment des notions suivantes :

- La cooccurrence est mesurée dans les textes en linguistique en fonction de la proximité des mots mais également par les psychologues en fonction de la probabilité de production libre d'associés à un mot. Cette notion est dans les deux approches très liée à celle de force d'association entre les mots, qui permet la description de réseaux sémantiques entre unités lexicales/conceptuelles.
- La notion d'association permet de décrire la structure associative locale d'une unité lexicale/conceptuelle, en termes de champ sémantique de taille variable et de « *step* » des associations directe vs. indirecte entre les unités, et ainsi de proche en proche de prendre en compte le réseau sémantique plus global.

Ce projet propose d'étudier par l'expérimentation et par la modélisation neuronale les modalités de codage du sens des concepts en mémoire humaine, en fonction notamment des simples associations en mémoire mais aussi de propriétés linguistiques permettant de définir la structure des champs sémantiques des concepts.

La modélisation de propriétés fondamentales du cerveau humain est indispensable à toute élaboration de modèles de traitements cognitifs et linguistiques particuliers. L'activation d'items conditionnée par un contexte fait partie des mécanismes de base de la cognition, comme dans les règles de combinatoire. Les réseaux de neurones artificiels et les modélisations connexionnistes qui s'en inspirent peuvent être dotés de cette capacité d'activation conditionnelle. La modélisation biologiquement réaliste d'un réseau de neurones, si elle rend bien compte des liens associatifs codés dans la matrice synaptique (mémoire à long terme), peine en revanche à rendre compte de cette dernière. Les récents travaux en neurobiologie mettent en évidence que des traitements élémentaires peuvent avoir lieu dans l'arbre dendritique des neurones. Cette propriété permet d'envisager l'activité cérébrale déterminée par les stimulations endogènes et exogènes, permettant au cerveau

de représenter non seulement les items en mémoire mais aussi leurs combinaisons. Cette propriété est essentielle pour la compréhension des problématiques communes aux linguistes et aux psychologues, concernant les effets de contexte sur la sélection sémantique en temps réel. La méthode électrophysiologique permet, dans la mesure où les potentiels évoqués sont étroitement liés aux courants synaptiques, d'établir un lien entre processus cognitifs et neuronaux. L'étude des potentiels liés à l'évènement s'avère ainsi particulièrement pertinente pour d'examiner en temps réel les différentes étapes du traitement de l'information afin d'appréhender l'organisation lexico-sémantique (Kalhaoui et al., 2007).

En parallèle, des recherches seront menées dans le cadre d'un autre paradigme des sciences cognitives : l'énaction (Stewart, Gapenne & Di Paolo, 2010), qui rappelle par bien des aspects la théorie piagétienne de la connaissance. Dans ce cadre théorique, le cerveau est considéré comme une structure globale et autonome dont la logique de fonctionnement est essentiellement déterminée par les règles régissant au niveau local les activations ou inhibitions permanentes et réciproques des neurones entre eux. L'interaction gestuelle avec le milieu ne constitue qu'une source de variation externe, de nature sensori-motrice, modifiant localement une activité cognitive globale qui obéit quant à elle à une logique dynamique non linéaire (Thelen & Smith, 1994). Dans ce contexte, nous nous intéresserons à la production des unités lexicales du langage, soit effective, soit mentalement simulée et aux liens qu'entretient cette production avec la perception auditive du langage. Rappelons, par exemple, que dans le cadre de la théorie motrice de la perception du langage (Liberman, 1996), les objets de la perception du langage sont des anticipations des gestes phonatoires du locuteur. En collaboration avec le laboratoire CNRS de linguistique de Paris Ouest, nous essayerons de montrer comment la perception auditive de voyelles s'appuie sur la simulation mentale de leur prononciation. Une première expérience a permis de recueillir des résultats allant dans le sens de cette hypothèse, de nouvelles expériences sont en cours. La rédaction d'un article s'achève (*How Seeing Yawning and Pouting Faces Orients Ambiguous Vowels Auditory Perception*, Olivier et al., en préparation).

Un deuxième ensemble de recherches, mené en collaboration avec le Collège de France, en particulier Alain Berthoz, aborde la question des liens existant entre langage et stratégie de navigation spatiale. La question posée est la suivante : que fait le sujet lorsqu'il doit refaire un trajet qu'il a mémorisé sur une carte ? Pour répondre à cette question, un dispositif expérimental appelé « tapis magique » est utilisé. Il s'agit d'un ensemble de dalles réparties sur le sol et qui envoient des signaux à un ordinateur quand le sujet marche sur elles. Ce dispositif permettra d'étudier l'éventuelle interférence entre la stratégie dite « à la 1<sup>ère</sup> personne », consistant à simuler les saccades exécutées lors de l'apprentissage du trajet sur la carte, avec une stratégie plus liée au langage consistant à se souvenir verbalement des étapes du trajet. Les logiciels permettant de piloter l'expérience sont en cours de conception et les expériences sont prévues au Collège de France lors du deuxième semestre 2010.

Sur le plan théorique, un modèle de fonctionnement du cerveau soulignant la « gestualité de la connaissance », non seulement au niveau du langage mais plus généralement à tous les niveaux de l'organisation cognitive, sera proposé dans le cadre d'une HDR soutenue par G. Olivier au mois de décembre 2010.

## A2. Organisation syntaxique

F. Lavigne, I. Negro, M. Oliiviéri, L. Dumercy,

Collaboration externe : N. Brunel (Laboratoire de Neurophysique et Neurophysiologie, CNRS – Paris)

Dans la plupart des modèles de production (i.e., Bock & Levelt, 1994), la syntaxe est envisagée comme un module spécifique et autonome chargé, d'une part, de l'attribution de la nature et de la fonction des items lexicaux ainsi que de leurs traits grammaticaux (genre, nombre, temps, personne...) et, d'autre part, de l'organisation de ces items. Ce module serait chargé d'effectuer un certain nombre de calculs en fonction de règles procédurales stockées en mémoire.

A l'inverse, selon une conception connexionniste (Seidenberg & McClelland, 1989), les connaissances syntaxiques seraient représentées dans le système par des connexions entre des unités selon différents poids. Ce type de réseau coderait non seulement les fréquences d'occurrence des différents types d'informations mais également des statistiques d'ordre plus haut concernant la combinaison des éléments entre eux. Dans ce cas, la nature de l'item ne serait pas portée directement par l'unité lexicale abstraite mais serait encodée statistiquement dans un réseau probabiliste dans la relation de combinatoire entre les unités.

Toutefois, les modèles biophysiquement réalistes du cerveau permettent difficilement de rendre compte du codage de règles syntaxiques dans le cortex cérébral (Rigotti et al., 2009). L'objectif de cet axe sera donc de déterminer comment la syntaxe est encodée dans le cortex. Un modèle neuronal en cours d'élaboration sera testé par la simulation de données expérimentales sur les traitements syntaxiques. Le croisement des travaux en B2 relatifs à l'acquisition de la syntaxe et des recherches en modélisation

conduites en C1 devraient nous permettre de préciser les types de codages en mémoire des connaissances syntaxiques, et en particulier de tester si des règles abstraites peuvent être codées sur la base d'associations entre items ou requièrent de nouveaux types de codage neuronal. Les différents types de codages neuronaux testés intégreront par ailleurs différents types de règles d'apprentissage synaptique en relation avec l'axe B2 'acquisition de la syntaxe'.

### **B1. Acquisition lexico-sémantique**

MN. Magnié-Mauro, F. Lavigne, L. Chanquoy, L. Dumercy

Collaborations externes : M. Habib (INCM, UMR 6193 CNRS-Université de la Méditerranée) et L. Vaivre-Douret (INSERM U-483 (Paris VI))

Les concepts en mémoire correspondant aux mots et objets sont codés par la structuration de larges populations de neurones. Les travaux en neurosciences (neurobiologie et modélisation des réseaux de neurones biologiques) ont permis d'établir un lien entre la structuration de ces items et le champ sémantique. L'analyse propre de la structuration du champ sémantique cherchera à établir des relations vers des grandeurs observables à l'expérimentation (amorçage, normes associatives). Ces relations seront utilisées pour aborder l'évolution et l'usage du champ sémantique dans le cadre d'études auprès d'enfants tout-venant et d'enfants intellectuellement précoces. Ce projet propose d'étudier les particularités d'acquisition lexico-sémantique chez ces enfants qui se caractérisent notamment par une précocité d'acquisition lexicale, mais aussi par un style cognitif en faveur d'un surinvestissement de l'hémisphère droit. Les données comportementales et électrophysiologiques suggèrent que, dans la précocité intellectuelle, le niveau d'usage sémantique n'est pas en adéquation avec l'acquisition du lexique. Or, les champs sémantiques seraient plus larges et diffus dans l'hémisphère droit, contenant des informations plus distantes sémantiquement que dans l'hémisphère gauche. Ce travail permettrait ainsi de préciser le rôle de la particularité de la spécialisation hémisphérique dans la précocité intellectuelle mais aussi de mieux comprendre la structuration du champ sémantique chez l'enfant grâce à une approche pluridisciplinaire. L'étude de processus décisionnels lors de la détection d'un changement (lors de la présentation de scènes visuelles, par exemple) permettra de préciser les relations entre d'une part le champ sémantique et d'autre part les observables (ce que l'on peut mesurer) et l'usage (les conditions de la mesure).

### **B2. Acquisition de la syntaxe à l'écrit, articulation entre grammaire et écriture**

I. Negro, L. Chanquoy, F. Lavigne, L. Dumercy, M.N. Magnié-Mauro, C. Hudelot et M. Oliviéri, A. Campigotto (doctorante), C. Lussion (doctorante), M.C. Hazard (post-doctorante), K. Palasis (post-doc interne)

Collaborations externes : I. Bonnotte (U. Lille 3), B. Lété (U. Lyon 2), M. Fayol (U. Clermont-Ferrand), B. De Cara (U. Nice), E. Veneziano (U. Paris 5)

La production d'un message oral ou écrit implique de gérer simultanément un très grand nombre d'opérations allant de la simple intention de parler ou d'écrire à la formulation de mots, de phrases, de discours. Chez les enfants, l'acquisition du langage est un mécanisme très lent car si un enfant commence à produire ses premiers mots entre 18 mois et 2 ans, il entre dans l'écrit à l'âge de 6 ans et n'en maîtrise l'ensemble des processus qu'autour de l'âge de 15/16 ans. D'ailleurs, à cet âge, tous les processus rédactionnels ne sont pas encore automatisés (certains le sont difficilement, comme l'organisation syntaxique par exemple).

L'objectif des travaux de cet axe est de comprendre comment l'enfant acquiert le système symbolique de l'écrit et les processus indispensables à la production des textes. Les recherches seront centrées, d'une part, sur le processus de formulation, via l'analyse de différents types de texte et, d'autre part, sur les types d'interventions visant l'amélioration des productions textuelles, formes d'interventions « conversationnelles » à l'oral, révision de texte et maîtrise de l'orthographe à l'écrit.

Le processus de formulation sera abordé *via* l'analyse, à l'oral comme à l'écrit, des connecteurs (causatifs, additifs, adversatifs...) utilisés dans des types de texte particuliers (narratifs, explicatifs, argumentatifs...), et dans des situations différentes (notamment en tenant compte du destinataire et de la nécessité d'adopter, ou non, un point de vue autre que le sien propre). En outre, à l'écrit, seront analysées conjointement avec les connecteurs les marques de ponctuation. Finalement, les productions recueillies devraient également participer à l'établissement d'une typologie « naturelle » des erreurs orthographiques rencontrées dans les productions spontanées (*cf.* ci-dessous).



Concernant la révision, les recherches ont à la fois une visée fondamentale et une visée appliquée. En effet, l'objectif est non seulement de mieux comprendre et d'interpréter les processus sous-tendant la révision, en fonction de la nature du matériel linguistique à réviser (notamment, orthographe lexicale *versus* orthographe grammaticale ; révisions profondes – ou sémantiques – *versus* révisions de surface ; etc.), mais aussi d'élaborer des outils favorisant la mise en œuvre du processus de révision chez les enfants d'école élémentaire (poursuite des travaux sur les grilles de révision).

Un autre ensemble de recherches (en collaboration avec E. Veneziano) porte sur les interventions conversationnelles (étayage). Ces recherches ont également une double visée fondamentale et appliquée. Une diversification des types d'intervention couplée à une approche interlangue et à la comparaison d'enfants à développement typique et présentant des troubles du développement permettra d'évaluer la part qui revient aux facteurs linguistiques (lexicaux, grammaticaux, phrastiques ou interphrastiques, syntaxiques), pragmatiques (adaptation à la situation d'interlocution) ou cognitifs (mémorisation, planification, théorie de l'esprit). Les données recueillies serviront de base pour la construction d'un outil d'évaluation et de comparaison.

L'objectif des travaux réalisés sur l'orthographe est de mieux comprendre comment les enfants acquièrent l'orthographe lexicale et grammaticale. En particulier, l'un des débats actuels en psychologie du langage est de déterminer si un recours à des règles de production est indispensable ou si des mécanismes associatifs peuvent rendre compte des formes produites. Les premières études, longitudinale et transversale, déjà réalisées par I. Negro en collaboration avec I. Bonnotte et B. Lété sur l'acquisition des formes du participe passé, suggèrent que les enfants les plus jeunes sont sensibles aux fréquences d'exposition des formes morphosyntaxiques alors que chez les adultes l'impact de cette fréquence tend à diminuer. De nouvelles études, longitudinales et transversales, sont destinées, d'une part, à répliquer les résultats obtenus avec d'autres formes morphosyntaxiques que le participe passé et, d'autre part, à étudier les interactions possibles entre fréquence d'exposition et consistance phonographémique. L'expertise de M. Olivieri aidera à faire le choix des constructions syntaxiques nouvelles qui seront examinées.

Une autre série de recherches vise à déterminer les relations entre orthographe lexicale et grammaticale. Si l'orthographe de formes lexicales morphologiquement simples ou complexes dépend de la fréquence avec laquelle un mot est rencontré à l'écrit, l'enfant devrait extraire progressivement un certain nombre de régularités tant au niveau lexical que morphosyntaxique. Le matériel présenté sous dictée aux enfants sera extrait de la base Manulex de façon à manipuler la fréquence de rencontre des formes et leur consistance orthographique.

Dans une perspective proche, un projet [en collaboration avec B. De Cara, M.C. Hazard (U. Nice) et M. Fayol (U. Clermont-Ferrand)] sera consacré à une typologie des fautes d'orthographe et à une analyse des processus qui les sous-tendent. Ce projet vise notamment, pour son versant appliqué, à apporter aux professionnels confrontés aux troubles orthographiques un outil de dépistage et de diagnostic des difficultés rencontrées. Une aide à la remédiation leur sera également apportée en proposant des pistes pour entraîner les habiletés déficitaires. Ce travail doit donc déboucher sur la création d'un test fiable permettant de mesurer le niveau orthographique des élèves (test financé par Ortho-Edition).

Enfin, un dernier axe, mené dans le cadre d'un contrat doctoral (C. Lusson), consiste à préciser les processus impliqués dans l'encodage grammatical chez les enfants. Si l'hypothèse d'arbre syntaxique est souvent avancée dans la littérature anglo-saxonne pour rendre compte des erreurs d'accord sujet-verbe en production orale chez l'adulte, une hypothèse sérielle semble plus pertinente chez l'enfant d'école élémentaire (Negro, Chanquoy, Fayol et Louis-Sydney, 2005). Différentes études *off line* (dictée, tâche de complètement) et *on line* (tablette graphique et analyse de mouvements oculaires) en français à l'écrit et à l'oral conduites dans une perspective développementale visent à préciser ce débat et à déterminer les processus mis en œuvre par l'enfant dès le CE1 pour accorder correctement un nom, un verbe et un adjectif en nombre et en genre.

### **C1. Traitements lexico-sémantiques (base textuelle)**

F. Lavigne, L. Dumercy, V. Ngo (doctorante), S. Mellet, D. Mayaffre, M.N. Magnié-Mauro  
Collaborations externes : F. Vitu (LPC, CNRS - Université Aix-Marseille)

Les traitements sémantiques sont hautement contextualisés, c'est-à-dire que la représentation du sens d'un mot dépend de son environnement textuel. Les analyses fines des types de contexte sémantique proposées par les collègues linguistes permettent d'élaborer des nouveaux types d'expérimentations et de modèles des mécanismes de sélection sémantique contextuelle. Les travaux d'analyse de larges corpus en collaboration avec les linguistes (logométrie), ainsi que les calculs numériques inhérents à la modélisation neuronale, permettent d'aborder les processus dynamiques en relation avec le fonctionnement de réseaux neuronaux biologiquement réalistes. L'ambition du projet est de proposer un modèle unifié à la fois des structures

contextuelles analysées par les linguistes, et des processus dynamiques tels que mis en œuvre dans le cortex cérébral.

Il s'agit donc ici d'une approche interdisciplinaire des concepts de représentation, du sens des concepts en contexte, impliquant la logométrie en linguistique, l'expérimentation en psychologie cognitive, et la modélisation par des réseaux neuronaux. Notamment, la notion d'activation en mémoire permet d'appréhender une définition dynamique des représentations du sens des concepts en temps réel, du cas simple entre paires associées au cas de triplets de mots puis de séquences plus complexes telles que rencontrées dans des textes. Les travaux électrophysiologiques ont souligné le rôle joué par l'intermodalité sur le traitement sémantique en contexte phrastique. Une problématique centrale concernera l'étude des modalités et dynamiques de représentation des concepts en fonction d'un contexte. Il s'agit notamment de comprendre comment une même structure sémantique codée en mémoire à long terme peut générer en mémoire de travail des processus de représentation variables en fonction du contexte.

Les dynamiques temporelles des traitements sémantiques en contexte phrastique sont extrêmement importantes lors de la compréhension du discours écrit pendant la lecture. Les traitements lexico-sémantiques font donc l'objet d'enregistrements du comportement oculomoteur et de potentiels évoqués pendant la lecture de phrases. Il s'agit notamment d'analyser le point d'atterrissage des saccades oculaires dans les mots en fonction du contexte sémantique. Ce projet est financé par un contrat ANR en cours et fait l'objet d'un nouveau projet déposé. Le comportement oculomoteur est influencé lors de la lecture par le lien sémantique entre les mots présentés et par la structuration entre mots présentés (mots isolés, phrases, textes). L'objectif est donc d'étudier le comportement oculomoteur naturel pendant la lecture de phrases afin de mieux comprendre ses déterminants visuels, oculomoteurs et sémantiques. Cette approche permettra de comparer les modèles classiques de traitements sémantiques de mots 'isolés' avec des modèles de traitements sémantiques dans les phrases. Elle pourra donc prendre en compte la dimension syntaxique dans la phrase et faire le lien avec les traitements syntaxiques (point 2 « syntaxe »).

## **C2. Traitements syntaxiques : Modélisation des règles syntaxiques**

Negro, L. Chanquoy, F. Lavigne, L. Dumercy, T. Scheer, M. Oliviéri  
Collaborations externes : B. Lété (Lyon 2) et I. Bonnotte (Lille 3)

Cet axe prolonge l'étude conduite sur l'acquisition des formes morphosyntaxiques simples et complexes (B2). L'objectif est de tenter de modéliser les résultats expérimentaux obtenus. Un premier modèle envisagé reposera sur le processus de Markov selon lequel toute lettre produite est conditionnée soit par la lettre qui la précède soit, s'il s'agit de la première lettre, par la terminaison du lemme plutôt: mot?? précédent. L'objectif ici sera de reproduire la suite de lettres la plus probable, qu'il s'agisse d'unités morphologiquement simples ou complexes.

D'une part, l'hypothèse est qu'il existe une corrélation entre le taux d'erreurs et la fréquence de la chaîne de Markov représentant la suite de lettres. D'autre part, l'hypothèse est que la chaîne de Markov n'est pas un prédicteur suffisant : les participants recourent aussi à leurs connaissances grammaticales pour écrire, et ces connaissances ne sont pas linéaires mais encodent des informations de manière hiérarchisée (qui sont exprimées dans un arbre syntaxique). L'objectif est donc de montrer que la production écrite est le résultat à la fois de procédés grammaticaux (non-linéaires) et de procédés non-grammaticaux (linéaires, chaînes de Markov). Ou, en d'autres termes, que l'événement futur ne peut être entièrement prédit par l'évènement présent. Un objectif à moyen terme est ensuite de découvrir quel est le poids relatif des processus grammaticaux et des autres processus dans la production écrite, et quels sont les facteurs qui le déterminent.

Les résultats expérimentaux du participe passé seront également mis à l'épreuve du modèle présenté en A1. L'intérêt de ce modèle est qu'il permet de manipuler l'hypothèse associative dans une perspective développementale. Cette action de recherche fera l'objet d'un projet ANR programmé en 2012.

## **A2 et C2. Architecture cognitive de la grammaire**

F. Lavigne, T. Scheer, M. Oliviéri, O. Rizzolo, X. Barillot, L. Dumercy

L'axe « Interface, modularité, Sciences Cognitives », bien que situé dans l'équipe DDP, concerne les deux équipes DDP et Langage et Cognition. Partant du modèle modulaire en T inversé de la grammaire générative où un module concaténatif (morpho-syntaxe) alimente deux modules interprétatifs (phonologie PF et sémantique LF), il s'agit d'une part de participer au débat en linguistique sur la place de la morphologie, qui peut ou non être un module indépendant de la syntaxe.

D'autre part, la théorie minimaliste qui a actuellement cours en linguistique est basée sur la théorie de la phase (*Phase Theory*) dont l'idée est qu'une phrase entière est trop grande pour être calculée dans la mémoire de travail. Elle est donc coupée en morceaux plus petits (des phases), et la computation morphosyntaxique de chaque phase est autonome. Aussi, le résultat de chaque computation est envoyé aux modules interprétatifs de façon autonome, ce qui génère, pour la construction d'une seule phrase, un trafic incessant d'aller-retour entre les modules morphosyntaxique et phonologique/sémantique.

Or ce scénario n'est pas compatible avec une propriété fondamentale de la théorie modulaire du système cognitif (Fodor, 1983) : l'encapsulation, selon laquelle la computation modulaire est autiste (une fois lancée, elle ne peut être ni influencée par des données absentes de son *input*, ni communiquer des résultats intermédiaires avant l'achèvement de son activité).

Du côté psychologique, il y a également des raisons de penser que l'encapsulation dans le fonctionnement cognitif impliquerait d'un côté un certain niveau de spécialisation fonctionnelle des modules de traitement, mais aussi des interactions fortes à travers des limites perméables entre modules. L'identification des interactions possibles ou pas entre modules tels que ceux traitant la grammaire et ceux traitant la sémantique est un enjeu important pour la psychologie cognitive expérimentale et la linguistique. De plus les modèles neuronaux du cortex cérébral seront utilisés pour appréhender les connectivités entre populations neuronales à la fois spécifiques de représentations mentales bien définies et en interaction forte.

L'objectif à terme de cet axe est donc de faire évoluer le modèle modulaire en maintenant son cœur (l'existence d'unités de calcul fonctionnelles et spécifiques à un domaine), tout en abandonnant l'idée qu'un module est encapsulé: il communique sans cesse avec d'autres modules.

## Références

- Bock, J.K. & Levelt, W. (1994). Language production: Grammatical encoding. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945-984). San Diego, Californie: Academic Press.
- Fodor, J.A. (1983). *The modularity of mind*. Cambridge, Mass : M.I.T. Press.
- Liberman, A. M. (1996). *Speech, a special code*, MIT Press, Cambridge, Mass.
- Kahlaoui K, Baccino T, Joannette Y, Magnié MN (2007). Electrophysiological bases of semantic processing of objects. *Med Sci (Paris)*, 23(2):193-7.
- Negro, I., & Chanquoy, L., Fayol, M., & Louis-Sydney, M. (2005). Subject-verb agreement in children and adults: Serial or hierarchical processing. *Journal of Psycholinguistic Research*, 34 (3), 233-258.
- Rigotti, M., BenDayanRubin, D., Wang, X.-J. and Fusi, S. (2010). Internal representation of task rules by recurrent dynamics: the importance of the diversity of neural responses. *Frontiers in Computational Neuroscience*.
- Seidenberg, M.S. & McClelland, J.L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.
- Stewart, J., Gapenne, O., & Di Paolo, E. (2010). *Enaction: towards a new paradigm for cognitive science*, MIT Press, Cambridge, Mass.
- Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. MIT Press. Cambridge, Mass.

## Equipe 6

### Linguistique et Didactique (LiDida)

(resp. Simona Ruggia, Laurent Rouveyrol)

#### 1. Composition

NOM	Statut	spécialité	Taux de participation
<b>Membres permanents</b>			
<b>HDR</b>			
HUDELOT Christian	DR (34 <sup>e</sup> CNRS)	psycho-linguiste	25% (75% langage et cognition)
<b>non-HDR</b>			
FOLLETTE Peter	MCF 11 <sup>e</sup>	angliciste	50% (50% logométrie)
RUGGIA Simona	MCF 7 <sup>e</sup>	linguiste, FLES	75% (25% énonciation)
ROUYEYROL Laurent	MCF 11 <sup>e</sup>	angliciste	75% (25% logométrie)
SCHNITZER Nathalie	MCF 12 <sup>e</sup>	germaniste	25% (75% énonciation)
<b>Membre adjoint</b>			
(cf. section 2.7 du projet général sur ce statut)			
WHYTE Shona	MCF 11 <sup>e</sup>	angliciste	
<b>Membres associés</b>			
(cf. section 2.7 du projet général sur ce statut)			
LEONCINI Stefano	MCF 14 <sup>e</sup>	italianiste	Université de Nice-Sophia Antipolis, CIRCPLES
BRANDT Stefanie	créatrice d'entreprise en FLES, ancienne ATER et post-doc interne, thèse soutenue en 2007	FLES	

L'équipe LiDida est composée d'enseignants-chercheurs rattachés à des filières de langues vivantes (anglais, allemand et italien) et de français langue étrangère et seconde, dont les travaux portent sur la linguistique et la didactique.

#### 2. Projet scientifique

Les relations entre la **linguistique** et la **didactique des langues** sont depuis toujours étroites car la linguistique et la didactique sont deux disciplines complémentaires. L'objectif de cette équipe est donc avant tout de renforcer ces liens, en proposant une approche originale qui s'appuie sur une **interpellation réciproque** des deux disciplines. En effet la didactique a trop souvent été perçue et construite comme une simple branche applicative de la linguistique ; mais elle peut et doit aussi interpellier la linguistique et, développant ses propres questionnements, les soumettre à la linguistique et lui offrir ainsi quelques-unes des pierres de touche qui permettront à celle-ci de valider ses hypothèses théoriques et d'affiner ses analyses explicatives. Réciproquement, la linguistique, parce que son objet est de décrire et de modéliser le fonctionnement des langues, se trouve particulièrement bien placée pour soumettre à la didactique des ensembles de données utilement structurées et des modèles d'analyse qui pourront inspirer la pratique et la réflexion des didacticiens. Il s'agit là d'une démarche dialectique qui a vocation à être portée par des **linguistes-didacticiens**, dont la pratique pédagogique s'accompagne d'une recherche fortement ancrée dans un laboratoire reconnu.

Nos travaux visent à améliorer l'enseignement d'une langue-culture et faciliter son apprentissage afin de faire acquérir une *compétence à communiquer langagièrement* selon l'approche actionnelle préconisée par le *Cadre Européen Commun de Référence pour les Langues* (2001).

Or la *compétence à communiquer langagièrement* comporte trois composantes à savoir : linguistique, sociolinguistique et pragmatique ; ceci conduit nécessairement à développer une dimension réflexive de l'enseignement / apprentissage, dans une approche qui dépasse le simple niveau grammatical pour englober toutes les strates de l'activité langagière telle qu'elle se développe au niveau cognitivo-discursif.

Pour autant, notre équipe ne cherchera pas à couvrir l'ensemble du champ, mais délimitera ses axes de recherche en fonction du paysage national de la recherche en didactique, de l'ensemble des compétences spécifiques présentes et reconnues à Nice et de la philosophie générale du laboratoire BCL dont nous faisons partie.

## 2.1. Démarche heuristique

Nos travaux auront pour spécificité de placer **les corpus** non seulement au centre de la recherche et de la pratique didactiques, mais plus précisément au cœur de chacune des étapes constitutives de notre démarche. Ce cercle heuristique pourrait être ainsi résumé : l'étude de corpus variés permet de cerner sur divers phénomènes particuliers le fonctionnement des langues cibles et de documenter les difficultés rencontrées par les apprenants. Les mêmes corpus, après avoir suscité l'analyse linguistique initiale, vont fournir le matériel exploitable en classe de langue ; l'utilisation de ces données comme support pédagogique en cours et dans les activités d'apprentissage va permettre à rebours d'évaluer leur pertinence dans la démarche didactique et l'efficacité de leur modélisation ; ce retour sur expérience permettra d'opérer un tri parmi les données, éventuellement d'en proposer une nouvelle structuration, de les enrichir par les productions de la classe de langue elle-même (qui sera enregistrée) et de lister les difficultés persistantes qui nécessitent une nouvelle analyse et une remédiation.

Cette boucle réflexive s'accompagnera simultanément d'un travail sur le **métalangage** linguistico-didactique et sur la transposition terminologique de l'une à l'autre des disciplines. On voit alors comment une telle démarche assure à la fois le dialogue équilibré entre les disciplines et le va-et-vient nécessaire entre données empiriques et modélisation théorique.

Les corpus ainsi constitués pourront être de tous ordres à partir du moment où ils ont une pertinence dans la classe de langue (plurimédiatiques : soit audiovisuels (débat télévisés, publicités, extraits de film, interactions enseignant/apprenants en classe de langue, etc.), sonores (émissions radiophoniques) ou textuels (comportant des manuscrits, leurs reproductions).

L'analyse dont ils feront l'objet se vaudra sémantico-pragmatique en contexte, dans la perspective d'une méthodologie d'enseignement par compétences ; étude de l'impact des genres de discours, toujours dans la même perspective ; utilisation et adaptation des outils d'exploration automatique et de traitement quantitatif développés à Nice. Nous envisageons donc de croiser, dès le départ une perspective qualitative avec une démarche quantitative, en parfaite cohérence avec l'épistémologie développée dans d'autres équipes de BCL avec qui l'équipe LiDida a des liens naturels. L'attention portée aux faits énonciatifs d'une part, aux contraintes génériques d'autre part, ainsi que le recours à des outils de traitement automatique mettra l'équipe LiDida en position de collaborer très étroitement avec l'équipe « Linguistique de l'énonciation » et l'équipe « Logométrie ».

## 2.2. Axes de recherche

Dans l'état actuel des collaborations envisagées, ces travaux se déploieront principalement à propos de l'enseignement du français langue étrangère et seconde (FLES) (S. Ruggia, S. Brandt), de la didactique de l'anglais (L. Rouveyrol, P. Follette) de l'italien (S. Leoncini) de l'allemand (N. Schnitzer) et autour de l'acquisition (C. Hudelot), mais la possibilité d'autres champs d'application reste ouverte.

Deux axes principaux se dessinent ; d'une part l'analyse se focalisera sur l'interaction verbale et d'autre part sur les productions langagières plutôt monogérées. Les supports de notre recherche seront des corpus variés et plurimédiatiques, différents genres de discours dont bien sûr le discours didactique (enseignant / apprenant). Ainsi des cours seront enregistrés et notamment les cours de langue qui ont lieu à l'Université de Nice-Sophia Antipolis, cours qui sont organisés selon l'approche actionnelle du *Cadre Européen Commun de Références pour les Langues*.

### *A propos d'interactions verbales*

L'un des problèmes langagiers majeurs auxquels nous consacrerons une part de nos travaux ainsi que notre expérience professionnelle est le fonctionnement et la maîtrise des Marqueurs de la Structuration de la Conversation (MSC) en français et en anglais.

Les MSC, nécessaires à la maîtrise des échanges dialogaux, comprennent des marqueurs exclusifs de l'oral : « *ah bon, tu vois, ...* », « *really ?* » ainsi que des marqueurs qui sont utilisés à la fois à l'oral et à l'écrit : « *enfin, en fait, ...* », « *actually, in fact* ». De plus, la plupart de ces marqueurs sont **polyvalents** et **polyfonctionnels**, ce qui est source de nombreuses difficultés pour des étudiants allophones.

A travers l'étude de plusieurs types d'interactions verbales, nous avons établi un classement (Ruggia : 2007), pour le FLE, qui distingue les différentes valeurs sémantico-pragmatiques des MSC. Ce classement s'appuie sur des nombreuses études des interactions verbales et l'analyse du discours, études qui, fatalement, font appel à un métalangage linguistique dont l'usage en classe de langue doit être réévalué et régulé.

L'analyse des nuances du fonctionnement des marqueurs ainsi que leurs différentes utilisations à l'oral et parallèlement l'étude des difficultés rencontrées lors de l'apprentissage du français langue étrangère, et plus généralement des langues étrangères nous permettra d'élaborer des méthodes didactico-linguistiques. Dans quels contextes discursifs précis (argumentation, description ...) ces marqueurs apparaissent-ils plutôt ? Quelle est leur orientation pragmatique privilégiée ? Telles sont les questions qui guideront en partie l'exploration.

### *A propos de productions langagières monogérées*

On sait l'importance des tâches de nature monologique dans la classe de langue (*cf* la prise de parole en continu). Il est donc essentiel de s'y intéresser pour améliorer les stratégies développées. Les concepts de *séquence discursive* (J-M Adam : 1992) permettant de modéliser la structuration interne des productions langagières seront envisagées et plus particulièrement *l'argumentation*. Une étude préalable (Rouveyrol : 2003, 2005) a montré comment la planification interne (structure en double pyramide inversée) mettait en valeur la conclusion vers laquelle tendait les interventions monologiques étudiées dans le cadre de débats-panels politiques médiatisés. Une étude plus récente (Rouveyrol : 2009) s'inscrivant dans ce prolongement a mis l'accent sur le marquage précis opéré par les différents locuteurs pour introduire les conclusions 1 et 2 et insérer les séquences descriptives entre les deux. Il conviendrait maintenant d'établir une modélisation précise du rapport entre planification discursive et marquage microlinguistique.

### *Prise en compte des certifications*

Les exemples développés plus haut ont naturellement vocation à prendre en compte les certifications linguistiques (notamment CLES et DALF) qui demandent aux apprenants de mettre en application des stratégies interactionnelles et argumentatives dans une perspective actionnelle.

Les outils, ressources à visée didactique développées permettront *in fine* d'élaborer des méthodes d'apprentissage des langues performantes et théoriquement cohérentes pour l'acquisition de véritables compétences discursives. Une réflexion sur l'enseignement à distance en rapport avec la prise en compte des TICE sera également menée.

## **3. Animation, diffusion**

- Organisation du 16<sup>ème</sup> congrès de RANACLES : « *Didactique et pratique des langues, quelles priorités ?* » 20-22 novembre 2008, Université de Nice-Sophia Antipolis.
- Organisation du 31<sup>ème</sup> colloque du GERAS : « *A l'intersection des discours de spécialité : hétérogénéité et unité* », 18-20 mars 2010, Université de Nice-Sophia Antipolis.