

**Laboratoire**  
**Modèles, Dynamiques, Corpus**

**Unité mixte de recherche 7114**  
**CNRS et Université Paris X**

**Dossier scientifique**

**2005-2008**

**30 novembre 2003**

# Table des matières

---

<b>DÉCLARATION DE POLITIQUE SCIENTIFIQUE POUR LA PÉRIODE (2005-2008) .....</b>	<b>3</b>
<b>ORGANIGRAMME PAR OPÉRATIONS SCIENTIFIQUES (2005-2008) .....</b>	<b>7</b>
<b>MEMBRES DU LABORATOIRE .....</b>	<b>8</b>
<b>1. MODÈLES.....</b>	<b>10</b>
1.1. Phonologie de la syllabe et phonologie du français.....	10
1.2. Morphologie et sémantique lexicale .....	18
1.3. Les connecteurs.....	28
1.4. Modélisation de données linguistiques dans la perspective du traitement automatique des langues (Metal) .....	35
1.5. Problèmes épistémologiques du TAL (Petal) .....	41
<b>2. DYNAMIQUES.....</b>	<b>51</b>
2.1. Acquisitions .....	51
2.1.1. Appropriation du français en situation diglossique.....	51
2.1.2. Représentation des événements : sémantique verbale, temporalité et développement des énoncés complexes .....	60
2.1.3. Apprentissage / enseignement précoce des langues étrangères .....	66
2.2. Didactique .....	69
2.2.1. Didactique de la syntaxe .....	69
2.2.2. Elaboration d'un glossaire de termes linguistiques.....	73
2.3. La variation stylistique.....	80
<b>3. TEXTES / CORPUS.....</b>	<b>84</b>
3.1. Genres et linguistique de corpus .....	84
3.2. Approche multidimensionnelle de l'écrit .....	95
3.2.1. Communication électronique .....	95
3.2.3. Ecrit et Ecriture scolaire.....	99
3.3. Problèmes diachroniques .....	109
3.4. Textes littéraires.....	115
3.3.1. Edition scientifique de Charles Nodier .....	115
3.4.2. Métrique de l'alexandrin au tournant des XIXe et XXe siècles.....	120

<b>4. RESSOURCES LINGUISTIQUES .....</b>	<b>122</b>
4.1. Grammaire de référence du français contemporain.....	122
4.2. Lexique syntaxique du français .....	126
4.3. <i>Terminalf</i> : ressources terminologiques en langue française.....	130
<b>ANNEXE 1 .....</b>	<b>134</b>
<b>ANNEXE 2 : LE SITE MODYCO .....</b>	<b>139</b>

## **Déclaration de politique scientifique pour la période (2005-2008)**

---

Le laboratoire Modèles, Dynamiques, Corpus, UMR 7114 du CNRS/Université de Paris X, a été créé en janvier 2001. Il prend la suite de l'Equipe d'Accueil (EA 372) du même nom. Le laboratoire s'inscrit ainsi dans la continuité thématique et méthodologique de la recherche linguistique menée à l'Université Paris X depuis plusieurs décennies.

Centrée sur la linguistique française, sans pourtant exclure a priori l'étude d'autres langues, cette unité thématique est marquée par un souci empirique qui adosse les analyses linguistiques à une base descriptive solide prenant en compte la réalité des usages qu'ils soient conversationnels, discursifs ou textuels. C'est donc sur la base de corpus de tous ordres et de toutes dimensions que sont proposées des analyses et des modélisations linguistiques s'inscrivant dans les cadres théoriques modernes les plus divers. L'unité thématique et méthodologique du laboratoire se double en effet d'une grande ouverture théorique et d'une polyphonie revendiquée pour ce qui concerne la modélisation. La spécificité des travaux du laboratoire MoDyCo consiste ainsi à intervenir dans les débats théoriques les plus fondamentaux à partir d'un ancrage solide en linguistique française et d'un souci empirique et descriptif constant.

Le laboratoire a participé à deux GDR : phonologie (directeur Bernard Laks) et morphologie (directeur Bernard Fradin, encore en cours). Plusieurs projets de recherche pluri-formations y sont actuellement coordonnés : ACI (directrice Colette Noyau), réseau Phonologie du Français Contemporain (directeur Bernard Laks).

Membre de l'Institut de Linguistique Française (ILF, Fédération CNRS, directrice Christiane Marcello-Nizia), le laboratoire entretient de nombreuses collaborations avec des équipes françaises CNRS et universitaires : IUFM Créteil, Lille 3, Paris III, Paris VII, Paris VIII, Toulouse, etc.

Il participe aussi à de nombreuses opérations internationales contractualisées avec notamment les universités suivantes : Chicago, Fès, Kenitra, Lisbonne, Munich, Oslo, Sousse, Stockholm, Uppsala.

La demande de renouvellement est motivée par la volonté de poursuivre les actions de recherche et la dynamique dont nous avons fait état dans le rapport scientifique concis. Elle est aussi en rapport avec le nombre croissant des étudiants en DEA et des thèses inscrites sous la direction des différents membres de l'équipe nouvelle (Paris X et CNRS). En effet, la diversité articulée des problématiques constitue un cadre précieux et efficace pour la formation des jeunes doctorants et post doctorants qui font partie de l'unité. L'approche diversifiée des recherches, des méthodes, des théories, des outils conceptuels, des terrains, et des modalités d'évaluation des résultats leur permet d'acquérir une vision représentative de ce que sont les sciences du langage actuellement.

**La thématique des recherches conduites au sein du laboratoire se décline selon trois grandes options :**

La perspective modélisatrice concerne au premier chef les recherches phonologiques, morphologiques, syntaxiques et sémantiques, ainsi que les travaux de formalisation en traitement automatique des langues.

La perspective dynamiques renvoie aux recherches menées en acquisition de langue maternelle ou seconde, en sociolinguistique, ainsi qu'en didactique.

La perspective des corpus, tout en étant plus spécifique des recherches en linguistique textuelle, diachronique ou discursive, et en linguistique de l'écrit traverse, en tant que méthodologie d'appui, les autres perspectives de recherche.

La perspective modélisatrice se donne comme programme de recherche la description explicite et formalisée des langues, qui se réalise par la définition et l'exploration de différents plans organisationnels du langage (à tout le moins : phonologie, morphologie, syntaxe, sémantique) et par la découverte des relations de dépendance et d'autonomie de ces plans. En effet un point de convergence pour la plupart des chercheurs peut se résumer en la formule « le langage n'est pas un », ou « le langage est structurellement hétérogène ».

La prise en compte des usages et des dynamiques contribue quant à elle à l'évaluation et à l'évolution des modèles linguistiques. Des études récentes sur l'acquisition des langues illustrent la compétition entre fonctions phrastiques, sémantiques, et discursives. Les recherches sur l'acquisition et sur les usages en situation obligent aussi à tenir compte de la variabilité des systèmes linguistiques à travers le temps, l'espace et les caractéristiques sociales des locuteurs, et à rendre compte de la façon dont chaque locuteur fait sienne une langue pour ses visées communicatives concrètes, régies par des contraintes cognitives, situationnelles et interactionnelles. Elles visent à décrire comment une langue est couramment utilisée par le locuteur, comment elle peut être construite par l'enfant devenant locuteur, ou par un sujet acquérant un nouveau système linguistique. Elles amènent enfin à examiner comment les comportements linguistiques et l'appropriation de la langue sont en partie sous-tendus par les représentations que les locuteurs se font de la langue, y compris celles induites ou construites par l'institution éducative.

L'étude des corpus en situation montre que le lexique, la morphologie, la syntaxe, la manière dont se posent les problèmes sémantiques de l'ambiguïté et de l'implicite, tout cela varie aussi avec les genres et les discours. Pour parvenir à des traitements automatiques spécifiques et efficaces de corpus, il convient de spécifier les fonctionnements propres aux différents discours et genres textuels pour adapter les stratégies. Par ailleurs, la typologie textuelle peut apporter de nouvelles données à la morphologie et à la syntaxe. Bref, l'étude des normes d'usage complète utilement celle des règles, et permet de préciser leurs conditions d'application.

L'analyse de grands corpus intervient alors tant dans la formulation des hypothèses que dans la recherche de contre-exemples ou de variations. On découvre à l'intérieur d'une même langue des normes de portée et de niveau différents (de discours, de genres, de styles). Ainsi, le statut même de l'empirique est-il en train de changer. Il contribue de plus en plus au débat proprement théorique et ceci d'autant plus que les linguistiques de corpus, dès lors qu'elles adoptent un point de vue réflexif à l'égard de leurs propres démarches, peuvent permettre de rompre avec certaines formes d'objectivisme.

La constitution de corpus informatisés de données orales associés à des techniques nouvelles d'étiquetage et de pré-analyse autorisant des traitements à grande échelle permet ainsi des avancées dans la validation des hypothèses sur le rôle respectif de différents niveaux d'analyse et niveaux d'organisation de la langue, et engage en outre à modifier le rapport entre théorie et pratique.

Ces trois dimensions de recherche structurent actuellement les projets en cours au sein du laboratoire. La diversité de ces projets au travers desquels un grand nombre de sous-disciplines des sciences du langage se trouvent convoquées montre comment se trouvent appliqués à des objets linguistiques les plus divers les mêmes soucis d'ordre méthodologique, descriptif et modélisateur : la diversité des objets linguistiques traités marque la richesse des recherches menées au sein du laboratoire, l'attention portée aux dynamiques langagières et acquisitionnelles, le souci descriptif et empirique dans la construction de corpus et la préoccupation de formalisation et de modélisation des hypothèses et des résultats soulignent les synergies et signent l'unité du laboratoire MoDyCo.

Pour répondre au souci de décloisonnement et à la mise en place de synergies trans-domaines dont nous avons parlé précédemment, le laboratoire a conduit une importante opération de restructuration interne autour de la notion de projet de recherche. Chacun des projets actifs à ce jour au sein du laboratoire correspond à la fois à une opération scientifique bien délimitée du point de vue des objets visés, des problématiques et des calendriers, et en même temps répond au souci de collaboration inter-domaines visant à dégager des dynamiques nouvelles.

Ces projets délimités dans le temps et dans l'espace ont donc tous une coloration « interdisciplinaire ».

\*\*\*

Achevée après la remise du contrat quadriennal recherche, reprise en tête du présent rapport, cette opération de restructuration interne de l'unité autour des trois thématiques que nous avons présentées ci-dessus débouche sur un organigramme des opérations scientifiques conduites dans le laboratoire en quatre grandes opérations :

- 1) *Modèles*, qui regroupe les opérations du thème 1 ci-dessus.
- 2) *Dynamiques*, qui regroupe les opérations du thème 2 ci-dessus.
- 3) *Textes / corpus*, qui regroupe les opérations du thème 3 ci-dessus.
- 4) *Ressources linguistiques*, qui donne une visibilité accrue à notre politique de développement et de mise à disposition de la communauté linguistique de ressources électroniques, de corpus balisés et d'outils de fouille. Cette dernière opération est conduite en relation étroite avec la FR Institut de linguistique française et les différents laboratoires qui la composent.

L'organigramme des opérations scientifiques est présenté ci-dessous. Les responsables des différents projets en italiques. Suit la structure de direction du laboratoire constituée de Bernard Laks (Directeur) et Sophie David (Directeur adjoint), et la liste de ses membres.

Les nouveaux projets sont identifiés par « Nouveau projet » au début de la description. Ils ne comportent pas bien évidemment de partie « Bilan ».

La production scientifique notée dans chaque projet est indicative et non exhaustive. On se reportera au rapport du 30 octobre 2003 pour disposer de la bibliographie complète.

# Organigramme par opérations scientifiques (2005-2008)

Modèles	Dynamiques	Textes / Corpus	Ressources linguistiques
<p><b>Phonologie de la syllabe et phonologie du français</b>  <i>Bernard Laks</i>  <i>Philippe Gréa</i>  <i>Marc Klein</i>  <i>Atanas Tchobanov</i>  <i>Richard Walter</i></p>	<p><b>Appropriation du français en situation diglossique</b>  <i>Colette Noyau</i>  <i>Maria Kihlstedt</i>  <i>Carole Tisset</i></p>	<p><b>Genres et linguistique de corpus</b>  <i>Denise Malrieu</i>  <i>François Rastier</i>  <i>Catherine Boré</i>  <i>Simon Bouquet</i>  <i>Evelyne Bourion</i></p>	<p><b>Grammaire de référence du français contemporain</b>  <i>Annie Delaveau</i>  <i>Jacques Anis</i>  <i>Marianne Desmets</i>  <i>Françoise Gadet</i>  <i>Danielle Leeman</i></p>
<p><b>Morphologie et sémantique lexicale</b>  <i>Françoise Kerleroux</i>  <i>Sophie David</i>  <i>Sarah Leroy</i>  <i>Sergueï Sakhno</i></p>	<p><b>Représentation des événements</b>  <i>Colette Noyau</i>  <i>Maria Kihlstedt</i>  <i>Danielle Leeman</i></p>	<p><b>Communication électronique</b>  <i>Jacques Anis</i>  <i>Françoise Gadet</i></p>	<p><b>Lexique syntaxique</b>  <i>Sylvain Kahane</i>  <i>Karine Baschung</i>  <i>Marcel Cori</i>  <i>Marianne Desmets</i></p>
<p><b>Les connecteurs</b>  <i>Danielle Leeman</i>  <i>Annie Bertin</i>  <i>Danièle Flament-Boistrancourt</i>  <i>Françoise Gadet</i>  <i>Colette Noyau</i></p>	<p><b>Apprentissage / enseignement précoce des langues étrangères</b>  <i>Maria Kihlstedt</i>  <i>Jean-Jacques Briu</i>  <i>François Muller</i>  <i>Colette Noyau</i>  <i>Sergueï Sakhno</i></p>	<p><b>Écrit et écriture scolaire</b>  <i>Catherine Boré</i>  <i>Jacques Anis</i>  <i>Marie-Laure Elalouf</i></p>	<p><b>Terminalf</b>  <i>Richard Walter</i></p>
<p><b>Modélisation de données linguistiques dans la perspective du TAL (METAL)</b>  <i>Marcel Cori</i>  <i>Marianne Desmets</i>  <i>Karine Baschung</i>  <i>Sophie David</i>  <i>Annie Delaveau</i>  <i>Françoise Kerleroux</i>  <i>Danielle Leeman</i>  <i>Anna Sörös</i></p>	<p><b>Didactique de la syntaxe</b>  <i>Danielle Leeman</i>  <i>Marie-Laure Elalouf</i>  <i>Carole Tisset</i></p>	<p><b>Problèmes diachroniques</b>  <i>Annie Bertin</i>  <i>Jacques Anis</i>  <i>Danielle Leeman</i>  <i>Isabelle Weill</i></p>	
<p><b>Problèmes épistémologiques du TAL (PETAL)</b>  <i>Marcel Cori</i>  <i>Sophie David</i>  <i>Karine Baschung</i>  <i>Marianne Desmets</i>  <i>Claude de Loupy</i>  <i>Denise Malrieu</i>  <i>Richard Walter</i></p>	<p><b>Elaboration d'un glossaire de termes linguistique</b>  <i>Jean-Jacques Briu</i>  <i>Anne Trévisse</i>  <i>Danièle Flament-Boistrancourt</i>  <i>Marie-Laure Elalouf</i>  <i>François Muller</i>  <i>Rémy Porquier</i>  <i>Sergueï Sakhno</i>  <i>Carole Tisset</i></p>	<p><b>Edition scientifique de Charles Nodier</b>  <i>Jean-François Jeandillou</i></p>	
	<p><b>La variation stylistique</b>  <i>Françoise Gadet</i>  <i>Danièle Flament-Boistrancourt</i>  <i>Christine Pauleau</i></p>	<p><b>Métrie de l'alexandrin au tournant des XIXe et XXe siècles</b>  <i>Jean-François Jeandillou</i>  <i>Michel Arrivé</i></p>	
			<p><b>Site PFC</b>  <i>Bernard Laks, Richard Walter</i>    infolang.u-paris10.fr/pfc</p>
			<p><b>Site MoDyCo</b>  <i>Bernard Laks, Sophie David, Richard Walter</i>    infolang.u-paris10.fr/modyco</p>



# Membres du laboratoire

---

Les listes sont légèrement différentes de celles qui ont été fournies le 30 octobre 2003. Elles tiennent compte des changements survenus depuis.

## Direction du laboratoire

Directeur : Bernard Laks

Directeur adjoint : Sophie David

Responsable des sites Web : Richard Walter

## Chercheurs

Sophie David	CR CNRS, Directeur adjoint
Sarah Leroy	CR CNRS
François Rastier	DR CNRS

## Enseignants-chercheurs

Jacques Anis	PR	Françoise Kerleroux	PR
Michel Arrivé	PR	Maria Kihlstedt	MC
Karine Baschung	MC	Marc Klein	MC
Annie Bertin	PR	Bernard Laks	PR, Directeur
Catherine Boré	MC	Danielle Leeman	PR
Simon Bouquet	MC	Claude de Loupy	MC PAST
Jean-Jacques Briu	PR	François Muller	PR
Marcel Cori	PR	Colette Noyau	PR
Annie Delaveau	PR	Christine Pauleau	MC
Marianne Desmets	MC	Rémy Porquier	PR
Marie-Laure Elalouf	MC	Sergueï Sakhno	MC
Danièle Flament-Boistrancourt	PR	Anna Sörös	MC
Françoise Gadet	PR	Carole Tisset	MC
Philippe Gréa	MC	Anne Trévisé	PR
Jean-François Jeandillou	PR	Isabelle Weill	MC
Sylvain Kahane	PR (2005)		

## Ingénieurs

Evelyne Bourion	IE CNRS
Denise Malrieu	IE CNRS
Atanas Tchobanov	Ingénieur Université Paris 10
Richard Walter	IE CNRS

**Membres associés**

Driss Ablali	MC	Michèle Perret	PR
Jacqueline Bacha	PR	Gérard Petit	MC
Cristina de Lorenzo Rosselló	MC	Wilfrid Rotgé	PR
Loïc Depecker	PR	Jérémi Sauvage	Prof. des écoles
Maria Jarrega	Prof. secondaire	Maxime Somé	Prof. secondaire
Sonia Gerolimich	Lecteur	Florence Villoing	MC
Pierre Patrick Haillet	MC		

**Jeunes chercheurs associés**

Ufoalè-Christine Afola		Sabine Lehman	
Fatiha Amouri		Marielle Lignereux	
Antonio Balvet		Céline Poudat	
Jin Nam Choi		Silvia Romeo	
Fidèle Diedhiou		Gema Sanz	
Christine Hénault-Sakhno		Mathieu Valette	
René Lavie		Cécile Vigouroux	

**Doctorants**

Mohamed Anis Abdelkader		Anne Lablanche	
Lamia Allal		Belinda Lavieu	
Ramon Alvarado		Chrystèle Lesselingue	
Chiraz Anane-Bahroun		Sylvain Loiseau	
Encarnacion Arroyo		Géraldine Mallet	
Allahverdi Azari Nadjaf Abad		Aboubacar Mamah	
Thomas Beauvisage		Izabel Marques da Silva	
Emilie Bellanger		Yvon Martin	
Mouna Ben Amor		Nicole Marty	
Corinne Berthery		Paul Memmi	
Mounia Bourrakadi		Rossitza Milenkova-Kyheng	
François Champion		Majid Moslemi	
Eliane Cheung		Agnès Muller	
Tae-Rin Cho		Valelia Muni-Toké	
Dominique Declercq		Aya Ono	
Fabrice Delumeau		Adriana Orlandi	
Ibrahima Diallo		Houda Ounis	
Dieynaba Diop		Ha Sung Park	
Smaïl Djaoud		Coralie Roger	
Monique Duboscq		Dana Saikali	
Khadija Elkandoussi		Madona Sakhokia	
Mauricio d'Escragolle Cardoso		Sophie Saulnier	
Caroline Facq-Mellet		Jamila Sebbar	
Kossi Souley Gbeto		Yasuo Shimokawa	
Hasna Ghoul		Takayoshi Suzuki	
Emmanuelle Guérin		Aboubakry Sy	
Nathalie Guezennec		Bayal Sy	
Solange Gularté		Gaëlle Tassin	
Baddredine Hamma		Ali Tifrit	
Sophie Hamon		Patrick Tomatis	
Florence Hochet		Henry Tyne	
Khadija Itrib		Céline Vaguer	
Kyung-Hwa Kang		Nora Vera	
Ichraf Khammari		Anamaria Villecco	
Savas Kilic		Theodora Vonitsanou	
Wanrudee Kurawan		Mehran Zendehboudi	

# 1. Modèles

---

- 1) Phonologie de la syllabe et phonologie du français
- 2) Morphologie et sémantique lexicale
- 3) Les connecteurs
- 4) Modélisation de données linguistiques dans la perspective du traitement automatique des langues (METAL)
- 5) Problèmes épistémologiques du TAL (PETAL)

## 1.1. PHONOLOGIE DE LA SYLLABE ET PHONOLOGIE DU FRANÇAIS

### **Participants**

*Membres rattachés à titre principal* : Bernard Laks (responsable), Marc Klein, Philippe Gréa, Atanas Tchobanov, Richard Walter

*Jeunes chercheurs associés* : Antonio Balvet, René Lavie

*Doctorants* : Muriel Azali, Géraldine Mallet, Allahverdi Azari Nadjaf, Ali Tifrit, Patrick Tomatis, Conrad Johansson (co-tutelle Université d'Uppsala), Elissa Sobota (co-tutelle Université de Munich), Katherine Ostby (co-tutelle Université d'Oslo)

*Collaborations extérieures* :

Ernesto d'Andrade (Université de Lisbonne)

Gabriel Bergounioux (Université d'Orléans)

Jacques Durand (Université de Toulouse)

Pierre Encrevé (EHESS)

John Goldsmith (Université de Chicago)

Chantal Lyche (Université d'Oslo)

Noël Nguyen (Université de Provence)

Sophie Wauquier-Gravelines (Université de Nantes)

### **Discipline(s) et /ou sous-discipline(s) concernée(s)**

Phonologie, morphologie, linguistique cognitive, réseaux connexionnistes, histoire de la linguistique.

## Problématique

Jusqu'à une date récente, les chaînes sonores étaient considérées comme strictement linéaires et leur organisation semblait pouvoir être décrite dans les termes d'une syntagmatique simple. On en prendra pour preuve le fait que le concept de syllabe était totalement absent de l'approche générative classique de SPE (Chomsky & Halle 1968). En effet, dans une conception dérivationnelle, les représentations phonologiques n'ont pas d'autres structures internes que leur organisation *unilinéaire*. Le champ phonologique contemporain (cf. Laks 1997 ; Goldsmith & Laks 2000 ; Laks 2001) s'est pour une part construit en réaction critique à cette conception dérivationnelle. Les approches les plus récentes ont mis l'accent sur l'architecture des représentations phonologiques et sur l'organisation interne des constituants, au premier rang desquels il faut placer la syllabe. L'approche autosegmentale (Goldsmith 1990), la géométrie des traits (Clements & Hertz 1996), la phonologie de laboratoire (Browman & Goldstein 1990), la phonologie computationnelle (Angoujard 1997), la phonologie du gouvernement (Kaye *et al.* 1988), la théorie de l'optimalité (Prince & Smolensky 1993) suggèrent toutes une approche très différente dans laquelle la syllabe joue un rôle organisateur central (cf. en outre Browman *et al.* 1998 ; Durand *et al.* 1998). L'architecture phonologique est en effet aujourd'hui conçue comme multilinéaire et la syllabe est vue comme le locus privilégié d'organisation de cette architecture complexe. Les phénomènes d'organisation temporelle des événements articulatoires, de phasage et de déphasage de gestes, de tempo et de rythme, les synchronicités partielles, et les désynchronisations, jouent dans ces approches un rôle particulièrement important. Les processus d'allongement, compensatoire ou non, et de quantité en sont la trace, le rythme et l'accentuation en dépendent. La phonologie contemporaine redécouvre ainsi les conceptions syllabiques et phonotactiques de Saussure (2002) et Grammont (1933).

Ces questions théoriques liées aux modélisations phonologiques se doublent d'une réactualisation des préoccupations proprement empiriques. La description phonologique du français considérée dans son ensemble est restée en jachère depuis les dernières grandes enquêtes structuralistes (Martinet 1945 ; Walter 1977, 1982). Les quelques travaux renouvelant l'empirie, d'orientation phonologique ou sociolinguistique (Encrevé 1988), sont restés limités dans leur ampleur descriptive. La plus grande partie des analyses et des modélisations de la phonologie du français s'appuie encore sur les descriptions normatives reçues de Fouché (1959). Il était donc nécessaire de renouveler nos connaissances concernant la variété des usages phonologiques du français et de documenter plus précisément les structures phonologiques telles qu'elles se livrent dans l'observation systématique.

Les modélisations phonologiques les plus contemporaines s'articulent à partir de thèses cognitives précises et mettent en œuvre des systèmes dynamiques allant des systèmes d'équilibration harmonique aux systèmes pilotés par des contraintes et aux approches neuromimétiques. Les préoccupations directement cognitives, tout comme les modèles de réseaux de neurones, ou la programmation par hiérarchie de contraintes constituent donc des pendants naturels du travail plus proprement phonologique.

### **Contexte international**

Au plan international, la phonologie connaît un développement rapide. De nombreux modèles apparaissent illustrant chacune des grandes tendances évoquées ci-dessus. Ces nouveaux modèles sont souvent peu ou mal connus au plan national. L'équipe a ainsi joué un rôle important en présentant et en diffusant largement les modèles phonologiques à contraintes dans leurs versions les plus récentes, ainsi que les modélisations connexionnistes appliquées à la phonologie et à la variation phonologique (Boersma 1998 ; Hayes 1999 ; Antilla & Cho 1998 ; Plaut & Kello 1999).

Dans cette même logique, et en liaison avec les suites de l'opération GDR Phonologies, l'équipe nanterroise est actuellement porteuse de l'organisation d'un réseau européen dédié à la phonologie. Ce réseau, soumis au soutien de la Fondation Européenne pour la Science au printemps 2004, aura pour objectif de structurer la communauté des phonologues des 25 pays européens en prenant appui sur le 3<sup>ème</sup> colloque international Current Trends in Phonology (CTIP III) que l'équipe co-organisera en 2006 à l'Abbaye de Royaumont.

### **Contexte national**

Si le champ phonologique français apparaît encore comme relativement dispersé, le champ international est en plein développement. Pour répondre à cette dispersion, l'équipe a joué un rôle central dans la construction et l'animation du GDR Phonologie (GDR 1954) jusqu'en 2003. Cette structure originale a permis de regrouper autour de thématiques communes l'entièreté des phonologues français soit plus d'une soixantaine de chercheurs appartenant à plus de 20 laboratoires ou équipes différentes. Cette structure a joué un rôle particulièrement important en faisant émerger la communauté des phonologues et en l'organisant autour de synergies précises (5 à 6 journées thématiques annuelles, colloque annuel, etc.).

### **Thèmes de recherche**

**1) Phonologie syllabique.** Le projet de recherche prend la suite des travaux menés au cours de la période précédente. Il s'agit tout d'abord d'interroger les primitives syllabiques et la structure minimale qui les réunit dans sa portée universelle. Un important travail est en cours sur ces questions (thèse d'Ali Tifrit, travaux de Marc Klein). L'organisation syllabique dans une langue comme le français conduit immédiatement à interroger la phonotactique et la syntagmatique phonologique (Bernard Laks). Enfin la question syllabique se trouve corrélée à des questions neuro-cognitives centrales : rôle de perception sur et dans la production, structuration syllabique minimale précablée au niveau directement neuronal (travaux d'Atanas Tchobanov, thèse de Patrick Tomatis).

**2) Liaison et interface phonologie / morphologie.** L'un des locus d'application des dynamiques syllabiques du français est constitué par la liaison et son interface avec la morphologie. Si la liaison a jusqu'alors été traitée dans une approche unitaire comme un processus strictement phonologique, les travaux récents menés au laboratoire suggèrent que se trouvent rassemblés sous un vocable unique des processus très différents : tendances dynamiques proprement phonologiques d'une part, mécanismes morphologiques d'inflection très anciens mais toujours actifs en français de l'autre (travaux de Bernard Laks, thèse de Géraldine Mallet).

**3) Phonologie du français contemporain : usages variétés et structure (PFC).** Ce projet, co-piloté par Bernard Laks, Jacques Durand et Chantal Lyche, consiste en une enquête de très grande envergure sur la phonologie du français. Plus de 50 points d'enquête répartis dans la francophonie, 500 locuteurs et 700 heures d'enregistrement sont programmés. A partir d'un protocole unique scientifiquement construit et testé, plus de 45 chercheurs différents se sont associés à ce projet aujourd'hui à mi-course. La base de données sonores en projet sera constituée de l'ensemble des données enregistrées, transcrites, étiquetées et organisées. Elle fournira un référent central pour tous les travaux de phonologie française. Un module prosodique est en cours de développement.

**4) Histoire.** Le développement d'une histoire de la phonologie fait partie intégrante des conditions de son développement futur. En effet, on peut montrer que la phonologie constitue une discipline réellement cumulative, pour laquelle les typologies de processus, les phénoménologies et les problématiques sont constantes et partagées. Au-delà des divergences de cadre théorique de référence, l'histoire de la phonologie doit permettre de relier les problématiques contemporaines aux riches tendances scientifiques dont, souvent sans le savoir, elles sont issues. Cet axe de recherche se concrétise essentiellement par l'organisation de colloques et de journées d'études (coopération Bernard Laks et Gabriel Bergounioux) et par une série de publications.

**5) Atelier de modélisation connexionniste.** Le développement de modélisations dynamiques suppose la maîtrise technique d'outils de programmation et de simulation. Un atelier de modélisation connexionniste a été mis en place qui poursuivra ses travaux dans la prochaine période (Ali Tifrit et Philippe Gréa responsables, travaux de Atanas Tchobanov et thèse de Patrick Tomatis).

## ***Bilan***

**1) Phonologie syllabique.** L'équipe s'est centrée sur la modélisation syllabique et sur les primitives neuro-cognitives (Thèse de Tchobanov, participation au projet Abispa dirigé par Jean-Louis Boé et Jean-Luc Schwartz à Grenoble : *Apprentissage bayésien intersensoriel des structures phonologiques*. Financement ACI Cognitive). Un modèle neuromimétique a été développé (Atanas Tchobanov).

**2) Liaison et interface phonologie / morphologie.** Un premier bilan sur les travaux concernant la liaison et sa modélisation a été effectué. Parallèlement au corpus PFC, un corpus spécifique portant sur le discours des hommes politiques dans la période 1908-1980 a été constitué, numérisé, balisé et étiqueté (collaboration Bernard Laks avec les Archives de la Parole – BNF, constitution et formatage de la base par Antonio Balvet). Ces résultats constituent la base de départ de travaux d'analyse en cours.

**3) Phonologie du français contemporain : usages variétés et structure (PFC).** Le projet PFC est actuellement à mi-course. Plus 50 enquêtes ont déjà été réalisées. Une quinzaine sont en cours de traitement et seront versées à la base au printemps 2004. Quatre colloques nationaux et quatre journées d'études techniques ont été organisés dans le cadre de cette opération qui a été marquée par l'édition d'un numéro spécial de *La Tribune des Langues*

*Vivantes* à laquelle 5 membres de l'équipe ont contribué (Bernard Laks, Ali Tifrit, Géraldine Mallet, Elissa Sobota, Richard Walter).

**4) Histoire.** Le thème « Histoire de la phonologie » a été marqué par l'organisation d'un colloque en 2002 (co-organisation Bernard Laks, Gabriel Bergounioux) et par des publications.

**5) Atelier de modélisation connexionniste.** L'atelier de modélisation connexionniste s'est réuni régulièrement ; il a produit une première ébauche de plate-forme dédiée à la modélisation des réseaux de neurones sous Matlab. Des outils de simulation et des outils de représentation graphique ont été élaborés (Philippe Gréa, Ali Tifrit, Atanas Tchobanov).

### ***Perspectives***

**1) Phonologie syllabique.** Analyse sur corpus des groupes consonantiques lourds en français et élaboration d'un indice de complexité structurale en relation avec une théorie de la marque. Relation entre complexité syllabique, perception auditive et pathologies. Elaboration d'un modèle neuronal de simulation syllabique.

**2) Liaison et interface phonologie / morphologie.** Analyse du corpus des hommes politique du point de vue diachronique et dynamique. La question des enchaînements et ses conséquences sur la resyllabation. L'analyse des contraintes morphologiques sur la liaison et la reprise de la tripartition obligatoire, facultatif interdit. Epenthèses consonantiques vraies et fausses en contexte de liaison. Organisation d'un colloque en mars 2004.

**3) Phonologie du français contemporain : usages variétés et structure (PFC).** Poursuite de l'analyse des enquêtes effectuées. Construction de la base de donnée balisée, développement d'outils. Trois volumes sont prévus à paraître chez Hermès et un chez Oxford University Press.

**4) Histoire.** Organisation d'un colloque en février 2004.

**5) Atelier de modélisation connexionniste.** Poursuite de la construction de la plate forme connexionniste. Développement et interfaçage des outils.

### **Production scientifique**

Gréa Ph. 2002. « Intégration conceptuelle et métaphore filée ». *Langue Française*, 134, 109-123.

Gréa Ph. 2003. « Les limites de l'intégration conceptuelle ». *Langages*, 150, 61-74.

Laks B. 2001. « Enquête sur la phonologie du français contemporain ». *Le Français dans le monde*, numéro spécial, 106-108.

Laks B. 2001. « Un siècle de phonétique et de phonologie : quelques questions permanentes ». *Modèles Linguistiques*, XXII-1, 75-103.

Laks B. 2001. « Matériaux pour une histoire de la phonétique et de la phonologie en France au XXe siècle ». *Modèles Linguistiques*, XXII-1, 3-9.

- Laks B. 2002. « Le comparatisme : de la généalogie à la génétique ». *Langages*, 146, 19-46.
- Laks B. 2002. « La description du français oral ». *L'information grammaticale*, 5-10.
- Laks B. 2003. « Linguistique cognitive et bilinguisme natif : quelques remarques ». *Traduire*, 196.
- Laks B. 2003. « L'oral et la variation ». In Durand J., Laks B., Lyche Ch. (éds), *La tribune internationale des langues vivantes* : « La prononciation du français dans sa variation », 33, 96-102.
- Laks B., Durand J., Lyche Ch. 2002. « The Phonology of Contemporary French : usages, varieties and structure ». *The Phonetician*.
- Laks B., Durand J., Lyche Ch. 2003. « La prononciation du français ». In Durand J., Laks B., Lyche Ch. (éds), *La tribune internationale des langues vivantes* : « La prononciation du français dans sa variation », 33, 1-3.
- Laks B., Victorri B. 2002. « Origine de l'homme, origine des langues : introduction ». *Langages*, 146, 1-7.
- Mallet G. 2003. « Quelques observations sur la liaison à Treize-Vents (Vendée) ». *La Tribune internationale des langues vivantes*, 33, 145- 150.
- Tchobanov A. 1997. « Apprentissage de la syllabation par des réseaux connexionnistes partiellement récurrents ». *Actes des Journées d'études «La voyelle dans tous ses états»*, Nantes, 5-7 décembre 1997.
- Tchobanov A. 2003. « Phonologie et Connexionnisme ». In Angoujard J.-P., Wauquier S. (éds), *Phonologie. Champs et Perspectives*. ENS éditions.
- Tifrit A., Tchobanov A., Laks B. A paraître. « Modélisation connexionniste de la syllabation ». *Actes du colloque «La Syllabe dans l'identification des mots écrits»*, Reims, 13-14 février 2002.

### **Thèses soutenues**

- Antonio Balvet. 2002. *Approches catégoriques et non catégoriques de la linguistique. linguistique des corpus spécialisés : application à un système de filtrage d'information*. Directeur B. Laks. Thèse CIFRE en collaboration avec l'entreprise Thalès.
- Philippe Gréa. 2001. *La théorie de l'intégration conceptuelle appliquée à la métaphore et à la métaphore filée*. Directeur B. Laks.
- René Lavie. 2003. *Le locuteur analogique ou la grammaire remise à sa place*. Directeur B. Laks.
- Atanas Tchobanov. 2002. *Représentations et apprentissage des primitives phonologiques : approches neuromimétiques*. Directeur B. Laks.

### **Habilitations à diriger des recherches**

- Anne Lacheret. 2001. *Prosodie du français contemporain*.
- Penou-Achille Somé. 2002. *Le Dagara et les variétés du groupe voltaïque (Burkina-Faso) : Linguistique fondamentale, ethno-linguistique, linguistique appliquée*.
- Odile Chasles. A venir. *Linguistique cognitive et langues de spécialités*.



### **Thèses inscrites**

- Muriel Azali. *Phonologie du malgache : le vocalisme*. Directeur B. Laks.  
Allahverdi Azari Nadjaf. *La structure accentuelle du persan*. Directeur B. Laks.  
Conrad Johansson. *La perception de l'espace sur internet et la théorie de l'intégration conceptuelle*. Directeur B. Laks en co-tutelle avec l'université d'Uppsala.  
Géraldine Mallet. *Phonologie de français : la liaison et l'enchaînement*. Directeur B. Laks.  
Katherine Ostby. *Dynamiques du e muet en français contemporain*. Directeur B. Laks en co-tutelle avec l'université d'Oslo.  
Elissa Sobota. *Variations phonologiques et identité linguistique*. Directeur B. Laks en co-tutelle avec l'université de Munich.  
Ali Tifrit. *La syllabe : description, modèles, théories*. Directeur B. Laks.  
Patrick Tomatis. *Apprentissage phonologique et perception multimodale : le rôle de l'oreille*. Directeur B. Laks.

### **Contrat / financement**

- 1) Abispa, dans le cadre de l'ACI Cognitive (terminé).
- 2) Contrat DGLFLF (2001 terminé ; 2002 terminé ; 2003 en cours).
- 3) Contrat ILF (2002-2003 en cours)
- 4) Contrat Région Midi-Pyrénées (terminé)
- 5) Contrat MEN « Ressources numériques » (en cours de signature).

### **Références générales sur le projet**

- Antilla A., Cho Y. M. Y. 1998. « Variation and change in optimality theory ». *Lingua*, 104, 31-56.
- Boersma P. 1998 *Functional phonology : Formalizing the interactions between articulatory and perceptual drives*. Amsterdam : LOT.
- Browman C. P., Goldstein L. 1990. « Tiers in articulatory phonology, with some implications for casual speech ». *Papers in laboratory phonology I : Between the grammar and physics of speech*.
- Browman C. P., Goldstein L., Honorof D. N., Jebbour A., Selkirk, E. (1998). « Gestural organization underlying syllable structure ». In Berber T. (éd.), *Current Trends in Phonology II : Models and Methods*.
- Chomsky N., Halle M. 1968. *Sound Pattern of English*. New York : Harper and Row.
- Clements G. N., Hertz S. R. 1996. « An Integrated Approach to Phonology and Phonetics ». *Current Trends in Phonology : Models and Methods*.
- Durand J., Laks B., Royauumont, Angoujard J.-P. 1998. *Théorie de la syllabe : rythme et qualité*. Paris : Editions du CNRS.
- Encrevé P. 1988. *La liaison avec et sans enchaînements : phonologie tridimensionnelle et usages du français*. Paris : Le Seuil.
- Fouché P. 1959. *Traité de prononciation française*. Paris : Klincksieck.
- Goldsmith J. 1990. *Autosegmental and Metrical Phonology*. Oxford : Blackwell.
- Goldsmith J., Laks B. 2000. « The History of phonology in the twentieth century ». *Folia Linguistica*, XXXIV/1-2.
- Grammont M. 1933. *Traité de phonétique*. Paris : Delagrave.

- Hayes B. 1999. « Phonetically-driven phonology. The role of optimality theory and inductive grounding ». In Darnell M. *et al.* (éds), *Functionalism and formalism in linguistics*, Amsterdam : John Benjamins, 243-285.
- Kaye J., Lowenstamm J., Vergnaud J.-R. 1988. « Constituent structure and government in phonology ». *Phonology*. 7.
- Laks B. 1997. *Langages* « Nouvelles phonologies », 125.
- Laks B. 2001. « Histoire de la phonétique et de la phonologie en France au XXe siècle ». *Modèles Linguistiques*, XXII-1.
- Laks B. 2003. Saussure's phonology. In Galberg H. *et al.* (éds), *Take danish for instance. Linguistic studies in honor of hans basboll*. Odense : University of Southern Denmark.
- Martinet A. 1945. *La prononciation du français contemporain : Témoignages recueillis en 1941 dans un camp d'officiers prisonniers*. Genève : Droz.
- Plaut D., Kello C. 1999. « The emergence of phonology from the interplay of speech comprehension and production ». *The emergence of language*. MacWhinney, Brian. Mahwah N.J. : Lawrence Erlbaum
- Prince A., Smolensky P. 1993. *Optimality Theory : constraints interaction in generative grammar*. New Brunswick - New Jersey : Ms. Rutgers University.
- de Saussure F. 2002. *Théorie des sonantes. Il manoscritto du genevra (bpu ms fr 3955)*. Florence : Università degli studi di Firenze, UNIPRESS.
- Walter H. 1977. *La phonologie du français contemporain*. Paris : Presses Universitaires de France.
- Walter H. 1982. *Enquêtes phonologiques et variétés régionales du français*. Paris : Presses Universitaires de France.

## **1.2. MORPHOLOGIE ET SÉMANTIQUE LEXICALE**

### **Participants**

*Membres rattachés à titre principal* : Françoise Kerleroux (responsable), Sophie David, Sarah Leroy, Sergueï Sakhno

*Membres associés* :

Sophie Aliquot (CNRS UMR 8528 - Silex - Université Lille 3)  
Florence Villoing (Université Paris VIII)

*Doctorants* : Chrystèle Lesselingue, Coralie Roger, Sophie Saulnier, Aboubakry Sy

*Collaborations extérieures* :

Carmen Kelling (Université de Konstanz, Allemagne)  
Judith Meinschäfer (Université de Konstanz, Allemagne)

### **Discipline(s) et/ou sous-discipline(s) concernée(s)**

Phonologie, morphophonologie, syntaxe, sémantique lexicale, morphologie flexionnelle et constructionnelle

### **Problématique**

1) La morphologie constructionnelle a pour objectif de décrire le plus explicitement possible les règles, principes et mécanismes qui gouvernent la construction des lexèmes. Autrement dit, les régularités qui organisent le sens et la forme des lexèmes non simples par rapport à leur base constituent un domaine qui se différencie de celui de la morphologie flexionnelle. Dans le premier cas, on étudie et on calcule les rapports entre deux lexèmes ; dans le second, on étudie le rapport entre un lexème et les formes fléchies qui constituent son paradigme. Les procédés formels sont, pour la plupart, communs aux deux domaines (affixation, conversion, apophonie, réduplication), mais les caractéristiques sémantiques en jeu sont complètement différentes. Les rapports d'interface entre la morphologie constructionnelle et la morphologie flexionnelle d'une part, et le lexique, la sémantique lexicale et la syntaxe, d'autre part, sont donc des objets de recherche.

2) La morphologie constructionnelle ainsi définie, et plus généralement le programme de la morphologie, identifiée comme composante autonome de la grammaire, ne remonte qu'à trois décennies (Aronoff 1976 ; Corbin 1987 ; Scalise 1984) ; ainsi les objectifs de la recherche sont à la fois la *description empirique des phénomènes et l'élaboration des unités et principes théoriques d'analyse*. En effet, la description des phénomènes est en un sens à peine commencée, dans la mesure où les cadres théoriques de la grammaire des mots construits ont évolué (cf. Kerleroux à par.). En particulier, les modèles de représentation concaténatifs à base morphème ont été mis en question et récusés, au profit de modèles dont les lexèmes fournissent les bases, et dont les opérations sont représentées en termes de processus ou de fonctions. L'organisation du lexique des langues naturelles est envisagée autrement que comme une liste de mots, analysables ou opaques, mais

comme une base de connaissances hiérarchisée (cf. *Langages*, 152, « Quoi de neuf en morphologie ? », Fradin & Kerleroux (éds)).

3) Les unités prises en compte par les opérations morphologiques sont donc les unités multidimensionnelles du lexique : un lexème, signe linguistique, abstrait puisque hors emploi, est caractérisé par les diverses propriétés que sont sa forme phonologique, sa catégorie, sa valence, son stock éventuel de radicaux allomorphiques ou supplétifs, son appartenance à une classe flexionnelle, sa représentation sémantique (cf. Fradin 2003 ; Kerleroux 2003 et à paraître ; Kerleroux & Fradin 2001 et 2003).

4) En sémantique lexicale, un important travail a été réalisé ces trente années (entre autres Katz 1972 ; Marandin 1984a, 1984b ; Lakoff 1987 ; Gross 1994 ; Pustejovsky 1995 ; Jayez & Godard 1995 ; Godard & Jayez 1996 ; Cadiot & Nemo 1997 ; Kleiber 1999). Ces approches, si elles se distinguent pour une part selon la manière de considérer (ou pas) et de traiter le sens référentiel, ont par ailleurs en commun de ne pas prendre en compte le calcul du sens dans les mots construits. Il importe cependant de les examiner en détail, afin de proposer à terme un modèle qui vaille pour les mots simples et les mots construits. En liaison avec cet examen, on se pose la question de la représentation sémantique des noms propres.

### **Contexte**

Dans le domaine de la morphologie, le cadre de travail esquissé ci-dessus ne constitue pas une position unitaire pour les morphologues contemporains. Dans les courants générativistes et post-générativistes, certains persistent à estimer que la grammaire des mots est homologue de la grammaire des syntagmes et des phrases.

Au sein de cette *approche réductionniste*, parmi les travaux de ces trente dernières années, on peut distinguer ceux qui revendiquent l'autonomie de la morphologie (Selkirk 1982), mais qui estiment que cette autonomie se limite à la reconnaissance d'unités spécifiques (les unités liées que ne voit pas la syntaxe, c'est-à-dire les affixes). Il y a ceux qui souhaitent et font en sorte que la morphologie ne soit pas autonome, par préférence explicite pour une grammaire moniste homogène (Lieber 1992). Enfin les travaux qui font expressément de la morphologie une sous-théorie de la syntaxe, au même titre que les autres sous-théories spécialisées (telles que la « théorie du liage » ou la « théorie du gouvernement ») et qui intègrent par le fait le niveau des mots et de leurs constituants dans la hiérarchie syntagmatique (telle qu'elle s'exprime dans la théorie dite X-barre), d'où les niveaux négatifs d'entités notées  $X^1$  (Baker 1985).

Parmi les travaux qui explorent la thèse de l'autonomie de la composante, et qui donc mettent à leur charge les trois objectifs de définir les primitives, les règles de la combinatoire, les principes sur lesquels sont fondées ces règles, on peut ranger Dell (1970), Aronoff (1976, 1994), Corbin (1987), Anderson (1992), Beard (1995).

Dans ce rapide panorama, nous n'avons pas fait figurer des recherches en morphologie qui ne jouent pas la carte de l'intégration dans une théorie grammaticale générale. La morphologie prise alors séparément peut être l'objet d'une recherche typologique et

cognitive, comme c'est le cas du courant appelé morphologie naturelle, illustré par les travaux de Dressler *et al.* (1987).

La recherche en morphologie en France nous est d'autant mieux connue que la majorité des membres du projet a appartenu ou appartient au GDR de morphologie dirigé par B. Fradin (2000-2004).

Trois principaux courants de recherche sont représentés en France en ce moment :

a) Des travaux portant sur la morphologie constructionnelle et la structuration du lexique, dans la lignée des travaux de Corbin (1987) (cf. David 1993 ; Dal 1997 ; Amiot 1997 ; Aurnague & Plénat 1997 ; Temple 1996). A ce courant se rattachent plusieurs des travaux des membres de l'équipe (Kerleroux 1996) et des thèses dirigées à Nanterre (Villoing, Lesselingue, Roger).

b) Des recherches en morphophonologie, menées en particulier à Toulouse par M. Plénat et M. Roché (Plénat 2000 ; Plénat & Roché, à par. ; Roché 1997 ; Roché 2003) ont rapporté des moissons de phénomènes formels, tels que la « bonne forme » est l'objectif visé par toutes sortes d'opérations comme l'haplologie, les contraintes de taille, les opérations dissimilatives, la présence d'interfixes à la jointure du radical et des affixes, etc.

c) Des recherches qui conjoignent morphologie lexicale et linguistique computationnelle, en mettant en œuvre les grands corpus (cf. Fradin, Hathout & Meunier 2003).

### **Thèmes de recherche**

**1) La morphologie constructionnelle comme révélateur de l'organisation sémantique du lexique.** Il s'agit de promouvoir l'étude de l'interface morphologie constructionnelle/sémantique lexicale, en parallèle avec celle qui porte sur l'interface syntaxe/sémantique. En effet depuis la position de l'« hypothèse lexicaliste » (Chomsky 1970), il existe tout un programme intégrant la syntaxe et la sémantique lexicale, fondé sur l'hypothèse que maints aspects des structures syntaxiques peuvent être expliqués comme le reflet de la structure argumentale des unités lexicales qui en sont les termes recteurs.

La morphologie constructionnelle revendique de participer à ce programme, en tant qu'elle constitue un mode d'accès, précis parce que grammaticalisé, à l'identité des propriétés sémantiques des lexèmes simples et construits. Autrement dit, c'est la relation morphologique elle-même qui sert à identifier les propriétés (en l'occurrence sémantiques) des deux termes L1 et L2 morphologiquement reliés, et des deux classes auxquelles ils appartiennent.

**2) L'héritage de la structure argumentale.** La morphologie constructionnelle a à charge de calculer si le lexème construit hérite intégralement de la structure argumentale du lexème qui lui sert de base, ou si la régularité morphologique en cause (procédé de dérivation affixale, conversion ou composition, et diversité des affixes) est caractérisée par une altération programmée de cette structure argumentale. Un objectif principal est d'étudier la transmission ou la non-transmission des structures argumentales entre les verbes et les noms déverbaux, que ces N soient des noms d'action et/ou de résultat (*construction, marchandage*), d'agent et/ou d'instrument (*aspirateur, réparateur*), de résultat seul (*blessure, lacis*), d'instrument et/ou de localisation (*sautoir, hachoir*).

**3) Les manifestations morphologiques de la polysémie des lexèmes.** La polysémie d'un lexème se manifeste dans le registre des constructions morphologiques de la façon la plus visible, puisque pour une forme phonologique donnée, associée à une flexion, on va constater le maximum de contrastes en terme de présence/absence de tel type de construction : ainsi le verbe *passer* est la base d'un nom d'agent *passeur*, quand il est caractérisé par une structure argumentale comportant un argument de type Patient (*passer des réfugiés de France en Belgique*) ; et quand il est caractérisable comme V inaccusatif à un seul argument (*passer souvent*), il ne peut fournir la base à un tel N d'agent.

La constructibilité de tels ou tels dérivés constitue donc un test crucial de l'appartenance des lexèmes à tel ou tel type sémantique, complémentaire des tests de distribution syntaxique (Godard & Jayez 1996).

**4) Les théories en sémantique lexicale.** Il s'agit d'élaborer une typologie des différentes théories en sémantique lexicale, et de développer une approche formelle, non référentielle, de la sémantique lexicale des noms, construite sur la notion de type de mots. Elle s'inscrit dans la continuité des propositions de Marandin (1984a et b, 1992) et de Jayez & Godard (1995), Godard & Jayez (1996). Sur le plan méthodologique, il s'agit d'identifier des propriétés strictement combinatoires, c'est-à-dire mettant en jeu des tests et des contrastes, à partir d'un ensemble de termes, et non pas terme par terme (Marandin 1992). On distingue deux types de tests : des tests fondés sur les restrictions sélectionnelles, des tests fondés sur l'interprétation des compléments.

**5) Nom propre et mots construits.** Ce thème constitue l'un des aspects de la recherche de S. Leroy. Il prend comme objet les dérivés de noms propres, dont l'existence va à l'encontre de l'idée selon laquelle le nom propre échapperait au système morphologique du lexique (alors qu'on observe des constructions par affixation et conversion). Il apparaît également que le nom propre a un fonctionnement particulier, notamment lorsqu'il sert de base à la construction de gentilés, bien qu'il s'aligne, la plupart du temps, sur les règles de construction concernant le nom commun. On s'aperçoit en outre que le sens des mots construits sur une base nom propre peut exploiter divers niveaux sémantiques du nom propre (sens dénominatif, sens descriptif, etc.). On éclaire ainsi certains des aspects sémantiques et discursifs du nom propre.

## ***Bilan***

### **Étapes passées**

Plusieurs descriptions empiriques et éléments de théorisation :

- 1) Analyse des mots composés de structure [VN] en français (thèse de F. Villoing, soutenance oct. 2002).
- 2) Analyse des mots suffixés en *-is* (DEA de C. Roger, 2001).
- 3) La question de l'éventuelle homonymie de plusieurs suffixes distincts (nombreux travaux de S. Aliquot sur *-ade*, cf. *balustrade*, *orangeade*, *promenade*).
- 4) Les noms composés de structure [NN] (thèse en cours de C. Lesselingue).
- 5) La construction de verbes de changement d'état à base nominale et adjectivale suffixés en *-iser* et en *-ifier* (thèse en cours de C. Roger).

- 6) Les formes morphologiques complexes à base de numéraux (thèse en cours de S. Saulnier).
- 7) La discrimination des suffixes dits nominalisateurs de verbe, l'héritage de la structure argumentale et les classes de verbe (F. Kerleroux 2002).
- 8) Application de la théorie des types sémantiques au domaine olfactif (S. David 2002).
- 9) Descriptions du nom propre et le procédé de l'antonomase (cf. notamment S. Leroy en préparation).

### **Résultats obtenus**

- 1) Les lexèmes entrant en composition ont le statut d'unités complètement spécifiées sémantiquement. La thèse de F. Villoing (2002) prouve les N de structure [VN], sont sémantiquement contraints (les V apparaissant dans cette structure doivent a) être des V dénotant un procès de type dynamique, à l'exclusion des V dénotant un état, b) comporter deux arguments dans leur structure argumentale, un de type Agent, et un de type Patient, et le N apparaissant en position 2 dans le composé correspond à l'argument de type Patient).
- 2) Les N composés de structure [N1 N2]<sub>N3</sub> permettent d'établir une ligne de démarcation morphologie/syntaxe. C. Lesselingue (thèse en cours, communication *Silexicales* 3) montre que l'interprétation canonique de ces N est celle d'une unité lexicale N3 hyponyme de N1, que le composant syntaxique ne peut engendrer à l'identique (*requin-marteau* par opposition à *Alsace-Lorraine*, *physique-chimie*, etc. où le N3 résultant n'est pas l'hyponyme du N1).
- 3) Les règles dérivationnelles ne sont pas regroupables en paradigme. Les travaux de C. Roger montrent, à propos qu'il faut remettre en cause la notion de Corbin (1987) « paradigme de procédés morphologiques » (à propos des unités présentant un rapport catégoriel A>V, glossées par « rendre (plus) Adj »). Réaffirmer le principe « un affixe = une règle de construction », constitue un point important dans le cadre de la distinction entre Flexion et Dérivation.
- 4) On a examiné les conséquences du fonctionnement sémantique des unités associées à l'olfaction dans le cadre d'expérimentations en sciences cognitives (David 2002). Sur le plan linguistique, l'approche en termes de types sémantiques appliquée à ce domaine laisse ouverte la question des objets versus propriétés. Les entités non matérielles, non informationnelles qui sont visées sont-elles à considérer comme des objets d'un autre type, comme des propriétés, comme relevant d'un autre type ?
- 5) Sur le nom propre, un important travail sur le nom propre et l'antonomase a été réalisé, qui va faire l'objet d'un ouvrage de synthèse (S. Leroy).

### **Actions réalisées**

- 1) Séminaire de doctorat, Université Paris X, années 2001 (10 séances), 2002 (12 séances), 2003 (une dizaine de séances prévues)
- 2) Journée d'étude de morphologie et syntaxe : 14 juin 2002, Nanterre, « Unités lexicales : propriétés morphologiques et propriétés sémantiques », sous la direction de F. Kerleroux et D. Leeman ; communications de F. Villoing, C. Roger et F. Kerleroux.
- 3) Participation au 3<sup>ème</sup> Forum de morphologie, Villeneuve d'Ascq, 19-21 sept. 2002, « Les unités morphologiques » : communications de S. Aliquot, C. Roger, C. Lesselingue,

F. Villoing, F. Kerleroux, publiées dans *Sillexicales*, 3, 2003, Publication de l'UMR 8528 du CNRS & Université Lille 3.

4) Participation aux Journées d'études de morphologie, 5-6 déc. 2002, Université Toulouse 2, ERSS (Communications de F. Villoing et de F. Kerleroux).

## ***Perspectives***

Sur le plan morphologique, la nouveauté consiste à pouvoir poser des questions qui traversent les deux domaines analysés comme séparés (cf. « La morphologie scindée (*split morphology*) », Anderson 1992) de la morphologie constructionnelle et de la morphologie flexionnelle : puisque le lexème constitue l'unité morphologique de base sont dans l'un et l'autre cas, on cherchera à mieux caractériser les propriétés formelles et sémantiques de l'individu lexémique, du point de vue formel et du point de vue de la représentation sémantique.

### **1) Flexion et dérivation**

1a) Des zones de chevauchement ? Comment rendre compte de l'existence de formes appartenant au système flexionnel et qui occupent une place spécifique dans le système dérivationnel :

1a1) Les N déverbaux à forme de participe passé féminin (*la traversée, la fuite, la prise*), cf. les travaux de S. Aliquot et la communication de F. Kerleroux mars 2002.

1a2) Les Adjectifs (et leurs antonymes préfixés en *-in*, bases d'adverbe en *-ment*) qui sont homonymes de participes passés (cf. *résolu/irrésolu, adapté/inadapté, attendu/inattendu, réfléchi, irréflechi*, etc.) et dont les N de propriété correspondants sont homonymes de N de procès déverbaux (*résolution, adaptation*), sauf à la forme négative (*irrésolution* vs *non-résolution, inadaptation* vs *non-adaptation*) (communication et publication de Kerleroux 2002 et 2003).

1a3) Les participes présents et les adjectifs en *-ant/ante*, fournisseurs de base à des adverbes en *-ment* (*plaisant, plaisante, plaisamment*). On se demandera en particulier si les phénomènes peuvent être saisis au moyen de l'hypothèse de « indirect mapping » (Aronoff 1999) et on évaluera les propositions de M. Haspelmath (1994, 1996).

1b) Examen du modèle « paradigmatique » (Stump 2001) qui permet de rendre compte des formes fléchies en postulant que les radicaux d'un lexème verbe sont mis en correspondance avec les « cases » du paradigme morphologique (Bonami & Boyé 2003, *Langages*, 152) et de caractériser la dérivation du français, comme mettant en œuvre un radical spécifique pour certaines opérations dérivationnelles (*récepteur, réceptif, réception* vs *receveur, accompagnateur* vs *accompagnement*). On se posera plus largement la question de la validité de ce genre de modèle pour le système nominal en français.

**2) Démarcation morphologie constructionnelle/morphologie flexionnelle/syntaxe.** Avec les thèses en cours de C. Lesseling (NN), S. Saulnier (mots de nombre), travaux de S. Sakhno.

**3) Sémantique lexicale.** On continuera d'examiner les différentes propositions théoriques en vue d'en élaborer une typologie et de développer une approche formelle, non référentielle, de la sémantique lexicale des noms, construite sur la notion de type de mots. On examinera en particulier les types « objet » et « propriété » (problème des objets non matériels et non informationnels ; proposition de tests identificatoires pour le second).



**4) Sur le nom propre.** La perspective s'élargit et prend en compte les contextes d'apposition et de modification.

### Actions prévues

- 1) Séminaire de morphologie, Université Paris 10, 2003-2004
- 2) 2<sup>èmes</sup> Journées d'études en morphologie « Morphologies non conventionnelles », Toulouse 2 & ERSS : 3/4 décembre 2004 - Communication F. Kerleroux.
- 3) Journée Conscila, 26 mars 2004, sur la compositionnalité organisée par D. Apotheloz et G. Boyé (Nancy 2)- Communication de Françoise Kerleroux « Objection(s) à un recours au principe de compositionnalité dans le domaine de la morphologie constructionnelle ».
- 4) Colloque international sur les Noms déverbaux : CNRS UMR 8528 - Silex - Université Lille 3 (Ph. Miller), 23-25 sept. 2004. Conférence invitée de Françoise Kerleroux.
- 5) Projet d'ouvrage collectif sur la morphologie. « Aperçus sur la morphologie lexicale du français » ou « Aspects de la théorie morphologique ». Y seront rassemblées des propositions de quelques démonstrations concrètes du renouveau théorique des travaux contemporains. Participent à ce projet C. Lesselingue, C. Roger, S. Saulnier ; F. Villoing ; F. Kerleroux.

### Production scientifique

- David S. 2002. « Linguistic expressions of odors in French ». In Rouby C., Schaal B., Dubois D., Gervais R., Holley A. (éds), *Olfaction, Taste and Cognition*. Cambridge : Cambridge University Press.
- Fradin B., Kerleroux F. (éds). A paraître. « Quoi de neuf en morphologie ? ». *Langages*, 152.
- Kerleroux. A paraître. « Sur quels objets portent les opérations morphologiques de construction ? ». *Lexique*, 16.
- Kerleroux F., Fradin B. 2003. « Quelle base pour les procédés constructionnels de la Morphologie ? ». In Fradin B., Dal G., Kerleroux F., Hathout N., Plénat M., Roché M. (éds), *Silexicales* « Les unités morphologiques / Morphological Units », 3, 76-84.
- Kerleroux F., Fradin B. A paraître. « Troubles with lexemes ». In Booij G., De Cesaris J., Scalise S., Ralli A. (éds), *Proceedings of the Third Mediterranean Meeting of Morphology*. Barcelona : IULA-Universitat Pompeu Fabra.
- Leroy S. En préparation. *Le Nom propre en français*. Gap - Paris : Ophrys.
- Leroy S. A paraître. « Les noms propres et la dérivation suffixale ». In Grass T., Humbley J., Vaxelaire J.-L. (éds), *Meta*, 49 (1).
- Lesselingue C. 2003. « Les noms composés [NN]N 'holonymiques' : illustration de la spécificité sémantique des unités construites morphologiquement ». In Fradin B., Dal G., Hathout N., Kerleroux F., Plénat M., Roché M. (éds), *Silexicales* « Les unités de la morphologie », 3.
- Roger C. 2003. « Derived change-of-state verbs in French : a case of semantic equivalence between prefixes and suffixes ? ». *Acta Linguistica Hungarica*, 50 (1-2), 187-199.
- Sakhno S. 2001. Semantičeskoe opisanie russkoj leksiki i problema sootnošenija sinxronii i diaxronii (Description sémantique du lexique russe : Problème du rapport entre synchronie et diachronie). In Guiraud-Weber M. (éd.). *Russkij jazyk : peresečaja granicy*. Dubna : Meždunarodnyj universitet, 189-200.

Sakhno S. 2003. « Les prépositions russes O(B) et PRO : objet et propos ». *Revue des études slaves*, 74 (2-3), 453-468.

### **Thèses soutenues**

Jai-Ho Leem. 5 mai 2001. *Les petites propositions en français contemporain. Syntaxe et Interprétation*. Directeur F. Kerleroux.

Fabio Montermini. 4 janvier 2002. *Le système préfixal en italien contemporain*. Directeur F. Kerleroux.

Florence Villoing. 11 octobre 2002. *Les mots composés [VN]<sub>N/A</sub> du français, réflexions épistémologiques et propositions d'analyse*. Directeur F. Kerleroux.

### **Thèses inscrites**

Coralie Roger, *Sens et structures des verbes suffixés en -iser et en -ifier*. Directeur F. Kerleroux.

Chrystèle Lesselingue, *Propositions pour un traitement différentiel des unités nominales polylexématiques en français et en anglais*. Directeur F. Kerleroux.

Sophie Saulnier, *Les nombres : grammaire et lexique*. Directeur F. Kerleroux.

Aboubakry Sy, *Morphologie nominale du peulh*. Directeur F. Kerleroux.

### **Références générales sur le projet**

Amiot D. 1997. *L'antériorité temporelle dans la préfixation en français*. Villeneuve d'Ascq : Presses universitaires de France.

Anderson S. R. 1992. *A-Morphous Morphology*. Cambridge : Cambridge University Press.

Aronoff M. 1976. *Word Formation in Generative Grammar*. Cambridge : The MIT Press.

Aronoff M. 1994. *Morphology by itself*. Cambridge : The MIT Press.

Aronoff M. 1998. « Isomorphism & Monotonicity : or the Disease Model of Morphology ». In Lapointe, Brentari, Farrell (éds), *Morphology and its Relation to Phonology and Syntax*. Stanford : CSLI, 411-418.

Aronoff M. 1999. « Le système malgré lui ». In Beninca P., Mioni A., Vanelli (éds), *Fonologia e Morfologia dell'italiano e dei dialetti d'Italia*. Roma : Bulzoni, 321-332.

Aurnague M., Plénat M. 1997. « Manifestations morphologiques de la relation d'attachement habituel ». In Corbin D. et al. (éds), *Sillexicales « Mots possibles et mots existants »*, 2, 15-24.

Baker M. C. 1988. *Incorporation. A Theory of Grammatical Function Changing*. Chicago : The University of Chicago Press.

Beard R. 1995. *Lexeme-Morpheme Base Morphology*. Albany : State University of New York Press.

Bonami O., Boyé G. A paraître. « Supplétion et classes flexionnelles dans la conjugaison du français ». *Langages « Quoi de neuf en morphologie »*, 152.

Cadiot P., Nemo F. 1997. « Propriétés extrinsèques en sémantique lexicale ». *Journal of French Language Studies*, 7, 127-146.

Carstairs McCarthy A. 1987. *Allomorphy in inflexion*. Londres : Croom Helm.

Chomsky N. 1970. « Remarks on nominalization ». In Jacobs R. A., Rosenbaum P. S. (éds), *Readings in English Transformational Grammar*. Waltham : Mass, 184-221.

- Corbin D. 1987. *Morphologie dérivationnelle et structuration du lexique*. Tübingen : Niemeyer.
- Dal G. 1997. *Grammaire du suffixe -et(te)*. Paris : Didier Erudition.
- David S. 1993. *Les unités polylexicales. Eléments de description et reconnaissance automatique*. Thèse de doctorat, np.
- Dell F. 1970. *Les règles phonologiques tardives et la morphologie dérivationnelle du français*. Thèse de doctorat, np.
- Dressler W. U., Mayerthaler W., Panagl O., Wurzel W. U. 1987. *Leitmotivs in Natural Morphology*. Amsterdam/Philadelphie : John Benjamins Publishing Company.
- Fradin B. 2003. *Nouvelles approches en Morphologie*. Paris : Presses universitaires de France.
- Fradin B., Hathout N., Meunier F. A paraître. « La suffixation en -ET et la question de la productivité ». *Langue française*, 141.
- Godard D., Jayez J. 1996. « Types Nominaux et Anaphores : le cas des objets et des événements ». *Chronos*, 1, 41-58.
- Gross G. 1994. « Classes d'objets et description des verbes ». *Langages*, 115, 15-30.
- Haspelmath M. 1994. « Passive Participles accross Languages ». In Fox B., Hopper P. J. (éds), *Voice, Form & Function*. Amsterdam : John Benjamins, 151-177.
- Haspelmath M. 1996. « Word-class-changing inflection and Morphological Theory ». *YOMO*, 43-66.
- Jayez J., Godard D. 1995. « Principles as Lexical Methods ». *Proceeding of AAAI Spring Symposium on the Representation and Acquisition of Lexical Knowledge : Polysemy, Ambiguity and Generativity*. Stanford University.
- Katz J. J. 1972. *Semantic Theory*. New-York : Harper & Row.
- Kerleroux F. 1996. *La coupure invisible*. Villeneuve d'Ascq : Presses Universitaires du Septentrion.
- Kleiber G. 1999. *Problèmes de sémantique*. Villeneuve d'Ascq : Presses universitaires du Septentrion.
- Lakoff 1987. *Women, Fire and Dangerous Things*. Chicago : University of Chicago Press.
- Marandin J.-M. 1984a. « Miniatures sentimentales. Syntaxe et discours dans une description lexicale ». *LINX*, 10, 75-95.
- Marandin J.-M. 1984b. « Distribution et contexte dans une description lexicale ». *Cahiers de Lexicologie*, 44 (1), 137-149.
- Marandin J.-M. 1992. « Il y a de la synonymie ». *Cahiers de lexicologie*, 61 (2), 39-57.
- Lieber R. 1992. *Deconstructing Morphology*. Chicago : Press of the University of Chicago.
- Plénat M. 2000. « Quelques thèmes de recherche actuels en morphophonologie française ». *Cahiers de Lexicologie*, 77 (2), 27-62.
- Plénat M., Roché M. A paraître. « La suffixation décalée ». *Lexique*, 16.
- Pustejovsky J. 1995. *The Generative Lexicon*. Cambridge : The MIT Press.
- Roché M. 1997. « Briard, bougeoir et camionneur : dérivés aberrants, dérivés possibles ». In Corbin D. et al. (éds), *Sillexicales « Mots possibles et mots existants »*, 1.
- Roché M. 2003. « La variation en morphologie dérivationnelle ». In Van Deyck R. (éd.), *Les quatre variations*. Gand : Communication et cognition.
- Scalise S. 1984. *Generative Morphology*. Dordrecht : Foris Publications.
- Scalise S. 1994. *Morfologia*. Bologna : Il Mulino.
- Selkirk E. 1982. *The Syntax of Words*. Cambridge : The MIT Press.
- Stump G. T. 2001. *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge : Cambridge University Press.

Temple M. 1996. *Pour une sémantique des mots construits*. Villeneuve d'Ascq : Presses universitaires du Septentrion.

**Contrat / financement**

Le projet d'ouvrage collectif sur la morphologie a reçu le soutien scientifique et financier de l'ILF.

### 1.3. LES CONNECTEURS

#### **Participants**

*Membres rattachés à titre principal* : Danielle Leeman (responsable), Annie Bertin, Danièle Flament-Boistrancourt, Françoise Gadet, Colette Noyau

*Membre associé* :

Jacqueline Bacha (Faculté des lettres de Sousse, Tunisie)

*Doctorants* : Mohamed Abdelkader, Ichraf Khammari, Badreddine Hamma, Sophie Hamon, Belinda Lavieu, Majid Moslemi, Houda Ounis, Céline Vaguer

*Collaborations extérieures* :

Monji Kahloul (Faculté des lettres de Sousse, Tunisie)

Ahmed Brahim (Faculté des lettres de Tunis)

Salah Mejri (Faculté des lettres de Tunis)

#### **Discipline(s) et/ou sous-discipline(s) concernée(s)**

Morphologie, syntaxe, lexique, sémantique

Dans l'intitulé de ce projet, le terme « connecteur » est à prendre en un sens relativement général et indépendamment de l'acception spécifique qu'il peut prendre dans telle ou telle théorie particulière : il a été choisi de manière à rassembler les thématiques traitées par les différents participants dans la diversité de leurs approches (synchronique ou diachronique, grammaticale, lexicale, sémantique ou discursive, etc.) et des cas précis étudiés : groupes prépositionnels, prépositions, conjonctions, adverbes...

#### **Problématique**

Le projet en cours et les actions passées qui y ont conduit sont centrées sur la préposition, d'une part en elle-même, d'autre part en tant qu'elle peut être mise en correspondance avec un préfixe d'un côté, des conjonctions ou des locutions de l'autre (cf. par exemple *en* : *engranger*, *en ce que*, *en effet*).

En effet, lorsque la préposition peut être rapprochée (étymologiquement et sémantiquement) d'un préfixe, on peut se demander s'il n'y a pas à se poser la question de la possibilité d'un morphème unique (en langue) actualisé par deux allomorphes (en discours) selon que la forme est libre (*mettre en cave*, *courir par les rues*) ou liée (*encaver*, *parcourir*). Cette hypothèse, déjà avancée pour *en* dans un travail relativement ancien de Franckel et Lebaud (1991) s'inscrivant dans le cadre de la linguistique culiolienne, est naturelle si l'on admet le postulat saussurien, relayé par Guillaume (1919), d'une indissociabilité du signifiant et du signifié (dit « principe de naturalité »), en y ajoutant celui d'une pertinence sémiologique des formes morphologiques telles qu'observables en discours (postulat propre à Guillaume) - contre celui du caractère relativement aléatoire, capricieux et non significatif des actualisations en parole.

Lorsque du moins on ne tient pas les prépositions pour des mots vides de sens qui n'auraient de rôle que purement syntaxique (« les mots outils »), se pose la question de la compositionnalité des expressions dites habituellement « figées » : dans cette dernière perspective, *par exemple* formerait un bloc inanalysable, globalement équivalent à *ainsi*. Or, lorsque l'on dispose d'une hypothèse sur l'identité de la préposition, on peut au contraire montrer que les combinaisons dans lesquelles elle entre (ou n'entre pas) obéissent à une logique linguistique reposant sur sa définition dans le système : *en exemple* n'a pas la même valeur ni le même emploi que *par exemple* (*X, par exemple Y* vs \* *X, en exemple Y* et *donner Y en exemple* vs \* *donner Y par exemple*). Et ce n'est pas non plus un hasard si *exemple* - mais non *preuve*, qu'on a dans des emplois similaires à *par exemple* mais sous la forme *X, à preuve Y* ou *X, la preuve Y* - se combine avec *par*, étant donné son identité lexicale propre.

Pour résumer l'axe directeur du projet, il s'agit donc :

- 1) de procéder à l'analyse de différentes prépositions d'une part en tant que telles (*à, en, dans, par, jusque...*) ;
- 2) d'autre part dans leur rapport avec d'autres unités grammaticales (comme le préfixe) ;
- 3) enfin, dans les combinaisons phraséologiques (locutions, mots composés) qu'elles contribuent à former.

Dans cette entreprise, outre le travail de description synchronique mené par Danielle Leeman et des doctorants qu'elle dirige, l'investigation diachronique a sa place, permettant d'établir l'évolution syntaxique et sémantique des termes retenus, et donc d'une certaine manière d'évaluer les hypothèses avancées pour le français contemporain (Bertin 2002). Il en va de même de l'observation de la langue orale dans diverses situations (Françoise Gadet), ou de celle des processus d'acquisition (Noyau *et al.* 2002). Danièle Flament étudie la pertinence des résultats dans leur transposition didactique dans l'enseignement du français comme langue étrangère.

## **Contexte**

Sur le plan théorique, le présent projet s'inscrit en amont des représentations ordinairement construites, qu'il s'agisse de formules à base logique du type de celles de Jackendoff (1983), de répartitions entre emploi central et emplois périphériques dans la suite de la théorie du prototype comme le fait Lakoff (1987) ou de définitions intégrant la notion de « ressemblance de famille » (Vandeloise 1986). Il s'agit plutôt de construire l'identité lexicale de la préposition de sorte à en permettre ensuite la description syntaxico-sémantique susceptible de justifier en retour les emplois observables et leur représentation dans le cadre de formalismes contemporains associant forme et sens, comme HPSG (Pollard & Sag 1987, 1994).

L'hypothèse théorique générale que nous retenons est que la langue constitue un système autonome sécrétant ses propres valeurs (Saussure 1916) et que les formes, seules manifestations observables de l'identité du signe, constituent le point de départ obligé pour tenter de saisir le sens tel qu'institué par la langue (Leeman 1996) ; la méthodologie adoptée ne s'appuie donc pas sur une analyse des référents (Kleiber 1999) ou des conceptualisations (Jackendoff 1983 ; Lakoff 1987) posés a priori comme la source des signifiés (voire les signifiés eux-mêmes) : ces derniers sont à l'inverse construits par un travail de raisonnement sur les possibilités et impossibilités formelles (syntaxiques,

morphologiques, distributionnelles) à partir du principe (heuristique) de naturalité (Milner 1989).

Toujours au nom de ce principe, le résultat de l'analyse doit procurer non seulement une identité sémantique du morphème ou du constituant étudié, mais aussi prédire ses compatibilités formelles ; or cette dernière exigence n'est aucunement satisfaite dans les tentatives contemporaines, qu'il s'agisse de formuler la définition sous la forme d'un signifié de puissance (Picoche 1986), d'une famille dont les membres sont reliés plus ou moins lâchement selon leur degré de ressemblance avec un prototype (Vandeloise 1986), d'une forme schématique (Franckel & Paillard 1997). Il faut en fait trouver le moyen d'associer à une définition intensionnelle de l'unité (Cadiot 1997) l'instruction ou l'ensemble d'instructions spécifiant ses rapports avec le cotexte ; de ce dernier point de vue, les tentatives existantes pèchent par leur trop grande abstraction ou généralité : l'hypothèse de définition, trop puissante, prédit des combinaisons inacceptables aussi bien qu'acceptables et à l'inverse - faute d'une investigation empirique préalable suffisante - ne rend pas compte de toutes les possibilités.

Le projet en cours a permis la mise en place d'une action CMCU (Comité Mixte de Coopération Universitaire) associant l'UMR 7114 et le Département de français de la Faculté des Lettres et Sciences humaines de Sousse (Tunisie).

### **Hypothèses**

Le thème général est celui de l'identité formelle et sémantique des mots dits « grammaticaux » (incluant éventuellement des morphèmes liés comme le préfixe). L'objectif est d'en fournir une description empiriquement satisfaisante (telle qu'elle s'applique au maximum d'emplois) et formalisable (par exemple dans les grammaires HPSG).

Les premières questions qui se posent concernent le cadre théorique et méthodologique à adopter dans la mesure où la linguistique générativiste des années soixante-dix ne se prêtait pas à ce type d'investigation : la résolution passait en effet davantage par une analyse distributionnelle précise que par un raisonnement sur la place structurale des constituants et leurs propriétés syntaxiques. D'un autre côté, l'option localiste des premiers travaux cognitivistes ne permettait pas non plus de voir le moyen de résoudre ce type de problème (Vandeloise 1986) ; les représentations à base logique de Jackendoff (1990) supposent le problème résolu, comme les schémas de Langacker (2000) ou le recours au prototype par Lakoff (1987).

L'hypothèse générale est que l'on peut remplir le programme de description et de formalisation si l'on adopte un cadre autonomiste. « Autonomiste » au sens saussurien et non pas chomskyen, en ceci que c'est la langue (et non seulement la syntaxe) qui est conçue comme un système autonome : si la forme comme le sens sont arbitraires (la langue étant globalement définissable comme de la forme indissolublement liée à du sens), ce dernier ne peut être assimilé à notre perception ou à notre conceptualisation du monde (perspectives référentialiste ou cognitiviste) et n'est saisissable qu'à travers l'observation des formes (Leeman 1997) ; certes ce que celles-ci permettent d'appréhender n'est pas le tout du sens (qui inclut du référentiel, du pragmatique, du culturel, du cognitif, etc.) mais la démarche grammaticale ne peut saisir du sens que ce qui en est véhiculé par les formes (et leurs propriétés) – nous suivons ici toujours la ligne harrissienne : la grammaire n'épuise

pas le « meaning » (qui n'entre pas dans son domaine ou ses compétences mais relève des savoirs encyclopédiques), elle ne peut capter que l'« information », c'est-à-dire précisément ce qui, dans le sens, est institué par le système linguistique et dont témoignent les propriétés formelles (Leeman 1996, 1999).

Fondamentalement, on pose que la langue forme un système tel que les interrelations entre les formes instituent une certaine structure sémantique qui lui est propre, et qui donc ne peut être réduite à un calque de l'organisation du monde (paradigme objectiviste ou référentiel), ni à un reflet de la structure conceptuelle responsable de nos représentations mentales (paradigme cognitiviste).

Ces hypothèses rejoignent une certaine tradition européenne (Saussure 1916 ; Guillaume 1919 ; Benveniste 1966), semblable dans ses principes à ceux du distributionnalisme américain : il s'agit d'étudier minutieusement les emplois des prépositions (d'observer les formes possibles, qu'il s'agisse de compatibilités distributionnelles ou de propriétés syntaxiques), de manière à tâcher d'en tirer une hypothèse lexicale sur la préposition (associant forme et sens) qui, en retour, permettra d'expliquer les emplois possibles ou impossibles et, éventuellement, de justifier les jugements portés sur les énoncés.

## ***Bilan***

### **Étapes passées**

Différents types de travaux ont été menés :

- 1) Travail de description empirique.
- 2) Cadrage méthodologique.
- 3) Confrontations des cadres et des hypothèses théoriques.
- 4) Première diffusion des résultats.

### **Résultats obtenus**

La démarche mise en place paraît d'autant plus féconde que l'on s'intéresse à des unités grammaticales comme la préposition, pour laquelle il n'y a pas nécessairement de correspondant référentiel ou conceptuel accessible. Le signifié ici ne peut être que construit (de manière hypothétique) à partir d'un raisonnement sur les formes. Or rendre compte de l'identité d'un complément prépositionnel (« circonstanciel » ou autre) suppose évidemment que l'on dispose d'une hypothèse sur celle de la préposition qu'il comporte.

### **Actions réalisées**

- 1) Journée d'études *Les mots lexicaux, morphologie et syntaxe*, 14 juin 2002, Université Paris X (co-organisée par F. Kerleroux et D. Leeman).
- 2) Colloque international *Des mots du discours aux mots grammaticaux : une continuité ?*, Sousse, Tunisie (co-organisée par J. Bacha et D. Leeman) ; actes à paraître dans la *Revue de Sémantique et Pragmatique*.
- 3) Direction par D. Leeman du numéro 46 de *LINX* (2002).



Les actions 2 et 3 notamment ont permis de confronter les résultats obtenus dans le cadre de notre projet à ceux d'autres chercheurs.

4) Journée Conscila, 31 mai 2002. *Modèles syntaxiques* : comparaison des différentes théories contemporaines en matière de syntaxe (Leeman D. (éd.) 2003).

## ***Perspectives***

### **Étapes à venir**

L'ambition de notre projet est de procurer des prépositions du français une description autant que possible complète (c'est-à-dire qui ne se borne pas à leur emploi spatial, comme cela a le plus souvent été le cas dans les années 80-90). Le travail de description empirique doit être poursuivi sur *en* (Ichraf Khammari), *dans* (Céline Vaguer), *à* - expression du moyen (Belinda Lavieu), *par* (Badreddine Hamma), *jusque* (Mohamed Anis Abdelkader), *à* et *de* introduisant un complément de verbe (Majid Moslemi).

### **Résultats attendus**

Les thèses engagées depuis 1999 devraient venir à soutenance entre 2004 et 2006. Elles fourniront des résultats sur les prépositions étudiées et permettront de préciser le cadre et les concepts théoriques retenus.

### **Actions prévues**

1) Bacha, J. Bertin, A. et Leeman, D. 2003. *Les mots grammaticaux, histoire et grammaticalisation*, Journée d'étude, Sousse, 27 novembre. Intervenants : J. Bacha, A. Bertin, M. Mokni, M. A. Zrida, D. Leeman, L. Ben Hamad.

2) Bacha, J., Bertin, A., Flaux, N. 2004. Colloque *Connecteurs*, Sousse, actes édités aux Presses de l'Artois en 2005.

### **Production scientifique**

Bertin A. 2002. « L'émergence du connecteur *en effet* en moyen français ». *LINX*, 46.

Noyau *et al.* 2002. « Construction des énoncés et connecteurs dans la structuration des récits enfantins en arabe tunisien et en français ». *LINX*, 46.

Hamma B., Haillet, P. P. 2002. « *Par contre* : un type particulier de dynamique discursive ». *LINX*, 46.

Hamon S. 2002. « Les conjonctions causales et la propriété d'enchâssement ». *LINX*, 46, 25-36.

Khammari I. A paraître. « *En* préposition et *en-* préfixe : une même unité linguistique ? ». *Revue de Sémantique et de pragmatique*, Actes du colloque « Des mots du discours aux mots grammaticaux : une continuité ? », Sousse, Tunisie, 4-5 octobre 2002.

Lavieu B., Desmets M., Hamon S. 2003. « Les grammaires HPSG ». *LINX*, 48, 57-77.

Leeman D. (éds). 2002. *LINX* « Les connecteurs », 46.

Leeman D. 2003a. « *Contre* introduisant un complément de verbe ». *Actes du colloque « La médiation »*, 7-9 décembre 2002. Rouen : Presses universitaires de Rouen.

- Leeman D. 2003b. « Nom et verbe au regard de la préposition *contre* ». *Actes du colloque « Nom et verbe : catégorisation et référence »*, septembre 2001. Reims : Presses universitaires de Reims.
- Leeman D. A paraître. « Hypothèse d'une autonomie du nom et du verbe morphologiquement apparentés ». *Modèles linguistiques*, Actes du 10<sup>ème</sup> colloque international de psychomécanique du langage, 3-5 juin 2003.
- Leeman D. (éd.). 2003. *LINX* « Approches syntaxiques contemporaines », 48.
- Vaguer C. A paraître. « La préposition dans : un vecteur d'approximation ? ». *Revue de Sémantique et de pragmatique*, Actes du colloque « Des mots du discours aux mots grammaticaux : une continuité ? », Sousse, Tunisie, 4-5 octobre 2002.

### **Thèses inscrites**

- Mohamed Abdelkader, *La préposition jusque*. Directeur D. Leeman.
- Badreddine Hamma, *La préposition par*. Directeur D. Leeman.
- Sophie Hamon, *L'expression de la cause*. Directeur D. Leeman.
- Ichraf Khammari, *La préposition en*. Directeur D. Leeman.
- Belinda Lavieu, *Le complément de moyen en à*. Directeur D. Leeman.
- Majid Moslemi, *Les prépositions à et de*. Directeur D. Leeman.
- Houda Ounis, *La métaphore dans le champ de l'amour*. Directeur D. Leeman.
- Céline Vaguer, *La préposition dans*. Directeur D. Leeman.

### **Références générales sur le projet**

- Benveniste E. 1966. *Problèmes de linguistiques générale*. Paris : Gallimard.
- Brøndal V. 1950. *Théorie des prépositions : introduction à une sémantique rationnelle*. Copenhague : Munksgaard.
- Brugman C. M. 1988. *The Story of over. Polysemy, Semantics and the Structure of the Lexicon*. New York - Londres : Garland Publishing.
- Cadiot P. 1997. *Les prépositions abstraites*. Paris : Armand Colin.
- Deane P. D. 1992. *Grammar in Mind and Brain*. Berlin - New York : Mouton de Gruyter.
- Franckel J.-J., Paillard D. 1997. « Prépositions et travail notionnel sur les termes mis en relation. Le cas de sous en français ». *La Notion*. Paris : Ophrys.
- Gougenheim G. 1939. *Système grammatical de la langue française*. Paris : D'Artrey.
- Guillaume G. 1919 (rééd. 1975). *Le problème de l'article et sa solution dans la langue française*. Paris - Québec : Nizet et Université de Laval.
- Jackendoff R. 1983. *Semantics and Cognition*. Cambridge (Mass.) : The MIT Press.
- Jackendoff R. 1990. *Semantic Structure*. Cambridge - Londres : The MIT Press.
- Jaeggi A. 1956. *Le rôle de la préposition et de la locution prépositive dans les rapports abstraits en français moderne*. Berne : Francke.
- Kleiber G. 1999. *Problèmes de sémantique*. Villeneuve d'Ascq : Presses universitaires du Septentrion.
- Lakoff 1987. *Women, Fire and Dangerous Things*. Chicago : University of Chicago Press.
- Langacker R. W. 2000. *Grammar and Conceptualization*. Berlin - New York : Mouton de Gruyter.
- Leeman D. 1996. « « Le « sens » et l' « information » chez Harris » ». *LINX*, numéro spécial « Hommages à Denise Maldidier » ».

- Leeman D. 1997. « Définir une préposition : hypothèses et perplexités ». *Revue de Sémantique et Pragmatique*, 2.
- Leeman D. 1999. « L'unité lexicale dans la perspective harrissienne ». *LINX*, 40.
- Milner J.-Cl. 1989. *Introduction à une science du langage*. Paris : Le Seuil. Collection « Des Travaux ».
- Picoche J. 1986. *Structures sémantiques du lexique français*. Paris : Nathan.
- Pollard C., Sag I. 1987. *Information Based Syntax and Semantics*. Chicago : University of Chicago Press.
- Pollard C., Sag I. 1994. *Head-Driven Phrase Structure Grammar*. Chicago : University of Chicago Press.
- Pottier B. 1962. *Systématique des éléments de relation*. Paris : Klincksieck.
- de Saussure F. 1916. *Cours de linguistique générale*. Paris : Payot.
- Spang-Hanssen E. 1963. *Les prépositions incolores du français moderne*. Copenhague : GEC Gad.
- Talmy L. 2000. *Towards a Cognitive Semantics*. Amsterdam : John Benjamins.
- Vandeloise Cl. 1986. *L'espace en français*. Paris : Le Seuil.

## **1.4. MODÉLISATION DE DONNÉES LINGUISTIQUES DANS LA PERSPECTIVE DU TRAITEMENT AUTOMATIQUE DES LANGUES (METAL)**

### **Participants**

*Membres rattachés à titre principal* : Marcel Cori (responsable), Marianne Desmets (responsable), Karine Baschung, Sophie David, Annie Delaveau, Françoise Kerleroux, Danielle Leeman, Anna Sörös

*Jeune chercheur associé* : Antonio Balvet

*Doctorants* : Éliane Cheung, Fabrice Delumeau, Anne Lablanche, Gaëlle Tassin, Céline Vaguer

*Collaborations extérieures* :

Bernard Fradin (Université Paris 7 et LLF, UMR 7110)

Jean-Marie Marandin (Université Paris 7 et LLF, UMR 7110)

Jesse Tseng (Langue et Dialogue, LORIA ; précédemment post-doctorant à MoDyco)

### **Discipline(s) et/ou sous-discipline(s) concernée(s)**

Traitement automatique des langues, syntaxe, morphologie

### **Problématique**

Le programme se décompose en les diverses sous-tâches complémentaires caractéristiques du TAL - à savoir le niveau descriptif de recueil et de systématisation des données, le niveau de la modélisation théorique, et enfin le niveau des spécifications linguistiques formelles en vue d'une implémentation. Il s'agit donc entre autres :

- 1) de produire des descriptions linguistiques qui soient susceptibles d'être représentées dans des modèles formels ;
- 2) d'examiner les limites formelles précises des systèmes morphologiques, une fois que le modèle classique de concaténation segmentale a été récusé ;
- 3) d'étudier des problèmes généraux posés par les représentations syntaxiques (discontinuité, phénomènes à distance, sous-catégorisation, etc.) ainsi que la possibilité d'en rendre compte dans les formalismes existants ;
- 4) de comparer les formalismes selon les possibilités qu'ils offrent et de tenter d'en dégager certaines propriétés ;
- 5) d'envisager les extensions, restrictions ou modifications des formalismes existants afin de les rendre plus adéquats à la représentation des phénomènes linguistiques et au traitement informatique ;
- 6) d'aborder l'écriture des algorithmes et des programmes qui découlent de ces recherches théoriques.

L'hypothèse de travail sous-jacente, contrairement à la plupart des approches dites « robustes », est qu'il y a fortement intérêt à se doter d'un niveau de représentation symbolique et d'opérations sur ces symboles - en d'autres termes, que jusqu'à preuve du

contraire c'est une bonne théorie linguistique qui reste susceptible *in fine* de produire une bonne technologie.

### **Thèmes de recherche**

Les différents travaux se placent aux différents échelons de la chaîne des traitements définie dans la problématique. Plusieurs de ces travaux reprennent différentes sous-tâches.

#### **1) Travaux de description linguistique**

1a) Des recherches, dans un cadre typologique, sur l'ordre des mots dans les langues romanes et en hongrois ont été développées dans deux directions :

1a1) La structure informationnelle : l'apparente liberté de l'ordre des mots en hongrois ne semblant pas pouvoir s'expliquer sans faire appel à cette notion.

1a2) Le hongrois semble avoir subi le même changement de l'ordre des mots que les langues romanes, à savoir SOV>SVO. Mais le déroulement du processus représente deux solutions opposées. Anna Sörös, avec Christiane Marchello-Nizia, ont proposé une hypothèse qui permettrait d'expliquer ces comportements différents. Par ailleurs, une description des constructions passives dans une perspective translinguistique a été ébauchée (Sörös 2003).

1b) Un travail de description du créole guadeloupéen a été commencé, en vue de la génération automatique d'énoncés en créole à partir d'énoncés en français. Ce travail a déjà donné lieu à publication (Delumeau 2002).

1c) Différents travaux de description du français ont été menés. Ainsi, une étude sur les adjectifs relationnels (Valassis 2001), une étude sur les phrases en *comme* (Desmets 2001), un travail sur les propositions infinitives (Lablanche 2003). Le travail descriptif auquel procède Danielle Leeman, ainsi que celui de certains des doctorants qu'elle dirige, constituent une première étape de « défrichage » des données et de mise au jour des propriétés, en vue d'un traitement dans le cadre HPSG sous la direction de Marianne Desmets : ainsi Sophie Hamon étudie l'expression de la cause (juxtaposition, coordination, subordination) et Belinda Lavieu celle du moyen par un groupe prépositionnel en *à* (de type *à l'eau claire* dans *laver son pull à l'eau claire*). Le début de cette collaboration s'est concrétisée par un article (Desmets *et al.* 2003).

#### **2) Sur la formalisation en syntaxe et en morphologie**

2a) Il est apparu nécessaire de voir un peu plus clair dans la diversité des démarches et/ou des modèles en jeu dans les travaux actuels. D'où l'organisation (par Danielle Leeman) d'une journée d'études dans le cadre de l'association Conscila intitulée « Modèles syntaxiques ». Cette journée d'étude s'est prolongée par la publication d'un numéro de *LINX* sur le même thème « Approches syntaxiques contemporaines ». Dans ce numéro, Marcel Cori pose le problème de la mathématisation des formalismes syntaxiques, et s'interroge sur un recul actuel de cette mathématisation.

2b) Les travaux de Marianne Desmets ont porté essentiellement sur la modélisation en HPSG, et notamment la modélisation des relatives en français.

2c) Une modélisation dans le cadre des Grammaires d'arbres polychromes est poursuivie, notamment dans le cadre des travaux d'Anne Lablanche sur les propositions infinitives.

### **3) Problèmes généraux et implémentation**

3a) Une approche générale du problème de la discontinuité, avec un panorama des modélisations proposées et une formalisation de l'approche de Espinal (1991) a été proposée, ainsi qu'un algorithme d'analyse (Cori 2002). Ce travail a été prolongé par une étude sur l'allemand menée par Éliane Cheung.

3b) Une réflexion sur l'implémentation des formalismes, et plus particulièrement sur HPSG, a été menée par Jesse Tseng.

## ***Bilan***

### **Actions réalisées**

1) Une journée d'études, organisée dans le cadre de l'association Conscila, intitulée « Modèles syntaxiques » s'est tenue le 31 mai 2002 à l'Université Paris 7, Institut Charles V. Cette journée a été organisée par Danielle Leeman. Intervenants : Patrick Bellier, Marcel Cori, Jacques François, Béatrice Godart-Wendling, Yannick Mathieu.

2) Un numéro de *LINX* (n° 48, 2003) dirigé par Danielle Leeman sur le thème « Approches syntaxiques contemporaines » a prolongé cette journée. Ont été passées en revue dans leurs dernières versions les formalismes suivants : Grammaires d'arbres polychromes, Grammaire de construction, Grammaires HPSG, *Role and Reference Grammar*, Grammaire générative, Lexique-grammaire, Grammaires catégorielles lambekiennes, Grammaire applicative et cognitive.

## ***Perspectives***

### **1) Travaux de description linguistique**

1a) Une première action consistera à effectuer la description des constructions passives dans une perspective translinguistique. Au-delà des descriptions dans les grammaires des langues qui connaissent des constructions passives, nous disposons actuellement d'une analyse de prototype du passif, proposée par (Shibatani 1985). Selon cette approche, le passif prototypique peut être décrit comme représentant une fonction pragmatique primaire, des propriétés sémantiques, des propriétés syntaxiques, ainsi que des propriétés morphologiques. Par rapport au prototype, il est possible de décrire des constructions qui ne sont pas prototypiques mais qui représentent néanmoins le passif, à travers des manifestations à l'un des niveaux de la description, en particulier pragmatique et sémantique. En ce sens, les constructions passives des langues indo-européennes modernes ne sont que quasi-prototypiques, dans la mesure où elles permettent l'expression de l'agent sous forme oblique. En hongrois, c'est le facteur morphologique qui rend toutes les constructions atypiques, puisque dans la langue actuelle il n'existe pas de forme verbal spécifique pour le passif.

Dans un premier temps sera élaborée une description du cas du hongrois. Ensuite, on tentera d'élargir le nombre des descriptions comparables, afin de continuer par les travaux de modélisation.

1b) Le travail de description du créole guadeloupéen en vue de la génération automatique d'énoncés en créole à partir d'énoncés en français devrait être mené à son terme. Une

formalisation, probablement en termes de grammaire d'arbres polychromes sera élaborée. Un lien sera effectué avec les travaux de description de phénomènes syntaxiques du français, notamment autour de la question de la présence/absence d'infinitif.

1c) Les différents travaux de description du français seront poursuivis et développés. Ainsi, le travail d'Anne Lablanche sur les propositions infinitives sera élargi à une description des adjectifs (à partir des adjectifs qui sous-catégorisent des infinitives). Le travail descriptif auquel procède Danielle Leeman, ainsi que celui de certains des doctorants qu'elle dirige, constituent une première étape de « défrichage » des données et de mise au jour des propriétés, qui donneront lieu à un traitement dans le cadre HPSG ou celui des grammaires de construction sous la direction de Marianne Desmets : ainsi Sophie Hamon étudie l'expression de la cause (juxtaposition, coordination, subordination) et Belinda Lavieu celle du moyen par un groupe prépositionnel en *à* (de type *à l'eau claire* dans *laver son pull à l'eau claire*). Plus globalement, les recherches seront axées sur la description des systèmes corrélatifs (l'expression de la cause, de la conséquence, etc.) et comparatifs en français.

## **2) Sur la formalisation en syntaxe et en morphologie**

2a) Les formalismes seront utilisés et mis à l'épreuve à travers le traitement des différents faits de langue étudiés. Seront ainsi testés le formalisme HPSG, les grammaires de construction, les grammaires d'arbres polychrome. La modélisation des systèmes corrélatifs sera effectuée en HPSG et en grammaire de construction.

2b) On étudiera la possibilité d'une formalisation des approches morphologiques non concaténationnelles.

## **3) Problèmes généraux et implémentation**

3a) Le travail sur la discontinuité sera prolongé par une recherche sur l'ordre des mots, ainsi que la notion de syntagme ou de constituant. Cette recherche s'appuiera sur différentes langues, notamment, en dehors du français et de l'anglais, l'allemand et le finnois.

3b) Différents travaux expérimentaux d'implémentation seront menés à la suite des études théoriques. En particulier sur la génération automatique du créole à partir du français, sur le traitement des infinitives et des adjectifs dans le cadre des grammaires d'arbres polychromes, sur la discontinuité.

## **Résultats attendus**

- 1) Description du passif en hongrois, et dans d'autres langues.
- 2) Eléments pour l'élaboration d'un système de génération du créole guadeloupéen à partir d'énoncés en français.
- 3) Description des propositions infinitives en français, ainsi que des adjectifs et formalisation en grammaires d'arbres polychromes.
- 4) Description des corrélatifs en français et formalisation en HPSG et/ou grammaires de construction.
- 5) Ébauche d'une formalisation de l'approche non concaténationnelle en morphologie.
- 6) Synthèse sur la discontinuité, l'ordre des mots ainsi que les notions de constituant et de syntagme.

## **Production scientifique**

- Cori M. 2002. « Processing Discontinuity, in Advances in Natural Language Processing ». In Ranchod E., Mamede N. J. (éds.), *Lecture Notes in Artificial Intelligence*, 2389. Berlin - Heidelberg : Springer Verlag, 83-86.
- Cori M. 2003. « La mathématisation des formalismes syntaxiques ». *LINX*, 48, 13-28.
- Delumeau F. 2002. « Quelle forme graphique pour *lésan trankil* ». *Actes du XXVI<sup>e</sup> colloque international de linguistique fonctionnelle*, Gosier (Guadeloupe).
- Desmets M. A paraître. « French Free Relative Phrases in HPSG ». *Actes du colloque HPSG' 2002*, Séoul, CSLI on-line Publications : [cslipublications.stanford.edu/HPSG/1/hpsg02-toc.html](http://cslipublications.stanford.edu/HPSG/1/hpsg02-toc.html).
- Desmets M., Hamon S., Lavieu B. 2003. « Les grammaires HPSG ». *LINX*, 48, 57-76.
- Lablanche A. 2003. « Les constructions infinitives dans les grammaires d'arbres polychromes hiérarchisées ». *LINX*, 48, 29-42.
- Leeman D. (éd). 2003. *LINX* « Approches contemporaines en syntaxe », 48.
- Roussarie L., Desmets M. A paraître. « Référence citationnelle dans les reportives en *comme* ». *Actes du colloque CSSP'2001*, Université Paris 7.
- Sóres A. 2003. « A propos du passif en français et en hongrois ». In Szende T., Máthé G. (éds.), *Frontières et passages*, Actes du colloque franco-hongrois sur la traduction. Berne : Peter Lang, 265-274.
- Sóres A. A paraître. « La structure informationnelle du hongrois ». In Fernandez-Vest J. (éd.), *Les langues ouraliennes aujourd'hui : approche linguistique et cognitive*. Paris : H. Champion, Bibliothèque de EPHE
- Tseng J. 2003. « Phrasal affixes and French morphosyntax ». In Jäger G., Monachesi P., Penn G., Wintner S. (éds.), *Proceedings of Formal Grammar 2003*, Vienne : Technische Universitaet. 177-188.

## **Thèses soutenues**

- Marianne Desmets, 2001. *Les typages de la phrase en HPSG : le cas des phrases en comme*. Directeur A. Delaveau.
- Ioannis Valassis, 2001. *Syntaxe positionnelle et épithètes : le cas des adjectifs relationnels*. Directeur A. Delaveau.

## **Thèses inscrites**

- Eliane Cheung. *Formalisation de la discontinuité syntaxique*. Directeur M. Cori.
- Fabrice Delumeau. *Etude des problèmes linguistiques liés à la génération automatique d'énoncés en créole guadeloupéen*. Directeur M. Cori.
- Anne Lablanche. *Développement d'un fragment de grammaire hiérarchisée du français*. Directeur M. Cori.
- Gaëlle Tassin. *Ellipses et structures parallèles : la constituance dans les nouvelles syntaxes*. Directeur A. Delaveau.
- Céline Vaguer, *La préposition dans*. Directeur D. Leeman.



## Références générales sur le projet

- Abeillé A., Godard D. 1998. « French word order and lexical weight ». In Borsley R. (éd.), *Syntactic categories, Syntax and Semantics*. New York : Academic Press.
- Abeillé A., Godard D., Sag I. 1998. « Two Kinds of Composition in French complex Predicates ». In Kathol A., Hinrichs E., Nakazawa T. (éds.), *Complex Predicates in non transformational Syntax, Syntax and Semantics*. New York : Academic Press, 1-41.
- Bonami O. 1998. « DI/PL, linéarisation, arbres polychromes : trois approches de l'ordre des mots ». *LINX*, 39, 43-70.
- Bonami O., Godard D. 2001. « Inversion du sujet, constituance et ordre des mots ». In Marandin J.-M. (éd.), *Cahier Jean-Claude Milner*. Paris : Verdier, 117-174.
- Bonami O., Godard D., Marandin J.-M. 1999. « Constituency and word order in French subject inversion », In Bouma G., Hinrichs E., Kruijff G.-J. M., Oehrle R. T. (éds.), *Constraints and resources in natural language syntax and semantics*. Stanford : CSLI Publications.
- Bresnan J. 1999. *Lexical-Functional Syntax*. Oxford : Blackwell.
- Espinal M. T. 1991. The representation of disjunct constituents. *Language*, 67 (4), 726-762.
- Fillmore C., Kay P. 1999. « Grammatical constructions and linguistic generalizations : the What's X doing Y ? construction », *Language*, 75 (1), 1-33.
- Goldberg A. E. 1995. *A Construction Grammar Approach to Argument Structure*. Chicago : University of Chicago Press.
- Kathol A. 1995. *Linearization-Based German Syntax*. PhD dissertation, Ohio State University.
- Lascarides A., Briscoe E. J., Asher N., Copestake A. 1996. « Order Independent Persistent Typed Default Unification ». *Linguistics and Philosophy*, 19 (1), 1-89.
- Miller P. H. 1999. *Strong Generative Capacity : The Semantics of Linguistic Formalism*. Stanford : CLSI Publications.
- Miller P. H., Torris T. 1990. *Formalismes syntaxiques pour le traitement automatique du langage naturel*. Paris : Hermès.
- Pollard C., Sag I. A. 1994. *Head-Driven Phrase Structure Grammar*. Chicago : University of Chicago Press.
- Reape M. 1994. « Domain union and word-order variation in German », In Nerbonne J., Netter K., Pollard C. (éds.), *German in Head-Driven Phrase Structure Grammar*, Stanford : CSLI, 151-197.
- Sag I. A., Wasow T. 1999. *Syntactic theory, a formal introduction*. Stanford : CSLI Publications.
- Shibatani M. 1985. « Passives and related constructions : a prototype analysis ». *Language*, 61 (4), 821-848.

## **1.5. PROBLÈMES ÉPISTÉMOLOGIQUES DU TAL (PETAL)**

### **Participants**

*Membres rattachés à titre principal* : Marcel Cori (responsable), Sophie David (responsable), Karine Baschung, Marianne Desmets, Claude de Loupy, Denise Malrieu, Richard Walter

*Jeune chercheur associé* : Antonio Balvet

*Collaborations extérieures* :

Muriel Amar (BPI)

Jean-Christophe Binetti (Consult & Marketing GmbH)

Jean-Daniel Gronoff (UMR 8557 CNRS - EHESS)

Jacqueline Léon (Laboratoire d'Histoire des Théories Linguistiques, CNRS UMR 7597 et Université Paris 7 - ENS Lettres et Sciences Humaines)

Jean-Marie Marandin (Laboratoire de Linguistique Formelle, CNRS UMR 7110 et Université Paris 7)

Rachel Panckhurst (Praxiling - Université Montpellier 3)

Entreprise Synapse Développements

### **Discipline(s) et/ou sous-discipline(s) concernée(s)**

Traitement automatique des langues, épistémologie, syntaxe formelle

### **Problématique**

Il a paru important de lancer un travail de réflexion sur l'épistémologie du TAL, et ce pour des raisons de deux ordres différents. Tout d'abord parce que le domaine, relativement jeune, peine à se définir et à se donner des délimitations. Les désaccords sur ce qu'il couvre sont nombreux. Les positionnements par rapport aux théories linguistiques apparaissent divers, voire contradictoires. Parce que les méthodes qui y ont cours sont sujettes à controverse. Ensuite parce que la réflexion épistémologique déjà engagée sur le domaine montre peu de systématisme (du côté des champs couverts, que l'on peut relier aux désaccords sur l'extension du TAL, aux objets et aux méthodes qu'il a privilégiés, différents selon les périodes ; mais aussi du côté des acteurs qui l'ont menée, en étant partie prenante ou non du domaine). Dans le même temps, le contenu de cette réflexion n'a pas fait émerger de lignes directrices, elle a plutôt mis en avant des positionnements divergents et des tensions. Les lieux de visibilité (communications et publications) sont restés quant à eux faiblement organisés (excepté dans le domaine de l'évaluation de logiciels, mais c'est tout récent).

La réflexion apparaît très lacunaire tout au long des cinquante ans d'histoire du TAL. Elle est d'abord à mettre au compte du domaine de la traduction automatique (TA), où des rétrospectives critiques ont été régulièrement produites. On pense notamment à la rétrospective de Josselson (1971) qui se voulait une suite du rapport Bar-Hillel (1960), à l'introduction des numéros spéciaux de *Computational Linguistics* consacrés à la TA (Slocum 1985), aux travaux historiques d'Hutchins (1986, 2000), enfin à la rétrospective

historique contenue dans le chapitre 2 de Newton (1992). En dehors de la TA, on peut mentionner l'introduction du numéro spécial de *Computational Linguistics* consacré aux grands corpus (Church & Mercer 1993), et celle du numéro spécial consacré à la désambiguïsation lexicale dirigé par N. Ide et J. Véronis (1998).

En France, très peu de réflexions épistémologiques ont été menées sur les méthodes du TAL. On rappellera néanmoins la démarche initiée par le numéro spécial de la revue (*TAL*, 1992, vol. 33, n° 1-2), consacré au trentenaire de l'ATALA et de la revue.

Corrélativement au développement du TAL, s'est posée la problématique de l'évaluation des dispositifs informatiques dès les débuts de la traduction automatique, avec le rapport de Bar-Hillel (1960) et celui de l'Alpac (1966). Elle s'est cependant constituée comme un champ à part entière depuis un peu plus d'une dizaine d'années, à travers notamment l'organisation de conférences et de compétitions internationales (cf. par exemple TREC (*Text Retrieval Conference*), MUC (*Message Understanding Conference*)), les travaux de l'ISKO (*International Society of Knowledge Organization*). C'est l'une des thématiques de l'agence ELDA (Agence pour l'évaluation et la distribution de ressources linguistiques). Ces développements ne vont pas bien entendu sans enjeux intellectuels, scientifiques et financiers importants, qu'ils importent de situer et de préciser.

## **Thèmes de recherche**

### **1) Définir et délimiter le TAL**

La première question qui s'est posée a été de clarifier ce qu'était le Traitement automatique des langues (ou du langage), connu sous l'acronyme TAL. Pour désigner ce champ d'investigations et d'applications à l'intersection de la linguistique, des mathématiques, de l'informatique et des sciences cognitives hérité des expériences pionnières en traduction automatique, plusieurs termes sont ou ont été en concurrence : *Computational Linguistics* ou *Natural Language Processing* dans le domaine anglo-américain, *Traitement automatique des langues*, *Informatique linguistique* ou *Linguistique informatique* en France. Nous nous sommes proposés, en retraçant le parcours historique de ces dénominations, de montrer que le flottement sur les termes était symptomatique des tensions à l'œuvre dans le domaine, sur le plan des enjeux institutionnels, économiques, théoriques et conceptuels.

La deuxième question cherche à délimiter les changements survenus en linguistique du fait de son contact avec l'informatique (la linguistique et l'informatique sont des disciplines « en contact » comme on dit que deux langues sont en contact : ignorance réciproque, rivalité plus ou moins belliqueuse, emprunt et collaboration). La période historique prise en compte va de la définition de la notion de grammaire par Chomsky, et en particulier de la caractérisation mathématique des grammaires possibles (la linguistique algébrique), à la conception de grammaires multidimensionnelles qui trouvent dans les systèmes de représentation des connaissances leur fondation formelle.

### **2) Evaluer des dispositifs**

Des travaux récents ont cherché à clarifier les différentes options : évaluation quantitative/qualitative, évaluation orientée concepteur/utilisateur, évaluation boîte noire/boîte transparente, utilisation de phrases de tests/corpus, utilisation de procédures automatiques/recours à des évaluateurs (experts/non-experts). La réflexion a aussi porté sur la liste des critères pertinents, sur le type de mesures à utiliser (par exemple, les taux de

rappel et de précision, mis au point dans le cadre de la recherche d'information), sur la constitution de référentiels de comparaison (par le projet TSNLP, Test suites for natural language processing), sur la constitution de standards (par exemple, les travaux de M. King dans le cadre de EAGLES), etc. (cf. entre autres King & Falkedal 1990 ; Sparck-Jones & Galliers 1996 ; Nübel & Seewald 1998 ; Lavenus & Lapalme 2002). A travers la mise en œuvre d'évaluations, nous nous proposons de réfléchir sur les méthodologies mises en œuvre et les différents concepts impliqués.

### **3) Observer le Web**

La mise à disposition de contenus extrêmement variés sur le Web rend celui-ci particulièrement intéressants à « observer ». Sans épuiser les thèmes qui y sont associés, on s'intéresse ici à la collecte d'informations, plus particulièrement en liaison avec la terminologie et les sciences de l'information, et à la réflexion autour de ses futurs développements.

## ***Bilan***

### **Résultats obtenus**

#### **1) Définir et délimiter le TAL**

##### **1a) Les lignes de tension dans l'histoire du TAL** (Marcel Cori, Sophie David, Jacqueline Léon)

Nous avons pu mettre en évidence deux lignes de tension constantes dans l'histoire du TAL : la cohabitation paradoxale et nécessaire des recherches théoriques et des applications à visée industrielle d'une part, les antagonismes entre le TAL et les différentes disciplines qui le constituent, voire entre ces disciplines elles-mêmes quand elles rentrent en interaction dans un problème de TAL.

Si l'on considère le TAL aux débuts de la TA, ce sont les applications à visée industrielle qui sont privilégiées. Les questions théoriques comme les tensions entre les disciplines sont très réduites puisqu'on fait appel à une linguistique pour la machine, totalement *ad hoc* et déconnectée des résultats de la linguistique théorique.

La première rupture au début des années 1960 a fait émerger la *Computational Linguistics*, comme tentative de séparer les applications à visée industrielle des recherches théoriques fondamentales (rendue possible par la naissance des algorithmes d'analyse syntaxique).

La seconde rupture, à la fin des années 1980, prend acte de l'irruption de l'Intelligence artificielle, en même temps que la nécessité, une fois de plus renouvelée, de séparer applications et recherche théorique. Devant la difficulté d'autonomiser les investigations théoriques en discipline à part entière, et face aux contraintes définissant une « bonne application », impossibles à satisfaire totalement, on tentera de séparer, parfois de façon tout à fait artificielle et au risque de la faire disparaître, la *Computational Linguistics*, considérée comme plus théorique, au profit du *Natural Language Processing*, nouveau champ délibérément orienté vers les applications.

En France, le TAL, qui a rencontré davantage de difficultés à s'implanter institutionnellement, semble avoir acquis un rôle fédérateur entre recherches théoriques, pour lesquelles on utilise encore parfois le terme de linguistique informatique, et applications. Quel statut alors donner à l'ingénierie linguistique ou aux industries de la langue, qui comme leur nom l'indique, désignent clairement les applications industrielles ?

Alors qu'on pouvait penser qu'un point de stabilité du TAL était constitué par une communauté de chercheurs d'origines disciplinaires diverses, l'essor des applications au détriment des travaux théoriques ne va-t-il pas conduire les acteurs, informaticiens et linguistes, à retourner dans leurs disciplines respectives en abandonnant le terrain aux seuls ingénieurs ? L'étude menée a bien montré la fragilité du TAL comme champ autonome lié aux recherches linguistiques (Cori & Léon 2002).

### **1b) Importance des recherches informatiques dans le développement de la linguistique** (Marcel Cori, Jean-Marie Marandin)

Dans Cori & Marandin (2001) a été mise clairement en évidence l'importance déterminante des recherches en informatique pour le développement de la linguistique entendue comme entreprise de confection de grammaires. Les emprunts (aux notions, concepts, arguments issus de la théorie de la compilation ; aux langages de programmation ; à la modélisation des hypothèses du programme cognitiviste ; aux dispositifs pratiques de traitement d'information en langue naturelle) ont servi aussi bien à motiver le rejet de certaines formes de grammaires qu'à l'invention de nouvelles formes (par exemple, Chomsky et le rejet des grammaires syntagmatiques à partir des premiers résultats de la linguistique algébrique ; les solutions élaborées dans les ATN, en programmation logique et dans les systèmes de représentations de connaissances, sont reprises et métamorphosées pour donner lieu à de nouveaux formalismes dont les caractéristiques sont la déclarativité et l'évolution vers une plus grande multidimensionnalité)

## **2) Evaluer des dispositifs**

### **2a) Comparaison de logiciels d'extraction d'expressions à caractère terminologique** (Muriel Amar, Sophie David).

Réalisé dans le cadre du projet « Evaluation d'outils d'aide à la constitution automatique de terminologie et de relations sémantiques entre termes à partir de corpus » (convention entre l'AUF et l'université Lille 3 : ARC n° X/7.10.04/llec.A3o/CNRS UMR 8529), le travail a porté spécifiquement sur l'évaluation des quatre extracteurs d'expressions à caractère terminologique (*Acabit* B. Daille ; *Ana* Ch. Enguehard ; *Lexter* D. Bourigault ; *TermFinder* Xerox) :

- à partir d'un corpus inscrit dans le domaine de la physiologie de la reproduction ;
- pour les champs d'application suivants : l'indexation, la traduction et la terminologie.

Les résultats de cette recherche ont été essentiellement de nature méthodologique :

- 1) Réflexion sur les conditions requises permettant d'assurer une rigueur scientifique à l'évaluation.
- 2) Bilan critique d'un protocole par listes de référence (protocole classiquement utilisé pour ces différents domaines d'application).
- 3) Nécessité d'analyser le caractère opératoire des résultats produits : sont-ils transparents, cohérents et systématiques d'un point de vue linguistique ? Ces trois conditions apparaissent en effet comme des conditions nécessaires à l'élaboration de méthodologies de travail. Alors qu'une comparaison expressions produites/listes de référence fournit des résultats similaires quels que soient les logiciels évalués, ce type d'analyse permet au contraire de les hiérarchiser.

La méthodologie mise en place et ces différents résultats ont été explicités dans un rapport de recherche (Amar & David 2001) et ont été exposés au cours de deux journées d'étude.

**2b) Programme européen TRUST : évaluation de la précision sémantique d'un moteur de recherche par les protocoles SHEP** (Jean-Christophe Binetti, Jean-Daniel Gronoff, Denise Malrieu ; entreprise Synapse Développement, Toulouse).

TRUST (Text Retrieval Using Semantic Technologies) est un projet européen IST (Information, Societies Technology), qui vise à mettre en commun les techniques et les ressources de quatre partenaires européens pour quatre langues, le français, l'italien, le portugais et le polonais pour réaliser un moteur de recherche unique multilingue. La performance de ces moteurs repose sur la capacité à prendre en compte l'analyse syntaxique de la question posée en langage naturel. C'est la mesure de l'efficacité de l'analyse syntaxique qui est l'objectif principal des tests que nous avons commencés à mettre en place.

La méthodologie d'évaluation choisie privilégie les tests qualitatifs. Les corpus de textes utilisés (source bibliographique, romans français, normes ISO sur l'environnement) relèvent chacun d'un domaine spécialisé. L'évaluation mise en place revient à identifier *a priori* et baliser dans un corpus donné les bonnes réponses à une question donnée. L'évaluation des réponses du moteur tient compte du paramétrage du moteur (nombre de réponses possibles 1 à 20), de la langue, du type de support, et compare les réponses du moteur aux bonnes réponses repérées *a priori*. Les bonnes réponses nouvelles trouvées par le moteur sont intégrées et balisées pour enrichir les bonnes réponses de l'expert. Pour le français, le moteur développé s'appuie sur l'analyseur syntaxique *Cordial*, développé par la société Synapse qui est aussi leader du présent programme TRUST.

Les développements déjà réalisés ont permis de comparer de façon automatique les réponses du moteur aux bonnes réponses définies *a priori* par l'expert, de fournir les statistiques sur le taux de bonnes réponses et les rangs des bonnes réponses parmi les réponses données par le moteur, qu'elles correspondent à celles trouvées *a priori* par l'expert ou validées *a posteriori* par lui.

Les statistiques comparatives des bonnes réponses partielles ou complètes ont été obtenues sur la version multilingue du moteur (italien, français, portugais, polonais) sur certaines parties du corpus

Les résultats ont montré des performances assez contrastées selon que le corpus comporte des textes fortement structurés (normes ISO, bibliographies) ou pas (articles de journaux, romans) et selon les langues. Ils ont montré la nécessité que le moteur sache prendre en compte la structure hiérarchique du texte, des problèmes concernant l'ontologie universelle du moteur et de synonymie trop lâche. Le traitement des anaphores, partiellement effectué par le moteur, ne semble pas influencer de façon drastique les résultats.

Ce travail a fait l'objet d'un pré-rapport final. Le rapport final est en cours de rédaction (décembre 2003).

### **3) Observer le Web**

**3a) Étude de *La mesure de la culture sur le Web*** (Richard Walter)

La connaissance des ressources culturelles mises à disposition sur le Web, en termes de volume, de nature, d'accessibilité et de relation entre elles, est un enjeu primordial pour les

acteurs du secteur culturel francophone. L'étude commandée par le Ministère de la culture et de la communication (département des études et prospectives), avait pour objectif de mesurer la place qu'occupe la culture sur le réseau Internet, à travers le fonctionnement des moteurs de recherche. Ce sont ces derniers qui déterminent la visibilité des contenus sur le Web.

L'étude a été basée sur l'interrogation de *Google* et d'autres moteurs de recherche avec une liste de mots en français (termes associés à la culture et termes issus du vocabulaire utilisé au sein du Ministère de la culture). Nous avons fourni des résultats quantitatifs (nombre de réponses) et qualitatifs (origine des sites, langue, etc.), et mesuré en particulier la part des réponses pour le Web francophone en regard de celle pour le Web mondial. La phase informatique de l'enquête a consisté à lancer des requêtes multicritères sur les moteurs de recherche sélectionnés, à extraire et à regrouper les résultats obtenus. Les requêtes ont été activées automatiquement (« à la volée ») et les résultats stockés sous une forme homogène dans une base de données.

Nous avons pu ainsi analyser et comparer les différentes techniques d'indexation et de classement utilisées par les moteurs (« crawler », catégorisation, « ranking », etc.). On aboutit par ailleurs à la définition du champ de la culture dans les moteurs de recherche étudiés, et d'un point de vue méthodologique, à la définition d'indicateurs statistiques pertinents et à la mise au point de protocoles de lancement de requêtes. Enfin, nous avons effectué différentes recommandations visant à améliorer la visibilité, l'indexation et le catalogage des sites du secteur culturel francophone, notamment celui du Ministère de la culture et de la communication.

Cette recherche a été financée par le Ministère de la culture et de la communication (département des études et prospectives) et va faire l'objet d'un rapport (décembre 2003).

### **3b) Mise au point du module français du site du Réseau international francophone d'aménagement linguistique (Richard Walter)**

Le secteur de l'ingénierie linguistique accompagne l'essor d'Internet, du multimédia et de la bureautique. Nous avons besoin d'outils qui nous aident à écrire correctement, à traduire ou résumer un texte, à rechercher facilement des informations sur Internet, à dialoguer avec la machine, etc. Face à ces enjeux, il est important d'avoir une vision globale de ce secteur, plus particulièrement de ce qui se produit en France pour le traitement automatique du langage. Un site de référence dans ce domaine a été élaboré par la Délégation générale à la langue française et aux langues de France (DGLFLF, ministère de la culture et de la communication), qui en a confié la réalisation et la mise à jour au laboratoire MoDyCo.

La Délégation est en effet chargée de recenser l'offre française d'ingénierie linguistique, pour le compte du Réseau international francophone d'aménagement linguistique (RIFAL)<sup>1</sup>, réseau piloté par l'Office de la langue française (OLF, Montréal)<sup>2</sup>.

---

<sup>1</sup> Ce réseau a succédé au Réseau international des observatoires du français sur les inforoutes et le traitement informatique des langues (RIOFIL).

<sup>2</sup> Le site de la DGLFLF a été longtemps le seul site central de l'offre d'ingénierie linguistique en langue française. Depuis, le site d'ELDA ([www.elda.fr](http://www.elda.fr)) diffuse, dans le cadre d'un projet européen (Euromap), un répertoire des entreprises et des outils d'ingénierie linguistique. A partir de 2004, le site portail [www.technolangue.fr](http://www.technolangue.fr), résultant d'un appel d'offres des ministères de la culture, de l'industrie et de la recherche, aura les mêmes fonctions que le module français du RIFAL, avec des extensions plus nombreuses (organismes de recherche, thèses, annonces de nouveautés et de colloques, définitions du domaine, dépêches de presse, etc.).

Nous avons effectué deux campagnes (une campagne de vérification et de recherche d'éléments documentaires ; une campagne d'entretiens auprès des principaux experts de l'ingénierie linguistique).

Ces deux campagnes ont permis :

1) d'élaborer les thématiques permettant de structurer la présentation du domaine de l'ingénierie linguistique (les ressources linguistiques, les outils et services de traduction assistée par ordinateur, les outils de gestion documentaire (résumé et indexation automatiques), le filtrage et l'indexation, la reconnaissance et la synthèse vocales, l'aide à la rédaction, la correction orthographique et grammaticale, l'aide à l'apprentissage des langues, les ontologies et les outils de gestions de connaissance) ;

2) d'établir et de mettre à jour des annuaires des entreprises, des groupements professionnels, des institutions et associations et des laboratoires de recherche dans le domaine de l'industrie linguistique, avec une fiche descriptive pour chaque produit (support, prix, technologies, spécificités des bases linguistiques, interface, etc.).

La recherche a été financée par la Délégation générale à la langue française et aux langues de France (DGLFLF, ministère de la Culture et de la Communication).

### **3c) Le Web sémantique : enjeux linguistiques et documentaires** (Muriel Amar, Sophie David)

Si le Web met à disposition de nombreuses informations, force est de constater que les moyens pour trouver, comparer, évaluer ces informations restent limités. Les moteurs de recherche, en utilisant des ressources lexicales (partielles) et statistiques, fournissent trop ou trop peu de résultats sur le plan quantitatif, résultats dont la qualité n'est pas assurée. Le Web sémantique, impulsé par Tim Berners-Lee en 1998, développé dans le cadre du W3C (Consortium World Wide Web) depuis 2001, se veut une extension du Web actuel, qui vise à rendre les contenus, non plus uniquement accessibles et affichables, mais aussi exploitables et interprétables par des machines. En cela, les travaux autour du Web sémantique impliquent en premier chef des aspects linguistiques formels, et plus particulièrement sémantiques.

Le travail qui a été mené fait suite à celui qui a été réalisé par des étudiants de la filière information-documentation d'entreprise de l'IUT de Paris V, où l'on a montré l'intérêt de construire des actions de diffusion et de réflexion autour des différents concepts impliqués (ontologie, métadonnée, annotation, etc.). Ce travail a débouché sur l'organisation d'une journée d'études « Le Web sémantique : de nouveaux enjeux documentaires ? » (14 octobre 2003), co-organisée par Muriel Amar, Sophie David, l'ADBS (Association de professionnels de l'information) et l'IUT de Paris V.

Les synthèses et comptes-rendus, réalisés par Annabel Chevaux et Laurie de Temmerman, coordonnés par Muriel Amar et Sophie David, seront disponibles sur le site Internet de l'ADBS ([www.adbs.fr](http://www.adbs.fr)), courant décembre.

### ***Perspectives***

Le domaine du TAL a besoin d'un débat sur ses pratiques et ses finalités, plutôt que de continuer à avancer dans un certain flou conceptuel et avec des conflits souterrains qui restent inexprimés au grand jour. Actuellement, on peut dégager deux ensembles de



problèmes à explorer, qui touchent aux évolutions du domaine, aux déplacements de son centre de gravité :

1) Le premier concerne les rapports entre le TAL et la linguistique formelle, notamment la question de l'appartenance des formalismes de représentation syntaxique au domaine du TAL. Les analyseurs syntaxiques fondés sur les CFG ont été pendant longtemps au cœur de la discipline. Devant l'insuffisance de la couverture grammaticale des CFG, et devant les difficultés techniques posées par le traitement automatique des grammaires transformationnelles, il y a eu tout d'abord la création des ATN, puis la définition de nouveaux formalismes de représentation syntaxique, DCG, LFG, FUG, GPSG, TAG, HPSG, ou le perfectionnement de formalismes plus anciens tels que les grammaires catégorielles ou les grammaires de dépendance. Les ATN étaient un outil hybride, à la fois formalisme et dispositif d'analyse, dont le rattachement au TAL ne faisait aucun doute. Ces différents formalismes ont ainsi été créés soit clairement dans des perspectives de traitement automatique, soit dans des perspectives plus strictement linguistiques. Leur regroupement dans les années 1980 sous le terme de « grammaires d'unification » était une façon de les inclure globalement dans le domaine du TAL. On a eu là une prise de position délimitant des objets spécifiques pour le TAL. Or il y a lieu de s'interroger sur ce regroupement qui n'allait pas entièrement de soi, étant donné, au moins, la double origine de ces objets. Quelque vingt années plus tard, sont-ils à considérer comme des objets relevant uniquement de la linguistique formelle ? Leur double origine, dont on n'aurait pas pris suffisamment la mesure, permet-elle d'expliquer que la question se pose à nouveau ? Ou bien est-ce l'apparition de nouvelles méthodes, dites de TAL robuste, qui a rejeté les formalismes linguistiques à la périphérie du TAL ?

2) On rejoint là un second ensemble de problèmes. Il porte tout d'abord sur les méthodes de TAL robuste : méthodes de désambiguïsation fondées sur des techniques probabilistes, méthodes d'analyse partielle qui ne cherchent pas à obtenir des analyses des phrases mais simplement à délimiter certains constituants, méthodes d'acquisition sémantique qui ignorent la syntaxe, etc. On est en droit de s'interroger sur le statut et l'apport de ces méthodes. Mettent-elles en œuvre un modèle du texte, du langage, des langues ou bien ne font-elles qu'utiliser un outil ? Et, en ce cas, quel outil et pour quel usage ?

La question se pose également d'identifier les recouvrements et délimitations entre TAL robuste et linguistique(s) de corpus, le TAL robuste s'appuyant sur l'existence de grandes masses de données écrites ou orales, la (les) linguistique(s) de corpus empruntant certains outils du TAL robuste, notamment les méthodes probabilistes.

Quant aux propositions des linguistiques de corpus, fondent-elles une nouvelle linguistique, ou n'opèrent-elles qu'un retour à des conceptions anciennes ? Dans les réflexions déjà élaborées, les problèmes évoqués, qui dessinent un champ d'études se posant comme une alternative aux travaux menés depuis quarante ans, mériteraient des discussions et des analyses plus approfondies. Les positionnements s'articulent, souvent en filigrane, sur une critique des positions de Chomsky (1957), considérées en bloc, sans que l'on discute précisément des arguments qu'il a avancés, ni d'ailleurs que l'on distingue programme de recherche et théorie particulière, ni enfin qu'on remette en perspective ses positions dans le contexte de l'époque. La mise à disposition de très grands corpus suscite certes de nouvelles questions, mais on n'a pas circonscrit précisément celles qui ne seraient pas purement artefactuelles, celles dont on peut faire l'hypothèse qu'elles sont liées à un état de la technique, etc.

## Étapes à venir

**1) Linguistique de corpus et linguistique formelle** (Antonio Balvet, Karine Baschung, Marcel Cori, Sophie David, Marianne Desmets, Richard Walter)

Une réflexion doit être effectuée sur les linguistiques de corpus, sur les différences entre linguistique formelle et linguistique de corpus. Un projet d'édition scientifique d'un numéro d'une revue est en cours d'élaboration.

**2) Outils de TAL robuste** (Marcel Cori, Claude de Loupy)

Une étude précise d'outils de TAL robuste sera effectuée, en lien avec l'entreprise Sinequa qui réalise de tels outils. On examinera :

(a) des étiqueteurs probabilistes à base de chaînes de Markov, de la méthode de Brill ou d'arbres de décision à partir d'un apprentissage sur un corpus étiqueté en étiquetage des parties du discours ou en désambiguïsation sémantique ;

(b) des systèmes à base de transducteurs pour la reconnaissance des entités et l'analyse syntaxique ;

(c) l'utilisation conjointe de modèles probabilistes et de transducteurs dans les étiqueteurs. L'examen de ces outils, du point de vue des méthodes mises en œuvre et des résultats qu'ils produisent permettra d'en cerner les limites et, éventuellement, de proposer des améliorations.

### **3) Evaluation**

**3a) Projet Trust** (J.-C. Binetti, J.-D. Gronoff, Denise Malrieu)

L'évaluation menée pour le projet TRUST montre qu'il faut prendre en compte de façon plus précise les lexiques de domaines mais aussi ajouter des connaissances sur les genres textuels afin de mieux caractériser la complexité des questions et des bonnes réponses. On travaillera plus particulièrement sur les propriétés linguistiques :

- des questions (mots composés, anaphores, relations argumentales non réciproques etc.) ;
- des bonnes réponses dans les textes sources.

On cherchera en outre à évaluer de façon plus fine le traitement des anaphores et les équivalences synonymiques du moteur de recherche.

Un séminaire et la publication des premiers résultats sont prévus au cours du premier semestre 2004.

**3b) Concours EASA** (Muriel Amar, Sophie David)

L'activité d'expertise que Sophie David mène depuis 1998 pour le concours européen de logiciels EASA (European Academic Software Award), (pour laquelle elle est responsable de la section « Langues et linguistique » depuis 2002), en collaboration avec Rachel Panckhurst (Praxiling - Université Montpellier 3), va déboucher sur :

- l'édition d'un ouvrage *Evaluation in e-learning : the European Academic Software Award* (Panckhurst R., David S., Whistlecroft L. (éds)) ;
- l'organisation d'une journée d'études à Montpellier au premier semestre 2004 « EASA : nouvelles perspectives ».

## Production scientifique

Amar M., David S. 2001. *Evaluation de logiciels d'extraction dans les champs de l'indexation, la traduction et la terminologie. Corpus INRA. Rapport de recherche. Action de recherche concertée (n° X/7.10.04/llec.A3o), AUF et CNRS UMR 8529 (Cersates, Université Lille 3), 109 p.*

- Binetti J.-C., Gronoff, J.-D., Laurent D., Malrieu D. 2003. *TRUSTSearch V.1.3. : Test Report*. Rapport de contrat.
- Cori M., Marandin J.-M. 2001. « La linguistique au contact de l'informatique : de la construction des grammaires aux grammaires de construction ». *Histoire, Épistémologie, Langage*, 23 (1), 49-79.
- Cori M., Léon J. 2002. « La constitution du TAL. Étude historique des dénominations et des concepts ». *TAL « Problèmes épistémologiques »*, 43 (3), 21-55.
- Cori M., David S., Léon J. 2002. « Pour un travail épistémologique sur le TAL ». *TAL « Problèmes épistémologiques »*, 43 (3), 7-20.
- Cori M. 2003. « La mathématisation des formalismes syntaxiques ». *LINX*, 48, 13-28.
- Panckhurst R., David S., Whistlecroft L. (éds). A paraître. *Evaluation in e-learning : the European Academic Software Award*.
- Walter R. 2003. *La mesure de la culture sur le Web*. Rapport d'études. Ministère de la Culture, Dépt des Etudes et Prospectives, 40 p.

### **Références générales sur le projet**

- Bar-Hillel Y. 1960. « The Present Status of Automatic Translation of Languages ». *Advances in Computers*, New-York - Londres : F. C. Alt ed. Academic Press, 1, 91-141.
- Church K. W., Mercer R. L. 1993. « Introduction to the Special Issue on Computational Linguistics Using Large Corpora ». *Computational Linguistics*, 19 (1), 1-25.
- Hutchins W. J. 1986. *Machine Translation, Past, Present, Future*. Chichester : Ellis Horwood Ltd.
- Hutchins W. J. 2000. *Early Years in Machine Translation*. Amsterdam : John Benjamins.
- Ide N., Veronis J. 1998. « Introduction to the Special Issue on Word Sense Disambiguation : the State of the Art ». *Computational Linguistics*, 24 (1).
- Josselson H. H. 1971. « Automatic translation of Languages since 1960 : a Linguist's View ». *Advances in Computers*, 11, 1-58.
- Kay M. 1982. « Machine Translation ». *American Journal of Computational Linguistics*, 8 (2), 74-78.
- King M., Falkedal K. 1990. « Using Test Suites in Evaluation of Machine Translation Systems ». *Coling*, 2, 211-216.
- Lavenus K., Lapalme G. 2002. « Évaluation des systèmes de question réponse. Aspects méthodologiques ». *TAL « Problèmes épistémologiques »*, 43 (3).
- Nübel R., Seewald-Heeg U. (éds). 1998. *Evaluation of the Linguistic Performance of Machine Translation Systems*. St-Augustin : Gardez ! Verlag.
- Slocum J. (éd.). 1985. « Two Special Issues on MT ». *Computational Linguistics*, 11 (1-2-3).
- Sparck-Jones K., Galliers J. R. 1996. *Evaluating Natural Language Processing Systems : an Analysis and Review*. Berlin : Springer-Verlag.

## 2. Dynamiques

---

- 1) Appropriation du français en situation diglossique
- 2) Représentation des événements : sémantique verbale, temporalité et développement des énoncés complexes
- 3) Apprentissage / enseignement précoce des langues étrangères
- 4) Didactique de la syntaxe
- 5) Elaboration d'un glossaire de termes linguistiques
- 6) La variation stylistique

### 2.1. ACQUISITIONS

#### 2.1.1. APPROPRIATION DU FRANÇAIS EN SITUATION DIGLOSSIQUE

##### **Participants**

*Membres rattachés à titre principal* : Colette Noyau (responsable), Maria Kihlstedt, Carole Tisset

*Membre associé* : Maxime Somé

*Jeunes chercheurs associés* : Ufoalè-Christine Afola, Cécile Vigouroux

*Doctorants* : Innocent Adouna (en cours d'inscription), Mayé Diouf, Kossi Souley Gbeto

*Collaborations extérieures* :

Enseignants-chercheurs de l'équipe togolaise (séjours de recherche au sein de l'UMR) : Edith Bedou-Jondoh, Alilou Cissé, Issa Takassi (cf. 2.1.2. Représentation des événements : sémantique verbale, temporalité et développement des énoncés complexes)

##### **Discipline(s) et/ou sous-discipline(s) concernée(s)**

Linguistique de l'acquisition, psycholinguistique, sociolinguistique, didactique

## **Problématique**

Ce projet analyse la genèse du sujet bilingue francophone en situation diglossique, à travers la scolarisation en français et l'expérience langagière dans l'environnement en langues premières, et examine le fonctionnement du français dans ce cadre.

Il s'agit d'étudier de façon coordonnée l'appropriation et l'utilisation du français langue seconde au long de la scolarisation :

- depuis l'entrée à l'école jusqu'à l'issue de la classe de 3<sup>ème</sup>, d'où doivent sortir de jeunes adultes bilingues ;
- parallèlement au développement des capacités langagières en langue première ;
- en milieu urbain et en milieu rural.

On vise ainsi à mieux comprendre les processus en jeu chez les élèves dans l'apprentissage simultané de la langue de scolarisation et des connaissances dans les matières scolaires dans un environnement bilingue français / langue autochtone. En s'appuyant sur les processus d'apprentissage analysés et sur les caractéristiques spécifiques de ces situations d'acquisition, on espère contribuer ensuite à améliorer les démarches d'enseignement du français langue seconde et d'enseignement de l'ensemble des matières scolaires à travers elle.

Il s'agit d'un projet novateur, car il étudie l'état du français en francophonie dans sa dynamique de reproduction à partir d'une nouvelle constellation de disciplines scientifiques, incluant psycholinguistique et linguistique de l'acquisition.

Les chercheurs de MoDyCo travaillant en acquisition constituent une composante fondatrice du GDR 113 du CNRS sur l'Acquisition des Langues (1988-2000) qui a implanté ce domaine de recherche en France, à l'issue du programme européen sur l'acquisition des langues par des adultes.

Notre projet est le premier parmi ceux menés dans les dernières années sur l'apprentissage du français en Afrique (Burkina, cf. Batiana & Prignitz 1998 ; Guinée, cf. Holtzer 2002) à étudier la dynamique du français en francophonie à partir des processus acquisitionnels auxquels il donne lieu, en s'appuyant sur la linguistique de l'acquisition et la psycholinguistique.

Depuis 2002 un volet spécialisé s'est adjoint à ce projet, portant sur la construction de connaissances via le français langue de scolarisation, dans le cadre de l'ACI Cognitive, volet « Acquisition des langues : dysfonctionnements et remédiation ».

Notons que depuis l'ouvrage fondateur de S. Lafage de 1985 (thèse de 1978), très peu de choses avaient été publiées sur la situation linguistique et la dynamique du français au Togo.

Cette opération se déroule en collaboration avec une équipe de linguistes de l'Université de Lomé (Togo), constituant une Action de Recherche Partagée (ARP) de l'AUF, et un projet bilatéral du programme CORUS de coopération universitaire nord-sud du MAE.

## **Thèmes de recherche**

Ce projet a adopté une approche nouvelle des processus qui conduisent les jeunes enfants des pays considérés francophones à devenir francophones via la scolarisation. Plus précisément, on associe aux nécessaires aspects sociolinguistiques sur le statut du français parmi les langues parlées par ces apprenants (cf. Calvet & Dumont 1999 ; Dumont &

Maurer 1995) et sur l'exposition à la langue (input langagier, nature des occasions de communiquer) des méthodes permettant d'observer directement et indirectement les parcours d'acquisition du français et de construction des connaissances en français dans ce cadre, au long des années qui vont de l'entrée à l'école - entrée en francophonie en général - à l'entrée dans l'âge adulte au sortir du collège.

Nous avons souhaité faire servir les connaissances produites dans les quinze dernières années par la recherche psycholinguistique sur l'acquisition des langues dans le milieu social en Europe à une analyse en profondeur des parcours d'appropriation du français par les jeunes scolarisés, nous distinguant ainsi des approches antérieures : par exemple, les approches sociolinguistiques (Batiana & Prignitz 1998 ; Holtzer 2002) ; l'approche didactique (Verdelhan-Bourgade 2002).

Ainsi, nous espérons mieux comprendre :

- 1) comment se bâtit la francophonie et le bilinguisme français - langue(s) autochtone(s) chez les jeunes locuteurs ;
- 2) comment certains traits acquisitionnels des lectes individuels de français, se stabilisant chez des adultes, contribuent à alimenter l'expérience du français donc l'acquisition de la langue par de nouveaux locuteurs, contribuant ainsi à modeler la langue française (cf. Chaudenson 1997) ;
- 3) on s'interroge également sur la meilleure façon de faire une place à l'école sur le bilinguisme de fait et sur les connaissances de la L1 et en L1 chez les élèves (cf. Da Silveira & Hamers 1990 ; Bearth 1997).

La linguistique de l'acquisition des langues a contribué à ré-envisager radicalement la nature des processus cognitifs en jeu dans l'acquisition linguistique, et ce à partir d'études longitudinales et translinguistiques, portant sur l'appropriation d'une nouvelle langue par des adultes dans le milieu social, au cours des années 80. Les aspects principaux de cette conception sont les suivants :

- *autonomie du processus* de construction de la langue par rapport à l'exposition à la langue et au « programme » fourni par l'input langagier scolaire ;
- *lectes* structurés par des *principes d'organisation* de l'information, par des *hypothèses* évolutives sous-jacentes sur les liens formes-fonctions (ces hypothèses de structuration se manifestent au mieux dans la production de discours suivis, d'où le recours nécessaire à une approche *textuelle*) ;
- dimension *interactionnelle* (cadres discursifs du dialogue, négociation du sens et des formes de l'interlocution, etc.) ;
- échanges *exolingues* ;
- évolutions entre états de langue (lectes) successifs relevant de *facteurs endogènes* et *exogènes*.

Les études acquisitionnelles récentes ont porté leur regard, selon une perspective similaire, d'une part sur l'acquisition scolaire et enfantine des langues, aussi bien en milieu isolé (le seul contact avec la langue étrangère étant la classe de langue) qu'en situation mixte (contact par la classe de langue et dans l'environnement extérieur), d'autre part sur les caractéristiques linguistiques des lectes d'apprenants très avancés. Ces dernières réflexions peuvent être adaptées à l'examen des lectes stabilisés de français des locuteurs compétents que sont les enseignants, servant de référence à l'acquisition de la langue française par les élèves.

## ***Bilan***

### **Étapes passées**

1) Enquêtes sur le terrain (Togo)

a) Enquête transversale : enregistrement de groupes d'enfants pris à cinq paliers d'acquisition : CP1, CE2, CM2, 5<sup>ème</sup>, 3<sup>ème</sup>, et de L1 gengbe-éwé, dans une gamme étendue d'activités langagières, orales et pour quelques-unes, écrites, en français et pour certaines activités, également en L1, en contrastant zone urbaine (Lomé) et zone rurale.

b) L'enquête en zone urbaine (dans la capitale) est terminée en septembre 2003, l'enquête en zone rurale, plus compacte car s'appuyant sur les résultats de celle de Lomé pour être plus sélective, se déroule de novembre 2003 à juin 2004.

c) Enquêtes sur les situations didactiques et les activités de classe, pour l'enseignement de la langue française et les autres matières, en vue d'analyser l'acquisition de connaissances via la langue française.

d) Questionnaires et entretiens de nature métalinguistique sur les représentations du français, avec des maîtres et des élèves (analysés du point de vue méthodologique par Vigouroux nov. 2003).

e) Questionnaires et entretiens avec des parents et des enseignants sur les conceptions de l'intelligence enfantine et de l'apprentissage.

2) Exploitation des données et interprétation des résultats

La transcription-saisie des données est effectuée à mesure que l'enquête progresse, à Lomé (étudiants-enquêteurs) et à Paris X (mémoires de maîtrise et DEA sur le projet).

### **Résultats obtenus**

#### **1) Sur les productions orales d'élèves en français et en L1**

1a) Diverses techniques de sollicitation de productions narratives ont été essayées lors de la pré-enquête (1996-2000), en français et en L1, permettant de préciser la méthodologie de comparaison des productions. Ainsi, nous sommes en mesure de distinguer dans les productions en français langue seconde ce qui relève de l'étape d'acquisition de la L2, d'une éventuelle influence de la L1 sur la production en L2, et ce qui relève du degré de maîtrise des capacités langagières tel qu'il se manifeste en langue première (Noyau 1998, version remaniée sous presse). Sur la démarche comparative dans ces données en général, cf. Noyau, novembre 2003.

1b) Une étude menée à partir d'une technique de reprise de récit entendu examine les reformulations des procès par les enfants de CE2 et de CM2, en français et en L1 gengbe (Bedou-Jondoh & Noyau 2003).

1c) Le domaine linguistique des verbes de procès a fait l'objet de plusieurs études consécutives. Noyau (à paraître en 2003) s'est penchée sur les principes cognitifs et linguistiques du développement des verbes dans l'acquisition primaire du langage et dans l'acquisition d'une L2. Kihlstedt (mars 2003), Noyau (mars 2003), et Kihlstedt et Noyau (octobre 2003) reprennent cette même question, à partir de deux corpus de restitutions de récits oraux en français et en L1 (CE1 - CM1 et CM2 - 5<sup>ème</sup>).

1d) K. S. Gbeto (nov. 2003) a analysé l'utilisation des connecteurs dans des instructions (textes procéduraux), produits en français oralement, et y dégage la progression

linguistique et cognitive du point de vue de la cohésion et de la connectivité, et pour ce qui est des énoncés complexes.

1e) A. Cissé (nov. 2003) analyse les récits oraux en FL2 à partir d'une séquence d'images chez des élèves de CE1, et y dégage les faits acquisitionnels liés à la cohésion et à la connectivité.

## **2) Sur les interactions en classe et le français scolaire**

2a) Sur la question du français de référence, pour les élèves qui acquièrent le français via la scolarisation, Noyau (2001) montre à partir de l'analyse de 3 séquences de classe comment au niveau du premier degré (l'école de base) la classe est structurée par des échanges question-réponse dominés par le maître et orientés vers la formulation écrite normée des contenus présentés.

2b) Le statut de l'oral et de l'écrit dans l'apprentissage scolaire du français est examiné dans Noyau & Cissé (sous presse). L'étude conclut à ce que, contrairement à l'environnement culturel fondé sur l'oralité, le milieu scolaire est tourné vers l'écrit jusque dans les activités orales, ce qui rend difficile la régulation des échanges oraux par la communication. Ce thème du rapport à l'écrit a été approfondi par Noyau (2002 et chapitre à paraître).

2c) Tisset (à paraître) étudie l'initiation à la lecture au CP à partir de l'examen des manuels de langue et de séquences de classe de « langage », et constate que l'enseignement de la lecture du français, langue seconde, ne repose pas sur de véritables activités langagières, sur une acculturation à l'écrit, sur une réflexion linguistique contrastive. Mais il repose plutôt sur le déchiffrage, sans recherche de la compréhension littérale ou inférée, ce qui fait que pour parvenir au sens, le coût cognitif est extrêmement élevé.

2d) La classe (français et autres matières) comme environnement écologique d'acquisition de la L2 est étudiée par Noyau & Quashie (2003), qui montrent notamment comment la « bifocalisation » de l'interaction maître / élèves (sur le message / sur la langue) se résout bien souvent au détriment des connaissances présentées et au bénéfice d'un travail sur la « bonne forme » écrite, ce qui a des incidences sur la structuration des connaissances et sur l'appropriation de la langue comme outil cognitif.

2e) Une étude des relations langage-connaissances mathématiques (Noyau & Vellard, sous presse) dans le primaire a analysé des séquences de mathématiques de différents niveaux, et a montré que la bifocalisation aboutit à laisser de côté les apprentissages conceptuels au profit de l'attention à la langue (écrite).

2f) Afolo (2003) a examiné l'usage des verbes de représentation de procès dans des cours de sciences d'observation au primaire, et Noyau (2003) a analysé l'expression linguistique de notions temporelles et les représentations du temps dans des séquences d'Histoire. Noyau & Koudossou (à paraître) mettent en évidence les contraintes sur l'enseignement et donc sur l'apprentissage que font peser les pratiques d'évaluation certificative tendant à ignorer la dimension des contenus cognitifs et textuels en y substituant des critères de correction phrastique et surtout orthographique.

2g) Noyau (mars 2003, contribution aux états généraux sur l'enseignement du français en Afrique à Libreville) effectue un diagnostic des difficultés suscitées par l'actuelle doxa pédagogique en acte pour l'appropriation du français par les enfants, et trace des pistes de renouvellement du système d'enseignement pour potentialiser la situation plurilingue et pluriculturelle et l'immersion francophone offerte par l'école.



Enfin, le volet « Cognitique » du projet sur les conceptions de l'apprentissage et de l'intelligence infantine dans la culture du sud-Togo dans une perspective de psychologie interculturelle du développement a donné lieu à une étude onomasiologique des termes et expressions caractérisant l'intelligence en éwé et à l'étude d'entretiens auprès d'enseignants et de parents menés en français et en L1 sur ces questions (Noyau & Gbeto, juin 2003).

### **Actions réalisées**

1) Ateliers de formation scientifique de l'équipe togolaise

Ces sessions, prévues dans les programmes CORUS, avec l'apport de membres de l'équipe de MoDyCo, ont été au nombre de quatre : septembre 2001 (techniques d'enquête psycholinguistiques et sociolinguistiques), septembre 2002 (méthodes d'analyse de productions d'apprenants), février 2003 (bilan de l'enquête en zone urbaine, constitution de bases de données), septembre 2003 (analyses acquisitionnelles des données).

2) Les membres du projet ont participé à de nombreux colloques.

### ***Perspectives***

#### **Étapes à venir**

##### **1) Enquêtes**

1a) 2004 : fin des enquêtes transversales : réalisation de l'enquête en zone rurale, auprès d'élèves aux 5 paliers de scolarisation retenus.

1b) Enquête auprès de 2 groupes de contrôle d'adultes : a) enseignants de l'école de base ; b) adultes scolarisés jusqu'au collège étant entrés dans la vie active depuis plus de 10 ans.

##### **2) Exploitation des données**

2a) Transcription, saisie et codage des enregistrements et des données de productions écrites.

2b) Analyses des corpus obtenus selon les thèmes retenus.

2c) Comparaisons acquisitionnelles entre paliers de scolarisation, L1/L2, oral/écrit, entre données équivalentes en zone urbaine vs rurale, entre les données d'élèves et le langage des maîtres, entre les données des élèves et le corpus de contrôle d'adultes ayant été scolarisés jusqu'au collège.

#### **Résultats attendus**

##### **1) Résultats sur le plan acquisitionnel**

On vise à caractériser les étapes du développement linguistique entre 7 ans (CP1) et 15 ans (3<sup>e</sup>) en français L2 ainsi qu'en L1 à l'oral pour les grands domaines d'acquisition :

- représentation des événements, a) sémantique verbale, b) condensation des prédications en énoncés complexes

- référence aux entités (personnes et objets) et a) structure du SN ; b) pronominalisation et chaînes anaphoriques

dans des productions relevant de genres textuels déterminés : récits, descriptions, instructions de faire (procédures).

## **2) Résultats sur le plan sociolinguistique**

- Caractérisation fine du français des maîtres qui constitue le français de référence de cette situation de transmission de la francophonie, et du français des jeunes adultes ayant été scolarisés jusqu'au collège, qui forment la population pertinente pour l'évaluation de la dynamique de la francophonie dans ce secteur géographique.

- Evaluation de l'impact des deux types d'environnement linguistique (monolingue avec absence de français hors de l'école en zone rurale ; plurilingue avec présence du français hors de l'école en zone urbaine) en ce qui concerne l'appropriation de la langue de scolarisation et de construction de connaissances par les enfants.

- Mise au jour des représentations sur la langue française et sur son apprentissage qu'entretiennent les sujets cible de notre enquête.

## **3) Résultats sur le plan didactique**

- Caractérisation des pratiques pédagogiques liées à la langue dans l'enseignement du français et de l'ensemble des matières scolaires en français, en relation avec les programmes et instructions, avec les modes de transmission de compétence pédagogique, et avec les modalités d'évaluation sélective en vigueur annuellement à tous les niveaux.

- Formulation de suggestions de modifications à apporter aux pratiques pédagogiques, aux outils d'enseignement, et de contenus, pour la sensibilisation et la formation des maîtres à une didactique se fondant sur les processus d'acquisition, sur les capacités disponibles et la prise d'autonomie des apprenants, sur des objectifs globaux de capacités communicatives et cognitives et d'apprendre à apprendre.

## **4) Résultats sur le plan typologique**

Les parlers gbe de cette zone recèlent des différences typologiques cruciales avec le français dans les domaines de la sémantique verbale et la représentation des procès (notamment des constructions verbales sérielles), la référence aux entités et l'anaphore, la condensation de prédications en énoncés complexes pluri-propositionnels, qui n'ont pas toujours été décrites d'un point de vue fonctionnaliste et à partir de données textuelles. Ces domaines méritent de bénéficier au passage, à l'occasion du traitement des données en L1 des enfants et des adultes, de descriptions constituant des apports à une linguistique typologique, comme on a déjà commencé à le faire (Noyau & Takassi à paraître ; Noyau à paraître).

## **Actions prévues**

1) Communication au congrès ISSBD à Gand en juillet 2004 (C. Noyau, C. Sabatier, K. S. Gbeto « Emic definition of intelligence and a good learner in the Ewe culture »).

2) Séminaire conjoint au Togo sur les analyses comparatives à visée acquisitionnelle en vue des résultats d'ensemble (septembre 2004).

3) Colloque à Paris X en 2004-2005 : Etudes sur l'appropriation du français langue de scolarisation en Afrique et sur la construction de connaissances en français L2 (avec l'apport de chercheurs d'autres pays et sur d'autres pays d'Afrique francophone).

4) Colloque final au Togo à destination d'un large public d'enseignants (fin 2004-2005).

5) Publications :

Volume final sur les résultats acquisitionnels.

Volume final sur les implications didactiques des résultats, 2005.

Volume collectif sur les aspects typologiques des études menées : représentation des événements et structuration des énoncés complexes en langues gbe et en français, 2006.

### **Production scientifique**

- Kihlstedt M. A paraître. « Stratégies compensatoires dans l'emploi du lexique verbal en français L2 ». *Actes du colloque « Lexique : apprentissage et enseignement »*, Grenoble, mars 2003.
- Kihlstedt M., Noyau C. A paraître. « Reprise du lexique verbal dans des restitutions orales de récits – mémorisation conceptuelle et textuelle en français L2 chez des élèves en Afrique de l'ouest ». In Rispaïl M. (éd.), *Le langage oral de l'enfant scolarisé : acquisition, enseignement*. Paris : L'Harmattan.
- Noyau C. 2001. Le français de référence dans l'enseignement du français et en français au Togo ». In Francard M., Geron G., Wilmet R. (éds), *Cahiers de l'Institut de Linguistique de Louvain « Le français de référence. Constructions et appropriations d'un concept »*, 27 (1-2), 57-73.
- Noyau C. 2003. « Processus cognitifs de la construction du lexique verbal dans l'acquisition (L1 et L2). *Revue d'Intelligence Artificielle*, n° spécial « Analogie et cognition ». Paris : Hermès.
- Noyau C. A paraître. « De l'hégémonie de l'écrit et de l'écrit scolaire : point de vue critique ». In Barré-De Miniac Ch., Brissaud C. (éds), *La littérature : conceptions théoriques de la lecture et de l'écriture et pratiques d'enseignement*. Grenoble : PUG.
- Noyau C. A paraître. « Grammaticalisation et variation : constructions verbales sérielles et périphrases aspectuelles ». *Actes du colloque « Les périphrases verbales »*, Université de Caen.
- Noyau C., Bedou-Jondoh E. 2003. « Restitution de récits en langue première et en français langue seconde chez des enfants du Togo et du Bénin : formulations et reformulations des procès. In Ibrahim A., Martinot Cl. (éds), *La Reformulation : un principe universel d'acquisition*. Paris : Kimé, 357-386.
- Noyau C., Cissé A. A paraître. « L'oral et l'écrit dans la construction de connaissances via le français langue seconde à l'école ». In Bearth T. et al. (dir.), *African languages in global society / Les langues africaines à l'heure de la mondialisation*. Cologne : Koepe.
- Noyau C., Quashie M. 2003. « L'école et la classe comme environnement écologique d'acquisition du français en Afrique de l'ouest. » In Defays J.-M., Delcomminette B., Dumortier J.-L., Louis V. (éds), *Les didactiques du français, un prisme irisé*. Fernelmont : Editions modulaires Européennes, 12 p.
- Noyau C., Vellard D. A paraître. « Construction de connaissances mathématiques dans la scolarisation en français langue seconde ». *Cahiers du Français Contemporain : « Pratiques et représentations langagières dans la construction et la transmission des connaissances »*, 9, 14 p.
- Tisset C. A paraître. « L'apprentissage de la lecture au Togo. Analyse des manuels d'apprentissage du français : entre l'enseignement et l'acquisition ». *Actes du Colloque La didactique des langues face aux cultures linguistiques et éducatives*, Université Paris 3, décembre 2002, 12 p.

### **Thèse soutenue**

Ufoalè-Christine Afola. Janvier 2002. *La temporalité et le récit en ikposso, à partir d'un corpus de récits oraux*. Directeurs C. Noyau (Université Paris X) et L.-Ph. Bolouvi (Université de Lomé).

### **Thèses inscrites**

Innocent Adouna, *Le développement des énoncés complexes dans les productions orales et écrites chez les éwé apprenants du français au Togo*. Directeur C. Noyau.

Mayé-Mboul Diouf, *Le français de référence et les normes endogènes dans la parole des enseignants togolais*. Directeur C. Noyau.

Kossi Souley Gbeto, *Les discours procéduraux et explicatifs : représentation des procès et développement des énoncés complexes en français et en éwé chez des enfants scolarisés au Togo*. Directeur C. Noyau.

### **Contrat / financement**

AUF\* : ARP sur 2 ans, 1<sup>ère</sup> dotation 2002, 2<sup>ème</sup> dotation en attente.

CORUS : 1<sup>ère</sup> dotation 2001 Paris-X (2/3), 2<sup>ème</sup> dotation (1/3) en attente.

ACI Cognitive\* : 1<sup>ère</sup> dotation (3/4), 2<sup>ème</sup> dotation (1/4) en attente.

\* Les montants gérés par Paris X sont destinés en partie à l'équipe de Lomé.

### **Références générales sur le projet**

Batiana A., Prignitz G. (éds). 1998. « Francophonies africaines ». *Dyalang*. Rouen : Publications de l'Université de Rouen, 125 p.

Bearth T. (éd.). 1997. *TRANEL* « Langues et éducation en Afrique noire », 26.

Calvet L.-J., Dumont P. (éds). 1999. *L'enquête sociolinguistique*. Paris : L'Harmattan.

Chaudenson R. (éd.). 1997. *L'évaluation des compétences linguistiques en français. Le test d'Abidjan*. CIRELFA-ACCT : Didier-Érudition, 203 p.

Da Silveira Y., Hamers J. 1990. « Scolarisation et bilinguisme en contexte africain : un défi ? ». *Langage et Société* « Le français en Afrique », 52, 23-58.

Dumont P., Maurer M. 1995. *Sociolinguistique du français en Afrique francophone : gestion d'un héritage, devenir d'une science*. Vanves : EDICEF / AUPELF, 224 p.

Holtzer G. (dir.). 2002. *Recherches sur le français en Guinée*. Besançon : Presses universitaires franc-comtoises, 164 p.

Lafage S. 1985. *Français écrit et parlé en pays éwé (Sud-Togo)*. Paris : SELAF, 605 p.

Verdelhan-Bourgade M. 2002. *Le français de scolarisation. Pour une didactique réaliste*. Paris : PUF.

## **2.1.2. REPRÉSENTATION DES ÉVÉNEMENTS : SÉMANTIQUE VERBALE, TEMPORALITÉ ET DÉVELOPPEMENT DES ÉNONCÉS COMPLEXES**

### **Participants**

*Membres rattachés à titre principal au laboratoire* : Colette Noyau (responsable), Maria Kihlstedt, Danielle Leeman

*Membres associés* : Sonia Gerolimich, Jérémie Sauvage

*Jeunes chercheurs associés* : Fatiha Amouri, Jin Nam Choi, Gema Sanz

*Doctorants* : Lamia Allal, Innocent Gbandi Adouna (en cours d'inscription), Chiraz Anane, Dieynaba Diop, Kossi-Souley Gbeto, Ha Sung Park, Monique Sénémaud-Dubosq, Nora Vera, Maria Elena Villecco

*Collaborations extérieures* :

Urszula Paprocka

Ake Viberg, université de Lund (Suède)

- Equipe de Lomé du projet « Appropriation du français en situation diglossique » :

Karine Duvignau, PRESCOT Toulouse

Bruno Gaume, PRESCOT Toulouse

### **Discipline(s) et/ou sous-discipline(s) concernée(s)**

Psycholinguistique, psychologie cognitive, linguistique, acquisition des langues, bilinguisme (successif/simultané).

Sémantique lexicale et textuelle, syntaxe (des énoncés complexes), temporalité, grammaticalisation, morphologie, typologie des langues, temps et aspect, mode d'action, structuration du récit

### **Problématique**

Dans nos recherches d'orientation cognitive sur l'acquisition des langues, il s'agit de caractériser les processus de traitement du langage pour produire, pour comprendre et pour acquérir une langue : structurer ses connaissances en un système de communication linguistiquement organisé. La description linguistique des productions d'apprenants se relie à des modélisations conceptuelles adossées à la psychologie cognitive et à la psychologie du développement. L'étude des contenus sémantiques mobilisés permet de rendre compte des processus cognitifs sous-jacents chez les sujets. Nous explorons ces caractéristiques du traitement des langues en situation d'acquisition et de contact de langues (plurilinguisme émergent ou stabilisé) à partir de domaines privilégiés subsumables comme *la représentation des événements et leur placement dans le temps*. Les thèmes récurrents de ces travaux sont :

- la temporalité dans le récit (oral) ;
- la grammaticalisation de l'information temporelle ;

- le développement de la cohérence et de la connexité (relations temporelles, causales, téléologiques entre événements) ;
- la sémantique des procès (développement du lexique verbal pour les événements, granularité temporelle et sémique) ;
- le développement des énoncés complexes (condensation de l'information dans le discours).

Les trois premiers thèmes ont déjà fait l'objet d'un grand nombre de travaux et de mémoires et thèses depuis une dizaine d'années. Les travaux qui en émanent sont maintenant des bilans de recherches menées à terme ou des synthèses théoriques. Les deux derniers sont les thèmes vedettes des travaux coordonnés récemment et à venir.

Les travaux synthétisés ci-dessous sont soit des études longitudinales, soit des études transversales comparant différentes populations d'apprenants (tranches d'âge, niveau scolaire, langues diverses, langue 2 et langue 1 des mêmes locuteurs, groupes d'apprenants / de locuteurs natifs).

### **Contexte international**

Les chercheurs en acquisition de MoDyCo s'inscrivent depuis longtemps dans diverses configurations (programme européen ESF 1981-88, ATP du CNRS 1985-88, GDR CNRS 1988-2000). Ils participent à des réseaux internationaux de recherche en Europe : ESF, MSH, Max-Planck Institut für Psycholinguistik depuis les années 80, actuellement réseau européen « Dynamics of learner varieties », en Amérique Latine (réseau constitué à partir d'anciens doctorants de Paris X), dans les deux Afriques (animation de projets nord-sud). Ce qui favorise des études comparatives interlangues, incluant ainsi la dimension typologique, ce qui permet alors de voir comment les structures des langues particulières infléchissent les processus d'acquisition chez les apprenants, comment la conceptualisation des situations (« thinking for speaking », cf. Slobin 1995) contraint la construction des énoncés et des discours : la cognition tournée vers le langage est influencée par les langues (réactualisation de l'hypothèse Sapir-Whorf informée par les modèles psycholinguistiques récents du traitement du langage).

### **Contexte national**

Les configurations nationales regroupant des chercheurs en acquisition des langues, dont ceux de MoDyCo, sont liées à la dimension internationale ci-dessus. L'Université Paris X est l'un des deux grands pôles nationaux de recherche sur l'acquisition des langues non maternelles, l'autre étant l'université Paris 8. Ces deux universités ont été à l'initiative du GDR 113 « Acquisition des langues » du CNRS.

### **Thèmes de recherche**

Trois grands volets dans les études comparatives et translinguistiques menées sur la représentation des événements :

- a) le développement du lexique verbal et la sémantique des procès - prérequis pour b) ;
- b) la morphologisation temporo-aspectuelle et la mise en perspective des événements ;

c) la mise en relation entre événements et le développement des énoncés complexes dans le discours. Nous noterons dans les sections qui suivent par a), b), c) que les travaux cités portent sur l'un ou l'autre de ces trois grands thèmes.

## ***Bilan***

### **Étapes passées**

- 1) Etudes menées sur l'acquisition de L2 en situation institutionnelle, avec les mêmes méthodes fonctionnalistes et textuelles que lors des études antérieures sur le développement initial de L2s dans le milieu social par des adultes. Etapes initiales à intermédiaires (Brum, des étapes avancées aux stades quasi-natifs : Sanz, De Lorenzo).
- 2) Thèses de Brum de Paula : étude longitudinale sur le développement de la morphologie temporelle en français par des étudiants brésiliens b).
- 3) Etudes portant sur le développement précoce de L2 ou de 2 L1 chez l'enfant bilingue : thèses de Choi b), c) ; thèses en cours ayant déjà donné lieu à des présentations de travaux : Allal, Anane, Park, Sénémaud-Dubosq.

### **Résultats obtenus**

- 1) Synthèses de Noyau (2002), Kihlstedt (2002) sur les processus de morphologisation temporelle.
- 2) Synthèse collective de Noyau *et al.* (sous presse) sur la sémantique verbale et les énoncés complexes (Granularity and Condensation).

## ***Perspectives***

### **Étapes à venir**

1) Le volet a) « Sémantique verbale » étudie les processus de développement du lexique verbal du point de vue sémantique et cognitif dans une gamme de situations d'acquisition : acquisition primaire du langage en situation monolingue et plurilingue, acquisition précoce et tardive de langues étrangères, remédiation des retards de langage, et ce dans une variété de langues, ce qui nécessite de conférer à l'étude une dimension typologique (cf. Talmy 1985).

Méthodes utilisées : analyse de corpus de productions narratives expérimentales ; dans un deuxième temps, production sur supports animés numériques élaborés par le laboratoire Lordat du PRESCOT de Toulouse.

2) Le volet b) « Développement de la morphologisation temporo-aspectuelle », qui a fait l'objet du plus grand nombre d'études dans notre équipe et depuis longue date, se bornera à des écrits de synthèse si l'occasion s'en présente (volumes thématiques comme Salaberry & Shirai (éds) 2002, dont nous sommes partie prenante ou des conférences invitées). On notera en outre les études empiriques sur corpus sur la base de données togolaises

transversales en français L2 et en gbe L1 du projet « Appropriation du français en situation diglossique » (Kihlstedt en préparation ; sur la LI, Bedou & Noyau, en préparation).

3) Le volet c) « Développement des énoncés complexes dans le discours » s'appuie également en première ligne sur des corpus narratifs, et en second sur des productions orales et écrites de procédures (dire-de-faire), dans une visée acquisitionnelle : évolution entre tranches d'âge en L1, développement en L2 au fil de l'acquisition, et ce dans une visée typologique. C'est par l'analyse de productions textuelles contextualisées comparables dans diverses langues qu'on pourra mettre en lumière les tendances de conceptualisation liées à ce domaine propres aux langues particulières, et traquer les différences subtiles des apprenants très avancés, souvent qualifiés de « quasi-natifs », confrontés aux locuteurs natifs, différences occasionnées par le poids de leurs habitudes de « penser-pour-parler » dans la conception et la formulation de leurs messages en L2.

### **Résultats attendus**

Les comparaisons entre des travaux menés selon des méthodes similaires sur des corpus et des groupes d'apprenants, dans des situations déterminées contrastables permettront d'avancer dans la caractérisation des processus d'évolution des deux domaines : la sémantique verbale et les énoncés complexes.

Le domaine de la sémantique verbale est amené à jouer un rôle moteur dans la conception de l'acquisition adossée à la psychologie cognitive que nous faisons avancer. Les résultats antérieurs des études suédoises de Viberg (notamment 2002) mettant en évidence le processus universel des verbes nucléaires sur plusieurs langues européennes seront prolongés par nos études impliquant plusieurs langues extra-européennes (arabes maghrébines, coréen, langues de l'ouest africain). Par ailleurs, l'alliance avec le PRESCOT de Toulouse (projet Gaume et Duvisseau de l'ACI Cognitive « Apprentissages des Langues ») permettra de s'appuyer sur des méthodes statistiques récentes pour représenter les réseaux sémantiques entre verbes comme propriétés émergentes.

Le domaine des énoncés complexes sera travaillé avec Danielle Leeman d'une part, et, d'autre part avec les partenaires du réseau européen « Dynamics of learner varieties » travaillant sur les phénomènes des « clauses combining » dans l'acquisition, dans une perspective typologique (cf. Chini, Giacalone, Ramat, notamment).

### **Actions prévues**

1) Journée d'études internationale sur la sémantique verbale dans l'acquisition (perspectives translinguistiques) 2005, avec notamment des participants du réseau européen « Dynamics of learner varieties », du projet Togo-Bénin, d'Espagne (Jin Nam Choi, équipe de Nuria Sebastian, Barcelone), du PRESCOT de Toulouse. Parution d'un ouvrage.

2) Journée d'études sur le développement des énoncés complexes en 2007-2008. Parution d'un ouvrage.



## **Production scientifique**

- Allal L., Amouri F. A paraître. « La reformulation dans les productions en arabe et en français chez des enfants issus de l'immigration en France ». In Ibrahim A., Martinot Cl. (éds), *La Reformulation : un principe universel d'acquisition*. Paris : Kimé.
- Allal L., Anane Ch., Sénémaud M., Noyau C. 2002. « Construction des énoncés et connecteurs dans la structuration des récits enfantins en arabe tunisien et en français ». *LINX* « Les connecteurs », 46, 133-151.
- Choi J. N., Lee H.-Y. 2002. « La frontière de l'unité prédicative dans le discours narratif en coréen ». In Leroy S., Nowakowska A. (éds), *Aspects de la prédication*. Montpellier : Publications Montpellier III, 315-333.
- Kihlstedt M. 2002. « Reference to past events in dialogue. The acquisition of tense and aspect by advanced learners of French ». In Salaberry R., Shirai Y. (éds.), *The L2 Acquisition of Tense-Aspect Morphology*. Amsterdam/Philadelphia : John Benjamins, 323-362.
- Kihlstedt M. 2002. « L'emploi de l'imparfait dans le dialogue chez des locuteurs suédophones et des locuteurs natifs ». *Revue Française de Linguistique Appliquée*, II, 7-16.
- Noyau C. 2002. « Temporal relations in learner varieties : grammaticalisation and discourse construction ». In Salaberry R., Shirai Y. (éds.), *The L2 Acquisition of Tense-Aspect Morphology*. Amsterdam/Philadelphia : John Benjamins, 107-128.
- Noyau C. 2003. « Processus cognitifs de la construction du lexique verbal (L1 et L2) ». In Duvignau K., Gasquet O., Gaume B. (éds), *Revue d'Intelligence Artificielle* « Regards croisés sur l'analogie », 17 (5-6), 799-812.
- Noyau C. A paraître. « Typologie et acquisition primaire du langage / acquisition de langues secondes : le cas des constructions verbales sérielles ». *AILE (Acquisition / Interaction en Langue Etrangère)*, 19, (2).
- Noyau C., Kihlstedt M. (éds). 2002. *Revue Française de Linguistique Appliquée* « Acquisition des langues : tendances récentes », 2.
- Noyau C., de Lorenzo C., Kihlstedt M., Paprocka U., Sanz G., Schneider R. A paraître. « Two dimensions of the representation of complex event structures : granularity and condensation. Towards a typology of textual production ». In Hendriks H. (éd.), *The structure of learner language. From pragmatics to syntax : organizational principles of second language acquisition*. Berlin : De Gruyter.
- Noyau C., Takassi I. A paraître. « Catégorisation et recatégorisation : les constructions verbales sérielles et leur dynamique dans trois familles de langues du Togo ». *Verbum* « Typologie des langues et universaux linguistiques ».

## **Thèses soutenues**

- Fatiha Amouri. Directeur C. Noyau.
- Jin Nam Choi. Directeur C. Noyau.
- Sonia Gerolimich. Directeur C. Noyau.
- Cristina de Lorenzo. Directeur C. Noyau.
- Urszula Paprocka. Directeurs C. Noyau, M. Kashmarkowski.
- Silvia Romeo. Directeur C. Noyau.
- Gema Sanz Espinar. Directeurs C. Noyau, M. Tordesillas.
- Ricarda Schneider. Directeur C. Noyau.

### **Thèses inscrites**

Innocent Adouna, *Le développement des énoncés complexes dans les productions orales et écrites chez les éwé apprenants du français au Togo*. Directeur C. Noyau.

Lamia Allal. Directeur C. Noyau.

Chiraz Anane. Directeur C. Noyau.

Dieynaba Diop. Directeur C. Noyau.

Mayé-Mboul Diouf, *Le français de référence et les normes endogènes dans la parole des enseignants togolais*. Directeur C. Noyau.

Kossi Souley Gbeto, *Les discours procéduraux et explicatifs : représentation des procès et développement des énoncés complexes en français et en éwé chez des enfants scolarisés au Togo*. Directeur C. Noyau.

Ha Sung Park. Directeur C. Noyau.

Monique Sénémaud-Dubosq. Directeur C. Noyau.

Nora Vera. Directeur C. Noyau.

Maria Elena Villecco. Directeur C. Noyau.

### **Références générales sur le projet**

Berman R., Slobin D. 1994. *Different ways of relating events in narrative : a cross-linguistic developmental study*. Hillsdale : Lawrence Erlbaum.

Duvignau K. A paraître. « De la métaphore analogique à l'acquisition précoce du lexique : l'analogie pour l'apprentissage des verbes ». *Actes du colloque « Early Lexicon Acquisition (ELA 01) »*.

Ibrahim A., Martinot Cl. (éds). 2003. *La reformulation, un principe universel d'acquisition*. Paris : Kimé.

Hickmann M. 1998. *Déterminants fonctionnels de l'acquisition du langage : études comparatives de l'organisation discursive chez l'enfant*. Habilitation à diriger des recherches, Université de Paris V.

Levelt W. J. 1989. *Speaking : From intention to articulation*. Cambridge (Mass) - Londres : Bradford Books MIT Press.

Stutterheim C. von, Nüse R., Murcia Serra J. 2002. « Cross-linguistic differences in the conceptualisation of events ». *Revue Française de Linguistique Appliquée « Acquisition des langues : tendances récentes »*, 2, 89-105.

Viberg A. 1999. « Age-related and L2-related features in bilingual narrative development ». In Verhoeven L. Strömquist S. (éds), *Narrative development in a multilingual context*. Amsterdam : John Benjamins, 87-128.

### 2.1.3. APPRENTISSAGE / ENSEIGNEMENT PRÉCOCE DES LANGUES ÉTRANGÈRES

#### Nouveau projet

#### Participants

*Membres rattachés à titre principal au laboratoire* : Maria Kihlstedt (responsable), Jean-Jacques Briu, François Muller, Colette Noyau, Sergueï Sakhno

*Membre associé* :

Cristina de Lorenzo (Centre international des Langues, UFR Langues, Université de Nantes)

*Jeune chercheur associé* : Jin Nam Choi

*Doctorants* : Lamia Allal, Chiraz Anane

*Collaborations extérieures* :

Marlène Dolitzky (IUFM de Paris)

Gisela Håkansson (Dépt de linguistique, Université de Lund, Suède)

Maya Hickmann (Laboratoire Cognition et Développement, Université René Descartes)

Catherine d'Humières (IUFM de Versailles)

Claudine Quatreuille (Inspecteur, Académie de Versailles)

Suzanne Schlyter (Institut d'Etudes Romanes, Université de Lund, Suède)

#### Problématique

Ce projet est né d'une interrogation de partenaires extérieurs sur le bien-fondé du système mis en place tout récemment et généralisé à la rentrée 2002 pour l'enseignement des langues étrangères à l'école primaire. Il est de nature exploratoire et se fera en deux étapes :

1) bilan des environnements acquisitionnels offerts par différents dispositifs pour enseigner une première langue étrangère au primaire : immersion partielle via l'enseignement de certaines matières ou activités dans la langue étrangère / enseignement « traditionnel » de langue étrangère / éveil aux langues, etc.

2) mise en relation de ces options avec le potentiel d'acquisition linguistique reconnu pour ces âges par l'examen des études psycholinguistiques disponibles et, dans un deuxième temps, par l'élaboration des études expérimentales dans les écoles étudiées dans la première étape, permettant de voir l'impact des différents dispositifs sur le développement linguistique en langue étrangère chez les élèves.

Le projet s'effectuera dans un cadre européen (collaboration avec un projet en cours sur l'apprentissage/l'enseignement des langues et la compétence plurilingue des jeunes enfants en situation d'immersion scolaire, mené par Suzanne Schlyter et Gisela Håkansson)

Grâce à la participation de Maria Kihlstedt dans un des projets effectués dans le cadre de l'action Ecole et Sciences Cognitives (Action Cognitive), financée et lancée par le

ministère de la recherche entre 2000-2001, le projet s'inscrira dans la ligne directe des résultats récemment obtenus sur l'apprentissage des langues dans le contexte scolaire français. Il constitue une suite logique au projet *Grammaticalisation et bilinguisme chez l'enfant et l'adulte - aspects didactiques et dysfonctionnements* (sous la direction de Maya Hickmann, directrice de recherche au laboratoire Cognition et Développement, CNRS UMR 8605). Pour l'aspect plus expérimental, le projet pourra profiter des stimuli informatisés adaptés aux enfants déjà élaborés et testés dans le cadre du projet Cognitique.

### **Thèmes de recherche**

Le but principal du projet est d'examiner comment sont exploitées les capacités acquisitionnelles du jeune enfant dans l'enseignement des langues vivantes à l'école primaire. Il est à noter que les études en acquisition en contexte francophone ont majoritairement concerné l'acquisition du français par des apprenants étrangers, alors que l'acquisition des langues étrangères par des élèves français est un domaine peu exploré, à l'exception des quelques études pilotes pionnières (par ex. Love 1997).

De nature exploratoire, le projet vise à répondre aux questions générales suivantes : quel niveau de maîtrise en langue étrangère est-il possible/réaliste d'obtenir en contexte scolaire ? Quel rôle joue la méthode adoptée à cet égard ? Pourquoi certaines structures ne sont-elles jamais acquises ? Pourquoi, au contraire, certaines sont-elles acquises rapidement et/ou sans enseignement explicite ?

Les questions seront examinées à la lumière des études antérieures sur différentes situations d'apprentissage scolaire des jeunes enfants :

- études de la capacité narrative (cf. par ex. les travaux de Amouri, Garat, Paprocka) ;
- études sur les rythmes d'acquisition en fonction du contexte (cf. par ex. les travaux de Allal, Choi, Gerolimich) ;
- études sur l'enseignement par immersion (cf. par ex. Dalgalian 2000 ; Genesee 1994 ; Hagège 1996 ; Petit 2001).

### **Étapes à venir**

1) Janvier-juin 2004 : bilan des environnements acquisitionnels offerts par différents dispositifs pour enseigner une première langue étrangère au primaire (immersion partielle via l'enseignement de certaines matières ou activités dans la LE / enseignement « traditionnel » de langue étrangère / éveil aux langues), notamment l'examen des programmes officiels (réalisé avec Cl. Quatreuille).

Cette phase exploratoire débouchera sur une phase plus expérimentale :

2) Septembre 2004-juin 2005 : examen des niveaux de maîtrise obtenus dans la langue étrangère à différentes étapes du développement linguistique (après un, deux ou trois ans d'enseignement). Il s'agira d'enregistrer des groupes d'élèves dans différentes tranches d'âge selon certains paramètres préétablis, permettant d'évaluer aussi bien leurs capacités perceptives (compréhension) que leurs capacités productives dans diverses tâches. Les résultats de ces analyses seront mis en relation avec le contenu de l'enseignement ainsi qu'avec les méthodes adoptées, à condition que ces dernières se différencient suffisamment systématiquement entre elles pendant le temps de l'observation.

## **Résultats attendus**

En fonction des conclusions obtenues dans la deuxième étape, nous espérons pouvoir :

- 1) proposer des ajustements dans les enseignements afin d'améliorer/optimiser le potentiel d'acquisition des élèves dans le contexte donné ;
- 2) favoriser les échanges entre chercheurs et didacticiens concernés par l'enseignement des langues vivantes à l'école ;
- 3) construire des outils pour l'enseignement des langues vivantes et proposer des outils de diagnostic, prévention et remédiation pour les professeurs des langues.

## **Actions prévues**

- 1) Séminaire : « La situation actuelle de l'enseignement des langues vivantes ». Claudine Quatreuille (Académie de Versailles). Prévu pour le printemps 2004
- 2) Participation au colloque international « Les enjeux d'une sensibilisation très précoce aux langues étrangères en milieu institutionnel (crèche, école maternelle, deux premières années de l'école primaire) », 2004, Centre International des Langues, Université de Nantes.
- 3) Organisation de journées de formation continue pour professeurs des écoles ; colloque final, 2005-2006.

## **Références générales sur le projet**

- Bailly D., Luc C. 1992. *Approche d'une langue étrangère à l'école*. Vol. 2 « Etude psycholinguistique et aspects didactiques ». Paris : INRP.
- Comblain A., Rondal J. 2001. *Apprendre les langues ; Où, quand, comment ?* Liège : Pierre Mardaga.
- Dalgalian G. 2000. *Enfances plurilingues. Témoignage pour une éducation bilingue et plurilingue*. Paris : L'Harmattan.
- Genesee F. 1994. *Educating second language children, the whole child, the whole curriculum, the whole community*. Cambridge : Cambridge University Press.
- Groux D. 1996. *L'enseignement précoce des langues : des enjeux à la pratique*. Paris : Chronique sociale.
- Hagège C. 1996. *L'enfant aux deux langues*. Paris : Odile Jacob.
- Kihlstedt M. 2001 « La recherche en acquisition des langues secondes et l'enseignement – quels rapports ? ». *Moderna Språk*, numéro spécial, 62-71.
- Lietti A. 1994. *Pour une éducation bilingue. Guide de survie à l'usage des petits Européens*. Paris : Payot / Rivages.
- Love R. 1997. *Le développement de la capacité communicative en anglais chez des enfants du CE2 en classe d'initiation à l'anglais*. Mémoire de maîtrise de sciences du langage, Université Paris X.
- Luc C. 1992. *Approche d'une langue étrangère à l'école*. Vol. 1 : « Perspectives sur l'apprentissage ». Paris : INRP.
- Petit J. 2001. *L'immersion, une révolution*. Colmar : J.-D. Bentzinger Editeur.

## **2.2. DIDACTIQUE**

### **2.2.1. DIDACTIQUE DE LA SYNTAXE**

#### **Nouveau projet**

#### **Participants**

*Membres rattachés à titre principal* : Danielle Leeman (responsable), Marie-Laure Elalouf, Carole Tisset

*Membres associés* :

Pierre Patrick Haillet (UMR Métadif, dir. Jean Pruvost)

Serge Meleuc (Université Paris 10)

Gérard Petit (UMR RES-SYLED Paris 3)

*Doctorants* : Badreddine Hamma, Sophie Hamon, Ichraf Khammari, Belinda Lavieu, Houda Ounis, Madona Sakhokia, Céline Vaguer

*Collaborations extérieures* :

Volker Fuchs (Université de Greifswald)

#### **Discipline(s) et/ou sous-discipline(s) concernée(s)**

Syntaxe, morphologie, sémantique, lexicque : domaine du verbe

#### **Problématique**

La didactique s'est constituée comme domaine à part entière dans l'investigation linguistique au cours des années quatre-vingt, et n'est plus conçue comme une simple application (vulgarisée), à l'enseignement, des résultats de la recherche théorique : des travaux précis ont montré, en particulier, d'une part que l'enseignement du français comme langue étrangère s'adresse à un public spécifique auquel l'on ne peut simplement servir les mêmes contenus (ou les mêmes exercices) qu'au public dont la langue maternelle est le français ; d'autre part que, dans les deux cas, les éventuels résultats de la recherche théorique ne sont pas également pertinents dans l'élaboration de méthodes d'apprentissage ; enfin, qu'il faut nécessairement tenir compte, dans les propositions pédagogiques, des savoirs, de l'expérience et des conditions concrètes de travail des enseignants, des objectifs ministériels (instructions et programmes), des possibilités cognitives et des motivations des élèves. Le projet « Didactique de la syntaxe » est axé sur le premier degré et a pour objectif de réfléchir à des propositions concernant le verbe.

Le problème est de savoir jusqu'à quel point on peut intégrer les recherches linguistiques dans la pédagogie de la langue, d'une part, et à travers quelles transpositions didactiques, d'autre part. On peut en effet considérer qu'il est essentiel, dans l'entreprise de développement de la compétence des élèves, de livrer des règles et descriptions que l'on

suppose exactes, ou en tous cas plus appropriées que les règles et descriptions dites « traditionnelles » (par exemple on sait que les critères ontologiques, de type « Le verbe exprime l'action accomplie par le sujet », ne sont ni généraux, ni spécifiques, et par conséquent scientifiquement contestables). Ou bien on peut considérer que, compte tenu du mode de constitution des connaissances chez les enfants (qui peut varier selon l'âge, le niveau, et les compétences acquises), il est plus efficace de procéder par approximations successives et cumulatives, même si les descriptions et règles que l'on procure à tel ou tel moment de la scolarité ne sont pas des plus adéquates à l'aune scientifique (par exemple la question *Qui est-ce qui / Qu'est-ce qui*, pour révéler la fonction sujet, n'est pas générale - ne s'appliquant pas, entre autres, au sujet impersonnel dans *Il est 5 heures* - mais n'en reste pas moins opératoire dans le cas de l'accord : un élève qui a écrit *Les oiseaux chante* peut comprendre par ce biais qu'il faut mettre le verbe au pluriel). La réflexion est donc double : d'une part établir la doxa contemporaine telle qu'elle ressort des recherches purement scientifiques, d'autre part voir ce qui peut en être retenu, dans les cadres didactiques et pédagogiques.

La collaboration de Serge Meleuc avec Volker Fuchs (Université de Greifswald, Mecklembourg-Poméranie) dans le cadre d'une action de formation permanente destinée aux futurs enseignants de français (Universités de Berlin, Postdam, Leipzig, Iéna, Erfurt, Halle et Greifswald), a permis d'établir des liens et de comparer les hypothèses entre Allemands et Français concernant l'enseignement / apprentissage des langues (maternelles, secondes, étrangères). Il en va de même de Gérard Petit, coordinateur scientifique de l'action n° P1.2003RR302 dans le cadre du 6<sup>ème</sup> appel à collaboration du Réseau LTT (Lexicologie, Terminologie, Traduction), qui associe Français, Tunisiens et Grecs. Ajoutons que Gérard Petit, en tant qu'il participe à la conception et à la réalisation du *DAFLES (Dictionnaire d'Apprentissage du Français Langue Etrangère et Seconde)*, a des contacts réguliers avec l'équipe d'enseignants-chercheurs de Leuven (Belgique).

Le projet ici défendu constitue une exception - pour ne pas dire une forme de provocation - dans le contexte national. La plupart des didacticiens et pédagogues, comme la plupart des responsables institutionnels, considèrent en effet que l'heure n'est plus à l'enseignement/apprentissage de la « grammaire » (entendue comme « grammaire de phrase », autrement dit « syntaxe »), tout au plus utile dans la mesure où elle permet de mettre l'orthographe, l'axe essentiel étant celui du discours et du texte (cf. les programmes officiels publiés par le MENRT au CNDP, 1999 et 2002 par exemple). Nous considérons au contraire que la réflexion sur la phrase constitue un préalable à la compréhension et à la production des textes et que de surcroît le travail grammatical représente l'un des moyens d'apprendre à raisonner - à condition toutefois que la démarche pédagogique ne consiste pas à imposer des règles toutes faites susceptibles d'être contestables, ni à réduire les exercices à la simple répétition ou application de ce qui a été appris par cœur dans la leçon.

### **Thèmes de recherche**

Les actuels textes ministériels fixent à l'Ecole élémentaire deux thématiques en matière de langue : le nom et le verbe ; le projet étudiera le domaine verbal. Il s'agit de mettre au jour l'ensemble des propriétés avancées en linguistique pour caractériser le verbe sous ses

différentes facettes, et de voir comment procéder à leur transposition didactique. Le projet est trop récemment défini pour qu'on puisse se risquer à produire des hypothèses pour l'instant.

Il existe de multiples travaux, dont il s'agit d'opérer le bilan critique et d'étudier l'adaptation possible dans les classes, concernant d'une part les propriétés morphologiques des verbes, le classement de ces derniers et les règles qui concernent leurs flexions ; concernant d'autre part leurs propriétés syntaxiques, distributionnelles et sémantiques, avec en particulier la quadripartition qui s'opère entre verbes distributionnels (dits « à sens plein »), verbes supports (sortes d'auxiliaires construits avec des noms prédicatifs et en permettant le regroupement en classes sémantiques), verbes semi-auxiliaires (suivis de l'infinitif, et véhiculant des valeurs temporelles, aspectuelles ou modales), et verbes auxiliaires ; concernant enfin l'identité en langue et les valeurs en emploi des modes (en particulier subjonctif vs indicatif) et des temps (présent, futur, conditionnel, imparfait, passé simple, etc.).

## ***Perspectives***

### **Actions prévues**

Journée d'études « Didactique de la syntaxe », 29 janvier 2004, Université Paris X. Contributions de P. P. Haillet, B. Hamma, S. Hamon & M. Sakhokia, I. Khammari, B. Lavieu, S. Meleuc, G. Petit, C. Tisset, C. Vaguer

### **Thèse inscrite**

Madona Sakhokia, *Identité et variation du verbe voir*. Directeur D. Leeman.

### **Références générales sur le projet**

- Boilleau O., Tisset C. 2003. « *Les Fleurs du Mal* » : éléments de mise en perspective grammaticale et stylistique. Paris : Atlande.
- Curat H. 1991. *Morphologie verbale et référence temporelle en français moderne*. Genève-Paris : Droz.
- Gosselin L. 1996. *Sémantique de la temporalité en français*. Louvain-la-Neuve : Duculot.
- Haillet P.-P. 2002. *Le conditionnel en français : une approche polyphonique*. Paris : Ophrys, 172 p.
- Lavieu B. A paraître. « Les résultats de la recherche en linguistique sont-ils utilisables en didactique de la grammaire ? », *Actes du colloque Langue et étude de la langue*, IUFM Marseille-Canebière, 3-5 juin.
- Le Goffic P. 1997. *Les formes conjuguées du verbe français*. Paris : Ophrys.
- Leeman D. 2001. « Pourquoi ne peut-on combiner *si* et le conditionnel ? ». In Dendale P., Tasmowski L. (dir.) *Le conditionnel, Recherches linguistiques de Metz*, 25.
- Leeman D. 2002. *La phrase complexe : les subordinations*. Bruxelles : de Boeck, 176 p.
- Leeman D. 2003. « Identification d'un nouveau semi-auxiliaire ». *Actes du colloque « Périphrases verbales »*, Caen, 25-27 juin 2003.



- Leeman D. 2003. « Le passé simple et son co-texte : examen de quelques distributions ». *Langue française*, 138.
- Leeman D. 2004. *Syntaxe et sémantique des déterminants du français*. Paris : Hachette, 210 p.
- Meleuc S., Fuchs V. 2003. *Linguistique française pour le français langue étrangère*, vol. I : *La communication en français*. Berne : Peter Lang, 220 p.
- Petit G. et al. (éds). 2001. « Linguistique de la dénomination ». *Cahiers de Praxématique*, 36.
- Pinchon J., Couté B. 1981. *Le système verbal du français*. Paris : Nathan.
- Soutet O. 2000. *Le subjonctif en français*. Paris : Ophrys.
- Tisset C. 2001. « Attendus, sous-entendus, mal-entendus dans le discours des enseignants », *Actes du Colloque international « La Médiation, marquages en langue et en discours »*, IRED, Université de Rouen.
- Tisset C. 2002. « Les énoncés a-typiques dans *Les voyageurs de l'impériale* d'Aragon ». *L'Information grammaticale*, 93, mars 2002.
- Touratier C. 1996. *Le système verbal français*. Paris : Colin.
- Vetters C. (éd.). 1993. *Le temps, de la phrase au texte*. Lille : Presses universitaires de Lille.

NB : un grand nombre de numéros de revues diverses sont tout ou partie consacrés aux verbes, aspects, modes, temps (*Chronos, Langages, Langue française, Le Gré des langues, Syntaxe & Sémantique*, etc.) : nous n'en énumérons pas ici toutes les références.

## 2.2.2. ELABORATION D'UN GLOSSAIRE DE TERMES LINGUISTIQUES

### **Participants**

*Membres rattachés à titre principal* : Jean-Jacques Briu (responsable), Anne Trévisse (responsable), Marie-Laure Elalouf, Danièle Flament-Boistrancourt, François Muller, Rémy Porquier, Sergueï Sakhno, Carole Tisset

*Membres associés* :

Maria Jarrega

Wilfrid Rotgé (Groupe de recherche CREA, Institut d'Etudes anglo-américaines, Université de Paris X).

*Doctorants* : Dominique Declercq, Agnès Muller

### **Discipline(s) et/ou sous-discipline(s) concernée(s)**

Linguistique, didactique, acquisition et apprentissage des langues (maternelles et étrangères)

### **Problématique**

Descriptions théoriques et explications pour l'enseignement utilisent nécessairement une métalangue. Or on connaît les problèmes d'hétérogénéité des descriptions dans les deux domaines (*LINX*, 36 et 37, 1997). Les terminologies trouvées dans les manuels ou dans les Instructions Officielles pour le français langue maternelle ou pour les langues étrangères en témoignent. Les approximations ou les flottements que l'on y rencontre au niveau des terminologies de description sont à l'origine de beaucoup des difficultés métalinguistiques des apprenants scolarisés en langues étrangères en France.

Il semble qu'à l'heure actuelle il manque une harmonisation globale des termes de métalangue employés dans l'enseignement des langues étrangères. Les apprenants se voient donc confrontés à des termes différents suivant qu'on leur enseigne le français langue maternelle, l'anglais, l'allemand ou le russe, par exemple. La prise de conscience des difficultés éprouvées de ce fait par les enseignants et les apprenants a donné naissance à ce projet qui interroge du coup la ou les théories linguistiques et leurs appareils métalinguistiques (Gobert 2002 ; Combettes 1998 ; Zemb 1981 ; Vassant 1979 ; Martin 1999).

Le projet s'articule autour de la constitution progressive d'un glossaire (partiel) de métatermes linguistiques de description du langage et des langues, adapté donc à la description de plusieurs langues, dont le russe, l'allemand, l'anglais et le français (Zemb 1968 ; Simon 1998).

Les premiers termes définis seront, par exemple, situation, contexte, énonciation, coordonnées énonciatives, repérage, deixis, embrayeur, sujet, anaphore, pronom, coréférence, préposition, syntaxe « pure » ou « mécaniste », hypotaxe, parataxe, subordination, complément, discours rapportés. Seront également abordées, parallèlement,

ou comme développements particuliers, certaines réalisations matérielles de notions dites plus « sémantiques » comme hypothétique, irréel, potentiel, cause, concession, etc.

Les objectifs sont pluriels :

- remises en cause de certaines catégories et étiquettes traditionnelles ;
- établissement de plusieurs approches théoriques ;
- définition de plusieurs niveaux au nombre de trois au moins : recherche, formation des enseignants, et didactisation (pour les apprenants de différents niveaux).

L'exhaustivité n'est pas un des objectifs.

Le projet est conçu en deux étapes :

- durant les deux premières années, rédaction des points déjà travaillés par l'équipe et addition de quelques nouvelles entrées ;
- durant les deux années suivantes, rédaction des notes concernant la didactisation.

Chaque entrée de glossaire se conforme à un gabarit identique qui se compose de 6 rubriques :

1. *Titre : nom du terme et définition courante (modalisée)*
2. *Corrélat en français (ie : renvois à d'autres entrées du glossaire)*
3. *Histoire en quelques lignes : rapide histoire du terme en linguistique*
4. *Niveau didactique (connaissances stabilisées)*

Si l'on constate que les fonctionnements sont identiques dans les 4 langues : exemples dans les 4 langues. Cette rubrique comporte des manipulations permettant d'identifier le fait de langue étudié, une définition (plus rigoureuse que la définition courante), une mention des spécificités, une typologie, des emplois en textes et en discours. S'y ajoutent certaines spécificités des diverses langues étudiées.

5. *Niveau recherche*

Cette rubrique expose quelques pistes de recherche actuelles, des approches différentes ou évoque des questions non tranchées.

6. *Bibliographie*

Quelques titres des auteurs qui ont marqué l'évolution de la recherche sur le thème évoqué par le titre de l'entrée du glossaire.

### **Contexte international**

Dans les pays francophones, une réflexion sur la terminologie grammaticale et liaison avec l'évolution de la didactique du français a conduit à un renouvellement institutionnalisé : rénovation de l'enseignement en Suisse romande (Genevay 1993 ; Martin 1999 ; Béguelin 2000), refonte des programmes canadiens (Chartrand *et al.* 1996 ; Chartrand *et al.* 1999) ; nouveaux programmes et terminologie grammaticale de 1997 en France (Legrand 1998 ; Tomassone 1996). Pour autant, une confrontation des choix opérés dans des contextes nationaux différents reste à faire (Elalouf 1998, 1999). Certains travaux de recherche pointent les difficultés, mais il reste beaucoup à faire dans le seul domaine francophone (Branca *et al.* 1995).

## **Contexte national**

En France, peu d'équipes réunissent des linguistes didacticiens de plusieurs langues autour des questions d'acquisitions métalinguistiques et rares sont celles qui allient la perspective de l'enseignement de la langue nationale à celle d'une langue étrangère. Il y a donc lieu de confronter les travaux faits dans des secteurs différents : pour l'enseignement du français (Combettes 1998 ; Garcia-Debanc *et al.* 2001), pour l'enseignement des langues étrangères, rarement réunis dans des synthèses (Colombat & Savelli 2001).

En dehors des membres de cette équipe, un certain nombre de chercheurs ont travaillé dans le domaine de la terminologie, pour la théorie et pour la didactisation (Branca *et al.* 1995, Willens 1995 ; Zemb 1972, 1981 ; Neveu 2000 ; Combettes 1998 notamment), mais un travail portant sur la description et l'enseignement de plusieurs langues, maternelle et étrangères, n'a jamais été poursuivi systématiquement.

## **Thèmes de recherche**

Les chercheurs travaillant à ce projet sont des linguistes de différentes langues (français langue maternelle, et langues dites « étrangères » : français langue étrangère, anglais, allemand, russe et chinois). Ce sont aussi des didacticiens de ces langues, c'est-à-dire qu'ils s'intéressent à l'articulation entre la description théorique des langues et l'enseignement de ces mêmes langues (« savoir enseigné »).

La réunion de chercheurs en langues différentes implique un volet comparatiste ou interlinguistique autant qu'un volet de linguistique générale, en termes d'invariants sémantico-cognitifs notamment (cf. notamment à propos des préoccupations évoquées dans Anscombe 1992 ; Bronckart 1996 ; Zemb 1968 ; Tyl 1970 ; Rey-Debove 1998 ; Wierzbicka 1996 ; Lerat 1999 ; Masson 1999 ; Sweetser 1990).

Nos recherches ont toujours deux volets liés :

- 1) une réflexion sur les instructions officielles et la métalangue officielle de description des subordonnées ;
- 2) qui mène à une réflexion théorique, puis ultérieurement et de façon assez fonctionnelle, à une tentative de clarification et d'harmonisation des métalangues pour l'enseignement du français langue maternelle et des langues étrangères.

## ***Bilan***

### **Étapes passées**

A la suite des travaux menés dans le champ sémantico-syntaxique de l'hypothétique (cf. LINX, 41, 1999), notre recherche s'articule toujours autour de la problématique de la métalangue de description et d'enseignement des langues. Les membres du groupe représentent les langues étudiées et travaillent en commun pour l'établissement d'une métalangue commune, parfois difficile à déterminer étant donné les cultures différentes en ce domaine entre anglicistes par exemple ou germanistes ou francisants (langue maternelle et français langue étrangère). Les discussions font se confronter les diverses prises de

position théoriques, et les font expliciter afin de les exposer clairement dans un tel glossaire.

La bibliographie des membres du projet montre les centres d'intérêt divers et convergents des différents participants : cadres théoriques différents, langues différentes (allemand (Muller et Briu), anglais (Trévisé), français (Elalouf et Tisset), FLE (Flament), russe (Sakhno)), mais souci de didactisation de repères métalinguistiques solides pour les futurs enseignants et pour les apprenants. Tous les membres du projet sont à la fois des chercheurs en linguistique et des didacticiens.

### **Résultats obtenus**

Deux entrées du glossaire sont presque achevées. Il s'agit de : pronom (avec un historique de la notion et des comparaisons interlinguistiques), étymologie.

D'autres entrées sont actuellement en cours d'élaboration : thème/rhème, anaphore, deixis, nom, phrase complexe, sujet.

### **Actions réalisées**

1) Participation au séminaire du laboratoire Modyco le vendredi 13 juin 2003.

2) Présentation de deux entrées du glossaire (travail en cours) : Jean-Jacques Briu : pronom ; Sergueï Sakhno : étymologie

3) R. Porquier et E. Rosen ont proposé une série régulière de réunions de travail et de rencontres ayant pour thème les situations de communication et d'acquisition en langue étrangère, de communication exolingue, d'acquisition des langues et de contextes et situations d'apprentissage / enseignement des langues, avec le triple objectif : d'établir et développer des échanges et débats entre chercheurs ; d'alimenter leurs activités de recherche, individuelles et collectives ; de discuter et préparer des projets de communication et de publication.

En 2002-2003, l'activité a porté :

- Sur l'acquisition et l'apprentissage des langues et les interactions exolingues dans des contextes diversifiés.
- Sur la méthodologie de recueil de données dans des contextes diversifiés d'appropriation / enseignement des langues.
- Sur les dimensions didactiques (tâches, activités, exercices) de l'enseignement / apprentissage des langues.
- Sur la mise au point notionnelle et terminologique de repères-outils.

### ***Perspectives***

#### **Étapes à venir**

Continuer sur les prochaines rubriques prévues : thème/rhème, sujet et deixis pour 2003-2004. Chaque rubrique nécessite de longues discussions pour intégrer les particularités des langues dans une problématique commune.

### **Actions prévues**

- 1) Participation au séminaire de Modyco le 2 avril 2004 : Maria Jarrega (titre non précisé) ; Danièle Boistrancourt « Thème / rhème = présentation d'une entrée du glossaire ».
- 2) Journée d'étude prévue le 11 juin 2004 à l'Université de Paris X « Approches de l'opposition thème / rhème ». Des personnalités extérieures à l'équipe seront invitées à présenter leur travail. Des membres de l'équipe participeront également.
- 3) Publication du numéro 49 de la revue *LINX*, à paraître en janvier 2004, rassemblant les travaux du séminaire (Porquier, Rosen), années 2000 à 2003.

### **Production scientifique**

- Boilleau O., Tisset C. 2003. « *Les Fleurs du Mal* » : éléments de mise en perspective grammaticale et stylistique. Paris : Atlande.
- Briu J.-J. 2000 et 2001. « Les groupes syntaxiques en allemand, un cours polycopié inédit de Jean Fourquet (1966, Paris, Sorbonne) ». *Histoire Epistémologie Langage*, tome XXII, 2, 2000 : 133-164 et tome XXIII, 1, 2001 : 153-182.
- Elalouf M.-L. A paraître. « Constitution d'un grand corpus de textes d'élèves : problèmes méthodologiques et premiers résultats ». *LINX*, Actes de la journée « Quand les pratiques scolaires interrogent les théories et modèles de l'écriture ».
- Elalouf M.-L., Keraven J. 2003. « Ecrits exploratoires, écrits à contraintes. Quelle complémentarité en classe de 5<sup>ème</sup> ? ». *Littératures et langages, Ecritures*, 1, 149-456.
- Elalouf M.-L., Keraven J. A paraître. « L'acquisition du lexique à l'épreuve d'un grand corpus d'élèves ». In Calaque E., David J. (éds.), *Didactique du lexique : contextes, démarches, supports*. Bruxelles : De Boeck.
- Muller F. 2003. « Le pronom et le groupe relatif ». *Hommage à Nicole Fernandez-Bravo*. Paris : L'Harmattan.
- Porquier R., Py B. 2003. *L'apprentissage des langues étrangères : contextes et discours*. Paris : Didier. Collection « Essais ».
- Sakhno S., Chicouène M. 2001 (Réédition 2002). *Parlons russe : une nouvelle approche*. Paris : L'Harmattan, 333.
- Tisset C. 2003. « Le contre-exemple, un outil didactique ». Colloque « L'exception, entre le théorique et l'empirique linguistiques ». Paris, 10-12 juin 2003.
- Trévisé A. 2003. « A propos des relations inter-énoncés : parataxe, hypotaxe, asyndète et construction du sens. Tentatives de clarification ». *Enoncé complexe et énonciation*. Toulouse : Presses Universitaires du Mirail.
- Trévisé A. 2003. « Les représentations du moment de l'énonciation : hétérogénéités, ambivalences et reconstruction du sens ». *Anglophonia*, 14.

### **Thèses soutenues**

- Evelyne Rosen. 2001. Directeur R. Porquier.
- Sébastien Piotrowski. 2002. Directeur R. Porquier et Kaczmarkowski (Université de Lublin).

### **Thèses inscrites**

Encarnacion Arroyo. Directeur R. Porquier.

François Champion. Directeur R. Porquier.

Dominique Declercq, *Expression et représentation de la subjectivité dans des textes de fiction en anglais*. Directeur A. Trévisé.

Agnès Muller, *Une approche de l'ordre des mots et de la thématization en anglais*. Directeur A. Trévisé.

### **Références générales sur le projet**

Anscombe J.-Cl. 1992. « Imparfait et passé composé : des forts en thème/propos ». *L'information grammaticale*, 55, 43-53.

Béguelin M.-J. 2000. *De la phrase aux énoncés : grammaire scolaire et descriptions linguistiques*. Bruxelles : De Boeck-Duculot

Branca S., Piquer A., Willens D. 1995. « Vers une terminologie grammaticale européenne ? Nécessités et obstacles ». *Travaux de linguistique*, 31.

Bronckart J.-P. 1996. *Activité langagière, textes et discours. Pour un interactionisme socio-discursif*. Lausanne : Delachaux & Niestlé.

Chartrand S. 1996. *Pour un nouvel enseignement du français*. Montréal : Editions logiques.

Chartrand S. 1999. *Grammaire pédagogique du français*. Montréal : Graficor.

Colombat B., Savelli M. 2001. « Métalangage et terminologie linguistique ». *Actes du colloque international de Grenoble Orbis / Supplementa 17*. 2 Vols. XLIX-1086. Leuven : Peeters.

Combettes B. 1983. *Pour une grammaire textuelle : la progression thématique*. Bruxelles-Paris : Danes.

Combettes B. 1998. « Analyse critique de la nouvelle terminologie grammaticale des collèges et des lycées ». *Pratiques*, 97-98.

Elalouf M.-L. 1998a. « De la nomenclature de 1975 à la terminologie de 1997, évolutions et problèmes dans le choix du métalangage grammatical ». *Le Français aujourd'hui*, 124, 115-121.

Elalouf M.-L. 1998b. « De la nomenclature de 1975 à la terminologie de 1997, évolutions et problèmes dans le choix du métalangage grammatical ». *Le Français aujourd'hui*, 125, 122-128.

Elalouf M.-L. 1999. « De la maîtrise de la langue à la maîtrise des discours en France et dans la francophonie ». *Pratiques*, 101-102, 193-204.

Garcia-Debanc C., Confais J.-P. Grandaty M. (éds). 2001. *Quelles grammaires enseigner à l'école et au collège ? Discours, genres, textes, phrase*. Paris : Delagrave-CRDP Midi-Pyrénées.

Genevay E. 1993. *Ouvrir la grammaire*. Lausanne : LEP

Gobert F. 2002. « La dénomination étymologie populaire ou l'utopie d'une terminologie non ambiguë ». *Cahiers de lexicologie*, 81 (2), 5-37.

Legrand G. 1998. *Pour l'enseignement de la grammaire*. Arras : IUFM et CRDP du Nord pas de calais.

Lerat P. 1999. « L'offre en sémantique lexicale ». *Cahiers de lexicologie*, 75 (2), 5-22.

Martin D. 1999. « La terminologie à l'école : facilitateur ou obstacle aux apprentissages ? L'exemple de la suite du verbe ». *Travaux neuchâtelois de linguistique*, 31, 13-35.

- Masson M. 1999. *Matériaux pour l'étude des parallèles sémantiques*. Paris : Presses de la Sorbonne Nouvelle.
- Neveu F. 2000. *Lexique des notions linguistiques*. Paris : Nathan. Collection « 128 ».
- Rey-Debove J. 1998. *La linguistique du signe : Une approche sémiotique du langage*. Paris : Colin.
- Simon J.-P. 1998. « Réflexion sur la terminologie grammaticale en primaire ». *Callimaque*, 4.
- Sweetser E. 1990. *From Etymology to Pragmatics*. Cambridge : Cambridge University Press.
- Tomassone R. 1996. (Rééd. et augmenté 2002). *Pour enseigner la grammaire*. Paris : Delagrave.
- Tyl Z. 1970. *A tentative bibliography of studies in functional sentence perspective (1900-1970)*. Prag.
- Vassant A. 1979. « A propos de la nomenclature grammaticale. Quelle grammaire enseigner ? ». *L'information grammaticale*.
- Wierzbicka A. 1996. *Semantics : Primes and Universals*. Oxford et New York : Oxford University Press.
- Zemb J.-M. 1968. *Les structures logiques de la proposition allemande*. Paris : OCDL.
- Zemb J.-M. 1972. *Métagrammaire. La proposition*. Paris : OCDL.
- Zemb J.-M. 1981. « De quelques précautions terminologiques en matière de thème, de rhème et de phème ». *Cahiers du CISL*, 3, 7-49.



## 2.3. LA VARIATION STYLISTIQUE

### **Participants**

*Membres rattachés à titre principal* : Françoise Gadet (responsable), Danièle Flament-Boistrancourt, Christine Pauleau

*Jeune chercheur associé* : Cécile Vigouroux

*Doctorants* : Tae-Rin Cho, Hasna Ghoul, Emmanuelle Guérin, Harry Tyne

*Collaboration extérieure* :

Marie-Louise Moreau (Université de Mons, Belgique)

### **Discipline(s) et/ou sous-discipline(s) concernée(s)**

Acquisition, didactique, linguistique des textes, TAL, syntaxe, morphologie

### **Problématique**

La sociolinguistique est, dans notre perspective, conçue comme un certain point de vue, orienté vers la pratique du locuteur, sur des phénomènes relevant des champs de la phonologie, de la morphologie, de la syntaxe, du lexique, du discours. On conduit ainsi une réflexion sur les objets de la linguistique générale (la sociolinguistique est-elle neutre par rapport aux modélisations offertes ? quels sont les modèles les plus adéquats, et pourquoi ?), et sur les concepts spécifiques de la sociolinguistique (variation, variété, variétés, variabilité, changement ; répertoire verbal, langue/dialecte/créole, niveau de langue, standard et standardisation, norme, etc.).

Un des aspects de notre réflexion sur la variation concerne le « style ». Celui-ci est souvent réduit dans sa prise en considération dans le champ linguistique et grammatical français aux seuls « niveaux de langue » ou « effets de style », alors qu'il apparaît comme un phénomène jouant un rôle majeur dans la compréhension de la dynamique des langues. C'est au regard de la complémentarité d'approches différentes (constitution de corpus, diachronie, syntaxe du français parlé, traduction, acquisition et didactique des langues étrangères - en particulier FLE -, alternance codique, bilinguisme...) que sera étudié, en confrontant des linguistes d'horizons différents (sociolinguistes, didacticiens, syntacticiens, phonologues, diachroniciens), le rôle de ce phénomène. On tentera de définir le terme même de « style » dans son utilisation sociolinguistique.

### **Contexte**

La variété et l'étendue du champ s'occupant du style peuvent être appréhendées à partir des bibliographies extensives de Biber et Finegan (1994), Spillner (1987), Cheshire (1992), Gadet (1998), Sandig et Selting (1997) qui prennent en compte les travaux européens sur la question. Cependant les chercheurs travaillant sur le style ont encore à se mettre d'accord

sur le nom à donner à ce type de variation. Les sociolinguistes ont fréquemment essayé de définir le style par rapport à d'autres termes utilisés pour désigner des variétés de langue tels que *registre* (Dittmar 1995 ; Finegan & Biber 1994 ; Irvine 2001 ; Halliday 1978 qui travaille l'opposition style vs dialecte ; *genre*, qui suppose une stabilité sociale ou une institutionnalisation (Ferguson 1994) ; ou encore *répertoire* qui prend en compte tous les moyens linguistiques disponibles aux membres d'une communauté donnée.

L'examen de la littérature portant sur le style montre que cette dimension de la variation ne semble pas intéresser les linguistes français. Peu ou pas de travaux existent sur la question.

### **Thèmes de recherche**

Si l'on considère le « style » comme principe organisateur d'une réflexion sur la langue, il constitue une entrée possible pour la compréhension de différents phénomènes concernant le locuteur :

- 1) Pourquoi les locuteurs ne parlent-ils pas toujours de la même façon ?
- 2) Pourquoi y a-t-il diversification des façons de parler, plutôt qu'unité ?
- 3) Quels sont les facteurs qui participent à la variation stylistique ?

### ***Bilan***

### **Résultats obtenus**

Nous avons essayé de faire un point sur l'ampleur de ce que l'on peut saisir à travers la notion de style et les différentes orientations des usages de ce mot, en sociolinguistique et au-delà. Il a résulté de cette recherche plusieurs articles (Gadet 2002, 2003, à paraître ; Tyne 2002, 2003 ; Vigouroux 2001) ainsi qu'un ouvrage de vulgarisation à destination des étudiants (Gadet 2003). Celui-ci discute la question de la variation sociale à partir d'exemples tirés du français.

### **Actions réalisées**

1) Dans le cadre du projet Franco6 dirigé par Marie-Louise Moreau, une première série d'enquêtes de terrain a été réalisée, en décembre 2002, auprès d'étudiants de cycle supérieur. Nous nous sommes plus particulièrement intéressés à évaluer les variations stylistiques chez ces locuteurs natifs de français résidant dans la région parisienne. Cette enquête fait partie d'une vaste recherche menée dans plusieurs zones de l'espace francophone. Elle comporte une dimension comparative (Belgique, France, Québec, Sénégal, Suisse, Tunisie).

2) Le projet sur le style s'est concrétisé par la tenue d'un colloque *Le style comme point de vue sur la dynamique des langues*, le 25-26 avril 2003, organisé par Françoise Gadet et Harry Tyne. Il a réuni des sociolinguistes, des didacticiens, des syntacticiens, des phonologues et des diachroniciens.

## ***Perspectives***

### **Actions prévues**

1) La deuxième étape du projet Franco6 est en cours et devrait être achevée en janvier 2004. Il s'agit de poursuivre les enquêtes de terrain menées auprès des locuteurs natifs de français et diversifier l'échantillon de la population. Nous allons nous intéresser, dans cette seconde phase du travail à un groupe de locuteurs de faible niveau de scolarisation, exerçant prioritairement une activité d'apprenti.

2) Un numéro spécial de *Langage et société*, va paraître en juin 2004 : *Le style comme point de vue sur la dynamique des langues*, coordonné par Françoise Gadet et Henry Tyne. La variation stylistique sera envisagée à la fois chez des locuteurs natifs de français et chez des locuteurs de français langue seconde. Les thèmes développés recouvriront des perspectives différentes : acquisitionnelle, didactique et sociolinguistique.

### **Production scientifique**

Flament-Boistrancourt D. 2003. « L'activité questionnante chez natifs et non-natifs : deux styles différents pour des enjeux différents ». Colloque « Le style comme point de vue sur la dynamique des langues », Université Paris X, 25-26 avril 2003.

Gadet F. 2003. *La variation sociale en français*. Paris : Ophrys, 138 p.

Gadet F., Tabouret-Keller A. (éds). 2003. « Sociolinguistics in France : Theoretical Trends at the Turn of the Century ». *International Journal of the Sociology of Language*, 160.

Gadet F. A paraître. « Sociolinguistic research on style ». In Ammon U., Dittmar N., Mattheier K., Trudgill P. (éds), *Handbuch der Soziolinguistik / Handbook of Sociolinguistics*.

Gadet F. A paraître. « French in France ». In Ammon U., Dittmar N., Mattheier K., Trudgill P. (éds), *Handbuch der Soziolinguistik / Handbook of Sociolinguistics*.

Gadet F. A paraître. « Migrant languages in France ». In Ammon U. (éd.), *Encyclopedia of European Languages*.

Tyne H. A paraître. « Variation stylistique en L2 vue par le chercheur, vécue par le locuteur ». Danemark : Presses de l'Université de Roskilde.

Tyne H. 2003. « Accommodation en français L2 ». Communication au colloque de l'association AFLS, Université de Tours.

Tyne H. 2003. « L'appropriation du style chez des apprenants britanniques avancés ». Colloque « Le style comme point de vue sur la dynamique des langues », Université Paris X, 25-26 avril 2003.

Vigouroux C. 2001. « Je suis étranger en anglais c'est pas ma langue ». *Le français en Afrique*, 15, 9-28.

### **Thèse soutenue**

Cécile Vigouroux. 2003. *Réflexion méthodologique autour de la construction d'un objet de recherche : la dynamique identitaire des migrants africains francophones en Afrique du Sud*. Directeur F. Gadet.

### **Thèses inscrites**

Cho Tae-Rin Cho. Directeur F. Gadet.

Hasna Ghoul. Directeur F. Gadet.

Emmanuelle Guérin. Directeur F. Gadet.

Harry Tyne. *L'appropriation du style chez les apprenants britanniques avancés du français*. Directeur F. Gadet.

### **Références générales sur le projet**

Biber D., Finegan E. 1994. *Sociolinguistic Perspectives on Registers*. New York - Oxford : Oxford University Press.

Cheshire J. 1992. « Register and Style ». In Bright W. (éd.), *International Encyclopedia of linguistics*. New York - Oxford : Oxford University Press, 324-7.

Dittmar N. 1995. « Register ». In Verschueren J., Ostman J., Blommaert J. (éds), *Handbook of Pragmatics*. Amsterdam/Philadelphia : John Benjamins Publishing Company, 1-16.

Ferguson Ch. 1994. « Dialect, Register and Genre : Working Assumptions about Conventionalization ». In Biber D., Finegan E. (éds), *Sociolinguistic Perspectives on Registers*. New York - Oxford : Oxford University Press.

Gadet F. 1998. « Cette dimension de la variation que l'on ne sait nommer ». *Sociolinguistica*, 12, 53-71.

Halliday M. 1978. *Language as social Semiotics*. Londres : University Park Press.

Irvine J. 2001. « « Style » as Distinctiveness : The Culture and Ideology of Linguistic Differentiation ». In Eckert P., Rickford R. (éds), *Style and Sociolinguistic Variation*. Cambridge : Cambridge University Press.

Sandig B., Selting M. 1997. « Discourse styles ». In Van Dijk T. (éd.), *Discourse as Structure and Process*. Londres, 138-156.

Spillner B. 1987. « Style and Register ». In Ammon U, Dittmar N, Mattheier K. (éds), *Sociolinguistics/Soziolinguistik*. Berlin - New York : de Gruyter.

## 3. Textes / Corpus

---

- 1) Genres et linguistique de corpus
- 2) Communication électronique
- 3) Ecrit et Ecriture scolaire
- 4) Edition scientifique de Charles Nodier
- 5) Métrique de l'alexandrin au tournant des XIXe et XXe siècles

### 3.1. GENRES ET LINGUISTIQUE DE CORPUS

#### **Participants**

*Membres rattachés à titre principal au laboratoire* : François Rastier (responsable), Denise Malrieu (responsable), Catherine Boré, Simon Bouquet, Evelyne Bourion

*Membre associé* :

Driss Ablali (Université de Besançon)

*Doctorants* : Thomas Beauvisage, Smail Djaoud, Sylvain Loiseau, Céline Poudat, Rossitza Kyheng

*Collaborations extérieures nationales* :

Bénédicte Pincemin (LLI, Villetaneuse)

Mathieu Valette (INALCO)

CORAL, LIFO, Université d'Orléans

*Collaborations extérieures internationales* :

Jean-Michel Baudouin (Sciences de l'Education, Université de Genève)

Kjersti Fløttum (Université de Bergen, dépt d'études romanes et KIAP project, [www.hit.uib.no/kiap](http://www.hit.uib.no/kiap) : Cultural Identity in Academic Prose : national versus discipline-specific, Dépt Romance Studies, Université de Bergen, avec le concours du Research Council of Norway)

Ludovic Tanguy (Toulouse)

#### **Discipline(s) et/ou sous-discipline(s) concernée(s)**

Linguistique textuelle, sémantique textuelle, linguistique de corpus

## Problématique

Ce projet concerne l'exploration des apports d'une linguistique de corpus prenant en compte les genres et types de textes. Les objectifs d'analyse privilégiés dans ce projet sont les suivantes : caractérisation des variations morpho-syntaxiques selon certains domaines et genres (voir les corpus impliqués), selon les parties ou séquences textuelles à l'intérieur d'un genre, comparaison des genres à l'intérieur de l'œuvre d'un auteur, comparaison intra-genre et entre auteurs, analyse des différentes modalités de la narrativité, apports d'une prise en compte du genre à la désambiguïsation syntaxique et sémantique de morphèmes ou lexèmes (en particulier analyse comparée des valeurs de temps x personnes, de certains modaux, des marques de la polyphonie selon les genres).

Applications : profilage de textes, extraction de concepts, typologie narrative.

## Thèmes de recherche

Nous partons du constat qu'une langue n'est pas « une et unique », elle est constituée par l'ensemble des textes qui l'attestent. Reprenant la démarche de la linguistique de corpus qui veut induire les normes des usages, nous nous appuyons pour ce faire sur la notion de genre : comme tout système sémiotique, la langue s'élabore et évolue au sein de pratiques sociales différenciées à l'intérieur de domaines et de champs de pratiques. Le genre constitue le lieu de normativité privilégié. Il permet une observation raisonnée des variations linguistiques prenant en compte le contexte discursif. Notre projet concerne donc l'exploration des apports d'une linguistique de corpus prenant en compte les genres. De plus nous nous restreignons dans ce projet à certains genres écrits. Par rapport à la démarche de Biber (1988, 2000), nous adoptons des catégories d'analyse proches des siennes, centrées non sur le lexique mais sur des variables morpho-syntaxiques ; nous nous différencions sur des points de méthode : sa définition des genres est discutable dans son fonctionnalisme ; ses analyses différentielles ne portaient que sur des extraits de textes. Or notre point de vue est que l'unité d'analyse est le texte attesté, à prendre donc dans son intégralité. Et corrélativement il ne prenait pas en compte la structure textuelle, éminemment définitoire du genre.

Les objectifs d'analyse privilégiés par les membres du projet sont multiples : caractérisation des variations morpho-syntaxiques selon les domaines, les champs génériques, les genres, les sous-genres, selon les parties ou séquences textuelles à l'intérieur d'un genre, caractérisation du style d'auteur, comparaison des genres intra-auteur, comparaison intra-genre et inter-auteurs, analyse des différentes modalités de la narrativité, comparaison interculturelle des modes d'allocution, apports d'une prise en compte du genre à la désambiguïsation syntaxique et sémantique de morphèmes ou lexèmes.

Les recherches s'effectuent à différents niveaux d'observation des phénomènes et d'affinement des concepts théoriques ; les outils nécessaires (existants ou à développer au cours du projet) varient également selon ces paliers d'analyse. Elles se regroupent sous trois actions :

### **1) L'Action *Typologie des textes*** (S. Bouquet, E. Bourion, G. Hazem, S. Djaoud, C. Boré)

Les travaux d'approfondissement théorique ont commencé par :

- une critique des théories récusant la typologie des textes ou la déclarant impossible ;

- la proposition de catégories descriptives (les composantes textuelles : thématique, dialectique, dialogique, tactique) ;
- la (re)définition en termes linguistiques d'unités textuelles restées intuitives (les thèmes, notamment) ainsi que de méthodologies de construction de ces thèmes.

Nous distinguons trois niveaux : les *discours* (ex. juridique vs littéraire vs scientifique), les *champs génériques* (ex. théâtre, poésie, genres narratifs), les *genres* proprement dits (ex. comédie, roman « sérieux », roman policier, nouvelle, conte, récit de voyage). Les *sous-genres* (ex. roman par lettres) constituent un niveau encore subordonné.

La sémantique des textes numériques allie les savoirs traditionnels des « arts et sciences des textes » (philologie, herméneutique, en particulier) à des méthodes récentes, permises par les NTIC. Elle étend les principes du comparatisme aux discours, genres et textes, tout d'abord en perspective unilingue. Les domaines choisis relèvent des Lettres et Sciences humaines. Ils l'ont été notamment pour les liens interdisciplinaires que leur étude permet de nouer avec les disciplines concernées, la philosophie par exemple et pour les débouchés didactiques des recherches (ex. enseignement du français ; cf. corpus Didactique).

Les travaux déjà conduits ont donné lieu à divers contrats, notamment :

- Semindex, sur l'indexation des documents multimédia (avec l'EnstB) 1998-2001.
  - Un autre projet utilisant la même méthodologie a porté sur l'indexation de séquences audiovisuelles (Télécom-Rennes, 2002).
  - Princip, contrat européen sur la caractérisation et détection automatique de sites racistes (2002-2004 ; consortium Inalco, Lip6, Université de Dublin, Université de Magdebourg).
- D'autre part, un colloque *Sémantique et archéologie* a été organisé avec l'Ecole française d'Athènes (nov. 2001, actes sous presse).

## **2) L'Action Unités textuelles** (E. Bourion, F. Rastier, R. Kyheng, M. Valette, S. Loiseau, G. Hazem, S. Djaoud)

Selon nos options théoriques, on ne peut définir d'unités textuelles qu'au sein d'un genre déterminé. Par unités textuelles, on entend aussi bien des unités de l'expression, comme le paragraphe, que des unités du contenu, comme les thèmes. On cherchera naturellement à les corrélérer, en gardant à l'esprit que les deux plans du langage ne se correspondent pas terme à terme, et notamment que les unités sémantiques sont diffuses.

Ces recherches sont menées dans des corpus de discours et genres variés, où pourront être repérées des particularités alimentant aussi bien l'approfondissement des aspects théoriques présentés ci-dessus, que la réflexion menée dans les autres actions.

### **2a) TextIpsé** (E. Bourion, J.-M. Baudouin)

Il s'agit de contraster successivement la construction du *Je* dans les pages personnelles autobiographiques, avec :

- La construction du protagoniste comme sujet dans cinq romans d'apprentissage du XIXe.
- La construction du *Je* dans le dialogue littéraire (dialogue romanesque, dialogue théâtral, roman par lettre).
- La construction du *Je* dans des récits de vie.

On se propose de donner une caractérisation linguistique et sémiotique aussi précise que possible des différences et analogies entre la construction du personnage et celle de la personne textualisée. Outre les perspectives psychologiques et sociologiques, la construction du *Je* dans des corpus cliniques est envisagée ultérieurement. Une coopération

avec l'Université de Genève (département de Sciences de l'Education) est en cours, sur la définition et la numérisation du corpus de récits de vie.

**2b) *Thèmes théoriques*** (F. Rastier, R. Kyheng, S. Loiseau, M. Valette, S. Djaoud, G. Hazem, L. Tanguy)

Les méthodes de la thématique assistée sur corpus numériques peuvent s'appliquer aux discours théoriques, qu'ils soient scientifiques ou philosophiques. Comment les concepts se lexicalisent-ils ? Leur lexicalisation s'accompagne-t-elle d'un appauvrissement de leur richesse contextuelle ? L'étude diachronique et comparative (dans diverses disciplines : philosophie, linguistique, sociologie, politique) tient compte des différences notables des « concepts » (répartitions et contextes) selon les genres (article, traité, essai, etc.) et les discours, y compris à l'intérieur d'œuvres singulières (ex. Guillaume et Deleuze). Les retombées de cette recherche intéressent l'extraction assistée de concepts, la représentation des connaissances, et permettent de cerner certaines difficultés repérées dans les alignements de corpus.

Nous entendons comparer-conjuguer la méthode d'extraction de candidats termes par patrons constructionnels et la méthode des corrélats statistiques par écart réduit, cela dans un corpus d'articles de linguistique et un corpus de textes de philosophie. Un financement a été demandé à l'ILF.

**3) L'Action *Méthodologie de corpus et les genres*** (D. Malrieu, S. Loiseau, B. Pincemin)

Depuis trois ans, nous avons développé une méthodologie qui vise à intégrer la qualification d'un corpus, d'un texte de ce corpus dans sa structure (utilisation d'XML et de la TEI), et l'exploration des variations tant morpho-syntaxiques que lexicales en fonction de ces différents niveaux de localité. Ce projet implique que se développent de concert une méthodologie de balisage intégré des différentes sémiotiques du texte et une définition des variables morpho-syntaxiques et sémantiques intéressant nos objectifs, qui seront constituées à partir des sorties d'analyseurs et autres outils.

**3a) Méthodologie**

**3a1) Méthodologie de balisage** (S. Loiseau)

Notre démarche consiste à constituer des corpus de textes raisonnés quant aux critères de domaines, champs génériques, genres, auteurs, périodes. Par opposition à la banque textuelle qui engrangerait des textes non qualifiés, le corpus permet de maîtriser les critères de comparaison, dans la mesure où la méthodologie portée par le balisage XML permet de qualifier et commenter tout niveau de la hiérarchie de l'arborescence, du corpus au morphème. Il ne s'agit plus d'utiliser une banque de textes pour vérifier si telle forme est attestée ou pas, mais de mesurer les variations de cette forme selon le niveau de l'arborescence.

Le balisage est devenu incontournable dans les linguistiques de corpus, et plus généralement dans le traitement des données culturelles. Outre les gains pratiques en termes de pérennité et d'interopérabilité, l'enjeu scientifique porte notamment sur la capacité à décrire des singularités, et à unifier la description de systèmes sémiotiques :

1) L'ajout d'annotation sous forme d'encodage structuré et non typé renouvelle les moyens de traiter le rapport entre singularité et type : l'encodage permet à la fois de décrire avec un niveau de détail potentiellement infinis les singularités du texte (les occurrences), et de rapprocher tous les éléments de l'encodage dans des procédures de calcul et de



quantification. Face à cette capacité de description et de calcul, la question de la stratégie interprétative et le rapport critique aux méthodologies mises en œuvre reviennent au premier plan.

2) Représentation unifiée de systèmes sémiotiques hétérogènes : la plasticité descriptive de l'encodage, permet de marquer de façon uniformisée des systèmes sémiotiques hétérogènes. Par exemple, un corpus note le plus souvent la structure logique du texte (articulation en parties). Il peut également baliser un dépouillement morpho-syntaxique, ajouter des informations prosodiques, phonétiques, sémantiques (thématiques, isotopies, etc.). C'est donc en ce sens un format d'unification, proposant non seulement une standardisation industrielle mais aussi une rencontre entre différents modèles descriptifs. Les rapprochements rendus possibles, notamment dans la corrélation entre systèmes sémiotiques sont encore largement à explorer.

Ces deux aspects rendent XML particulièrement pertinent pour un projet de profilage de textes, indissociable d'une stratégie de qualification des corrélations de variables génériques.

Une première étape consiste à baliser la structure hiérarchique du texte et les segments graphiquement marqués. Pour ce faire il faut définir une typologie des marquages (qui connaissent des variantes selon les genres, les éditions et les périodes) et développer des routines de balisage automatique adaptées à cette typologie. Une collaboration avec le LIFO d'Orléans permettra les développements concernant l'automatisation d'un balisage minimal, goulot d'étranglement actuel dans la mise en œuvre d'une linguistique de corpus. Une seconde étape concerne la génération de variables qui qualifient le texte et la définition de leur mode de représentation en vue de leur intégration dans la représentation XML/TEI du texte. Une troisième étape concerne l'intégration de l'étiquetage de l'analyseur et des statistiques sur les variables dans le corpus.

### **3a2) La définition des variables morpho-syntaxiques pertinentes (D. Malrieu, B. Pincemin)**

Dans le travail effectué depuis deux ans sur une analyse morpho-syntaxique différentielle des genres, nous sommes partis de l'observation d'un corpus de 2600 textes intégraux classés en genres enrichis par Cordial Analyseur de la société Synapse-Développement. Cette approche statistique de grandes masses textuelles, doit permettre un repérage de traits stylistiques contrastés selon les genres, à travers la mise en évidence de faisceaux de variables corrélées. Les résultats obtenus (Malrieu & Rastier 2001 ; Malrieu 2002) par traitement de 250 variables morpho-syntaxiques fournies par Synapse montraient de fortes différences entre domaines, champs génériques et certains genres et nous ont confirmés dans l'intérêt d'approfondir l'approche morpho-syntaxique des genres.

Cependant les résultats obtenus et l'examen des variables nous ont amenés aux conclusions suivantes :

- le traitement statistique de textes non balisés peut entraîner des biais statistiques en cascades ;
- l'absence d'un parcours méthodique des dimensions de variation fait que des catégories sont redondantes, d'autres manquantes : la conception des variables est donc à revoir. Notre choix est donc de nous rendre autonomes par rapport aux variables statistiques fournies par Cordial et de parser les sorties étiquetées. Cette redéfinition est en chantier sur trois fronts :

1) Repérer dans les sorties de l'analyseur les étiquettes non fiables (déterminants définis/indéfinis, erreurs sur certains temps verbaux etc.); goulot d'étranglement : dépendance / état actuel des analyseurs ; confrontation des performances de différents analyseurs (C. Poudat) ; éviter les biais statistiques inhérents au calcul actuel de certaines variables (Malrieu 2001). Un descriptif des erreurs repérées dans les sorties de l'analyseur a été défini en format XML.

2) Redéfinir des variables comme objets linguistiques (regroupements discutables de catégories comme la catégorie « mots-outils » ; absence de certaines catégories (types de relatives, d'infinitives, forme passive, formes impersonnelles, types de modaux, de verbes (modaux, locutoires, croyance, jugement, croisement systématique des paradigmes temps x personnes etc.)). Cette redéfinition est conçue comme essentiellement évolutive pour servir les besoins ponctuels de l'analyse (cf. ci-dessous). Environ trois cents variables soit existantes soit dérivables des sorties de l'analyseur ont été pour l'instant définies.

3) Définir non seulement les variables pertinentes mais aussi leur mode de représentation formelle pour faciliter leur traitement en aval et leur manipulation dans la description du corpus. Quant aux outils statistiques, si l'ACP, l'analyse factorielle et la CHA se sont avérées opératoires sur des corpus très contrastés, elles semblent moins opératoires sur des corpus plus homogènes pour générer des classes interprétables et un travail parallèle sur les méthodes de clustering s'avère nécessaire (C. Poudat en collaboration avec le LIFO à Orléans).

### **3b) L'approche qualitative : comment articuler les niveaux de contraintes pour la désambiguïsation ?**

Si l'approche purement quantitative peut aider à la classification des textes, et si les variations de corrélations de certains morphèmes selon les domaines et genres peuvent être des indices de valeurs sémantiques divergentes, leur spécification passe par une analyse des contraintes portées par le genre à des paliers inférieurs du texte. Il paraît donc utile d'associer à une approche purement quantitative sur les textes globaux, une approche qui prend en compte les spécificités du genre dans son mode énonciatif et dans la structure du texte.

Ce travail se fonde sur une analyse morpho-syntaxique comparée des parties du texte pour les genres étudiés :

a) Articles de linguistique : analyse comparée des textes intégraux, des exemples, des notes, des textes sans notes et sans exemples et sans bibliographie.

b) Essais de philosophie : comparaison selon les sous-genres des textes intégraux, citations, exemples, des notes, des textes sans citations, sans exemples et sans notes

c) Roman : comparaison des parties récit du narrateur selon le type de narrateur, des types de discours rapportés (dialogues, discours indirect, monologue intérieur, etc.), du psychorécit (discours du narrateur sur la vie intérieure d'un personnage), descriptions focalisées.

Cette première étape doit aider à analyser les modes de contraintes interprétatives portées par le genre à un niveau local : dans un premier temps nous explorerons les variables discriminantes sur les différents types de parties du texte (philosophie, linguistique) ou de discours rapportés (dans le roman) ; cela permettra de décrire de façon plus fine les formes de dialogisme ou de polyphonie, selon les domaines, les genres ; cela permettra ainsi d'avancer dans la qualification linguistique de familles narratoriales à l'intérieur du genre

romanesque ; pour ce faire nous analyserons plus particulièrement certaines variables : temps x personnes, pronoms personnels, verbes modaux, négation, évaluation, certains types d'adverbes (analyse des fréquences, des corrélations) ; ce qui doit nous amener dans un second temps à mettre en évidence des valeurs sémantiques différentes de ces morphèmes et grammèmes selon les parties ou types de séquences du texte-genre.

### **Corpus impliqués dans le projet :**

- Corpus Philosophie (Sylvain Loiseau) : Deleuze, Macherey, autres philosophes du XXe, numérisé.
- Corpus Linguistique (Céline Poudat) : 250 articles de linguistique francophones, corpus comparable en cours de constitution sur des revues de linguistique anglophones, numérisé.
- Corpus Romans d'apprentissage et récits de vie (Evelyne Bourion), partiellement numérisé.
- Corpus Nouvelles (Rossitza Kyheng) : 360 nouvelles en français, bulgare, russe numérisées ; corpus à enrichir.
- Corpus Duras (Denise Malrieu) : 60 œuvres numérisées.
- Corpus Didactique (Catherine Boré) : cours et exercices d'écoliers.
- Corpus Sociologie (G. Hazem).
- Corpus Histoire (S. Djaoud).
- Corpus Critique littéraire (D. Ablali).

## ***Bilan***

### **Résultats obtenus**

#### **1) Recherches en méthodologie d'analyse thématique de gros corpus**

Les travaux effectués dans les années précédentes ont permis d'élaborer une méthodologie d'analyse thématique fondée sur une théorie sémantique et sémiotique des textes, contrastant corpus de travail et corpus de référence en utilisant des tests statistiques (en particulier le logiciel Hyperbase d'E. Brunet), qui permettent de mettre en évidence la variation des thématiques selon les genres textuels dans le champ littéraire et les corrélations entre thématiques et formes de l'expression.

#### **2) Recherches sur les liens entre typologie des textes et variables linguistiques**

2a) La collaboration mise en place de façon informelle avec la société Synapse a permis des échanges (évaluation logicielle *vs* fourniture de données statistiques sur gros corpus à partir des sorties de l'analyseur Cordial). Ces données (2 541 ouvrages) ont fait l'objet d'une reclassification des textes en domaines, champs génériques et genres, qui a permis d'effectuer une analyse comparative selon cette dimension sur l'ensemble des variables morpho-syntaxiques (250 environ) disponibles dans Cordial-Analyseur.

2b) Des analyses statistiques univariées et multivariées ont permis de caractériser les différences morpho-syntaxiques entre domaines, champs génériques et certains genres (cf. Malrieu & Rastier 2001 ; Beauvisage 2001). Denise Malrieu (2001) a mis en évidence, sur les mêmes corpus, une évolution diachronique de la fréquence de l'adjectif antéposé, ainsi qu'une variation forte selon les genres textuels. Le travail sur les « pronoms de dialogue » dans le théâtre classique pose la question de l'ambiguïté sémantique des grammèmes selon

les genres textuels, discute les indices stylistiques de Muller et met en évidence des différences pronoms de dialogue / nominalisations entre comédie et tragédie, et à l'intérieur de la comédie entre comédie en prose et en vers (Malrieu 2002).

Un travail plus approfondi sur l'analyse des corrélations entre temps et personne a ainsi été mené, qui montre bien des systèmes d'opposition des temps différents dans deux corpus contrastés (genres narratifs *vs* articles de linguistique).

2c) Ce travail s'est poursuivi en prenant en compte les familles narratologiques dans les genres narratifs. Nous avons montré que les statistiques morpho-syntaxiques issues de textes narratifs balisés différencient de façon significative le discours direct représenté du récit (le poids des dialogues dans le roman est source importante de variance importante à l'intérieur d'un genre ; l'examen des temps et personnes montre que les différences sont rapportables aux prédications thétiqes liées au dialogue sur l'état du monde). Si certaines des intuitions de Benveniste sont ainsi vérifiées (concernant le discours direct représenté, qui n'est pas assimilable cependant à sa notion de discours), il semble que le poids des temps dans le discours direct représenté soit sensible au temps dominant de la narration qui l'englobe, ainsi qu'aux différences de champ générique (roman *vs* comédie) : ce qui confirme la contrainte du local par le global et le fait que le texte n'est pas la juxtaposition de types discursifs prédéfinis

Les travaux en cours poursuivent ces analyses contrastives des genres et sous-genres en mettant en place la méthodologie de balisage décrite ci-dessus.

### **Actions réalisées**

#### **1) Numérisation de nombreux corpus**

#### **2) Animation d'un atelier de formation à la méthodologie de corpus** (S. Loiseau, C. Poudat, D. Malrieu).

Initié au printemps 2003 et ouvert à l'ensemble du laboratoire, cet atelier vise à aider ses membres à définir une méthodologie concertée de constitution et de balisage de corpus, à réfléchir ensemble sur les méthodes statistiques adéquates et à s'outiller pour un traitement informatique des corpus balisés et étiquetés. Il comporte trois modules (formation à XML, XSLT et à la TEI ; statistiques descriptives et multivariées ; initiation à la programmation Java).

### ***Perspectives***

#### **Étapes à venir**

1) La numérisation de textes est maintenant bien avancée grâce à un financement sur un contrat géré par F. Rastier ; cependant la collaboration d'autres ressources existantes pour la constitution de corpus mieux raisonnés concernant le roman s'avère nécessaire ; de plus un gros travail de normalisation des textes numérisés avant balisage XML reste à effectuer.

2) Développement de procédures de balisage semi-automatique des genres textuels dans leur structure et leurs séquences marquées.

Sur les deux premiers points, une collaboration avec l'ATILF semble possible et est en cours d'examen.

3) La définition d'un vaste jeu de variables morpho-syntaxiques liées aux phénomènes linguistiques explorés est d'ores et déjà disponible ; certaines variables requièrent l'établissement de listes de lexèmes (verbes modaux, de jugement, de perception, verbes d'incises de dire, adverbess modaux, etc.) qu'il reste à élaborer ; enfin, la définition de routines de génération des variables et des modalités de calculs à partir des sorties d'analyseur comme la représentation XML (mode de calcul, attributs formels) des variables sont en cours.

### **Résultats attendus**

- 1) Bilan sur les variables discriminantes, en particulier les variables temps personnes, à partir des différentes actions menées (TextIpsé, corpus traités dans l'action méthodologies de corpus et genre notamment).
- 2) Identification des thèmes et lexicalisation des concepts mises en relation avec les genres.
- 3) Descriptions de corpus : caractérisation sémantique du fonctionnement des concepts philosophiques (thèse S. Loiseau) ; étude des variations interculturelles du discours académique (thèse C. Poudat) ; caractérisation du genre romanesque (variables pertinentes, familles narratoriales, identification des thèmes, etc.).

### **Actions prévues**

- 1) Animation d'un atelier de formation à la méthodologie de corpus (S. Loiseau, C. Poudat, D. Malrieu). Poursuite de l'atelier initié au printemps 2003.
- 2) Mise sur pied d'un site dédié à l'échange de compétences en matière de linguistique de corpus. Ce site doit :
  - fournir des informations sur les outils disponibles sur les formats de fichiers, modes d'encodage, compatibilité d'outils, sites utiles ;
  - mettre à disposition et discuter des exemples de balisage, traitement et interprétation des résultats sur des corpus balisés.

Ces lieux d'échanges d'expériences et de discussions nous semblent importants tant pour l'aspect didactique (en direction des étudiants en littérature, linguistique ou plus généralement en sciences humaines) que pour l'élaboration collective de normes de philologie électronique.

### **Production scientifique**

- Assadi, H., Beauvisage T., Bourion E., de Charentay F., Cloarec T., Lupovici C., Micault L., Savary J.-P., Ziemlicki C. 2003. *Parcours des usagers dans les bibliothèques électroniques*. Rapport du projet BibUsages. France Télécom R&D, BnF, MoDyCo.
- Bouquet S. (éd.). 2004. *Langages* « Linguistique des genres », 152.
- Bourion E. 2002. « Doxa et écarts numériques dans un corpus électronique de romans d'apprentissage ». *Actes du colloque de Chambéry* « Littérature et Linguistique », 14-16 novembre 2002.
- Fløttum K., Rastier F. (éds). *Academic Discourse. Multidisciplinary Approaches*. Oslo : Novus.
- Malrieu D. 2002. « Stylistique et statistique textuelle. À partir d'un article de C. Muller sur les pronoms de personnes ». *Actes JADT*.

- Malrieu D. A paraître. « Linguistique de corpus, genres textuels, temps et personnes ». *Langages*, 152.
- Malrieu D., Rastier F. 2001. « Genres et variations morpho-syntaxiques ». *TAL*, 42 (2), 547-577.
- Poudat C., Loiseau S. 2003. « Evaluation of Authorial Presence in Academic Genres ». Colloque international « Evaluation in Academic Discourse », Sienna, juillet.
- Rastier F. 2001. *Arts et sciences du texte*. Paris : PUF.
- Rastier F. « Semantic approaches to theoretical texts ». In Fløttum K., Rastier F. (éds), *Academic Discourse. Multidisciplinary Approaches*, 15-35.
- Valette M. « Conceptualization and evolution of concepts. The example of French linguist Gustave Guillaume ». In Fløttum K., Rastier F. (éds). *Academic Discourse. Multidisciplinary Approaches*, 55-73.

### **Thèses inscrites**

- Thomas Beauvisage, *Sémantique des parcours des utilisateurs sur le Web*. Directeur F. Rastier.
- Smail Djaoud, *L'orientalisme en sciences humaines - Etude de thématique*. Directeur F. Rastier.
- Rossitza Kyheng, *L'interprétation en linguistique textuelle : des variations aux invariants des textes*. Directeur F. Rastier.
- Sylvain Loiseau, *Sémantique du discours philosophique*. Directeur F. Rastier.
- Céline Poudat, *L'article de linguistique : étude contrastive de corpus*. Directeur F. Rastier.

### **Références générales sur le projet**

- Adam J.-M. 1999. *Linguistique textuelle. Des genres de discours aux textes*. Paris : Nathan.
- Adam J.-M. et al. 1998. « Pour en finir avec le couple récit/discours ». *Pratiques*, 100, 81-98.
- Barlow M., Kemmer S. (éds). 2000. *Usage-based models of language*. Stanford : CSLI Publications.
- Authier J. 1979. « Problèmes posés par le traitement du discours rapporté dans une grammaire de phrases ». *Linguisticae Investigationes*, 2, 211-228.
- Authier-Revuz J. 1996. « Remarques sur la catégorie de l'îlot textuel ». *Cahiers du français contemporain*, 3, 91-115.
- Authier-Revuz J. 2001. « Le discours rapporté ». In Thomassone R. (éd.). *Une langue : le français*. Paris : Hachette, 192-201.
- Benveniste E. 1966. *Problèmes de linguistique générale*. Paris : Gallimard.
- Biber D. 1988. *Variation across speech and writing*. Cambridge : Cambridge University Press.
- Biber D. 2000. « Investigating language use through corpus-based analyses of association patterns ». In Barlow M., Kemmer S. (éds), *Usage-based models of language*. Stanford : CSLI Publications.
- Biber D., Johanson S., Leech G., Conrad S., Finegan E. 2000. *The Longman grammar of spoken and written English*. Londres : Longman.
- Cohn D. 1981. *La transparence intérieure*. Paris.

- Genette G. 1972. *Figures III*. Paris : Le Seuil.
- Genette G. 1983. *Nouveau discours du récit*. Paris : Le Seuil.
- Genette G. 1991. *Fiction et diction*. Paris : Le Seuil.
- Karlgren J., Cutting D. 1994. « Recognizing text genres with simple metrics using discriminant analysis ». *Proceedings of COLING 94*, Kyoto.
- Kessler B., Nunberg G., Schütze H. 1997. *Automatic detection of genre*. Palo Alto Research Center.
- Langue française* « La Parole intérieure ». 2001. 132.
- Lintvelt J. 1989. *Essai de typologie narrative. Le « point de vue »*. *Théorie et analyse*. Paris : Corti.
- Lips M. 1926. *Le Style indirect libre*. Paris : Payot.
- Marnette S. 2001. « Du discours insolite : le discours indirect sans que ». *French Studies*, 55 (3), 297-313.
- Marnette S. 2001. « The French Théories de l'Énonciation and the Study of Speech and Thought Presentation ». *Language and Literature*, 10 (3), 261-80.
- Marnette S. 2002. « Étudier les pensées rapportées en français parlé : Mission impossible ? ». *Faits de Langues*, 19, 211-20.
- Marnette S. 2002. « Aux frontières du discours rapporté ». *Revue Romane*, 37 (1), 3-30.
- de Mattia M., Joly A. 2001. *De la syntaxe à la narratologie*. Paris : Ophrys.
- Mellet S., Vuillaume M. (éds.). 2000. *Le style indirect libre et ses contextes*. Amsterdam - Atlanta : Rodopi.
- Rabatel A. 1998. *La construction textuelle du point de vue*. Lausanne et Paris : Delachaux & Niestlé.
- Rastier F. 1989. *Sens et textualité*. Paris : Hachette.
- Rastier F. et al. 1994. *Sémantique pour l'analyse*. Paris : Masson.
- Rivara R. 2000. *La langue du récit*. Paris : L'Harmattan.
- Rosier L. 1999. *Le discours rapporté. Histoire, théories, pratiques*. Paris - Bruxelles : Duculot.
- Sperberg-McQueen C.-M. et al. 2001. *TEI P4, Guidelines for Electronic Text Encoding and Interchange*. Text Encoding Initiative : [www.tei-c.org/Guidelines/index.htm](http://www.tei-c.org/Guidelines/index.htm).

## **3.2. APPROCHE MULTIDIMENSIONNELLE DE L'ÉCRIT**

### **3.2.1. COMMUNICATION ÉLECTRONIQUE**

#### **Participants**

*Membres rattachés à titre principal* : Jacques Anis (responsable), Françoise Gadet

*Jeune chercheur associé* : Antonio Balvet

*Doctorant* : Nicole Marty

#### *Collaborations extérieures :*

Hassan Atifi (Université de Technologie de Troyes ; Laboratoire Tech-CICO)

Hillary Bays (Université de Cergy Pontoise ; CELITH-EHESS)

Brigitte Chapelain (Université Paris XIII ; CRIS-SERIES)

Sérgio Roberto Costa (Universidade Federal de Juiz de Fora, Brésil)

Oriane Deseilligny (Paris X, CRIS-SERIES)

Michel de Fornel (EHESS, CELITH)

Béatrice Fraenkel (EHESS ; Centre d'étude de l'écriture et de l'image, CNRS/Paris 7)

Susan C. Herring (Indiana University, Bloomington, USA)

Heidi Gautschi (Université Paris X, CRIS-SERIES)

Brigitte Juanals (Université Charles de Gaulle, CRIS-SERIES)

François Mangenot (Université de Franche-Comté)

Michel Marcoccia (Université de Technologie de Troyes ; Laboratoire Tech-CICO)

Thierry Meyer (Paris X, Psychologie : processus cognitifs et conduites interactives. Déterminants culturels, sociaux et individuels)

Florence Mourlhon-Dallies (Université de la Sorbonne Nouvelle-Paris III ; EA 2290 SYLED-CEDISCOR)

Jacques Perriault (Université Paris X, CRIS-SERIES)

Carole Rodon (Université Paris X, Psychologie : processus cognitifs et conduites interactives. Déterminants culturels, sociaux et individuels).

#### **Discipline(s) et/ou sous-discipline(s) concernée(s)**

Sciences du Langage, en interdisciplinarité avec Sciences de l'Information et de la Communication, Anthropologie, Psychologie.

Sous-disciplines des Sciences du Langage : Sémiolinguistique de l'écrit, Sociolinguistique, Linguistique interactionnelle et analyse conversationnelle, Analyse du discours, Didactique des langues, Lexicologie, morphologie, syntaxe, sémantique, Linguistique de corpus, Traitement automatique des langues naturelles



## **Problématique**

Le projet contribue au développement des recherches linguistiques sur la communication électronique, au sein d'un champ pluridisciplinaire. Il se fixe non seulement de cerner les spécificités de la communication électronique à partir de plusieurs disciplines des sciences du langage, notamment la sémiolinguistique de l'écrit, la linguistique discursive et la sociolinguistique, mais aussi d'affiner certains concepts des disciplines grâce à l'analyse des corpus. On en attend aussi l'élaboration de méthodologies et d'outils de traitement.

Le domaine de la communication médiée par ordinateur (domaine que nous sommes amenés à élargir aux réseaux de la téléphonie mobile, d'où le terme que nous proposons de communication électronique) a été relativement peu exploré par les linguistes et étudié plutôt par les psychologues et les sociologues. On observe par ailleurs un certain retard des études francophones sur l'internet, en partie pour des raisons objectives liées au « retard français » dans l'implantation de ce réseau. Pourtant, le phénomène du Minitel avait suscité en France un certain nombre de travaux. La participation au projet de S. Herring, qui a développé aux USA des recherches sur l'analyse du discours médié par ordinateur, est un élément de la dynamique que nous voulons créer, tendant à fédérer les différentes approches proposées par les linguistes de divers horizons. Une synergie entre la sémiolinguistique de l'écrit, l'analyse du discours, la sociolinguistique sera mise en œuvre pour l'analyse des corpus électroniques.

## **Contexte international**

Recherches sur la Communication médiée par ordinateur (Computer-Mediated Communication) développées surtout en psychologie, sociologie et sciences de l'information et de la communication. Courant nord-américain de l'Analyse du Discours médié par ordinateur (Computer-Mediated Communication Discourse Analysis) lancé par Herring.

## **Contexte national**

A côté de nombreux travaux en psychologie, sociologie et sciences de l'information et de la communication, développement plus limité des recherches linguistiques : sur le Minitel à partir de 1985 ; sur l'internet francophone depuis les dernières années du XXe siècle (cf. Anis *et al.* 1999 ; Fraenkel et Marcoccia 2003 ; Mourlhon-Dallies & Reboul-Touré 2004).

## **Thèmes de recherche**

Les hypothèses mises à l'épreuve se situent à différents niveaux. La plus générale est que les conditions matérielles de la communication ont une forte influence sur les formes et les processus langagiers. Les autres sont plus spécifiques et se formulent à la fois dans les différentes sous-disciplines et dans leurs interrelations :

1) Le fonctionnement et le statut de l'écrit sont modifiés en profondeur par la communication électronique ; celle-ci introduit une nouvelle gestion de l'espace et du temps et développe la variation dans les systèmes et les usages graphiques ; les relations entre l'oral et l'écrit en deviennent beaucoup plus complexes.

2) La communication électronique fait émerger de nouveaux modes d'interaction langagière, de nouveaux genres, même si ces modes et genres s'inspirent largement des modes et genres intérieurs et en sont en un certain sens des simulations.

3) Des configurations discursives inédites, notamment celles qui surgissent dans les hypertextes, obligent à retravailler des notions aussi fondamentales que la polyphonie, la textualité, l'intertextualité...

La prise en compte des aspects cognitifs exige aussi un dialogue fécond avec d'autres disciplines, comme la psychologie et les sciences de l'information et de l'éducation. La construction de savoirs à travers la communication électronique, dans des environnements scolaires et extrascolaires sera un axe important du programme.

## ***Bilan***

### **Étapes passées**

Ce projet fait suite à un précédent projet intitulé « Analyse sémiolinguistique d'échanges écrits électroniques », menée par Jacques Anis, où il s'agissait d'étudier les échanges écrits électroniques : analyser des formes langagières, des formes textuelles et des modes d'interaction spécifiques. Le domaine de la communication médiée par ordinateur se révèle être peu étudié par des linguistes, même aux Etats-Unis (travaux de Herring). Nous avons travaillé à partir de l'hypothèse suivante : les échanges électroniques étudiés relèvent de la conversation écrite. Ils comportent des formes linguistiques spécifiques et doivent faire l'objet d'une sémiolinguistique de l'écrit pluridimensionnelle.

Les résultats obtenus sont de trois ordres :

- 1) Consolidation de la notion de conversation écrite.
- 2) Description de la dimension interactionnelle et des investissements énonciatifs des sujets.
- 3) Analyse de mécanismes formels produisant des néographies.

### **Actions réalisées de « Communication électronique »**

Séminaire « Communication électronique » :

- Prise de contact le 1<sup>er</sup> octobre 2003.
- Séance du 13 novembre : SMS (J. Anis, Modyco) ; « (we)blogs » (H. Gautschi, CRIS-SERIES) ; outils de traitement (A. Balvet, Modyco).

## ***Perspectives***

### **Résultats attendus**

Les résultats attendus portent sur la méthodologie et sur l'affinement et l'approfondissement de l'analyse de la communication électronique.

### **Actions prévues**

- 1) 2004 : colloque international « La communication électronique : Approches linguistiques et anthropologiques », 5 et 6 février à la Maison des Sciences de l'Homme à Paris, à l'initiative de Jacques Anis, Béatrice Fraenkel et Michel de Fornel.
- 2) 2004 : colloque de l'Ecole doctorale « Connaissance et Culture » de l'Université Paris X « Pratiques d'écriture et d'édition numériques dans la médiation des savoirs », projet en cours d'élaboration avec J. Perriault, B. Chapelain et B. Juanals.
- 3) 2005 : numéro de *Langages* « Communication électronique : approches linguistiques », en codirection avec Michel de Fornel, avec la participation de Susan C. Herring (Indiana University, Bloomington), Florence Mourlhon-Dallies, Françoise Gadet, Michel Marcoccia, Hillary Bays.
- 4) 2006 : publication collective à préciser.

### **Production scientifique**

#### **pour « Analyse sémiolinguistique d'échanges écrits électroniques »**

- Anis J. (éd.). 2001. *Parlez-vous texto ?* Paris : Le cherche-midi.
- Anis J. 2001. « Approche sémiolinguistique des représentations de l'ego dans la Communication Médiée par Ordinateur ». In Vivier J. (éd.), *Langages* « Psycholinguistique et intelligence artificielle », 144, 20-38.
- Anis J. 2001. « Approche sémiolinguistique d'interactions écrites en temps réel sur l'internet ». In Bouchard R., Mangenot F. (éds), *Interactivité, interactions et multimédia*, ENS Fontenay-St Cloud. Collection « Notions en question », 147-168.
- Anis J. 2003. « Communication électronique scripturale et formes langagières : chats et SMS ». *Actes des Quatrièmes Rencontres « Réseaux Humains / Réseaux Technologiques »*, Université de Poitiers, 31 mai et 1er juin 2002.
- Anis J. 2004 « Les abréviations dans la communication électronique (en français et en anglais) ». In Andrieux-Reix N., Branca S., Puech C. (éds), *Faits de langue*.

### 3.2.3. ECRIT ET ECRITURE SCOLAIRE

#### **Participants**

*Membres rattachés à titre principal* : Catherine Boré (responsable), Jacques Anis, Marie-Laure Elalouf

*Doctorant* : Nicole Marty

*Collaborations extérieures* :

Christian Bouillon (IUFM de Versailles)

Eduardo Calil (Université fédérale de l'Etat d'Alagoas (Brésil))

Pascale Hannoun (Université Paris VI)

Serge Herreman (IUFM de Versailles)

#### **Discipline(s) et/ou sous-discipline(s) concernée(s)**

Ecrit, didactique, linguistique des textes

#### **Problématique**

L'objectif à long terme de ce projet est de mieux caractériser l'écriture scolaire et les genres de discours par lesquels elle s'exprime.

On considère ici un continuum qui va de l'écrit - segmentation des énoncés, mise en graphie, disposition dans l'espace etc. à l'écriture entendue comme production de textes.

Ce qui donne une unité à ce projet est l'homogénéité des données qu'il rassemble, si l'on veut bien considérer comme homogène le fait que les participants travaillent tous sur des corpus *scolaires*. Ils sont nombreux aussi à travailler sur les *processus*, c'est-à-dire que sont incluses la dimension temporelle et le caractère progressif de la constitution de ces corpus.

L'écriture scolaire n'est devenue l'objet de recherches systématiques que depuis une vingtaine d'années, au cours desquelles elle a été étudiée selon des points de vue différents et divergents.

Considérée sous l'angle de l'écrit, elle a donné lieu à des développements privilégiant la genèse des traces écrites, et l'interprétation qu'en donnent les jeunes scripteurs. Ces recherches (Besse 1990 ; Gombert 1987 ; Jaffré 1994) inspirées au départ par les travaux d'E. Ferreiro, se sont maintenant affranchies des références piagétienne de celles-ci ; tout en gardant à l'esprit la visée ontogénétique de ces devanciers, elles empruntent aussi à d'autres voies : d'une part en mettant en relation « écrit et écritures », elles penchent vers les recherches anthropologiques issues des travaux de Goody (1994) (cf. aussi travaux sur les « écritures inventées » à l'école maternelle de David & Jaffré 1998) ; d'autre part surtout, elles rejoignent les courants psycholinguistiques et, plus récemment, les approches cognitivistes du problème.

L'étude de l'utilisation des signes de ponctuation, de la mise en espace graphique, par exemple, relève de plus en plus d'une approche psycholinguistique et psychologique

(Fayol 1997 ; Garcia-Debanc 1990) dans la mesure où ce qui intéresse ces chercheurs est davantage la mesure de la compréhension et de la mémorisation des textes produits par les élèves, et par suite les traitements formels et cognitifs qu'ils en font en production.

On relève ainsi un glissement de la genèse de l'écrit comme produit, à la genèse des processus engendrant la production écrite. Un modèle comme celui de Levelt (1989, initialement consacré au langage oral) est utilisé comme référence par des psychologues comme Fayol et regardé avec intérêt par une partie des linguistes théoriciens de la génétique des textes (Grunig 2002 ; Lebrave 1983).

La génétique des textes ou critique génétique (Hay 1979 ; Grésillon 1994), initialement tournée vers un traitement de l'établissement progressif des produits (manuscripts littéraires), devient de plus en plus une linguistique des processus d'écriture, avec un intérêt particulier pour les travaux de neuropsycholinguistes (Besson et Kutas, cités par Grunig 2002) qu'il s'agit de croiser avec la théorie linguistique.

Les travaux cités ci-dessus ne permettent pourtant pas à ce jour de proposer une analyse spécifique et ciblée de l'écriture scolaire en général : l'évolution de l'enfant et celle de l'élève ne sont pas une seule et même chose. Il est évident que la variable « apprentissage » de la didactique apporte un éclairage différent aux théories générales de la production graphique et écrite. Par ailleurs, il semble clair que les observations menées doivent l'être en contexte scolaire plutôt qu'en laboratoire ou sur un mode expérimental.

Dans un premier temps, il s'agit de constituer un corpus :

a) chacun des participants du groupe part d'un corpus de référence dont la constitution et les objectifs ne sont pas identiques, ce qui - en pratique - revient à se poser la question de la finalité et de la sélection de ces données.

b) le statut des données elles-mêmes doit être interrogé : ce sont des textes d'élèves complets qui forment ces corpus et non des échantillons.

Cela pose plusieurs questions :

1) La plupart sont recueillis à l'état de brouillons qui se transforment peu à peu en « écrits », « versions », « recopies », les termes sont variables et à interroger. Selon les corpus, le statut du brouillon est différent : soit il est vu comme texte transitoire servant lui-même de contexte, d'intertexte pour les versions suivantes ; soit il est vu comme instrument, outil psychologique, incité, suscité par l'enseignant pour permettre à l'élève des progrès cognitifs et prend le nom d'« écrit intermédiaire » (qui suppose la médiation de l'enseignant), mais l'écrit intermédiaire n'est pas non plus tout le texte intermédiaire (Bernié 2001).

2) Ils sont accompagnés d'éléments contextuels développés qui sont nécessaires à leur interprétation ; ces éléments contextuels sont loin d'être unifiés. Ce qu'on peut dire, cependant, c'est qu'il se dégage une caractéristique commune : la plupart de ces corpus sont hybrides ; il existe en effet un continuum oral / écrit dans le processus qui les a constitués, si bien que l'on peut parler des données orales aussi bien comme contexte que comme corpus propre.

3) Le point de vue est différent selon que ces données sont recueillies pour observer le travail de l'enseignant ou celui de l'élève. Le projet de l'enseignant s'observe dans le corpus oral (ex. : consignes, conseils, etc.) et dans les commentaires éventuellement portés

sur les copies et/ou à l'oral. Le point de vue de l'élève est rarement explicite, il est sollicité, et se marque par l'ensemble des modifications apportées sur les copies, en lien ou pas avec les remarques de l'enseignant.

On souhaite développer la première année une réflexion épistémologique, issue en partie des questions nées de la recherche dans le domaine de l'écriture scolaire de Elalouf *et al.* (à paraître).

Enfin, la nécessité d'interpréter tout observable - fût-il obtenu par des voies technologiques et scientifique complexes (imageries cérébrales, enregistrements automatiques des données temporelles corrélées à la production etc.) - invite à adopter une démarche critique.

### **Contexte international**

Sur le plan international, le groupe se réfère aux travaux de l'école suisse de Bronckart (1985, 1996), qui a inspiré toute la recherche française en typologie des textes, et contribué à une classification des textes et des discours à la fois psycho-sociale et linguistique (différente de celle d'Adam 1990). La référence à Bronckart est majeure pour les travaux du groupe, à cette réserve que les genres décrits ne recoupent pas forcément les genres inédits produits par l'école.

Par ailleurs, un échange est amorcé avec des chercheurs nord-américains (NYU en particulier) sur le « contexte dans la classe » : les travaux empiriques de Nancie Atwell sur les genres scolaires pratiqués en classe (12-15 ans) et les processus d'écriture mis en œuvre dans les classes occupent le versant pédagogique. Une recherche est en cours sur les modèles australiens du contexte, par le biais d'une correspondance avec Frances Christie, professeur à l'Université de Sydney.

D'une manière générale, les recherches européennes semblent se distinguer des recherches anglo-saxonnes et australiennes par l'importance accordée à la dimension sémiotique qui suppose une composante linguistique et énonciative dans la définition des genres. Le modèle communicationnel de Hymes est massivement utilisé dans les recherches américaines comme dans toutes les approches de type sociologique ou ethnolinguistique dans la classe. Ce n'est pas du tout l'option théorique retenue pour nos recherches.

### **Contexte national**

Trois courants, parfois antagonistes, se partagent le champ et inspirent diversement les travaux du groupe :

- Travaux de l'ITEM/CNRS (Institut des textes et Manuscrits modernes, Lebrave et Grésillon, dir.), qui travaille surtout sur les processus dans les manuscrits littéraires (genèse de l'écrit).

- Travaux de l'INRP de Lyon (Institut National de Recherche Pédagogique) : les équipes de didactique du français travaillent depuis longtemps (Garcia-Debanc, groupe Eva) dans le domaine de la production scolaire écrite et de la réécriture. Un colloque en mars 2002 organisé par S. Plane (actes à paraître dans *Repères*, 26-27, 2002-2003) a fait le point sur les apprentissages de la production d'écrits en didactique de l'écriture. Beaucoup de travaux de l'INRP se sont longtemps inspirés de travaux cognitivistes nord-américains (Flower & Hayes 1980) et des recherches de didactique générale au Québec.

- Travaux de J.-P. Bernié (Bordeaux 2, l'IUFM d'Aquitaine), de D. Bucheton (IUFM de Montpellier) : inspirés de Vygotsky (1997) et de Bakhtine (1984), ils se fondent sur l'interactionisme social du premier et les thèses génériques du second (associées à ses théories de l'hétéroglossie) pour proposer un modèle de la production écrite construit sur les interactions langagières.

Les travaux de ce dernier courant pourraient fournir une référence pour le groupe. Mais le modèle proposé n'articule aucune théorie linguistique à son cadre. Inversement les modèles linguistiques de l'ITEM ne relient pas la langue à une « pratique » ni à un genre. On ne dispose donc pas d'un cadre de référence unifiant.

Les travaux du projet « Genres et linguistique de corpus » ont en commun avec ceux de « Ecrit et écriture scolaire » la référence à Bronckart (1985, 1996). On peut considérer notre recherche descriptive de l'écriture scolaire insérée dans des genres scolaires à décrire, comme ayant des objectifs convergents.

### **Thèmes de recherche**

La recherche s'appuie sur les conclusions de Elalouf *et al.* (à paraître 2004), ce qui a permis de circonscrire deux types de problèmes :

1) Extension et définition de la notion de « contexte », liée à la production de types de textes scolaires

Il est clair que l'approche empirico-inductive dans laquelle se situe le projet s'appuie sur des corpus contextualisés. Le problème que l'on se pose est de sortir du caractère totalisant, et par là même, vague et inopérant du « contexte » quand celui-ci désigne des éléments non sélectionnés de l'environnement de la tâche. Toutes les variables d'une situation ou d'un contexte ne peuvent être maîtrisées : plus on prend en compte de variables (notamment contextuelles et qualitatives), plus le corpus est réduit en quantité... et inversement, pour de simples raisons de « faisabilité ». Cela est cohérent avec le projet de la « linguistique de terrain » (Blanchet 2000) qui est :

- d'une part de décrire et de comprendre des situations ponctuelles (ex. : analyse d'une situation sociolinguistique, d'une conversation, etc.) ;
- d'en proposer des interprétations qualitatives, sans prétendre à l'exhaustivité ni à la théorisation totale.

Une direction intéressante pourrait être fournie par la recherche du labo GRIC de Lyon 2 conduite par M.-M. de Gaulmyn (2001), qui a porté sur la rédaction conversationnelle ; commencée en 1991, elle se situait dans le domaine mal connu des pratiques conjointes de l'oral et de l'écrit en situations oralo-graphiques, avec l'hypothèse que ces formes hybrides favorisent une explicitation des savoirs et un transfert des savoir-faire chez les apprenants (des adultes de FLE).

L'écriture y était envisagée d'emblée comme une pratique sociale contextualisée. Le contexte y était défini comme se composant de la situation, des objets intermédiaires (croquis, plans etc.) et des événements objectifs vécus subjectivement pendant la rencontre enregistrée, des expériences, habitudes, apprentissage des partenaires, des modèles culturels, des goûts et préférences des sujets observés.

Il est évident que des éléments comme les stéréotypes structurels scolaires dans l'interprétation de la situation véhiculée par la consigne, les préconstruits, jouent un rôle de premier plan dans le processus d'écriture. Mais les problèmes qui se posent sont considérables, en particulier : quelle part du contexte faut-il considérer comme

« agissante » (et sur quelle base théorique) ? Quelle part serait au contraire « inerte » ? Il ne saurait y avoir de toute manière de corrélation simple entre contexte de production et processus d'écriture constaté.

Bernié (2001) a proposé de définir le contexte comme constitué par le genre et la situation dans lesquelles sont produites les actions langagières, à savoir une part essentielle et toujours inédite construite en situation scolaire (cf. fictionalisation). A ce titre, corpus oraux et écrits peuvent difficilement être clairement distingués, ils se nourrissent l'un de l'autre : le corpus oral n'est pas un contexte extérieur au corpus écrit, il en fait partie et donne au genre sa configuration.

2) Le genre est donc l'autre élément de réflexion important, notamment la question des « genres scolaires »

Du point de vue de la démarche, on peut dégager deux manières d'utiliser les typologies :

- soit on part de typologies existantes fonctionnelles ou situationnelles et on examine les textes en cherchant à établir des corrélations entre types et traits linguistiques.

- soit on fait émerger des types de textes grâce à un traitement statistique de textes étiquetés. On part alors des textes concrets. C'est par exemple le cas de la typologie de Biber (citée par Habert *et al.* 1997), qui a l'intérêt de présenter certains points communs avec celle de Bronckart (1985, 1996) et qui, privilégiant certaines associations, aboutit à des prototypes, des pôles de fonctionnement opposés. Pour Biber, 8 types sont trouvés, qui ne correspondent pas forcément aux grandes distinctions connues puisque par exemple, il y a des textes à interaction personnelle, d'autres à interaction informationnelle ; ou encore, il y a deux types d'exposés, « savant » et « scientifique ». On voit bien dans le cadre de ce projet l'intérêt d'une démarche inductive.

Le texte scolaire constitue un type de texte intermédiaire. De nombreux chercheurs en didactique et linguistique comme Bernié, s'appuyant sur les travaux de Bakhtine mettent en évidence l'existence de genres « intermédiaires », genres « secondaires » pour Bernié qu'il définit comme « des textes composites d'un point de vue générique mais discursivement organisés à partir d'un principe de coordination qui transmute leurs composantes » (Bernié 2001). Ils sont « caractérisés par leur fusion, qui fait souvent disparaître toute possibilité d'identification typologique, toute pertinence des préférences typologiques ». Les « textes intermédiaires » sont caractérisés par une construction de modes de points de vue orchestrateurs (cf. aussi Elalouf *et al.* à paraître 2004) ont révélé.

## ***Bilan***

### **Etapes passées**

La recherche de Elalouf *et al.* (à paraître) a été riche d'enseignements méthodologiques :

**1) Dans le recueil des données :** les écrits scolaires ne peuvent être recueillis qu'en contexte. Dans la mesure où les jeunes scripteurs répondent à une commande extérieure d'écriture (ce qui est le propre de la situation scolaire), la mise en route du processus d'écriture dépend de conditions externes à l'acte scriptural proprement dit. Le recueil des productions seules est donc insuffisant pour une interprétation des données, celles-ci ne pouvant se réduire à l'avant-texte, série de brouillons relevés lors des phases d'écriture. Cela ne signifie pas, bien entendu, que l'écriture scolaire soit la réponse à un stimulus



fourni par l'environnement. Sa mise en route répond à des considérations complexes, c'est pourquoi la définition de ce qui est « contexte » apparaît décisive dans la recherche. Son articulation avec la notion de « genres scolaires » reçus/produits à l'école et par l'école apparaît fondamentale. Les travaux de Bronckart (1985, 1996) sont à cet égard une référence incontournable.

Une définition du contexte empirique a été donnée dans la précédente recherche : elle s'appuyait sur le relevé des consignes, leçons orales, interactions langagières diverses entre les élèves et l'enseignant, et prenait aussi en compte les représentations de l'enseignant (entretiens). Elle est actuellement très sérieusement révisée à l'aide des travaux de Bernié (2001) et Bronckart (1985, 1996), entre autres.

Par ailleurs, le contexte se manifeste dans un format : la séquence, unité d'enseignement de plusieurs séances, propre au collège, transposable au primaire sous le terme de « module ». Il est acquis que les méthodes de recueil sous forme de vidéo sont préférables aux seuls enregistrements audio, même si ces derniers ont fourni l'essentiel des documents analysés lors de cette 1<sup>ère</sup> phase de recherche.

**2) Dans l'exploitation des différents sous-corpus :** jusqu'à présent les indicateurs utilisés ont emprunté aux théories de l'énonciation, les différentes positions énonciatives du sujet constituant l'indicateur dominant capable de différencier les étapes de l'écriture et d'unifier la diversité des genres rencontrés. Un obstacle majeur a été rencontré avec l'articulation des analyses d'oraux et d'écrits, très différenciées selon les corpus.

### **Résultats obtenus**

1) La plupart des modèles et théories (génétique textuelle, psychologie cognitive, linguistique textuelle, etc.) échouent isolément à rendre compte de l'écriture scolaire, dans la mesure où ils constituent des modèles partiels dont l'unification apparaît plus comme une juxtaposition de disciplines que comme un ensemble d'hypothèses fécondes. Accessoirement, elles ont entraîné par le passé toutes les « dérives applicationnistes » souvent dénoncées en didactique.

2) La théorie des genres scolaires et celle des contextes de production des corpus sont posées comme fondamentales. Elles déterminent en amont le projet de recherche « écrit et écriture scolaire » (Bronckart, Bernié) et le réorientent partiellement.

Les productions des membres ont rendu compte d'analyses partielles mettant l'accent tantôt sur la constitution de genres scolaires comme la description (Elalouf & Keraven 2002), tantôt sur le contexte oral articulé avec la production d'un écrit (Boré 2003).

### **Actions réalisées**

Journée Scientifique du 23 mai 2003, organisée par Jacques Anis et Catherine Boré : « Quand les pratiques scolaires interrogent les théories et modèles de l'écriture » ; une soixantaine de participants comportant de nombreux chercheurs venus de France mais aussi de Suisse, Belgique, Brésil, etc.

### ***Perspectives***

#### **Étapes à venir**

La question des genres et des typologies est donc centrale pour cette recherche. Avec elle se pose non seulement la question de la démarche typologique à adopter, mais aussi le problème des outils descriptifs utilisés pour rendre compte des corpus recueillis. Comme on l'a vu plus haut, il est nécessaire d'articuler :

- des micro-analyses recueillies en situations de dyades (Calil 2003) qui font émerger la question du sens pour le sujet scolaire ;
- des analyses quantitatives, sur la base d'un traitement automatique. Ce qui suppose de réfléchir aux éléments à privilégier pour une analyse automatique.

En effet, les textes scolaires - dans la mesure où ils ne sont pas normés - ne peuvent être traités directement par la plupart des logiciels existants.

Plusieurs types de difficultés surgissent :

- 1) norme et acceptabilité du texte : les différents modes de représentation et/ou de transcription du texte montrent assez la difficulté qu'il y a à transcrire des textes différents, par définition, de textes d'adultes experts de l'écrit. Bien souvent le chercheur est contraint d'interpréter le sens de la « phrase » qui, dans les textes d'élèves, n'est justement pas une unité syntaxiquement stable ;
- 2) codage des corrections / hésitations de l'élève : les ratures ;
- 3) codage des corrections de la maîtresse.

D'où la difficulté consistant à numériser un « texte » reformulé en cours de route et hétéro-reformulé par un tiers (corrections apportées par la maîtresse). Un traitement autre que manuel suppose la possibilité de faire identifier pour un même « texte » deux sources d'écriture différentes produites en des temporalités différentes, mais co-présentes.

Une enquête préalable sur les logiciels existants, adaptés aux types de corpus sur lesquels travaille l'équipe peut être intéressante. Mais surtout, la création d'outils descriptifs apparaît cruciale, dans la mesure où elle peut entraîner / permettre une caractérisation spécifique des textes/écrits scolaires.

### **Actions prévues**

- 1) Séminaire réunissant les membres du projet : un calendrier en a fixé les séances tous les deux mois.
- 2) Une collaboration est en cours avec les membres du projet « Genres et linguistiques de corpus » (François Rastier, Denise Malrieu) pour étudier la possibilité d'un traitement automatique des données visant à caractériser linguistiquement les genres scolaires. En

préalable, il faut étudier si la spécificité non normée des textes scolaires en cours d'élaboration constitue un obstacle à la fois épistémologique et technique à l'assimilation pure et simple d'un projet par l'autre.

### **Production scientifique**

- Anis J. 2002. « Le nom éponyme comme matrice textuelle dans le dossier des *Cinq Sapates* de Francis Ponge ». In Fenoglio I., Boucheron-Pétillon S. (éds), *Langages : « Processus d'écriture et marques linguistiques »*, 147, 98-112.
- Boré C. 2003. « Activités langagières en géométrie et en grammaire : de la définition à la notion ou l'inverse ? ». *Actes du Colloque* (sur cédérom), Université Victor Segalen CIRFEM / DAEST, Bordeaux, 3-5 avril 2003.
- Boré C. Sous presse. 2003. « Reformulations orales et écrites : part du cognitif, part du linguistique ». *Repères*, 26-27.
- Boré C. A paraître. « Brouillons et genres scolaires : contribution à la connaissance de l'écriture scolaire ». *Le Français Aujourd'hui*, 144.
- Boré C., Huynh J.-A. (éds). 2004. *Le Français Aujourd'hui : « Réécritures »*, 144.
- Elalouf *et al.* A paraître. *Des pratiques des enseignants aux effets sur les élèves : le cas de l'écriture à la charnière école/ collège*.
- Elalouf M.-L., Keraven J. 2002. « Une banque de données de textes d'élèves à l'épreuve ». *Pratiques*, 115-116, 107-124.
- Elalouf M.-L., Tomassone R (éds). 2001. *Grands repères culturels pour une langue, le français*. Paris : Hachette.
- Marty, N. A paraître. « Les nouvelles technologies de l'écriture : un rapport nouveau à l'écriture ? ». *Actes de la Journée d'études « Quand les pratiques scolaires interrogent les théories et modèles de l'écriture »*, mai 2003.

### **Thèse inscrite**

Nicole Marty, *Pratiques d'écriture à l'école primaire : l'apport des nouvelles technologies*. Directeur J. Anis.

### **Références générales sur le projet**

- Adam J.-M. 1990. *Eléments de linguistique textuelle*. Bruxelles : Mardaga.
- Anis J. 1994. *Thought, Language and Handwriting : What Can We Guess from Literary Drafts ? Advances in Handwriting and Drawing : a Multidisciplinary Approach*. In Faure C., Keuss P., Lorette G., Vinter A. (éds). Paris : Europia, 531-545.
- Bakhtine M. 1984. *Esthétique de la création verbale*. Paris : Gallimard, Collection « Bibliothèque des idées ».
- Bernié J.-P. 2001. *Apprentissage, développements et significations*. Bordeaux : PUB.
- Bernié J.-P. 2001. « Problèmes posés par la co-construction d'un contexte commun aux partenaires d'une activité rédactionnelle ». In de Gaulmyn M.-M., Bouchard R., Rabatel A. (éds), *Le processus rédactionnel, écrire à plusieurs voix*. Paris : L'Harmattan, 147-168.
- Bernié J.-P. 2001. « Genres scolaires et conception dialogique des conduites langagières ». *Actes du colloque IUFM de Toulouse*. Paris : Delagrave, 107-118.

- Besse J.-M. 1990. « L'enfant et la construction de la langue écrite ». *Revue française de Pédagogie*.
- Besson M., Kutas M. 1997. « Manifestations électriques de l'activité de langage dans le cerveau ». In Fuchs C., Robert C. (éds), *Diversité des langues et représentations cognitives*. Paris : Ophrys, 250-271.
- Blanchet Ph. 2000. *La linguistique de terrain, méthode et théorie*. Rennes : Presses Universitaires de Rennes.
- Bonnet C., Corblin C. Elalouf M.-L. 1998. *Les procédés d'écriture chez les élèves de 10 à 13 ans, un stade de développement*. Versailles et Lausanne : IUFM de Versailles, Loisirs et Pédagogie, Centre vaudois de recherches pédagogiques.
- Boré C. 2000. « Le brouillon, introuvable objet d'étude ». *Pratiques*, 105, 23-49.
- Boré-Lamothe C. 1998. *Choix énonciatifs dans la mise en mots de la fiction, le cas des brouillons scolaires*. Thèse de doctorat (Sciences du langage), Université Grenoble III.
- Bronckart J.-P. et al. 1985. *Le fonctionnement des discours*. Lausanne - Paris : Delachaux et Niestlé.
- Bronckart J.-P. 1996. *Activités langagières, textes et discours*. Lausanne - Paris : Delachaux et Niestlé.
- Bucheton D. 1995. *Ecriture, réécritures, récits d'adolescents*. Berne - Paris : Peter Lang.
- Calil E. 2003. « Processus de création et ratures ». *Langages et société*, 103, 31-55.
- David J., Jaffré J.-P. 1998. *La naissance de l'enfant scripteur*. Paris : ESF.
- Eva (coll.) 1996. *De l'évaluation à la réécriture*. Paris : Hachette /INRP
- Fabre Cl. 1990. *Les brouillons d'écoliers ou l'entrée dans l'écriture*. Grenoble : Céditel / L'Atelier du texte et Université de Grenoble.
- Fabre-Cols Cl. et al. (éd.). 2000. *Apprendre à lire des textes d'enfants*. Bruxelles : De Boeck-Duculot.
- Fabre-Cols Cl. 2002. *Réécritures à l'écrit et au collège : de l'analyse des brouillons à l'écriture accompagnée*. Paris : ESF.
- Fayol M., 1997. *Des idées au texte*. Paris : PUF.
- Flower L., Hayes J. R. 1980. *Cognitive Processes in Writing*. Hillsdale : LEA.
- François F. 1998. *le discours et ses entours. Essai sur l'interprétation*. Paris : L'Harmattan.
- de Gaulmyn M.-M., Bouchard R., Rabatel A. (éds). 2001. *Le processus rédactionnel, écrire à plusieurs voix*. Paris : L'Harmattan, 147-168.
- Garcia-Debanc Cl. 1990. *L'élève et la production d'écrits*. Metz : CRESAF.
- Gombert J.-E. 1987. « Processus rédactionnels et développement des méta-connaissances dans le domaine langagier ». *Actes du 3<sup>ème</sup> colloque international de didactique du français*. Bruxelles : de Boeck.
- Goody. 1994. *Entre l'oralité et l'écriture*. Paris : PUF.
- Grésillon A. 1994. *Eléments de critique génétique*. Paris : PUF.
- Grunig B.-N. 2002. « Linguistique et brouillons, dynamique et synchronisation ». *Langages*, 147, 113-123.
- Habert B., Nazarenko A., Salem A. 1997. *Les linguistiques de corpus*. Paris : Armand Colin.
- Hannoun P. 2003. *Les écrits intermédiaires à l'école primaire, sur les traces des fonctions cognitives exploitées par les enseignants, observations et analyse des pratiques effectives*. Mémoire de DEA (Sciences de l'éducation), Université Paris 5.
- Hay L. 1979. *Essais de critique génétique*. Paris : Flammarion.
- Jaffré J.-P. 1994. « Une approche génétique de l'écriture : de l'invention à l'acquisition ». *LINX*, 31.

- Lebrave J.-L. 1983. « Lecture et analyse des brouillons ». *Langages*, 63, 11-23.
- Levelt W. J. M. 1989. *Speaking : from intention to articulation*. Cambridge : The MIT Press.
- Marty N. (éd.). 1991. « Genèse de l'écrit et activités métalinguistiques dans des dialogues de jeunes scripteurs. L'écrit dans l'oral ». *Etudes de linguistique appliquée*, 81, 57-70.
- Schneuwly B., Dolz J. 1997. « Les genres scolaires. Des pratiques scolaires aux objets d'enseignement ». *Repères*, 15, 27-40.
- Vygotsky L. 1997. *Pensée et langage*. Paris : La Dispute.

### **3.3. PROBLÈMES DIACHRONIQUES**

#### **Participants**

*Membres rattachés à titre principal* : Annie Bertin (responsable), Jacques Anis, Danielle Leeman, Isabelle Weill

*Membres associés* :

Jacqueline Bacha (Faculté des lettres de Sousse, Tunisie)  
Michèle Perret

*Jeunes chercheurs associés* : Fidèle Diedhiou, Sabine Lehman, Marielle Lignereux

*Doctorant* : Dana Sleiman-Saikali

*Collaborations extérieures* :

Valentine Fesenko (Université Linguistique de Kiev (Ukraine))

#### **Discipline(s) et /ou sous-discipline(s) concernée(s)**

Diachronie, syntaxe, morphologie, pragmatique, énonciation, systèmes graphiques, linguistique textuelle

#### **Problématique**

Le projet est lié de par ses activités d'édition et d'études énonciatives à la linguistique textuelle (du fait du recours à une documentation textuelle). En même temps, par les domaines abordés (syntaxe des connecteurs et des prépositions) et par certaines approches du changement linguistique, il a des contacts avec la syntaxe d'une part, avec la psycholinguistique (phénomènes de grammaticalisation qui s'inscrivent dans la mouvance cognitive) et la sociolinguistique (rôle des « styles » dans la dynamique diachronique).

Ce projet s'est également inscrit dans le renouveau des études diachroniques amené par le cadre théorique de la grammaticalisation à la suite des travaux de Traugott, Lehman, Hopper relayés en France par Combettes et Marchello-Nizia. Sans sacrifier l'interface linguistique/littérature - portant sur le texte médiéval au sens large (XIIe-XVe siècles) -, on a cherché également à développer des travaux sur le changement linguistique, au croisement de déterminations cognitives, communicationnelles, sociales et culturelles. La diachronie s'entend comme étude des processus dynamiques dans l'histoire de la langue aussi bien qu'étude d'un état de langue ancien.

On s'inscrit donc dans un cadre pragmatique et de linguistique de l'énonciation pour observer les actes de parole (promesse, serment, insulte), le fonctionnement des connecteurs, le discours rapporté, la pertinence de la distinction histoire/discours à partir des marques du narrateur (temps et personnes verbales, connecteurs), la construction de la référence du récit (détermination, anaphore).

Par ailleurs, on a cherché à utiliser, mais aussi mettre à l'épreuve des corpus informatisés du moyen français, les processus de grammaticalisation pour l'évolution des connecteurs. On a cherché à montrer les limites de certains phénomènes donnés comme définitoires

dans ce modèle (unidirectionnalité, passage du concret à l'abstrait, différence entre lexicalisation et grammaticalisation), en mettant en valeur dans le changement l'interaction des « styles », principalement la dialectique de l'oral et de l'écrit, le rôle des mutations culturelles (passage d'une « syntaxe pour l'oreille » à la syntaxe de l'écrit, selon les termes de Shlieben-Lange), le rôle de l'iconicité et du besoin de flou. On voudrait utiliser à la fois les grands corpus constitués sur l'ancien et le moyen français, et les données fournies par les traductions intralinguales du XIIe au XVIe siècle.

Le projet a été centré dans un premier temps sur la diachronie des connecteurs, en fonction des collaborations internes et externes de l'équipe (voir *infra*) en relation avec la réflexion sur la grammaticalisation. Il veut aussi mettre à profit les faits de réécriture, particulièrement les traductions intralinguales (français-français) qui permettent d'observer *in vivo* la variation et le changement.

Les activités sont poursuivies selon les axes suivants :

- 1) diachronie des mots grammaticaux du français, en relation avec d'autres projets du laboratoire, en particulier celui sur les connecteurs ;
- 2) réflexion sur la grammaticalisation à rattacher à une réflexion plus large sur la catégorisation (rapports entre lexicalisation et grammaticalisation) ;
- 3) réflexion sur ce qui fonde la périodisation en histoire de la langue : la périodisation des historiens de la langue correspond-elle aux points de rupture perçus par les locuteurs au point de nécessiter une traduction ?
- 4) selon une direction propre aux diachroniciens, qui souhaitent à terme produire une grammaire de référence du Moyen Français, projet qui vise une description syntaxique mais aussi morphologique voire graphique de la tranche synchronique XIVE-XVIe siècles. Ce second volet ne peut s'entendre que dans le cadre d'une collaboration entre UMR voire d'une fédération d'UMR. Il est cependant vivement souhaité.

### **Contexte international**

Certains travaux de linguistique textuelle s'inscrivent dans la mouvance pragmatique des travaux sur les actes de parole, principalement ceux de Searle. On a travaillé avec le groupe *Ci-dit* sur le Discours Rapporté (Sophie Marnette, Cambridge, Juan-Maria Lopez, Cadix, L. Rosier, ULB) et avec une collègue de Madrid (Amalia Rodriguez Somolinos) sur les connecteurs du point de vue pragmatique. Pour le volet concernant la diachronie comme processus dynamique, on a utilisé les travaux sur la grammaticalisation de Traugott, Lehman, Klausenburger, les travaux de Sweetser rattachant le changement à des mécanismes cognitifs, la synthèse de Kortmann sur l'évolution de la subordination dans les langues européennes. On a entretenu un dialogue particulier et personnel avec Hava Bat-Zeev Shyldkrot.

Les coopérations internationales institutionnalisées sont les suivantes :

- 1) Le projet s'est associé au travail sur « les connecteurs » mené par les universités de Paris X Nanterre et de Sousse (Tunisie), dans le cadre de l'accord de la Convention Mixte de Coopération Universitaire France-Tunisie (coordonné par Danielle Leeman et Jacqueline Bacha).

2) Le projet doit s'intégrer dans la coopération avec l'Université Linguistique de Kiev dans le cadre du Programme « Dnipro » sous l'égide de la Commission Mixte franco-ukrainienne relevant du Ministère des Affaires Étrangères.

Le projet de coopération, coordonné par Annie Bertin et Valentine Fesenko, dans le cadre du Programme « Dnipro » porte sur « La catégorisation en sciences humaines ».

Dans le domaine linguistique, on abordera le thème de la catégorisation du point de vue de (i) la typologie et la dialectologie ; (ii) la périodisation ; (iii) les catégories grammaticales ; (iv) les genres.

Ce sera donc l'occasion d'envisager le problème de la grammaticalisation et de la lexicalisation à travers des langues différentes, le programme « Dnipro » devant permettre l'organisation d'une journée d'étude sur ce thème (projet soumis au Conseil Scientifique de l'Université Linguistique de Kiev en décembre 2003 puis à la Commission Paritaire en mai 2004).

### **Contexte national**

On a travaillé principalement en relation avec le GDR « Diachronie du français », dirigé par B. Combettes et C. Marchello-Nizia. On a bénéficié des corpus fournis par l'ATILF à partir de la base de données textuelles du *DMF*.

### **Thèmes/questions/hypothèses**

1) La description de l'évolution des mots grammaticaux (principalement connecteurs et prépositions) invite à confronter différents modèles proposés pour rendre compte du changement linguistique : grammaticalisation ; besoin contradictoire d'iconicité et de catégories floues ; implication subjective de l'énonciateur ; rôle de l'oral dans le changement linguistique ; rapports de l'oral et de l'écrit dans le changement ; évolution typologique et faits culturels.

2) On part d'analyses micro-linguistiques empiriques sur les distributions contextuelles, avec l'aide des corpus informatisés. On observe comment interagissent dans le changement des régularités d'ordre cognitif et typologique, des faits culturels, des besoins communicationnels, des facteurs de « style », parmi lesquels les spécificités de la communication orale, celle d'un écrit lié à l'oralité, réglé par une « syntaxe pour l'oreille » auquel se substitue un écrit coupé de l'oralité, réglé par d'autres spécificités syntaxiques.

3) On s'interroge sur la pertinence des catégories grammaticales : connecteur et adverbe ; connecteur et interjection ; lexèmes et mots grammaticaux.

4) On s'interroge, à partir des réécritures et traductions, sur la catégorisation d'une langue dans le temps et l'espace : variation dialectale et changement diachronique, et éventuellement sur leur interrelation.

### ***Bilan***

### **Actions réalisées**

1) Différentes publications et communications (voir bibliographie ci-dessous et bibliographie générale).



- 2) Participation aux travaux du GDR « Diachronie » sous la direction de C. Marchello-Nizia, B. Combettes et S. Prévost (conférences de A. Bertin au colloque Diachro I, janvier 2002 ; 28/3/2002 : « Processus métaphoriques et changement linguistique »).
- 3) Participation à différents travaux collectifs d'édition (*Anthologie* du théâtre médiéval pour la collection de la Pléiade sous la direction de J.-P. Bordier ; traduction dans le cadre du groupe de l'École Normale supérieure (Fontenay-Lyon) de la *Cronica* de Salimbene de Adam (900 pages ; XIIIe siècle).

## ***Perspectives***

### **Résultats attendus**

- 1) Contribution à une description exhaustive d'une synchronie sur laquelle on ne dispose pas d'ouvrage de référence.
- 2) Réflexion, en parallèle, sur les paramètres du changement linguistique

### **Actions à venir**

- 1) Colloque à Sousse, automne 2004, « Les connecteurs, diachronie-synchronie » (participants en cours d'évaluation). Organisation : J. Bacha, A. Bertin, N. Flaux.
- 2) Publication des *Actes* du colloque de Sousse aux Presses Universitaires d'Artois (2005).
- 3) Coordination du n° 52 de la revue *LINX* A. Bertin, J. Bacha : « Les connecteurs - approche diachronique ».
- 4) Table ronde sur une grammaire du moyen français et colloque sur « Les mots grammaticaux en moyen français ».
- 5) Communication au colloque « Diachro II » organisé par C. Marchello-Nizia, B. Combettes, S. Prévost, janvier 2004 à l'ENS, « Variation et changement » ; A. Bertin : « Récrire en *français* au 15<sup>e</sup> siècle : variation et/ou changement ? ».
- 6) Participation au numéro de *Langue Française* coordonné par Amalia Rodriguez-Somolinos, D. Lagorgette et E. Oppermann sur « Énonciation et pragmatique : approche diachronique » (2<sup>e</sup> semestre 2005) ; A. Bertin : « Interjection et connecteur- *E(t)* en ancien et moyen français ».
- 7) Concernant plus spécifiquement, le Projet « Grammaire de référence du Moyen Français » :

Le projet voudrait produire une référence aux chercheurs, linguistes et utilisateurs de textes, pour la période XIV<sup>e</sup>-XVI<sup>e</sup> siècle. La communauté des chercheurs en histoire de la langue convient que cette période est laissée dans l'ombre entre l'ancien français et le français classique, qui ont fait l'objet de synthèses actualisées dans les dernières années (ouvrages de C. Buridant et N. Fournier), alors même que le moyen français, au sens large (XIV<sup>e</sup>-XVI<sup>e</sup>), est une époque charnière où se jouent des changements décisifs, à une période où joue la « grammatisation » mise en lumière par S. Auroux. De plus, on peut en faire une approche renouvelée par les banques de données textuelles comme celles de l'ATILF.

Le moyen français a souvent été considéré du point de vue d'une désintégration de l'ancien français. Il s'agirait au contraire, tout en décrivant les spécificités de cette synchronie

(plutôt de ces synchronies), de la replacer dans la dynamique de l'évolution qui aboutit aux états ultérieurs.

On se propose deux types de travail :

- a) Monographies : bilan et étude renouvelée par documentation informatisée par types de problèmes (déterminants, connecteurs, ordre des constituants, l'adjectif ; morphologie et sémantique du système verbal etc.).
- b) Synthèse : rédaction d'un ouvrage de synthèse décrivant la langue de l'époque dans sa spécificité et sa variation tout en la replaçant dans les processus diachroniques généraux et l'histoire du français.

Plus précisément, il conviendra d'établir :

- a) une bibliographie et un bilan des études antérieures ;
- b) la liste des points d'étude à renouveler, en particulier :
  - b1) description des systèmes morphologiques et syntaxiques
  - b2) description de la variation ; paramètres dialectaux, textuels, diaphasiques, chronologiques
  - b3) évaluer le rôle de la grammaticalisation, de l'iconicité, de la typologie dans l'évolution
  - b4) évaluer le rôle de l'énonciation et de la subjectivité
  - b5) incidences de la grammatisation
  - b6) évaluer la part respective de la stabilité et du changement
  - b7) variation et changement
  - b8) apprécier la vitesse de l'évolution

### **Production scientifique**

- Anis J. 2002. « Petite contribution à l'étude diachronique de l'orthographe à partir d'éditions anciennes des Odes de Ronsard (1550-1584) ». *LINX*, n° spécial, 45-62.
- Bertin A. 2001. « *Maintenant* : un cas de grammaticalisation ? ». *Langue Française*, 130, 42-64.
- Bertin A. 2002. « L'émergence du connecteur *en effet* en moyen français ». *LINX*, 46,
- Bertin A. A paraître. *La mise en scène de la parole dans le Roman de Renart*. Paris : L'Harmattan.
- Bertin A. A paraître. *Courtois d'Arras. L'aveugle et son valet*. Paris : Gallimard, Pléiade.
- Combes A., Bertin A. 2001. *Ecritures du Graal*. Paris : PUF.
- Leeman D. A paraître. « Des jumeaux presque indistinguables : différents et divers ». Mélanges offerts à G. Kleiber.
- Lehman S. 2002 « Le chevalier et sa chevalière ou chevaleresse ? Le développement d'un suffixe spécifique pour les représentants féminins des estats au Moyen Age ». *LINX*, n° spécial, 187-196.
- Lignereux M. 2002. « Anaphore et promesse en ancien français ». *LINX*, n° spécial, 209-220.
- Weill I. 2002. « Dieu qui fit parler l'image. Un Problème d'emploi de l'article dans une formule épique médiévale ». In Anis J., Kleiber G. (éds), *LINX « Du sens au sens. Hommage à Michel Galmiche »*, 47, 169-174.

## Thèse inscrite

Dana Sleiman. Directeur A. Bertin.

## Références générales sur le projet

- Bart-Zeev Shyldkrot H. 1997. « Grammaticalisation et évolution de la langue : théories et systèmes ». *Modèles linguistiques*, 34 (XVII : 2), 27-36.
- Hopper P., Traugott. 1993. *Grammaticalization*. Cambridge : Cambridge University Press.
- Klausenburger. J. 2000. *Grammaticalization. Studies in Latin and Romance Morphosyntax*. Amsterdam : John Benjamins.
- Kortmann B. 1997. « The evolution of the adverbial subordinators in Europe ». In Schmid M., Austin J. R., Stein D. (éds), *Historical Linguistics*. Amsterdam : Benjamins, 213-228.
- Lehmann C. 1995. *Thoughts in Grammaticalization*. Munich - Newcastle : Lincom Europa.
- Marchello-Nizia C. 2000. « Les Grammaticalisations ont-elles une cause ? ». *L'Information grammaticale*, 87, 3-9.
- Marchello-Nizia C. 1997. « Evolution de la langue et représentations sémantiques : du « subjectif » à l'« objectif » en français ». In Fuchs C., Robert S. (éds). *Diversités des langues et représentations cognitives*. Paris : Ophrys.
- Marchello-Nizia C. 2001. « La linguistique diachronique ; grammaticalisations et sémantique du prototype ». *Langue française*, 130.
- Ramat A., Hopper P. (éds). 1998. *The Limits of Grammaticalization*. Amsterdam : John Benjamins.
- Schlieben-Lange B. 1992. « The history of subordinating conjunctions in some Romance languages ». In Gerritsen M., Stein D. (éds), *Internal and external factors in syntactic change*. Berlin : Mouton de Gruyter, 341-354.
- Sweetser E. 1990. *From Etymology to Pragmatics, Metaphorical and cultural aspects of semantic structure*. Cambridge : Cambridge University Press.

### **3.4. TEXTES LITTÉRAIRES**

#### **3.3.1. EDITION SCIENTIFIQUE DE CHARLES NODIER**

##### **Participants**

*Membre rattaché à titre principal* : Jean-François Jeandillou (responsable)

*Doctorant* : Solange Gularté

*Collaborations extérieures* :

Daniel Bilous (Université de Toulon)

Jacques-Philippe Saint-Gérand (Université de Clermont-Ferrand)

Daniel Sangsue (Université de Neuchâtel)

##### **Discipline(s) et/ou sous-discipline(s) concernée(s)**

Linguistique textuelle, histoire de la poétique et de la stylistique, sémiotique littéraire, titrologie

##### **Problématique**

Il s'agit de comparer la classification des genres élaborée par Nodier avec, d'une part, les typologies de son temps, d'autre part les recherches linguistiques développées dans ses *Questions de littérature verbale* regroupant ses travaux sur l'onomatopée (1808-1828) et sur l'histoire des langues (1810-1824) ainsi que ses *Notions élémentaires de linguistique* (1834).

##### **Contexte**

Ce programme de recherche, engagé dans le cadre de l'Institut universitaire de France, avait pour objectif premier l'édition scientifique des *Questions de littérature légale* (1812 et 1828) de Charles Nodier. L'établissement du texte (210 p.), confrontant les modifications (ajouts, retraits, reformulations ou rectifications, etc.) entre les deux états de 1812 et de 1828, est précédé d'une étude d'ensemble (LX p.), qui situe le livre dans son contexte historique et scientifique. Quelque 500 annotations exégétiques sont intégrées en bas de page. Le volume a paru aux éditions Droz, en 2003.

Cette réalisation s'inscrit par ailleurs en rapport avec l'édition en cours des *Œuvres de Nodier* (Les Belles-Lettres), qui laisse de côté les écrits techniques et scientifiques.

Ce programme éditorial entre en relation avec l'actualité des recherches sur l'origine des langues (CNRS, notamment), sur l'histoire de la linguistique naturaliste au XIXe siècle (Desmet 1996), sur le développement de la dialectologie et de la lexicographie (ou dictionnaire ; cf. Vaulchier 1984).

## **Thèmes de recherche**

1) Les théories de l'imitation, de la citation, de l'allusion ou encore du plagiat sont fondatrices des approches modernes en vigueur dans les champs juridiques et épistémologiques. La typologie des textes élaborée par Nodier - l'une des premières en son genre, en Europe, après la mise en place de la législation sur le droit d'auteur et la protection des œuvres - prolongeait les très anciennes recensions de La Croix du Maine et de La Monnoye (1584) puis celles de Bayle (1697). Mais elle préfigurait surtout nombre de travaux ultérieurs sur la théorie de l'intertextualité (Genette 1981). En tant que grammairien et lexicographe, l'auteur appuie de fait son argumentaire sur une appréhension linguistique des textes ; c'est même une dimension rigoureusement énonciative voire pragmatique qui apparaît dans les analyses relatives à la signature (comprise comme acte performatif) et au pastiche. Il convient donc d'en cerner l'efficace, rétrospectivement, et aussi les limites au vu de l'évolution de la jurisprudence en la matière.

2) Dans la mesure où les thèses génétiques de Nodier se fondent en partie sur la mimologie et l'onomatopéisme (Nodier 1808), on s'interrogera sur l'héritage proprement cratylien qu'assume l'auteur (De Brosses 1765 ou Court de Gébelin 1775), et aussi sur la place qu'occupe ce dernier dans le champ de la linguistique naturaliste, encore vivace jusqu'à la fin du XIXe siècle (Desmet 1996). Le rapport à la philosophie du langage, distinct de celui de Condillac (1746) ou de Rousseau (1721), s'oppose également au théisme de Bonald (1817) en prenant acte de la diversité des idiomes *ab origine* : d'où l'intérêt crucial que Nodier porte aux patois, s'opposant fermement au centralisme jacobin de Grégoire (de Certeau 1975). Résolument modernes, ses prises de position sur l'étymologie donnent toute leur place aux découvertes de Raynouard (1836) et préfigurent le développement de la dialectologie, de l'onomastique (anthroponymique et toponymique) à travers toute l'Europe.

Il importera donc de se demander comment Nodier articule son Histoire de la parole et de l'écriture sur son activité de lexicographe. Quel rôle - outre celui de la seule conservation de « l'étymologique immédiate » (Jeandillou 2002 et 2004) - il reconnaît aux questions d'orthographe ou de néologie (Paris 1868 ; Saint-Gérand 1976). Enfin, c'est toute la longue tradition des conceptions d'une langue universelle qui se trouve interrogée par les postulats naturalistes de Nodier (Couturat 1903).

## ***Bilan***

### **Étapes passées**

- 1) Numérisation des textes.
- 2) Confrontation critique des deux états de 1812 et 1828.
- 3) Établissement de l'appareil critique (500 apostilles)
- 4) Rédaction d'une introduction au volume (60 p.)
- 5) Réalisation de l'index exhaustif (800 entrées, 13 p.)
- 6) Articles et communications recensés ci-après.

## **Résultats obtenus**

- 1) Publication du volume des *Questions de littérature légale. Du plagiat, de la supposition d'auteur, des supercheries qui ont rapport aux livres* (1812 et 1828) de Charles Nodier (Genève, Droz, coll. « Histoire des idées », 2003, un vol. in-8°, LX-212 p.).
- 2) Publication de six articles scientifiques.

## ***Perspectives***

Dès les années 1805, Nodier projetait de publier :

- 1) un essai sur la *Philosophie des langues*, ensuite rebaptisé *Théorie des langues primitives* ;
- 2) une *Théorie de l'alphabet naturel* établissant « la disposition philosophique » de tous les sons employés dans le langage, leur attribuant « des figures fixes », instituant encore un alphabet - « propre par la quantité des figures et la précision des valeurs, à représenter distinctement toutes les consonances usitées » - et une « orthographe, non usuelle, mais scientifiquement et monumentalement bonne, qui a l'avantage de devenir applicable à toutes les langues connues, et d'en rendre l'étude et la prononciation excessivement faciles » ;
- 3) un *Archéologue* enfin, où l'on trouverait « sous chaque son radical toutes les dérivations qu'il a subies, avec un examen raisonné des causes qui ont influé sur la modification des sens ».

En février 1810 paraît un Prospectus de quelques pages esquissant les Prolégomènes de l'Archéologue, ou système universel et raisonné des langues. Onze années plus tard, un contrat sera signé avec le libraire Barba en vue d'un Essai sur la philosophie des langues, ou Théorie de l'alphabet naturel, qui n'aboutira pas.

Mises en sommeil pendant un quart de siècle, ces réflexions ne devaient voir le jour, sous une forme réduite mais néanmoins fidèle à leurs principes, qu'avec la publication des *Notions élémentaires de linguistique*. Signé alors par un académicien fraîchement élu (en octobre de l'année précédente), l'ouvrage réunit un ensemble de chroniques d'abord publiées dans le *Temps* (de septembre 1833 à juillet 1834), et fonde un examen diversifié du lexique et de l'alphabet, de l'étymologie et de la néologie, des noms propres, des patois ou encore des langues de convention sur la seule naturalité du signe, corollaire constant de la « langue organique ». C'est sans conteste à la motivation sémiotique que l'auteur s'intéresse, en s'appuyant sur les thèses de Court de Gébelin et du président de Brosses comme sur les pasigraphies de Leibniz, Dalgarno ou Wilkins, et sur les travaux entrepris par David de Saint-Georges.

L'édition de ce texte, qui resitue ces réflexions dans leur cadre synchronique (travaux de Domergues, de Volney, de Bonald, etc.), en montre aussi l'influence - le terme même de *linguistique* a connu la faveur que l'on sait par le biais crucial de cette *vulgarisation* - et le rayonnement, tant sur la 6<sup>e</sup> éd. du *Dictionnaire de l'Académie* que sur les travaux de Raynouard ou de Villemain. Les divers travaux lexicographiques de Nodier - contribution à plusieurs dictionnaires (de Boiste, de Verger, etc.), réalisation d'un important *Examen critique des dictionnaires* - seront ici remis en lumière grâce à la publication, en addenda à ces *Notions élémentaires*, de sept articles majeurs :

- Deux Suppléments aux *Notions élémentaires de linguistique* (nov. 1834, en réponse aux objections du sanscritiste Ferdinand d'Eckstein).
- *Archéologue ou Système universel et raisonné des langues* (1810).
- « Comment les patois furent détruits en France » (1835).
- « De quelques langues artificielles » (1835).
- « Du langage factice appelé macaronique » (1835).
- « Des nomenclatures scientifiques » (1835).
- « Langue française » (longue étude diachronique pour le *Dictionnaire encyclopédique* de Duckett, 1836 et 1854).

### **Étapes à venir**

- 1) Numérisation du volume de 1834.
- 2) Numérisation des divers articles de 1834 à 1836.
- 3) L'établissement du texte, confrontant les modifications (ajouts, retraits, reformulations ou rectifications, etc.) entre les deux états publiés d'abord dans le journal *Le Temps* puis en volume, sera précédé d'une étude d'ensemble (60 p.), situant le livre dans son contexte historique et scientifique.
- 4) Quelque 1100 annotations exégétiques sont prévues.

### **Production scientifique**

- Jeandillou J.-F. 2001. *La vie et l'œuvre des auteurs supposés*. Genève : Droz. Coll. « Titre courant ».
- Jeandillou J.-F. 2001. « Nodier et les Questions de littérature légale ». *Revue d'Histoire littéraire de la France*, 1, 57-75.
- Jeandillou J.-F. 2002. « Nodier ou la naïveté linguistique ». *Poétique*, 129, 91-104.
- Jeandillou J.-F. 2003. « Non bis in idem. Le plagiat au risque de la linguistique ». *Poétique*, 134, 183-192.
- Jeandillou J.-F. 2003. *Questions de littérature légale. Du plagiat, de la supposition d'auteur, des supercheres qui ont rapport aux livres, de Charles Nodier*. Genève : Droz. Coll. « Histoire des idées ».
- Jeandillou J.-F. 2003. « Un objet d'horreur pour les connaisseurs : le jugement de (Gaston) Paris ». *Revue des langues romanes*, 125-138.
- Jeandillou J.-F. 2004. « Aléas de la réception : Nodier et les romances pseudo-graphes de Clotilde de Surville ». *Fragmentos*. Florianopolis : Université fédérale de Santa Catarina (Brésil), n° spécial Charles Nodier.
- Jeandillou J.-F. 2004. « De l'étymologie immédiate selon Charles Nodier ». In Haßler G., Dietzel U. (éds), *Geschichte der Sprachwissenschaft*. Munster : Nodus Publikationen, 71-88.

### **Thèse inscrite**

Gularté Solange, *Le statut des typologies en linguistique textuelle. Étude historique et épistémologique*. Directeur J.-F. Jeandillou.

## Contrat / financement

Aide à la publication de l'ouvrage de la part de l'Institut universitaire de France (soutien de programme).

## Références générales sur le projet

- Bäckvall H. 1996. « Néologismes nodiériens ». *Revue romane*, 31-2, 9-32.
- Bayle P. 1695-1697. *Dictionnaire historique et critique*. Genève : Slatkine, 1995.
- Bonald L. G. A. 1817. « De l'origine du langage » et « De l'origine de l'écriture ». *Recherches philosophiques sur les premiers objets des connaissances morales. Œuvres complètes*, chap. II et III. Paris : Migne, t. III, 1859, 61-122 et 122-144.
- de Brosses Ch. 1765. *Traité de la formation mécanique des langues et des principes physiques de l'étymologie*. Paris.
- de Certeau M. et al. 1975. *Une politique de la langue : la Révolution française et les patois*. Paris : Gallimard.
- Condillac E. 1746. *Essai sur l'origine des connaissances humaines*. Rééd. Paris : Galilée, 1973.
- Court de Gébelin A. 1775. *Monde primitif, considéré dans l'Histoire naturelle de la parole, ou Origine du langage et de l'écriture*. Paris.
- Couturat L., Léau L. 1903. *Histoire de la langue universelle*, repr. Hildesheim, 2001.
- Desmet P. 1996. *La Linguistique naturaliste en France (1867-1922). Nature, origine et évolution du langage*. Louvain : Peeters, 1996.
- Genette G. 1981. *Palimpsestes*. Paris : Le Seuil.
- Jeandillou J.-F. 1994. *Esthétique de la mystification*. Paris : Éditions de Minuit.
- La Croix du Maine. 1584. *Bibliothèque ou Catalogue général de toutes sortes d'auteurs qui ont écrit en français depuis cinq cents ans et plus jusqu'à ce jour d'hui*. Ed. avec la Bibliothèque française d'Antoine Du Verdier et les Remarques de Bernard de La Monnoye, par Rigoley de Juvigny, 1772-1773, repr. Genève : Slatkine, 1969, 6 vol.
- Meschonnic H. 1985. « La nature dans la voix ». *Nodier. Dictionnaire raisonné des onomatopées françaises*. Reprint de la 2<sup>ème</sup> éd. Hildesheim : Olms, 11-104.
- Nodier Ch. 1808. *Dictionnaire raisonné des onomatopées françaises*. 2<sup>ème</sup> éd. augmentée, 1828, repr. Mauvezin : Trans-Europ-Repress, 1984.
- Nodier Ch. 1828. *Examen critique des dictionnaires de la langue française ou Recherches grammaticales et littéraires sur l'orthographe, l'acception, la définition et l'étymologie des mots*. Paris : Delangle.
- Nodier Ch. 1827-1834. *Œuvres complètes*. Paris : Renduel, 12 vol. (éd. très sélective).
- Nodier Ch. 1829. *Mélanges tirés d'une petite bibliothèque ou Variétés littéraires et philologiques*. Repr. Bassac : Plein Chant, 2000.
- Paris G. 1868. « De l'histoire de l'orthographe française ». *Bulletin du bibliophile*, 473-507.
- Raynouard F. 1816. *Grammaire romane*. Paris.
- Saint-Gérard J.-Ph. 1976. « La question de la réforme de l'orthographe entre 1825 et 1851 ». *Le Français moderne*, XLIV, 1, 28-56.
- Rousseau J.-J. 1781. *Essai sur l'origine des langues*. Rééd. Paris : Gallimard, 1990.
- Vaulchier H. de. 1984. *Nodier et la lexicographie française. 1808-1844*. Paris : INaLF-CNRS / Didier-Érudition.



### 3.4.2. MÉTRIQUE DE L'ALEXANDRIN AU TOURNANT DES XIXE ET XXE SIÈCLES

#### Nouveau projet

#### Participants

*Membres rattachés à titre principal au laboratoire :* Jean-François Jeandillou (responsable), Michel Arrivé

*Collaborations extérieures :*

Jean-Louis Aroui (Université Paris VIII)

Benoît de Cornulier (Université de Nantes)

#### Discipline(s) et/ou sous-discipline(s) concernée(s)

Métrie et prosodie, poésie et versification, linguistique textuelle

#### Problématique

Il s'agit d'examiner les contraintes majeures régissant la versification des alexandrins, dans la production poétique française, au tournant des XIXe et XXe siècles. En s'appuyant notamment sur la méthode métrico-métrique de Benoît de Cornulier (principalement exploitée, jusqu'à présent, pour la littérature classique et les années 1810-1880), on procédera spécialement à l'analyse des textes signés par Raymond Roussel entre 1897 et 1917 (soit près de trente mille vers).

C'est en particulier le statut de la césure et de la rime qui devra être reconsidéré, dans la mesure où les œuvres (publiées de son vivant ou posthumes) de Roussel présentent comme caractéristique exceptionnelle de placer en position 6 aussi bien des clitiques (C) que des prépositions (P) monosyllabiques, tout en évitant systématiquement le « -e » féminin posttonique de rang 6 ou 7 ainsi que l'enjambement d'un hémistiche à l'autre (critère MFs6 de Cornulier). A la différence des auteurs français qui, depuis les années 1850, ont commencé à mettre en question la césure par l'usage de CPMFs6, Roussel opère une bipartition radicale entre ces divers usages, afin d'opposer de façon sensible (et massive) la césure proprement dite (liée aux unités lexicales) et la coupe (déterminée par l'accentuation des fins de syntagmes). La position 12 (dévolue à la rime) se caractérise en revanche par un maintien de l'exclusion des critères CP (associés, évidemment à MF).

Via la numérisation des textes en question, dûment codés, on pourra dresser un bilan statistique d'un corpus manifestement spécifique, qui bouleverse peu ou prou les rapports (fussent-ils modernes) entre métrie et prosodie.

#### Contexte

1) Journée d'études Roussel dans le cadre de l'AIEF (juin 2003), organisée par Houppermans S. (Université de Leyde).

2) Edition des *Œuvres complètes* de Raymond Roussel. Paris : Fayard-Pauvert, en cours (6 vol. depuis 1997). Plusieurs volumes présentent l'édition originale des textes, pour la

plupart inconnus, découverts en 1989 (déposés à la BnF), ainsi que les variantes manuscrites des autres.

### **Thèmes de recherche**

Quatre thèmes seront plus particulièrement étudiés :

- 1) Histoire de la versification française.
- 2) Analyse linguistique du mètre.
- 3) Statut morpho-syntaxique de la césure et de la rime.
- 4) Prosodie accentuelle et syntaxe.

### **Étapes à venir**

- 1) Numérisation des textes de Raymond Roussel.
- 2) Codage métrico-métrique du corpus.
- 3) Établissement des bases de données.
- 4) Statistique des marques CPMFs6 / CPMF12 et des distributions syntagmatiques (la fin de chacun des deux hémistiches).
- 5) Repérage et évaluation d'éventuels trimètres (444) et de divers mètres de substitution au 66 dominant.
- 6) Analyse quantitative des résultats, avec prolongements stylistiques et théoriques.

### **Résultats attendus**

- 1) Numéro de la revue *SEMEN* (Labo LASELDI, Univ. Besançon) sur « L'ordre des mots » (2005).
- 2) Publications des résultats de la recherche engagée et mise à disposition de la communauté de la base de données informatique.

### **Références générales sur le projet**

- Beaudoin V. 2002. *Mètre et rythme du vers classique*. Paris : Champion.
- de Cornulier B. 1982. *Théorie du vers. Rimbaud, Verlaine, Mallarmé*. Paris : Le Seuil.
- Gouvard J.-M. 2000. *Critique du vers*. Paris : Champion.
- Murat M. (éd.). 2000. *Le vers français. Histoire, théorie, esthétique*. Paris : Champion.
- Roubaud J. 1978. *La vieillesse d'Alexandre*. Paris : Maspero.

## 4. Ressources linguistiques

---

- 1) Grammaire de référence du français contemporain
- 2) Lexique syntaxique du français
- 3) *Terminalf* : ressources terminologiques en langue française

### 4.1. GRAMMAIRE DE RÉFÉRENCE DU FRANÇAIS CONTEMPORAIN

#### **Participants**

*Membres rattachés à titre principal* : Annie Delaveau (responsable), Jacques Anis, Marianne Desmets, Françoise Gadet, Danielle Leeman

*Doctorant* : Gaëlle Tassin

*Collaborations extérieures* :

Anne Abeillé (Université Paris VII, UMR 7110)

Olivier Bonami (Université de Haute Bretagne, Rennes 2)

Danièle Godard (Université Paris VII, UMR 7110)

Elisabeth Delais (CNRS UMR 5610)

Georges Kleiber (Université Marc Bloch Strasbourg)

Jean-Marie Marandin (Université Paris VII, UMR 7110)

Michèle Noailly (Université de Bretagne occidentale Brest)

Liliane Tasmowski (Université d'Anvers)

Anne Zribi-Hertz (Université Paris VIII)

#### **Discipline(s) et/ou sous-discipline(s) concernée(s)**

Syntaxe, sémantique, pragmatique, prosodie, graphématique

#### **Problématique**

Le projet d'une grammaire de référence du français contemporain vise à enregistrer les savoirs acquis sur la langue française au cours des cinquante dernières années du XXe siècle. Ces savoirs appartiennent principalement à la description de la syntaxe : éclairée par des vues en phonologie, en morphologie, en sémantique, la description syntaxique constitue l'armature du projet, elle relève de plusieurs champs théoriques, comme la grammaire générative, la grammaire distributionnelle, la grammaire lexicale, la grammaire

fonctionnelle, qui servent ici d'instruments d'observation et d'analyse, mais ne sont pas le but de l'ouvrage. Un des problèmes à résoudre est de construire un ensemble compatible de descriptions. Les apports de la linguistique de corpus peuvent et doivent être intégrés de façon éclairante aux descriptions particulières afin de surmonter les apories de l'attitude normative qui conduit à l'exclusion de la description grammaticale des formes qui relèvent de l'hétérogénéité inhérente au fonctionnement linguistique.

### **Contexte international**

Plusieurs pays européens se sont déjà dotés d'une grammaire moderne de leur langue, l'Italie, l'Espagne, la Grande-Bretagne, le Portugal (cf. bibliographie). La France n'a pas encore fait ce travail, qui est donc d'actualité. Les collaborations à venir incluent des universitaires belges.

### **Contexte national**

Le travail engagé associe des universitaires et des chercheurs appartenant à l'équipe Modyco et à une équipe de Paris VII, l'UMR 7110. Les coordonnateurs de chapitre appartiennent à diverses universités et équipes de recherche française, et à une université belge.

### **Thèmes de recherche**

Regrouper les savoirs grammaticaux sur la langue française contemporaine, faire le point sur la compatibilité des descriptions obtenues dans des théories diverses, énumérer les formes de la langue française, cela constitue un enjeu pour mesurer l'apport des travaux faits à ce jour dans les domaines abordés.

## ***Bilan***

### **Étapes passées**

- 1) Présentation du projet : juin 2002.
- 2) Constitution de l'équipe de pilotage : juillet 2002.
- 3) Réunions de l'équipe de pilotage : septembre 2002 - juillet 2003.
- 4) Écriture du plan : septembre 2002 - juillet 2003.
- 5) Contacts avec les coordonnateurs de chapitre : juillet - novembre 2003.

### **Résultats obtenus**

Définition du plan de la grammaire, l'objectif étant un volume d'environ 2 000 pages. Le plan comporte 19 chapitres, précédés d'une introduction.

## ***Perspectives***

### **Étapes à venir**

- 1) Réunions des coordonnateurs de chapitres : novembre 2003 - juin 2004. Les coordonnateurs de chapitres ont été choisis, ce qui justifie le passage d'un comité de pilotage de 6 personnes à une équipe de travail de 15 personnes.
- 2) Choix des auteurs spécialistes des questions traitées dans chacun des chapitres.
- 3) Rédaction de l'introduction (novembre 2003 - janvier 2004).
- 4) Etablissement de la terminologie.
- 5) Définition des normes de rédaction pour les auteurs.

### **Résultats attendus**

A l'issue de l'année 2003-2004, les tâches de chacun auront été définies. Moyennant des concertations chapitre par chapitre, et des réunions de coordonnateurs de chapitre, la rédaction des chapitres, qui devrait débuter en 2004, devrait être achevée en 2006.

### **Thèses soutenues**

Marianne Desmets, 2001. *Les typages de la phrase en HPSG : le cas des phrases en comme*. Directeur A. Delaveau.

Ioannis Valassis, 2001. *Syntaxe positionnelle et épithètes : le cas des adjectifs relationnels*. Directeur A. Delaveau.

### **Thèse inscrite**

Gaëlle Tassin, *Ellipses et structures parallèles : la constituance dans les nouvelles syntaxes*. Directeur A. Delaveau.

### **Contrat / financement**

Ce projet est financé par l'ILF.

### **Références générales sur le projet**

Biber D., Stig J., Leech G., Conrad S., Finegan E. 1999. *Longman Grammar of Spoken and Written English*. Londres : Longman.

Bosque I., Delmonte V. 1999. *Gramática descriptiva de la lengua española*. Madrid : Espasa Calpe. 3 vol.

Grevisse M. 1993. *Le bon usage*. Bruxelles : Duculot (13<sup>ème</sup> édition).

Huddleston R., Pullum G. 2002. *The Cambridge Grammar of the English Language*. Cambridge : Cambridge University Press.

Renzi L., Salvi G. 1988-1991. *Grande Grammatica italiana di consultazione*. Bologne : Il Mulino.

Riegel M., Pellat J.-C., Rioul R. 1994. *Grammaire méthodique du français*. Paris : PUF.  
Wilmet M. 1997. *Grammaire critique du français*. Paris : Hachette et Bruxelles : Duculot.

## 4.2. LEXIQUE SYNTAXIQUE DU FRANÇAIS

### Nouveau projet

#### Participants

*Membres rattachés à titre principal* : Sylvain Kahane (responsable), Karine Baschung, Marcel Cori, Marianne Desmets

*Collaborations extérieures* :

Philippe Blache (LPL, Université de Provence)  
Didier Bourigault (ERSS, Toulouse 2)  
Lionel Clément (ATOLL, INRIA Rocquencourt)  
Eric de la Clergerie (ATOLL, INRIA Rocquencourt)  
Bertrand Gaiffe (Langue et Dialogue, LORIA)  
Kim Gerdes (ERSS, Bordeaux 2)  
Igor Mel'cuk (OLST, Université de Montréal)  
Piet Mertens (Université de Leuven)  
Alexis Nasr (Lattice, Paris 7)  
Alain Polguère (OLST, Université de Montréal)  
Azim Roussanaly (Langue et Dialogue, Nancy 2)  
Jesse Tseng (Langue et Dialogue, LORIA)

#### Discipline(s) et/ou sous-discipline(s) concernée(s)

Syntaxe, lexicologie

#### Problématique

On appelle lexique syntaxique un lexique des unités lexicales de la langue indiquant en particulier leur sous-catégorisation et leurs particularités syntaxiques (du point de vue de l'ordre des mots ou de la compatibilité avec des constructions syntaxiques particulières). Il s'agit d'une ressource différente d'un lexique morphologique, mettant en relation les formes fléchies avec les lemmes, ou d'un lexique sémantique donnant les définitions sémantiques des unités lexicales. Le développement d'un lexique syntaxique pose de nombreuses difficultés théoriques :

- a) l'identification des unités lexicales et notamment des expressions figées (locutions) et semi-figées (collocations) ;
- b) l'identification des différents compléments sous-catégorisés d'un lexie et la description des différents régimes ;
- c) la description des particularités syntaxiques de chaque lexie : ordre des mots, combinaison avec des éléments grammaticaux (voix, déterminant, copule, etc.), etc.

Le dernier point nous intéressera particulièrement. Il pose la question de l'interaction entre le lexique et la grammaire. Il n'est en fait pas possible de développer indépendamment un lexique et une grammaire. Prenons l'exemple de la passivation : attribue-t-on la fonction de

*complément d'objet direct* aux seuls compléments passivables ou bien donne-t-on une acception plus large au terme (incluant par exemple le complément de *valoir* : *un peu que ça les vaut*) et contrôle-t-on la passivation par un trait spécifique ? Quoi qu'il en soit, la grammaire ne pourra fonctionner correctement que si elle « sait » comment l'information est encodée dans le lexique.

L'importance de l'interaction lexique-grammaire a conduit au développement de grammaires lexicalisées (comme le lexique-grammaire du LADL ou les grammaires TAG lexicalisées) où lexique et grammaire font un tout indissociable. Si ces modèles sont séduisants et faciles à développer au départ, ils se sont avérés difficilement maintenables à grande échelle étant donné l'extrême redondance des informations qu'ils contiennent et leur survie n'a été possible que par une réorganisation du modèle et une séparation de fait du lexique et de la grammaire, les deux étant ensuite combinés pour obtenir la grammaire lexicalisée.

Ce projet se fixe plusieurs objectifs. Sur le plan pratique, il s'agit de fédérer les différents projets de développement de lexiques syntaxiques et d'œuvrer au partage des ressources actuelles et au développement d'une super-ressource libre. Une telle ressource est indispensable pour le développement d'un modèle linguistique du français, et elle ne peut être assumée par un seul chercheur ou même une seule équipe de recherche. Seul un travail coopératif peut permettre de développer un lexique syntaxique du français à grande couverture.

Sur le plan théorique, il s'agit d'évaluer la qualité des lexiques actuels, de réfléchir sur le type d'informations que doivent contenir ces lexiques et surtout de proposer un encodage de ces informations qui les rendent utilisables par une vaste panoplie de formalismes.

Le projet est initié par une collaboration actuelle entre Sylvain Kahane d'une part et Igor Mel'cuk et Alain Polguère d'autre part, sur la formalisation et la valorisation des dictionnaires développés depuis plus de 20 ans par l'Observatoire de Linguistique Sens-Texte de l'Université de Montréal (projet Dicobjet de l'équipe Lattice financé par une bourse Blaise Pascal attribuée à Igor Mel'cuk par la Fondation de l'Ecole Normale Supérieure). Il repose également sur les travaux de Sylvain Kahane sur l'architecture d'un modèle linguistique, l'interaction lexique-grammaire et la comparaison entre formalismes linguistiques tels que Théorie Sens-Texte, TAG, HPSG ou LFG.

Les autres partenaires du projet développent également des modèles linguistiques du français ou des analyseurs syntaxiques intégrant des modèles linguistiques (ces modèles sont généralement appelés grammaires bien qu'ils intègrent évidemment un lexique) : grammaire HPSG (Karine Baschung, Marianne Desmets, Jesse Tseng), grammaire d'arbres polychromes (Marcel Cori), grammaire de propriétés (Philippe Blache), méta-grammaire pour les TAG (Eric de la Clergerie, Bertrand Gaiffé, Azim Roussanaly), grammaire de dépendance (Didier Bourigault, Kim Gerdes, Alexis Nasr), approche pronominale (Piet Mertens), grammaire LFG (Lionel Clément). Le projet est à géométrie variable et d'autres partenaires sont attendus.

- 1) La première phase réunit essentiellement des spécialistes de la modélisation mathématique et informatique.
- 2) La deuxième phase, la constitution d'une super-ressource, fera appel à des lexicologues et des syntacticiens et pourra se raccrocher à d'autres projets du laboratoire, comme le projet de Grammaire de référence du français (cf. § 4.1.).



Il faut enfin noter qu'il n'existe pas de ressources libres de ce type connues, y compris pour l'anglais ; et que la communauté francophone est très active et en pointe dans le domaine.

### **Étapes à venir**

- 1) 2005 : inventaires des ressources existantes, questions théoriques liées au contenu des ressources.
- 2) 2006 : questions théoriques liées à la normalisation et l'utilisation des ressources par des formalismes grammaticaux.
- 3) 2007 : normalisation effective des ressources.
- 4) 2008 : constitution d'une super-ressource.

### **Résultats attendus**

- 1) Mise à disposition de ressources existantes.
- 2) Normalisation des ressources.
- 3) Constitution d'une super-ressource par la mise en commun des ressources existantes et par l'extension des ressources existantes.

### **Actions prévues**

- 1) Réunion bimensuelle des participants.
- 2) Réalisation d'un site Web avec les ressources évaluées (collaboration envisagée avec l'Atala pour l'hébergement).
- 3) Organisation d'une journée Atala.

### **Références générales sur le projet**

- Blache Ph. 2001. *Les grammaires de propriétés : des contraintes pour le traitement automatique des langues naturelles*. Paris : Hermès.
- Bourigault D., Fabre C. 2000. « Approche linguistique pour l'analyse syntaxique de corpus ». *Cahiers de Grammaire*, 25, 131-151.
- Candito M.-H. 1999. *Organisation modulaire et paramétrable de grammaires électroniques lexicalisées. Application au français et à l'italien*. Thèse de doctorat, Université Paris 7.
- Cori M., Marandin J.-M. 2001. « La linguistique au contact de l'informatique : de la construction des grammaires aux grammaires de construction ». *Histoire, Épistémologie, Langage*, 23 (1), 49-79.
- Crabbé B., Gaiffe B., Roussanaly A. 2002. « A New Metagrammar Compiler ». *Actes du colloque « TAG+6 »* (International Workshop on Tree Adjoining Grammars and Related Frameworks).
- van den Eynde K., Mertens P. 2003. « La valence : l'approche pronominale et son application au lexique verbal ». *French Language Studies*, 13, 63-104.

- Kahane S., Polguère A. 2001, « Formal foundation of lexical functions ». In Daille B., Williams G. (éds), *Workshop on Collocation*. ACL 2001, Toulouse.
- Kahane S. 2002 *Grammaire d'Unification Sens-Texte : vers un modèle mathématique articulé de la langue*. Habilitation à diriger des recherches, Université Paris 7, 82 p.
- Mel'cuk, I. A., Clas A., Polguère A. 1995. *Introduction à la lexicologie explicative et combinatoire*. Paris - Louvain-la-Neuve : Duculot.
- Polguère A. 2003. *Lexicologie et sémantique lexicale. Notions fondamentales*. Montréal : Presses de l'Université de Montréal.

### **Contrat / financement**

Des demandes de financement sont prévues auprès de l'ILF, ATIL, ACI Système Complexe, Projet Technolangue RNIL.

### **4.3. *TERMINALF* : RESSOURCES TERMINOLOGIQUES EN LANGUE FRANÇAISE**

#### **Nouveau projet**

#### **Participants**

*Membre rattaché à titre principal* : Richard Walter (responsable)

*Membre associé* :

Loïc Depecker (Université Paris III)

*Collaborations extérieures* :

Claudie Julliard (Université Paris VII)

John Humbley (Université Paris VII)

André Le Meur (Université de Rennes)

#### **Discipline(s) et/ou sous-discipline(s) concernée(s)**

Terminologie, linguistique appliquée à la terminologie, gestion des connaissances numérisées

#### **Problématique**

Le site et la base de données *Terminalf* ont pour objectif de diffuser et de faire connaître des ressources terminologiques en langue française. Ces ressources sont constituées de mémoires de terminologie présentant et structurant le vocabulaire technique ou scientifique d'un micro-domaine spécialisé. Ces inventaires sont susceptibles d'aider à la compréhension de la langue française utilisée dans les milieux de spécialistes. Le site veut offrir des données de travail et a le mérite quasi unique au monde de rendre accessibles les travaux de terminologie d'étudiants.

*Terminalf* est le résultat d'un ensemble d'idées issues directement de l'implantation en France d'internet en février 1994. Il prend en compte l'évolution de la théorie et des pratiques terminologiques confrontées à ce nouveau média. Les premières expérimentations de *Terminalf* datent de 1998. Avec le projet actuel, nous souhaitons porter un regard analytique sur les évolutions à venir en terminologie et en traitement informatique de données.

Projet pilote, *Terminalf* est orienté essentiellement vers la recherche et l'enseignement. À travers sa construction, il propose une réflexion méthodologique et épistémologique sur le traitement numérique de données structurées, en l'occurrence ici la terminologie. Le corpus de départ s'étend maintenant sur plusieurs années et sera appelé à être en constante évolution. Nous pourrions alors saisir l'évolution de la structuration de données numériques : évolution des usages, des normes, des standards, des « us et coutumes ».

*Terminalf* sera donc amené à recevoir des développements tant sur le plan des informations traitées et des notions terminologiques en jeu, que sur le plan informatique avec une

nécessaire confrontation aux systèmes automatique de gestion et d'indexation des données ainsi qu'aux outils de TAL développés au sein du laboratoire comme ailleurs.

Le site *Terminalf* sera développé conjointement par le laboratoire MoDyCo, la Société française de terminologie, le Département de langues étrangères appliquées (LEA) de l'Université de Paris III, le DESS ILTS de l'Université de Paris VII.

### **Contexte international**

Aucun projet de ce type n'existe actuellement en francophonie ou pour le traitement numérique de données terminologiques. Des tentatives ont eu lieu (projets Balneo, RIFAL) mais elles ont été arrêtées par manque d'expérience et d'analyse des évolutions de la gestion de données via le réseau internet. La notion d'interopérabilité des systèmes et des données et celle de gestion dynamique de données sont maintenant communément acceptées et un préalable indispensable à la création de base de données diffusable sur le réseau internet.

*Terminalf* a anticipé les problématiques induites par ces deux notions et permet de faire évoluer le projet selon l'environnement scientifique ou technique actuel et futur. De par son expérience et son évolution, *Terminalf* doit donc combler ce vide. Le projet peut s'appuyer sur un début de reconnaissance internationale de l'état actuel de *Terminalf*.

Le site est notamment consulté par nombre d'universités françaises et étrangères qui ont demandé une aide à la création d'une telle base pour leurs propres besoins. Un intérêt certain a aussi été exprimé par les membres du réseau Realiter (Réseau panlatin de terminologie/Union latine qui compte une cinquantaine d'universités de pays néolatins) et par l'Agence intergouvernementale de la francophonie ; nous comptons transformer cet intérêt en une collaboration active dans les prochaines années et en direction des réseaux universitaires francophones, notamment vers les pays du sud.

### **Contexte national**

Les terminologies de langue française atteignent aujourd'hui une masse de plusieurs millions de termes, sans commune mesure avec ce que notre langue a connu dans son histoire. Il n'existe actuellement aucun projet de ce genre, permettant de fusionner et de comparer des données terminologiques venant de plusieurs universités, de plusieurs lieux différents ou de conceptions et de méthodes terminologiques différentes.

Le site sera inscrit comme l'un des produits centraux d'un réseau national de terminologie en cours de réalisation par la Société française de terminologie. Il permettra de centraliser et de rediffuser informations et données entre les différentes universités ou institutions produisant des données terminologiques ou ayant besoin de ces données.

### **Thèmes de recherche**

En effectuant la mise sous format homogène de données numériques, différentes questions seront abordées, touchant à plusieurs champs disciplinaires : fabrication et pérennité des webographies, traitement informatique des notions de domaine et d'arborescence, évolution des principes de l'indexation automatique et semi-automatique, statut comparé

d'un terme en terminologie, en TAL et en informatique. Des recommandations pour la création d'outils de diffusion de bases de données numériques pourront ainsi être produites.

La base de *Terminalf* ayant été développée, l'une des autres orientations immédiates sera de renforcer les outils et les pages pédagogiques pour l'apprentissage et l'enseignement de la terminologie (en direction des ingénieurs, techniciens, scientifiques, étudiants, enseignants, formateurs de formateurs, etc.). Un développement attendu est la mise en ligne d'un manuel et d'un cours de terminologie. *Terminalf* est en effet un outil pédagogique susceptible d'être utilisé à distance, mais également dans le cadre de travaux collaboratifs présentiels ou virtuels.

### **Étapes à venir**

- 1) Poursuite de la mise en ligne des nouveaux mémoires de Paris III et de Paris VII (2002-2004) et introduction et mise au format actuel d'un corpus plus ancien (1998-2002).
- 2) Application d'outils statistiques et de TAL sur les données actuellement diffusées, en utilisant les compétences et les outils en TAL disponibles au sein du laboratoire MoDyCo.
- 3) Mise au point d'un outil automatisant une mise en ligne graphique d'arborescence terminologique, en se basant sur les recherches actuellement en cours en ergonomie et en cartographie.
- 4) Création d'un espace pédagogique sur la terminologie et la traduction spécialisée pour l'apprentissage et la formation, avec des formulaires de mise en ligne à distance d'informations terminologiques.

### **Résultats attendus**

- 1) Création d'un corpus de plusieurs dizaines de milliers de termes, structurés en domaines et en mémoires de terminologie, en augmentant, via des procédures automatisées, le corpus déjà disponible.
- 2) Développement de *Terminalf* vers un site de recherche et d'enseignement en terminologie à l'adresse d'utilisateurs en ligne et d'enseignants en présentiel.
- 3) Augmentation de la coopération internationale pour la création de site de ce type dans d'autres langues.

### **Actions prévues**

Journée d'études à l'université Paris X sur le thème « Traitement automatique des données terminologiques » (dans un premier temps avec Rennes II, Paris III, et Paris VII, la Société française de terminologie, etc.)

### **Contrat / financement**

Une convention avec la Société française de terminologie est en cours d'élaboration et des demandes de financement sont prévues auprès de la Délégation générale à la langue

française et aux langues de France (Ministère de la culture), l'Union latine (Réseau panlatin de terminologie), l'Agence intergouvernementale de la francophonie

### **Références sur le projet**

- Depecker L. 1992. « Langues françaises et francophonie ». *Universalia* 1992,128-132.
- Depecker L. 1995. « L'aménagement des langues ». *Universalia* 1995, 133-137.
- Depecker L., Le Meur A. 1995. « Balnéo : un projet de réseau informatique pour la veille néologique ». *Terminologies nouvelles*, 14, 48-53.
- Depecker L. 1996. « L'inventaire français des banques de terminologie effectué en 1995 dans le cadre du Réseau international de néologie et de terminologie ». *Terminologies nouvelles*, 15, 132-134.
- Depecker L. 1997. « L'internet, un médium qui bouleverse les contenus linguistiques ». *Le français dans le monde*, numéro spécial « Recherches et applications, Multimédia, réseaux et formation », 165-182.
- Depecker L. 1999. « L'ère de la terminologie informationnelle ». *Revue française de linguistique appliquée*, III (2), 7-15.
- Depecker L., Tucac C. 2000. « Recherches néologiques sur Internet ». *Terminologies nouvelles*, 20, 4-11
- Depecker L. 2001. *L'invention de la langue : le choix des mots nouveaux*. Paris : Larousse/Armand Colin.
- Depecker L. 2002. *Entre signe et concept : éléments de terminologie générale*. Paris : Presses de la Sorbonne nouvelle.
- Depecker L. À paraître. *Une linguistique pour traducteurs*. Grenade (Espagne) : Université de Grenade.

## **Annexe 1**

---

Sylvain Kahane rejoindra le laboratoire au 1<sup>er</sup> janvier 2005. Sa fiche d'activité non incluse dans le rapport du 30 octobre 2003 suit.

## **Annexe 2 : le site MoDyCo**

---

### **Adresse : [infolang.u-paris10.fr/modyco](http://infolang.u-paris10.fr/modyco)**

Le laboratoire MoDyCo gère son propre serveur ([infolang.u-paris10.fr](http://infolang.u-paris10.fr)) qui a pour mission de stocker, de diffuser et de valoriser les projets et les corpus développés au sein du laboratoire.

Actuellement, le site du laboratoire constitue le mode d'accès privilégié à ce serveur. Il a été créé en établissant une forte corrélation entre les exigences de contenu des auteurs et les besoins propres au média utilisé.

### *Objectifs*

Disposer de ce site a demandé au préalable une réflexion sur son utilité et sur le public que l'on peut ou doit toucher par ce biais. Son contenu doit en effet se plier non seulement aux contraintes techniques du média (écran informatique, vitesse de transmission, etc.) mais aussi, voire surtout, à ses spécificités de communication (un média international, instantané, interactif, etc.)

Le site réalisé n'est donc pas un simple décalque des rapports d'activités ou une vitrine publicitaire. Il tend à être un espace de recherche et de valorisation et, de même, il s'adapte aux habitudes de consultation d'un site internet où on ne peut contrôler les visites des internautes. On s'adresse alors aussi bien au public scientifique qu'à un plus grand public intéressé par les sciences du langage. Cette spécificité d'internet (permettre une consultation à plusieurs niveaux) a été prise en compte, sinon le site ne remplirait pas ses trois objectifs :

- 1) Se faire connaître et susciter l'intérêt. Pour cela, il y a nécessité d'avoir plusieurs niveaux d'informations et une synthèse des axes essentiels.
- 2) Établir un dialogue avec des personnes que l'on ne touche pas ou difficilement par d'autres médias. Pour que les internautes continuent de consulter le site, il faut une mise à jour régulière (nouveaux contenus, actualisation des liens).
- 3) Créer un lien structurant au sein même du laboratoire en favorisant l'échange d'informations et la mutualisation des connaissances et des outils.

De nouveaux modes de présentation (les cadres, les animations interactives) s'instaurent sur le réseau ; des habitudes de consultation se créent. Il convient d'être attentif à ces évolutions, d'intégrer ou non les différentes nouveautés techniques. L'objectif final est que le site propose différents modes d'accès et de navigation, pour être accessible et attractif à partir de tous les systèmes de consultation.

La navigation sur le site a donc été simplifiée ; un accent particulier est mis sur la lisibilité des informations proposées, sur l'évidence de liens souples entre *toutes* ces informations.



Ce site veut aussi être inséré dans le réseau internet (le faire connaître et créer des liens avec d'autres organismes ou institutions œuvrant dans le même domaine). Il doit devenir un outil important pour fédérer et refléter la vitalité des études en sciences du langage.

Il ne faut pas aussi négliger les autres possibilités d'expression et de diffusion qu'offre le réseau internet (forums de discussion, listes de discussions, outils de l'internet en français) : centraliser toutes ces ressources offrira un tour d'horizon nouveau des ressources sur le français ou les sciences du langage, disponibles sur internet.

### *Méthodologie*

Le contenu, hétérogène par définition, a demandé le développement d'une structure de consultation et d'analyse des données. Cette structure est l'ossature du site internet et répond aux normes et standards actuellement en usage pour la gestion dynamique de données en ligne (XML, SQL, etc.).

Le site a été développé avec une base de données fortement structurée et relationnelle, accessible qui plus est avec un simple navigateur. L'interface d'interrogation permet des requêtes larges et fines sur ces données avec de multiples possibilités de croisements inédits. Il a donc pris l'option d'être dynamique et de diffuser une information structurée, facilement évolutive et permettant une exploitation poussée du contenu.

Ce site est un élément structurant du laboratoire et demande la participation active de chaque membre. Il est constitué de trois espaces :

1) *Espace membre* : chaque membre du laboratoire y dispose d'un espace pour présenter ses axes de recherches (présentation, objectifs, résultats, etc.), ses activités (publications, communications, etc.) et diffuser un certain nombre de documents, sous réserve d'en avoir les droits de diffusion électronique. Il peut naturellement les modifier.

2) *Espace projet* : pour chaque projet, le responsable de projet présente le travail associé au projet de recherche (problématique, étapes, production scientifique, participants, etc.). Il est autorisé à modifier les données de ce projet : ajout, suppression, modification des textes de présentation, de la liste des participants, des références et textes, des liens.

Les formulaires de saisie ont été simplifiés à l'extrême, tout en tenant compte des exigences scientifiques de diffusion de l'information. Pour « alimenter » le site, il n'est donc pas nécessaire d'avoir une connaissance préalable des langages internet.

Un autre parti-pris a été de faire accompagner chaque donnée (nom, référence, lien, etc.) de la possibilité de présenter ou de commenter cette information à travers des champs textuels à la saisie en « texte libre » (forme ou taille illimitées) : diffuser une donnée brute n'est pas satisfaisant et ne permet pas sa bonne exploitation.

3) *Espace administration* : dans sa partie accessible aux membres du laboratoire par mot de passe, le site diffuse :

- a) Des outils pour la mise en ligne des textes.
- b) Des documents internes (comptes-rendus du Conseil scientifique, appels d'offres, etc.).
- c) L'annuaire du laboratoire.
- d) Une aide et les protocoles de saisie des données.
- e) Des statistiques concernant aussi bien le laboratoire (membres, projets, publications, etc.) que la fréquentation du site.

Enfin une liste de diffusion par courrier électronique permet de faire circuler rapidement l'information au sein du laboratoire.

En externe, grâce à sa base de données, le site propose de multiples navigations et croisements : les secteurs d'activités, les projets, les membres du laboratoire, les références et les textes accessibles, la webographie.

Deux outils présents sur le site sont enfin des éléments dynamiques de la vie du laboratoire et de son inscription dans la communauté scientifique :

a) L'*Agenda* : un outil qui permet aux membres du laboratoire d'annoncer des colloques, des journées d'études, des séminaires et des soutenances de thèse, etc.

b) Le *Séminaire* : cet outil diffuse le calendrier du séminaire du laboratoire, avec des courtes présentations de chaque séance. Il permet aussi aux membres du laboratoire de gérer directement les propositions d'intervention, avec un processus dynamique de proposition, validation et mise en ligne

### *Développements*

Un certain nombre d'outils sont en cours de développement ou prêts à être activés selon les besoins :

1) Une base de données bibliographiques permettant la gestion en ligne de la bibliothèque du laboratoire.

2) Un module de vote électronique pour les consultations et les élections inhérentes à la vie scientifique. Ce module a déjà été opérationnel pour le renouvellement du conseil scientifique du GDR Phonologie.

3) Une base de données de consultation et d'exploitation de données sonores et textuelles. Cette base est développée dans le cadre du projet « Phonologie de la syllabe et phonologie du français » (§ 1.1.), qui doit permettre la consultation d'un vaste corpus sonore, plus développé qu'une simple lecture des données. Cette consultation devra se fonder sur les outils de l'Internet et les spécificités d'une consultation à distance (utilisation d'un système documentaire et consultation des enregistrements sous forme compressée), avec un croisement inédit entre les données documentaires textuelles et les données sonores numérisées.

4) Une base de données de structuration XML de données textuelles ou documentaires. Cette base est développée dans le cadre du projet « *Terminalf* : ressources terminologiques en langue française » (mise en ligne de mémoires de terminologie d'étudiants, § 4.3.), en collaboration avec le Département InfoCom de l'Université Paris X (Stéphane Chaudiron, Madjid Ihadjadene), qui travaille sur la même problématique avec le projet Codex (mise en ligne de mémoires d'étudiants en documentation électronique) : il s'agira de proposer une méthodologie de structuration et d'étiquetage d'un corpus en XML qui devra ensuite être inséré dans une structure de base de données. Celle-ci sera personnalisée aux types de documents du corpus de départ mais devra aussi respecter les normes et les obligations quant à la diffusion de travaux d'étudiants et à leur exploitation à des fins scientifiques et pédagogiques.